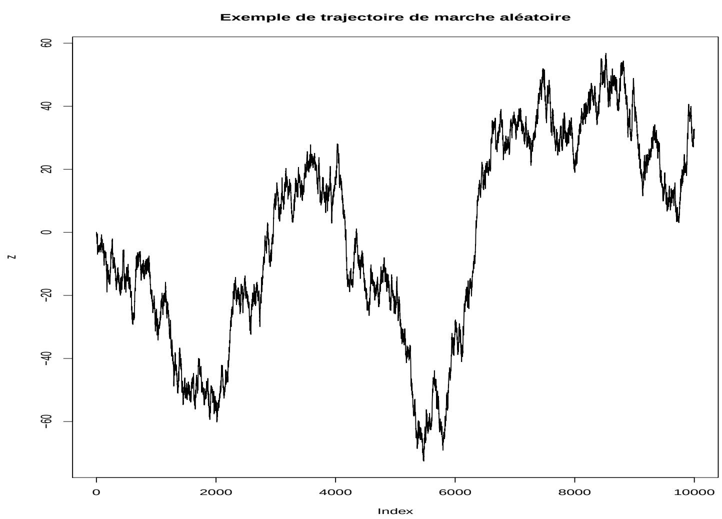
Université Paris I, Panthéon - Sorbonne

LICENCE M.I.A.S.H.S.

Cours de Probabilités

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)



U.F.R. 27 et SAMM (Statistique, Analyse et Modélisation Multidisciplinaire)
Université Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris.

Plan du cours

Introduction

1. Quelques rappels de statistiques descriptives
2. Espace de probabilité, événements et indépendance
3. Variables aléatoires
4. Théorèmes limite et statistique inférentielle

References

[1] Appel, W. (2015). Probabilités pour les Non-Probabilistes, Edition H&K.

[2] Dauxois, J. et Hassenforder, C. (2004). Toutes les probabilités et Statistiques. Cours et Exercices corrigés. Ellipses.

[3] Deledicq, JC. (2000). Faîtes vos jeux! 50 grands classiques du calcul des probabilités, ACL Les éditions du kangourou

[4] Engel, A. (1990). Les certitudes du hasard, Alea éd.

[5] Leboeuf, C., Guegand, J., Roque, J.L. et Landry, P. Cours de Probabilités et de statistiques, Ellipses

[6] Leboeuf, C.,Guegand, J.,Roque, J.L. et Landry, P. Exercices corrigés de probabilités, Ellipses

[7] Ross, S.M (2007). Initiation aux probabilités, Enseignement des Mathématiques. Presses polytechniques et universitaires romandes.

[8] Saporta, G. Probabilités, analyse des données et statistique (2nd édition), éditions Technip.

Documents accessibles librement sur internet

- Des cours sur le site STAFAV: `https://www.math.u-psud.fr/ stafav/`.
- Un site de Paris V: `http://helios.mi.parisdescartes.fr/ glaunes/Proba3/CoursProbasStats.pdf`.
- Le site du LPSM de Paris VI: `https://www.lpsm.paris/formation/supports-de-cours/`. Notamment le cours de J. Bertoin.
- Un poly de maths générales (avec probas) très bien fait: `https://www.math.univ-toulouse.fr/ barthe/L1mat`
- Le premier cours de probabilités de l'Ecole Polytechnique: `http://josselin-garnier.org/wp-content/uploa`

Introduction

- Les maths sont partout!
- La statistique et les probabilités: qu'est-ce et pour quoi faire?
- Nécessité de savoir aussi d'autres maths (algèbre, géométrie, analyse,...)

1 Statistique descriptive

1.1 Introduction

Contenu, buts et raisons d'être de la statistique.
Statistique descriptive et inférentielle.

1.2 Statistique unidimensionnelle

Définition.
- *Variable (caractère) X.*
- *Variable quantitative.*
- *Variable qualitative $\Longrightarrow$ modalités (classes).*

Définition. *Soit X une variable et une population de n individus.*

- *Répartition en k classes.*
- *Effectifs, fréquences.*
- *Diagramme à baton, diagramme circulaire ("camembert").*

Définition. *Soit X une variable quantitative.*

- *Mode, amplitude de classe, densité d'effectif, densité de fréquence.*
- *Histogramme.*
- *Fonction de répartition empirique, polygône des fréquences cumulées.*
- *Médiane empirique, quantiles empiriques.*
- *Moyenne empirique, variance empirique, écart-type.*

1.3 Statistique bidimensionnelle

- Effectifs joints, tableau de contingence.
- Nuage de points.
- Covariance empirique, corrélation empirique.
- Droite de régression.

