

UV Statistique pour l'ingénieur

Cours n° 1

Introduction

Rappels de probabilités

Variables aléatoires

Statistique pour l'ingénieur

- Objectif :
 - Prise en compte de l'aléatoire pour aider à prendre une décision
 - « Comment prévoir en présence du hasard »
- Domaine d'application :
 - Médecine, qualité, environnement, assurance, jeux, informatique, ...
- Signification ambiguë du terme :
 - La Statistique = ensemble de méthodes permettant d'analyser (de traiter) des ensembles d'observations (des données)
 - Une statistique = une donnée statistique (exemple : statistique du commerce extérieur français)

Statistique pour l'ingénieur

- Vocabulaire :
 - Population (limitée ou de très grande taille)
 - Individus appartenant à la population

 - Recensement = Interroger toute la population
 - Sondage = Interroger une partie de la population (un échantillon)

 - Variable = caractéristique définie sur la population
 - quantitative :
 - Discrète (exemple : age)
 - continue (exemple : taille)

 - qualitative :
 - Nominale (exemple : couleur des yeux = marron | vert | bleu)
 - ordinale (exemple : type de voiture = aucune | petite | moyenne | grande)

Démarche statistique

- Les méthodes statistiques peuvent se séparer (approx.) en deux familles :
 - Statistique *descriptive* (ou *exploratoire*)

 - Statistique *inférentielle* (ou *décisionnelle*)

Démarche statistique

- Statistique *descriptive* (ou exploratoire)
 - Rôle :
 - *ressortir des propriétés* de l'échantillon étudié,
 - *suggérer des hypothèses*
 - Méthodes d'analyse de données :
 - *représentation des données*
 - *classification* pour réduire la taille de l'ensemble d'individus (i.e. regrouper ceux qui se ressemblent en « classes »)
 - *factorielles* pour réduire le nombre de variables (i.e. analyse en composantes principales, analyse de correspondances)

Démarche statistique

- Statistique *inférentielle* (ou décisionnelle)
 - Rôle :
 - *étendre les propriétés* constatées sur un échantillon à toute la population
 - *Vérifier l'adéquation des hypothèses a priori* ou issues d'une phase exploratoire
 - Méthodes :
 - *estimation d'une moyenne*
 - *vérification d'une hypothèse* (ou test)
 - *modélisation et prévision statistique*

Rappels de Probabilité

- Expérience aléatoire :
expérience dont le résultat ne peut pas être prévu *a priori*
- Espace fondamental Ω : ensemble des résultats possibles
- Événement *élémentaire* ω : élément de l'espace fondamental
- Événement aléatoire ($\subset \Omega$) pouvant être VRAI ou FAUX suivant le résultat de l'expérience aléatoire

Rappels de probabilité

- Liens entre ensembles et probabilités

ω	point de Ω	événement élémentaire
A	sous-ensemble de Ω	événement aléatoire
$\omega \in A$	ω appartient à A	ω réalise A
$A \subset B$	A est contenu dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
\bar{A}	complémentaire de A	contraire de A
\emptyset	ensemble vide	événement impossible
Ω	ensemble plein	événement certain
$A \cap B = \emptyset$	A et B disjoints	A et B incompatibles

Rappels de probabilité

- A_1, \dots, A_n est un *système complet d'événements*, si les parties A_1, \dots, A_n de Ω constituent une partition de Ω (par exemple A et \bar{A} forment un système complet)
- \mathcal{A} est une *tribu* sur Ω si c'est un ensemble de parties de Ω stable par intersection et par union et dénombrable
- (Ω, \mathcal{A}) est un *espace probabilisable*, avec \mathcal{A} est une tribu de parties de Ω

Rappels de probabilité

- Définition d'une loi de *probabilité* sur (Ω, \mathcal{A}) :
 - C'est une application $P : \mathcal{A} \rightarrow [0, 1]$ telle que
 - $P(\Omega) = 1$
 - $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$, pour tout ensemble dénombrable d'événements incompatibles A_1, \dots, A_n
- (Ω, \mathcal{A}, P) est un *espace probabilisé*

