Chap 0 - Motivations et Rappels

Jordy Palafox

CY Tech - Ing 2 GSI 2023-2024

Statistiques Inférentielles

En statistique descriptive, ...

on analyse des données. On veut les résumer en utilisant :

- des représentaions graphiques (histogrammes de fréquence, courbe de pourcentages cumulés, etc...)
- des indicateurs numériques (moyenne, variance, écart-type, médiane, quantiles, coefficient de Skewness (symétrie de la distribution), coefficient de Kurtosis (applatissement de la distribution).

Tout cela a été vu en Data exploration en ING1 !

En probabilité, ...

on étudie des variables aléatoires.
Ces variables vont modéliser des phénomènes. On peut alors étudier un phénomène ponctuel ou la répétition (le plus souvent indépendante) de ces phénomènes donc des suites de variables aléatoires donc on veut connaître les comportements limites etc...

Lier les deux domaines ...

... grâce aux **Statistiques inférentielles** !

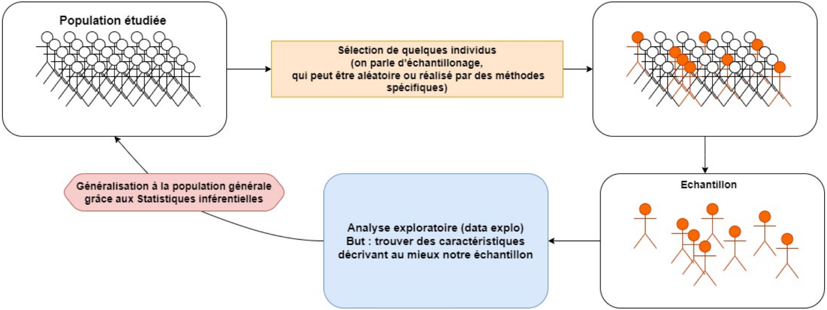
Le postulat

Le monde que nous percevons est régi par des lois de probabilités trop difficiles (car trop de paramètres gouvergent notre univers) pour permettre à l'humain de les percevoir. Ce que l'on voit est le résultat d'évènements quasi-certains, moyennisés. Ainsi on autorise la possibilité d'évènements inattendus, improbables...
Mais où sont les statistiques inférentielles ?

Les Stats Inf. en bref

Partant de l'observation d'un grand nombre de données, on veut essayer de trouver un modèle probabiliste qui colle le mieux aux données et généraliser à une population totale.

En résumé



Population étudiée
Sélection de quelques individus
(on parle d'échantillonage,
qui peut être aléatoire ou réalisé par des méthodes
spécifiques)
Echantillon
Analyse exploratoire (data explo)
But : trouver des caractéristiques
décrivant au mieux notre échantillon
Généralisation à la population générale
grâce aux Statistiques inférentielles

Pour quelles applications ?

Quelques exemples

- Etudier la propagation d'un virus sans tester l'ensemble d'une population.
- Valider la qualité de pièces industrielles produites à la chaîne sans les tester toutes.
- On modélise un phénomène physique par une loi gaussienne dont on doit déterminer l'espérance et la variance pour généraliser et prédire le comportement futur.

Bien garder à l'esprit que l'outil de base ici, c'est les probabilités !

Formalisme

Dans la suite, on considèrera un **échantillon** $(x_1, \ ..., \ x_n) \in \mathbb{R}^n$ **réalisation** d'un vecteur de variables aléatoires $(X_1, \ ..., \ X_n)$. On supposera toujours que $(X_1, \ldots, X_n)$ est un échantillon de variables aléatoires :

- **indépendantes** i.e $\mathbb{P}(X_1, ..., X_n) = \prod \mathbb{P}(X_i)$ et
- **identiquement** distribuées i.e de même loi μ inconnue.

On notera en abrégé i.i.d.

Attention

En statistique, on ne connait que les x_i !