1.4 Statistique multidimensionnelle

- Nuage de points.
- Réduction de dimension.
- Classification et clustering.

2 Espace de probabilité, événements et indépendance

2.1 Espace de probabilité

Définition.
- *Expérience aléatoire.*
- *Évènement élémentaire, ensemble fondamendal (univers).*
- *Événement et tribu.*

Définition.
- *Intersection, union.*
- *Évènements incompatibles.*
- *Évènement contraire.*

2.2 Mesure de probabilité d'un événement

Définition. *Une probabilité sur* $(\Omega, \mathcal{A})$ *est une application* $\mathbb{P} : \mathcal{A} \to [0,1]$*, qui à un événement* $E \in \mathcal{A}$ *associe le réel* $\mathbb{P}(E) \in [0,1]$ *et telle que:*

- $\mathbb{P}(\Omega) = 1$.
- *Pour* J *un ensemble inclus dans* $\mathbb{N}$, $(E_i)_{i\in J}$ *des événements incompatibles de* $\mathcal{A}$, $\mathbb{P}(\bigcup_{i\in J} E_i) = \sum_{i\in J} \mathbb{P}(E_i)$.

Propriété.
- $\mathbb{P}(\overline{E}) = 1 - \mathbb{P}(E)$ *pour* $E \in \mathcal{A}$.
- $\mathbb{P}(\emptyset) = 0$.
- *si* $A \subset B$, $0 \leq \mathbb{P}(A) \leq \mathbb{P}(B) \leq 1$.
- $\mathbb{P}(E \cup F) + \mathbb{P}(E \cap F) = \mathbb{P}(E) + \mathbb{P}(F)$*, pour* $(E,F) \in \mathcal{A}^2$.
- *Pour* $(E_i)_{i\in\mathbb{N}}$ *une suite d'événements de* $\mathcal{A}$*, telle que* $E_i \subset E_{i+1}$ *pour tout* $i \in \mathbb{N}$*, alors* $\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} E_i\right) = \lim_{n\to\infty} \mathbb{P}(E_i)$.

Définition.
- *Cardinal d'un ensemble fini.*
- *Equiprobabilité (ou probabilité uniforme) sur un ensemble fini.*
- *Equiprobabilité (ou probabilité uniforme) sur un ensemble fini.*

Propriété. *Si* Ω *est fini et si* $\mathbb{P}$ *est une probabilité uniforme sur* $(\Omega, \mathcal{A})$*, alors pour tout* $E \in \mathcal{A}$*,* $\mathbb{P}(E) = \dfrac{card(E)}{card(\omega)}$.

Remarque : Pour calculer le cardinal d'un ensemble, on peut utiliser les résultats combinatoires suivants: on considère un ensemble de n éléments et on tire k éléments:

- S'il y a remise, et que l'ordre compte, un tirage est un k-uplet, et le nombre total de k-uplets est: n^k.
- S'il n'y a pas remise, et que l'ordre compte, un tirage est un arrangement, et le nombre total d'arrangements est: $A_n^k = n(n-1)..(n-k+1) = n!/(n-k)!$.
- S'il n'y a pas remise, et que l'ordre ne compte pas, un tirage est une combinaison, et le nombre total de combinaisons est: $C_n^k = A_n^k/k! = n!/(k!(n-k)!)$.

2.3 Probabilité conditionnelle et indépendance

Définition. *Soit $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E, F) \in \mathcal{A}^2$. On appelle probabilité conditionnelle de B sachant A, la probabilité que B soit réalisé sachant que A est réalisé et* $\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ *si* $\mathbb{P}(A) \neq 0$.

Remarque : Calculer une probabilité sachant A revient à travailler avec une nouvelle probabilité sur l'ensemble fondamental A et la tribu qui lui est associée.

Définition. *A et B, événements de $(\Omega, \mathcal{A})$ sont indépendants pour $\mathbb{P}$ si $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.*

Conséquence : A et B, événements de $(\Omega, \mathcal{A})$ sont indépendants pour $\mathbb{P}$ si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Remarque : Ne pas confondre indépendance et incompatibilité! De la dernière conséquence, on montre facilement que l'on peut être incompatible sans être indépendant (puisqu'on a toujours $\mathbb{P}(A \cap \overline{A}) = 0$, différent de $\mathbb{P}(A)\mathbb{P}(\overline{A})$ sauf si l'un des deux est nul) et l'on peut être indépendant sans être incompatible (penser à deux lancers successifs d'une pièce équilibrée et aux évenèments A: le premier lancer est Pile et B: le second lancer est Pile: on a $A \cap B = \{(PP)\}$).