- Propriétés :

$$P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A) \leq P(B) \text{ si } A \subset B$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\cup_{i \in I} A_i) \leq \sum P(A_i)$$

$$0 \leq P(A_i) \leq 1$$

Rappels de probabilité

- Probabilité conditionnelle de l'événement A sachant B :
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
- Indépendance de 2 événements A et B :
$$P(A \cap B) = P(A)P(B)$$

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$
- Indépendance mutuelle des événements :
$$P(\cap A_i) = \prod_{i \in I} P(A_i)$$
 - Indépendance 2 à 2 des événements, mais *la réciproque est fautive !*
- Théorème de Bayes pour 2 événements A et B quelconques :
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Conditionnement et indépendance

- **Exemple :**
 - La législation de la marijuana aux USA

	Pour	Contre
Est	0.078	0.222
Autre	0.182	0.518

 - $P(\text{Est}) = 0.300$
 - $P(\text{Autres}) = 0.700$
 - $P(\text{Pour}) = 0.260$
 - $P(\text{Pour}|\text{Est}) = P(\text{Pour} \cap \text{Est}) / P(\text{Est}) = 0.078 / 0.300 = 0.260$

⇒ Pour et Est sont deux événements indépendants

Variables aléatoires

- Variable aléatoire :
application (fonction) mesurable $X : (\Omega, A, P) \rightarrow (\Omega', A')$
telle que $\forall A' \in A', X^{-1}(A') \in A$
- Classification :
 - Variable aléatoire réelle si $\Omega' = \mathfrak{R}$
 - Variable aléatoire discrète réelle si $\Omega' = V$,
où V est un sous-ensemble fini ou dénombrable de \mathfrak{R}
- Remarque:
 - ne pas confondre une variable aléatoire, notée X ,
avec la valeur prise par cette variable, notée x .

Variables aléatoires

- Exemples :
 - Tirer un dossier médical : le patient à la maladie ou n'a pas la maladie
 - Regarder deux caisses à Leclerc : nombre de personnes total dans les deux queues
 - Choisir une personne à une caisse : son temps d'attente
 - Lancer deux dés : somme des deux résultats
 - Un tour à la roulette, rien ne va plus : ses gains ou ses pertes...
 - Cueillir une pomme : son poids

Variables aléatoires

- Loi de probabilité de la variable aléatoire X :
mesure image de P par X

$$P_X(B) = P(\{\omega \in \Omega \mid X(\omega) \in B\}) = P(X^{-1}(B))$$

- Notations :
 - $P(X \in B)$ pour $P_X(B)$
 - $P(X < x)$ pour $P(X \in]-\infty, x])$

Variable aléatoire discrète

- Loi d'une variable aléatoire discrète X :

$$p(x) = P(\{\omega; X(\omega) = x\}) \text{ pour } x \in V$$

- Propriétés :
 - $0 \leq p(x) \leq 1$, pour tout $x \in V$
 - $\sum_{x \in V} p(x) = 1$
 - Valeur de la probabilité de toute partie de V : $P(X \in A) = \sum_{x \in A} p(x)$

Variable aléatoire continue

- Densité de probabilité d'une variable aléatoire continue X : $f(x)$

$$P(X \in I) = \int_I f(x) dx \text{ pour tout intervalle } I \text{ de } \mathbb{R}$$

- Propriétés :

- $f(x) \geq 0$, pour tout x

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

- Valeur approx. de la probabilité : $P(x < X < x + \Delta x) = f(x) \Delta x + o(\Delta x)$

- Remarque : la densité de probabilité $f(x)$ peut être supérieure à 1 (contrairement à la probabilité $p(x)$)

Variables aléatoires

- Fonction de répartition d'une variable aléatoire réelle X :

$$F : \mathbb{R} \rightarrow [0,1] \text{ définie par } \forall x \in \mathbb{R}, F(x) = P(X < x)$$

- Propriétés :