Rappel : x_i est une réalisation de X_i signifie $x_i = X_i(\omega)$ pour $\omega \in \Omega$ l'univers.

Définition

On appelle espace probabilisé un triplet $(\Omega, \mathcal{T}, \mathbb{P})$ tel que :

- Ω est un ensemble appelé **univers** ou **espace fondamental** de tous les évènements élémentaires,
- $\mathcal{T}$ est une **tribu** i.e un ensemble de parties de Ω vérifiant :
  1. $\emptyset, \ \Omega \in \mathcal{T}$,
  2. si $A \in \mathcal{T}$ alors $A^{\complement}A \in \mathcal{T}$,
  3. pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{T}$, on a $\underset{i\in\mathbb{N}}{\cup} A_i \subset \mathcal{T}$.
- $\mathbb{P}$ est une mesure de probabilités, c'est-à-dire une application $\mathbb{P} : \mathcal{T} \to [0,1]$ telle que :
  1. $\mathbb{P}(\Omega) = 1$ (évènement certain !)
  2. pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'évènements disjoints 2 à 2, $\mathbb{P}(\underset{i\in\mathbb{N}}{\cup} A_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$.

Cas discret i.e Ω fini ou dénombrable

- $\mathcal{T} = \mathcal{P}(\Omega)$ i.e toute partie de Ω est un évènement
- La mesure $\mathbb{P}$ est donnée par l'ensemble des $\mathbb{P}(w),\ w \in \Omega$ (et on a une formule directe pour $\mathbb{P}$)
- $\sum_{w \in \Omega} \mathbb{P}(w) = 1$ (ça on l'utilise tout le temps,**retenez-le.**

Cas continu i.e $\Omega = \mathbb{R}$ ou un intervalle de $\mathbb{R}$, μ mesure de Lebesgue

- $\mathcal{T} = \mathcal{B}(\mathbb{R})$ tribu borélienne, c'est la plus petite tribu engendrée par les intervalles,
- $\mathbb{P}$ est donnée par sa fonction de densité $f : \Omega \to \mathbb{R}$ poisitive, mesurable et telle que $\int_\Omega f(x)d\mu = 1$, la proba d'un évènement$A$ est donnée par :

$$\mathbb{P}(A) = \int_A f(x)d\mu$$

Définition

L'application $X : \Omega \to \mathbb{R}$ est une **variable aléatoire** si $X^{-1}(B) \in \mathcal{T}, \forall B \in \mathcal{B}(\mathbb{R})$ avec $\mathcal{B}(\mathbb{R})$ tribu borélienne.

- Si Ω discret, X est une variable aléatoire discrète,
- Si Ω "continu", X est une v.a. continue.

Loi de probabilité

Cas discret i.e Ω fini ou dénombrable

La **loi de probabilité** est donnée par la fonction de masse p_X de $\mathbb{P}$:

$$\forall w \in \Omega,\ p_X(x_k) = \mathbb{P}(\{w \in \Omega, X(w) = x_k\}).$$

Cas continu i.e $\Omega = \mathbb{R}$ ou un intervalle de $\mathbb{R}$, μ mesure de Lebesgue

La loi est définie par la densité de probabilité $f_X : \mathbb{R} \to \mathbb{R}$, continue presque-partout et vérifiant :

$$\int_{\mathbb{R}} f(x)dx = 1.$$

Fonction de répartition

Définition

La **fonction de répartition** d'une variable aléatoire X est une application $F_X : \mathbb{R} \longrightarrow [0,1]$ telle que :

$$\forall x \ R, \ F_X(x) = \mathbb{P}(X \leq x) = p_x(]-\infty, x]).$$

Cas discret

$$F_X(x) = \sum_{\substack{k \\ x_k \leq x}} p_X(x_k)$$

Cas continu

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Propriété

La fonction de répartition F_X vérifie :

- croissante,
- $\lim\limits_{x \to -\infty} F_X(x) = 0$ et $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Espérance, variance et écart-type

On notera $\mathbb{E}(X) = \mu$ et $\mathbb{V}ar(X) = \sigma^2$.