Définition. *Soit Ω un ensemble. On dit que $(E_i)_{i \in J}$ famille de sous-ensembles de Ω forme une partition de Ω si:*

- *Les E_i sont incompatibles deux à deux soit $E_i \cap E_j = \emptyset$ pour $i \neq j$.*
- *L'ensemble des E_i couvre Ω soit $\displaystyle\bigcup_{i \in J} E_i = \Omega$.*

Proposition. (Formule des probabilités totales) *Soit $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E_j)_{j \in J}$ des évenements de $\mathcal{A}$ constituant une partition de Ω. Alors pour tout $A \in \mathcal{A}$,*

$$\mathbb{P}(A) = \sum_{j \in J} \mathbb{P}(A \cap E_j).$$

Proposition. (Formule de Bayes) *Soit $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E_j)_{j \in J}$ des évenements de $\mathcal{A}$ constituant une partition de Ω. Soit A un événement de $\mathcal{A}$). On suppose que l'on connaît $\mathbb{P}(E_j)$ et $\mathbb{P}(A \mid E_j)$ pour tout $j \in J$. Alors, pour $k \in J$:*

$$\mathbb{P}(E_k \mid A) = \frac{\mathbb{P}(A \mid E_k)\mathbb{P}(E_k)}{\sum_{j \in J} \mathbb{P}(A \mid E_j)\mathbb{P}(E_j)}.$$

Définition. *Soit $(A_i)_{i \in I}$ une famille d'événements de $(\Omega, \mathcal{A})$. On dit que la famille $(A_i)_{i \in I}$ est une famille d'événements (mutuellement) indépendants si et seulement si pour tout $k \in \mathbb{N}$, pour tout $j_1, \cdots, j_k \in I^k$ distincts, $\mathbb{P}\left(A_{j_1} \cap \cdots \cap A_{j_k}\right) = \mathbb{P}(A_{j_1}) \times \cdots \times \mathbb{P}(A_{j_k})$.*

Remarque : Etre mutuellement indépendant est plus fort que d'être indépendant deux à deux. Prendre à ce sujet l'exemple précédent de deux lancers successifs d'une pièce équilibrée et considérer les évenèments A: le premier lancer est Pile et B: le second lancer est Pile et C: les deux lancers donnent la même chose. On aura bien A et B indépendants, A et C également, de même que B et C, et pourtant A, B et C ne sont pas mutuellement indépendants...

3 Variables aléatoires

3.1 Définitions et propriétés générales

Définition. *Soit* $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathcal{A})$. *On appelle variable aléatoire* X *dans* $I \subset \mathbb{R}$, *une application de* $\Omega \to I$ *telle que pour tout* $x \in I$, *l'événement* $\{X \leq x\} = \{\omega \in \Omega$ *tel que* $X(\omega) \leq x\}$ *soit un événement de* $\mathcal{A}$. *Deux cas particuliers importants sont à distinguer:*

- *Si* $I = \{x_i\}_{i \in J}$ *avec* $J \subset \mathbb{N}$ *(par exemple* $I = \{P, F\}$, $I = \mathbb{Z}, \ldots$*),* X *est appelée variable aléatoire discrète.*
- *Si* I *est un intervalle de* $\mathbb{R}$ *(par exemple* $I = [0, 1]$, $I = \mathbb{R}^+, \ldots$*),* X *est appelée variable aléatoire réelle.*

Définition. *Soit* $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathcal{A})$ *et* X *une variable aléatoire de* $(\Omega, \mathcal{A}, \mathbb{P})$ *à valeurs dans* I. *On appelle fonction de répartition de* X *la fonction* $F_X : \mathbb{R} \to [0, 1]$ *telle que* $F_X(x) = \mathbb{P}(X \leq x)$, *pour* $x \in \mathbb{R}$.

Propriété.

- F_x *est une fonction croissante sur* $\mathbb{R}$.
- $\lim\limits_{x \to -\infty} F_X(x) = 0$ *et* $\lim\limits_{x \to +\infty} F_X(x) = 1$.
- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$, *pour tout* $-\infty \leq a < b \leq +\infty$.

Définition. *Soit* $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathcal{A})$ *et* X *une variable aléatoire de* $(\Omega, \mathcal{A}, \mathbb{P})$ *à valeurs dans* I.