- si X est continue : $F(x) = \int_{-\infty}^x f(t) dt$

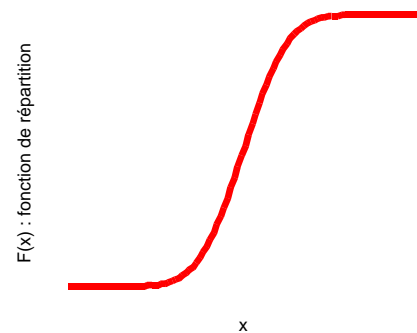
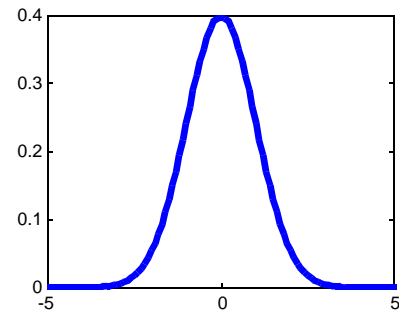
- F est une fonction croissante, continue à gauche

- $\lim_{x \rightarrow -\infty} F(x) = 0$ $\lim_{x \rightarrow +\infty} F(x) = 1$

- $P(a \leq X < b) = F(b) - F(a)$

Variables aléatoires

- Exemple de densité de probabilité
- Exemple de fonction de répartition



Fonction d'une variable aléatoire

- Loi d'une fonction d'une variable aléatoire réelle $Y = \varphi(X)$

- Variable aléatoire discrète

$$P(Y = a) = \sum_{\{x \mid \varphi(x) = a\}} P(X = x) = \sum_{\{x \mid \varphi(x) = a\}} p(x)$$

- Variable aléatoire continue

Supposons

- X continue avec une densité f et une fonction de répartition F
- φ dérivable

Si φ bijective et strictement croissante

$$P(Y < y) = P(\varphi(X) < y) = P(X < \varphi^{-1}(y)) = F(\varphi^{-1}(y))$$

Si φ bijective et strictement décroissante

$$P(Y < y) = 1 - F(\varphi^{-1}(y))$$

Moments des variables aléatoires

- Espérance mathématique d'une variable aléatoire discrète :

$$E(X) = \sum_{x \in V} xP(X = x) = \sum_{x \in V} xp(x)$$

- Espérance math. d'une variable aléatoire continue :

$$E(X) = \int_{\mathbb{R}} xf(x) dx$$

- Espérance math. d'une fonction d'une variable aléatoire :

$$E(\varphi(X)) = \int_{\mathbb{R}} \varphi(x)f(x) dx$$

Moments des variables aléatoires

- Remarques sur l'Espérance Mathématique :

- L'espérance peut ne pas exister !!!
- Valeur de l'espérance math. issue d'un calcul :
 - ponctuel pour f connue
 - théorique pour f inconnue !!!

- Propriétés :

$$E(a) = a$$

$$E(aX) = aE(X)$$

$$E(aX + bY) = aE(X) + bE(Y)$$

$$E(XY) = E(X)E(Y) \text{ si } X \text{ et } Y \text{ v. a. indépendantes}$$

Moments des variables aléatoires

- Variance d'une variable aléatoire : $var(X) = \sigma^2 = E([X - E(X)]^2)$
où σ est l'écart type de la v. a. X $var(X) = E(X^2) - (E(X))^2$

- Propriétés :

$$E((X - a)^2) = var(X) + (E(X) - a)^2 \text{ (Konig - Huyghens)}$$

$$var(X + a) = var(X)$$

$$var(aX) = a^2 var(X)$$

$$var(X) = 0 \Leftrightarrow X = a \text{ (presque sûrement)}$$

$$P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$$

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

$$cov(X, Y) = E(XY) - E(X)E(Y) = E((X - E(X))(Y - E(Y)))$$

$$var(X + Y) = var(X) + var(Y) \text{ si } X \text{ et } Y \text{ v. a. indépendantes}$$

Moments des variables aléatoires

- Moment centré d'ordre k : $\mu_k = E([X - E(X)]^k)$

- $\mu_1 = 0$ et $\mu_2 = var(X)$

- Pour une loi symétrique: $\mu_{2k+1} = 0 \quad \forall k$

- Moment non-centré d'ordre k : $m_k = E(X^k)$

- Remarques :

- Variance $Var(X) =$ moment centré d'ordre 2
- Espérance $E(X) =$ moment non-centré d'ordre 1

- Coefficient d'asymétrie : skewness $\gamma_1 = \frac{\mu_3}{\sigma^3}$
- Coefficient d'aplatissement : kurtosis $\gamma_2 = \frac{\mu_4}{\sigma^4}$