Cas discret

$$\mathbb{E}(X) = \sum_k x_k \mathbb{P}(X = x_k) = \sum_k x_k \mathbb{P}_X(x_k)$$

$$\mathbb{V}ar(X) = \mathbb{E}\left(\mathbb{E}(X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mu^2 = \sum_k x_k^2 \mathbb{P}(X = x_k) - \mu^2.$$

Cas continu

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$\mathbb{V}ar(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2$$

L'écart-type noté σ_X de X est la racine de la variance i.e $\sigma_X = \sqrt{\mathbb{V}ar(X)}$.

Propriétés de l'espérance et la variance

Linéarité de l'espérance

$\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$ pour $\alpha,\ \beta \in \mathbb{R}$ et $X,\ Y$ deux VA.

Propriété de la variance

- $\mathbb{V}ar(\alpha X) = \alpha^2\mathbb{V}ar(X)$,
- Si X et Y **indépendantes**, $\mathbb{V}ar(X+Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y)$.
- $\mathbb{V}ar(\alpha X + \beta Y) = \alpha^2\mathbb{V}ar(X) + \beta^2\mathbb{V}ar(Y) + 2\alpha\beta Cov(X,Y)$.

Covariance

$Cov(X,Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right).$

Loi uniforme (discrète) $\mathcal{U}_{\{1,\ldots,n\}}$

- $X(\Omega) = \{1, \ldots, n\}$,
- $\mathbb{P}(X = x_k) = \frac{1}{n}$,
- $\mathbb{E}(X) = \frac{\sum_{i=1}^{n} x_i}{n}$, $\mathbb{V}ar(X) = \frac{\sum_{i=1}^{n} x_i^2}{n} - \mu^2$
- *Modélisation* : un lancer de dé !

Quelques rappels sur les lois discrètes

Loi de Bernouilli $\mathcal{B}(p)$

Si $X \sim \mathcal{B}(p)$, $p \in [0,1]$ alors :

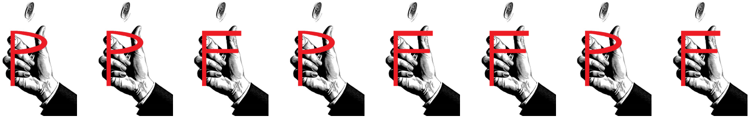
- $X(\Omega) = \{0,1\}$
- $\mathbb{P}(X=1) = p$, $\mathbb{P}(X=1) = 1-p$
- $\mathbb{E}(X) = p$, $\mathbb{V}ar(X) = p(1-p)$
- *Modélisation* : on parle d'épreuve de Bernouilli s'il n'y a que deux issues possibles (succès ou échec, pile ou face, etc)

Quelques rappels sur les lois discrètes

Loi binomiale $\mathcal{B}(n, p)$

Si $X \sim \mathcal{B}(n, p)$, $p \in [0, 1], n \in \mathbb{N}^*$ alors :

- $X(\Omega) = \{0, ..., n\}$
- $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- $\mathbb{E}(X) = np$, $\mathbb{V}ar(X) = np(1 - p)$
- *Modélisation* : c'est la somme de n épreuves de Bernouilli indépendantes et indentiquement distribuées, on compte le nombre de succés.



Quelques rappels sur les lois discrètes

Loi de Poisson $\mathcal{P}(\lambda)$

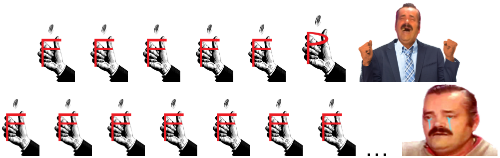
Si $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$ alors :

- $X(\Omega) = \mathbb{N}$
- $\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ pour $k \in \mathbb{N}$
- $\mathbb{E}(X) = \lambda, \ \mathbb{V}ar(X) = \lambda$
- *Modélisation* : Occurence d'un évènement sur un intervalle de temps donné. Par exemple, le nombre de voitures passant à un péage sur une journée.