- *Si* X *est une variable aléatoire discrète (*$I = \{x_i\}_{i \in J}$*), on appelle loi de probabilité de* X *l'application* $\mathbb{P}_X : \{x_i\}_{i \in J} \to [0, 1]$ *telle que* $\mathbb{P}(X = x_i) = \mathbb{P}_X(x_i)$.
- *Si* X *est une variable aléatoire réelle (*I *intervalle de* $\mathbb{R}$*), on appelle densité de probabilité de* X **si elle existe** *l'application* $f_X : I \to [0, +\infty[$ *telle que* $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(t)dt$. X *est alors appelée variable aléatoire absolument continue.*

Propriété. *Si* X *est une variable aléatoire discrète à valeurs dans* $I = \{x_i\}_{i \in J}$ *alors* $\displaystyle\sum_{i \in J} \mathbb{P}(X = x_i) = 1$.

Propriété. *Si* X *est une variable aléatoire absolument continue de densité de probabilité* f_X *alors:*

- f_X *vérifie* $\displaystyle\int_{-\infty}^{+\infty} f_X(t)dt = 1$.
- $F'_X(x) = f_X(x)$ *pour presque tout* $x \in \mathbb{R}$.
- $\mathbb{P}(a < X \leq b) = \displaystyle\int_a^b f_X(t)dt$, *pour tout* $-\infty \leq a < b \leq +\infty$.
- $\mathbb{P}(X = x) = 0$ *pour tout* $x \in \mathbb{R}$.

3.2 Moments d'une variable aléatoire

Définition. *Soit* $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathcal{A})$, X *une variable aléatoire de* $(\Omega, \mathcal{A}, \mathbb{P})$ *à valeurs dans* I *et* h *une fonction de* I *dans* $\mathbb{R}$.

- *Si X est une variable aléatoire discrète ($I = \{x_i\}_{i\in J}$), l'espérance de X est*

$$\mathbb{E}X = \sum_{i\in J} x_i \mathbb{P}(X = x_i) = \sum_{i\in J} x_i \mathbb{P}_X(x_i) \quad \textbf{si cette somme existe.}$$

  Plus généralement, $\mathbb{E}h(X) = \sum_{i\in J} h(x_i)\mathbb{P}(X = x_i)$ **si cette somme existe**.

- *Si X est une variable aléatoire aboslument continue à valeurs dans I et de densité f_X, l'espérance de X est*

$$\mathbb{E}X = \int_I x f_X(x)dx = \int_{-\infty}^{+\infty} x f_X(x)dx \quad \textbf{si cette intégrale existe.}$$

  Plus généralement, $\mathbb{E}h(X) = \int_I h(x)f_X(x)dx$ **si cette intégrale existe**.

Définition. *Soit X une variable aléatoire de $(\Omega, \mathcal{A}, \mathbb{P})$.*

- *La variance de X,* **si elle existe**, *est* $var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}((X - \mathbb{E}X)^2) \geq 0$.
- *L'écart-type de X,* **si la variance existe**, *est* $\sigma(X) = \sqrt{var(X)}$.

Propriété. *Soit X une variable aléatoire (dont l'espérance et la variance existent), et a, b deux réels.*

- $\mathbb{E}b = b$ *et* $\mathbb{E}(aX + b) = a\mathbb{E}X + b$.
- $var(b) = 0$ *et* $var(aX + b) = a^2 var(X)$.

3.3 Lois à connaître

Définition. *Les lois à connaître sont:*

- *Lois discrètes: loi uniforme, loi de Bernoulli, loi binomiale, loi de Poisson.*
- *Lois continues: loi uniforme, loi exponentielle, loi gamma, loi normale.*

3.4 Fonction d'une autre variable aléatoire

On suppose que $Y = h(X)$ où X est une variable aléatoire.

Propriété. *Soit X variable aléatoire sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et soit $h : \mathbb{R} \to \mathbb{R}$ une fonction continue par morceaux. alors $Y = h(X)$ est également une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$.*

Pour déterminer la loi d'une fonction d'une variable aléatoire $Y = h(X)$:

1. On détermine l'ensemble des valeurs prises par Y;
2. Pour les variables discrètes, on identifie directement la probabilité de Y;
3. Pour les variables continues, on détermine la fonction de répartition de Y en revenant à celle de X, puis, suivant les cas, on détermine la probabilité de Y ou la densité de Y (en dérivant la fonction de répartition).