Loi géométrique $\mathcal{G}(p)$

Si $X \sim \mathcal{P}(\lambda)$, $p \in [0,1]$ alors :

- $X(\Omega) = \mathbb{N}^*$
- $\mathbb{P}(X = k) = (1-p)^{k-1}$ pour $k \in \mathbb{N}^*$
- $\mathbb{E}(X) = \frac{1}{p}$, $\mathbb{V}ar(X) = \frac{1-p}{p^2}$
- *Modélisation* : succession d'épreuves de Bernouilli jusqu'au premier succés (potentiellement il n'arrive jamais !)



Quelques rappels sur les lois continues

Loi Uniforme $\mathcal{U}_{(a,b)}$

Si $X \sim \mathcal{U}_{(a,b)}$ alors :

- $X(\Omega) = [a, b]$
- sa densité est $f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$
- $\mathbb{E}(X) = \frac{b+a}{2}$, $\mathbb{V}ar(X) = \frac{(b-a)^2}{12}$
- *Modélisation* : toutes les issues dans $[a, b]$ ont la même probabilité (c'est la version continue d'une loi équiprobable dans un ensemble fini)

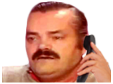
Rappels :

- $\mathbf{1}_A(x) = 1$ si $x \in A$ et 0 sinon, c'est la **fonction indicatrice.**

Loi Exponentielle $\mathcal{E}(\theta)$

Si $X \sim \mathcal{E}(\theta)$, $\theta > 0$ alors :

- $X(\Omega) = \mathbb{R}^+$
- Sa densité est $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \mathbf{1}_{x>0}(x)$
- $\mathbb{E}(X) = \frac{1}{\theta}$, $\mathbb{V}ar(X) = \frac{1}{\theta^2}$
- *Modélisation* : modélise une durée de vie (d'une ampoule, d'un atome radioactif ...) et cela peut aussi modéliser un temps d'attente (au téléphone par exemple).



Loi Normale $\mathcal{N}(\mu, \sigma^2)$

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors :

- $X(\Omega) = \mathbb{R}$
- Sa densité est $f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $\mathbb{E}(X) = \mu, \ \mathbb{V}ar(X) = \sigma^2$
- *Modélisation* : correspond au comportement limite d'une suite de variables aléatoires i.i.d correspondants souvent à des phénomènes naturels. Le terme de mesures physiques réalisées avec un instrument (pression, longueur etc...) est souvent modélisée par une loi normale (on dit aussi gaussienne).

On vient de parler de comportement limite, c'est quoi déjà ?

Définition

Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelle.

- $(X_n)_{n\in\mathbb{N}}$ converge **presque-sûrement** vers X quand n tend vers $+\infty$ s'il existe un évènement A tel que :

$$\mathbb{P}(A) = 1 \quad et \quad \lim_{n\to+\infty} X_n(w) = X(w) \quad \forall w \in A.$$

- $(X_n)_{n\in\mathbb{N}}$ converge **en moyenne** vers X quand n tend vers $+\infty$ si :

$$\lim_{n\to+\infty} \mathbb{E}(|X_n - X|) = 0.$$

Notions de convergence

Définition

Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles et X une variable aléatoire réelle.

- $(X_n)_{n\in\mathbb{N}}$ converge **en moyenne quadratique** vers X quand n tend vers $+\infty$ si :

$$\lim_{n\to+\infty} \mathbb{E}(|X_n - X|^2) = 0.$$

- $(X_n)_{n\in\mathbb{N}}$ converge **en probabilité** vers X quand n tend vers $+\infty$ si :

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Définition

Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles de fonctions de répartition F_{X_n} et X une variable aléatoire réelle de fonction de répartition F_X.

- $(X_n)_{n\in\mathbb{N}}$ converge **en loi** vers X quand n tend vers $+\infty$ si :

$$\forall x \in \mathbb{R},\ \ tel\ que\ \mathbb{P}(X = x) = 0,\ \ on\ a\ \lim_{n\to+\infty} F_{X_n}(x) = F_X.$$

		Moyenne quadratique		
		$\Downarrow$		
		Moyenne		
		$\Downarrow$		
p.s	$\Rightarrow$	Proba.	$\Rightarrow$	Loi

Le théorème centrale limite

La loi normale intervient dans de nombreux domaines, la justification théorique d'une telle utilisation provient du TCL :

Théorème

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d de carrés intégrables (i.e $|\mathbb{E}(X_i^2)| < +\infty$).

On note $\mathbb{E}(X_i) = m$ et $\mathbb{V}ar(X_i) = \sigma^2$. On pose $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Alors :

$$Z_n = \sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

En pratique

Dès que l'échantillon est assez grand ($n > 30$), on considère :

$$Z_n = \sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

Autres théorèmes de convergence

Loi faible des grands nombres

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d et de carrés intégrables. Alors $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge en moyenne quadratique vers $\mathbb{E}(X_1)$.

Loi forte des grands nombres

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d intégrables (i.e $\mathbb{E}(|X|)$ fini). Alors $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge presque-sûrement vers $\mathbb{E}(X_1)$.

Derniers rappels

Inégalité de Markov

Soit X une variable aléatoire réelle. Alors $\forall a > 0$, on a :

$$\mathbb{P}(|X| > a) \leq \frac{1}{a}\mathbb{E}(|X|).$$

Inégalité de Bienaymé-Tchébichev

Soit X une variable aléatoire réelle de carré intégrable. Alors $\forall a > 0$, on a :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{1}{a^2}\mathbb{V}ar(X).$$

Loi du Chi-deux χ^2_n

- C'est la loi d'une somme de n carrés indépendants de loi $\mathcal{N}(0,1)$:

$$X \sim \sum_{i=1}^{n} X_i^2$$

avec $X_i \sim \mathcal{N}(0,1)$.

- On remarque que $X \geq 0$.
- $\mathbb{E}(X) = n$, $\mathbb{V}ar(X) = 2n$.

Pour finir : Quelques nouvelles lois !

Loi de Student $\mathcal{T}(n)$ (ou Student-Fisher)

- C'est la loi dee la variable $X = \frac{U}{\sqrt{\frac{V}{n}}}$ avec $U \sim \mathcal{N}(0,1)$ et $V \sim \chi^2_n$ indépendantes.
- $\mathbb{E}(X) = 0$, $\mathbb{V}ar(X) = \frac{n}{n-2}$, $n \geq 2$.

Pour $n = 1$, c'est une loi de Cauchy, nous ne l'utiliserons pas en principe.

Pour finir : Quelques nouvelles lois !

Loi de Fisher $\mathcal{F}(p, q)$

- C'est la loi d'une variable $X = \frac{\frac{U}{p}}{\frac{V}{q}}$ avec $U \sim \chi^2_p$ et $V \sim \chi^2_q$ indépendantes.
- On remarque que $X \geq 0$.
- $\mathbb{E}(X) = \frac{q}{q-2}$, $\mathbb{V}ar(X) = \frac{2q^2(p+q-2)}{p(q-2)^2(q-4)}$.

Les références du cours

- *Mathematiques L3 appliquees : Cours complet avec 500 tests et exercices corriges*, sous la direction de Yger et Weil.
- Cours de Paul Rochet, `https://www.math.sciences.univ-nantes.fr/~rochet/enseignements.html`.
- Cours Abderrahim Bourhattas.
- `https://www.bibmath.net/`
- `http://exo7.emath.fr/`