Remarque. *Attention! Si une fonction d'une variable aléatoire discrète est toujours une variable aléatoire discrète, une fonction d'une variable aléatoire continue peut-être une variable aléatoire continue (par exemple si h est continue et bijective), une variable aléatoire (par exemple si h est constante sur un intervalle intérieur du support) ou bien ni continue ni discrète (par exemple $Y = X$ si $X \geq \varepsilon$ et $Y = 0$ sinon).*

On rappelle que:

Propriété.
- *Pour une variable discrète à valeurs dans* $\{(x_i)_{i\in I}\}$, $\mathbb{E}h(X) = \sum_{i\in J} h(x_i)\mathbb{P}(X = x_i)$ **si cette somme existe**.
- *Pour une variable aléatoire continue à valeurs dans* I *et de densité* f_X $\mathbb{E}h(X) = \int_I h(x)f_X(x)dx$ **si cette intégrale existe**.

4 Suites de variables aléatoires indépendantes, théorèmes limite et introduction à l'estimation

4.1 Variables aléatoires indépendantes

Définition. *Soit* $(X_i)_{i\in I}$ *une famille de variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$. *Ces variables sont indépendantes si et seulement si pour tout* $J \subset I$, *pour toute famille* $(B_j)_{j\in J}$ *de boréliens de* $\mathbb{R}$ *(prendre tous les intervalles* $]-\infty, x_j]$ *suffit),*

$$\mathbb{P}\Big(\bigcap_{j\in J} X_j \in B_j\Big) = \prod_{j\in J} \mathbb{P}\left(X_j \in B_j\right).$$

Propriété. (Autre caractérisation de l'indépendance) *Soit* $(X_i)_{i\in I}$ *une famille de variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$. *Ces variables sont indépendantes si et seulement si pour tout* $J \subset I$, *et pour toute famille de fonctions* $(g_j)_{j\in J}$ *telles que les espérances suivantes existent,*

$$\mathbb{E}\Big(\prod_{j\in J} g_j(X_j)\Big) = \prod_{j\in J} \mathbb{E}\big(g_j(X_j)\big).$$

Définition. *Pour* X *et* Y *deux variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$, *on définit, si elle existe, la covariance de* X *et* Y *par:*

$$cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}\big[(X - \mathbb{E}(X))\, Y - \mathbb{E}(Y))\big].$$

Propriété. *Pour* X *et* Y *deux variables aléatoires* **indépendantes** *définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$, *alors* $cov(X,Y) = 0$. *La réciproque est fausse en générale (mais vraie dans le cas de variables gaussiennes ou Bernoulli).*

4.2 Théorèmes limite

Définition. *Pour* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$,

- **Convergence en probabilité.** *On dit que* $(X_n)_{n\in\mathbb{N}}$ *converge en probabilité vers une variable aléatoire* X *définie sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *(*X *pouvant être une constante réelle), si pour tout* $\varepsilon > 0$,

$$\mathbb{P}\left(|X_n - X| \geq \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0, \quad \textit{ce qui est noté} \quad X_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} X.$$

- **Convergence en loi.** *On dit que* $(X_n)_{n\in\mathbb{N}}$ *converge en loi vers la loi d'une variable aléatoire* X *définie sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *(*X *pouvant être une constante réelle), si pour tout* $x \in \mathbb{R}$ *où la fonction de répartition* F_X *est continue,*

$$F_{X_n}(x) = \mathbb{P}\left(X_n \leq x\right) \underset{n\to+\infty}{\longrightarrow} F_X(x), \quad \textit{ce qui est noté} \quad X_n \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} X.$$

Dans la suite, on va s'intéresser à la convergence de la moyenne empirique définie par:

Définition. *Pour* $(X_1, \ldots, X_n)$ *une famille de variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$*, on appelle* **moyenne empirique** $\overline{X}_n$ *de* $(X_1, \ldots, X_n)$ *la variable aléatoire définie sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *par*

$$\overline{X}_n = \frac{1}{n}\big(X_1 + \cdots + X_n\big) = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

On commence par obtenir les propriétés suivantes:

Propriété. *Pour* $(X_1, \ldots, X_n)$ *une famille de variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$*,*

- $\mathbb{E}\big(X_1 + \cdots + X_n\big) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) \quad \Longrightarrow \quad \mathbb{E}\big(\overline{X}_n\big) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k).$
- $var\big(X_1 + \cdots + X_n\big) = var(X_1) + \cdots + var(X_n) \quad \Longrightarrow \quad var\big(\overline{X}_n\big) = \frac{1}{n^2}\sum_{k=1}^{n} var(X_k).$

Propriété. (Inégalité de Markov) *Soit* X *une variable aléatoire* **positive** *définie sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *dont l'espérance existe. Alors pour tout* $\varepsilon > 0$*,*

$$\mathbb{P}\,(X \geq \varepsilon) \leq \frac{\mathbb{E}X}{\varepsilon}.$$

Voici une conséquence directe de cette inégalité:

Propriété. (Inégalité de Bienaymé-Tchebitchev) *Soit* X *une variable aléatoire définie sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *dont l'espérance* m *et la variance* σ^2 *existent. Alors pour tout* $\varepsilon > 0$*,*

$$\mathbb{P}\,\big(|X - m| \geq \varepsilon\big) \leq \frac{\sigma^2}{\varepsilon^2}.$$

On va maintenant appliquer tout se qui précède à la moyenne empirique d'une famille de variables aléatoires indépendantes ayant toute la même loi:

Théorème. (Loi Faible des Grands Nombres) *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires* **indépendantes** *définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$ *et* **identiquement distribuées** *par rapport à une loi dont l'espérance* m *et la variance* σ^2 *existent. Alors :*

$$\overline{X}_n \xrightarrow[n\to+\infty]{\mathcal{P}} m = \mathbb{E}(X_1).$$

Il est possible d'être plus précis quant au comportement asymptotique de la moyenne empirique autour de son espérance: c'est ce que précise le théorème suivant, à savoir que ce comportement est gaussien et se resserre à vitesse $/\sqrt{n}$ autour de l'espérance!

Théorème. (Théorème de la limite centrale) *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires* **indépendantes** *définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$ *et* **identiquement distribuées** *par rapport à une loi dont l'espérance* m *et la variance* σ^2 *existent. Alors :*

$$\sqrt{n}\Big(\frac{\overline{X}_n - m}{\sigma}\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

4.3 Introduction à l'estimation d'un paramètre et à l'obtention d'un intervalle de confiance

Définition. *Si on dispose de* $(X_1, \ldots, X_n)$ *une famille de* n *variables aléatoires définies sur le même espace de probabilités* $(\Omega, \mathcal{A}, \mathbb{P})$*, un estimateur d'un vecteur* $\theta \in \mathbb{R}^d$ *est une fonction (mesurable... donc typiquement continue par morceaux) de* $(X_1, \ldots, X_n)$ *ne dépendant pas de* θ*.*

Cette définition est très générale et peut être reprécisée dans le cas particulier de l'espérance de la loi d'une famille de v.a.i.i.d.:

Estimation d'une moyenne et intervalles de fluctuations et de confiance:
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires **indépendantes** définies sur le même espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$ et **identiquement distribuées** par rapport à une loi dont l'espérance $m \in \mathbb{R}$ existe mais est **inconnue** (on suppose également que $\sigma^2 < \infty$). On suppose que l'on a **observé** $(X_1, \ldots, X_n)$, c'est à dire qu'il existe $\omega \in \Omega$ tel que l'on connaisse le vecteur $(X_1(\omega), \ldots, X_n(\omega))$. Alors, on peut estimer m par l'estimateur $\widehat{m} = \overline{X}_n = \frac{1}{n}\left(X_1 + \cdots + X_n\right)$ et on montre:

1. On a $\widehat{m} \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} m$ par la Loi Faible des Grands Nombres: l'estimateur est convergeant, il se rapproche du paramètre inconnu lorsque le nombre de données croît.

2. Plus précisément, on a $\sqrt{n}\big(\widehat{m} - m\big) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \sigma^2)$: on mesure ainsi comment cet estimateur se rapproche de m, d'où les deux intervalles suivant:

   - Pour n suffisamment grand, $\big[m - 1.96\sigma/\sqrt{n}\,,\, m - 1.96\sigma/\sqrt{n}\big]$ est l'intervalle de fluctuation de $\widehat{m}$ à 95%;
   - Pour n suffisamment grand, $\big[\widehat{m} - 1.96\sigma/\sqrt{n}\,,\, \widehat{m} - 1.96\sigma/\sqrt{n}\big]$ est l'intervalle de confiance de m à 95%.

   En pratique, il faut remplacer σ par un estimateur convergeant de σ (par exemple $\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^n \big(X_k - \overline{X}_n\big)^2$) et on peut montrer que cela ne change rien aux intervalles précdéents!