UV Statistique pour l'ingénieur

Cours n° 1

Introduction

Rappels de probabilités

Variables aléatoires

Principales lois discrètes et distributions continues

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 ambigüe 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
$\overline{A}$	complémentaire de A	contraire de A
$\varnothing$	ensemble vide	événement impossible
Ω	ensemble plein	événement certain
$A \cap B = \phi$	A et B disjoints	A et B incompatibles

Rappels de probabilité

- $A_1,\ldots,A_n$ est un *système complet d'événements* ,
  si les parties $A_1,\ldots,A_n$ de Ω constituent une partition de Ω
  (par exemple *A* et $\overline{A}$ 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

- $(\Omega, \mathcal{A})$ est un *espace probabilisable*,
  avec $\mathcal{A}$ est une tribu de parties de Ω

Rappels de probabilité

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

$$
\boxed{
\begin{aligned}
&P(\phi)=0\\
&P(\overline{A})=1-P(A)\\
&P(A)\leqslant P(B)\ si\ A\subset B\\
&P(A\cup B)=P(A)+P(B)-P(A\cap B)\\
&P(\cup A_i)\leqslant \sum P(A_i)\\
&0\leqslant P(A_i)\leqslant 1
\end{aligned}
}
$$

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 fausse !*

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

  $\Rightarrow$ Pour et Est sont deux événements indépendants

Variables aléatoires

- Variable aléatoire :
  application (fonction) mesurable X : $(\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$
  telle que $\forall A' \in \mathcal{A}'$, $X^{-1}(A') \in \mathcal{A}$

- Classification :
  - Variable aléatoire réelle si $\Omega' = \Re$
  - Variable aléatoire discrète réelle si $\Omega' = V$,
    où V est un sous-ensemble fini ou dénombrable de $\Re$

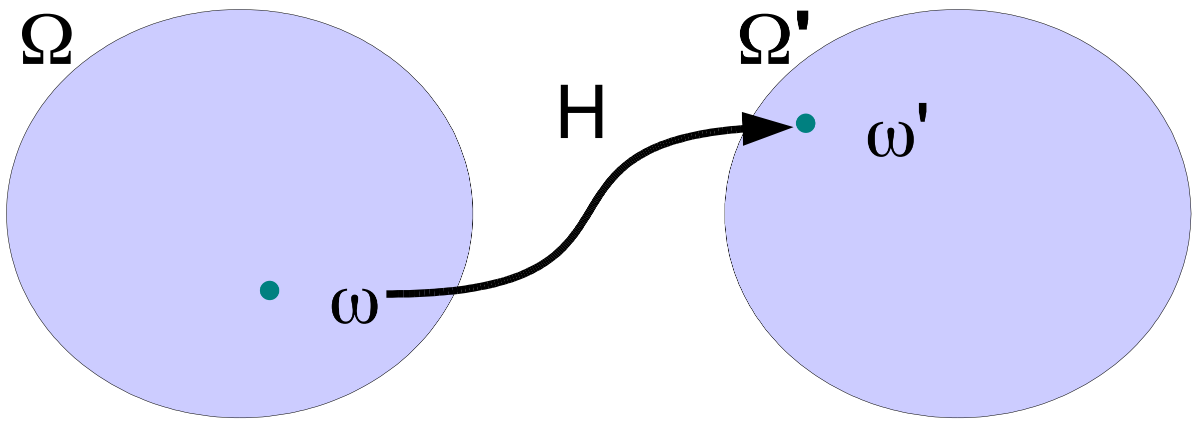
- Remarque:
  - ne pas confondre une variable aléatoire, notée X,
    avec la valeur prise par cette variable, notée x.

Variables aléatoires

- Expérience ω = tirer une carte...

Ω= {A♣, 2♣, ... ,R♣,A♦, 2♦, ... ,R♦,A♠, 2♠, ... ,R♠,A♥, 2♥, ... ,R♥}

- Exemple de variables aléatoires:
  - $H(\omega)$ = vrai si ω est un ♠ , faux sinon Ω'={vrai, faux}
  - $N(\omega) = n$ si ω est un n , 0 sinon $\Omega'=\{0, 1, ...10\}$
  - $F(\omega) = 1$ si ω est une figure , 0 sinon $\Omega'=\{0, 1\}$



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* $P(X \in]-\infty, x])$

- Exemple :
  - P(H = vrai) = ?

Variable aléatoire discrète

- Loi d'une variable aléatoire discrète X :

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

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

- Exemple :
  - P(N) = ?

Variable aléatoire continue

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

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

- Propriétés :
  - $f_X(x) \geqslant 0\ ,\ \ pour\ tout\ x$
  - $\int_{\mathbb{R}} f_X(x)\,dx = 1$
  - 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 :

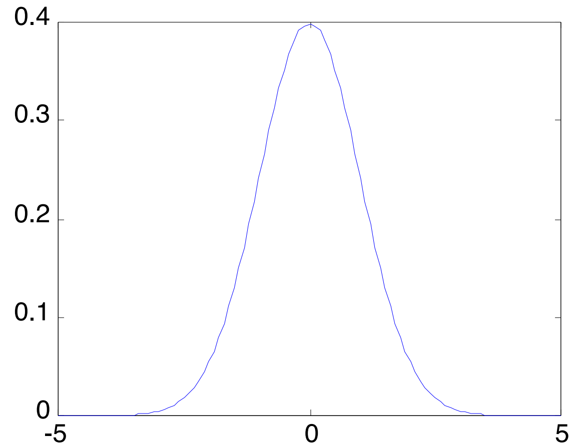
$$F : \mathbb{R} \rightarrow [0,1] \textit{ définie par } \forall x \in \mathbb{R},\ F_X(x) = P(X \leqslant x)$$

- Propriétés :
  - $\textit{si } X \textit{ est continue} : F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ $\boxed{F'(x)=f(x)}$
  - F est une fonction croissante, continue à gauche
  - $\lim_{x \to -\infty} F(x) = 0$ $\lim_{x \to +\infty} F(x) = 1$
  - $P(a \leqslant X < b) = F_X(b) - F_X(a)$

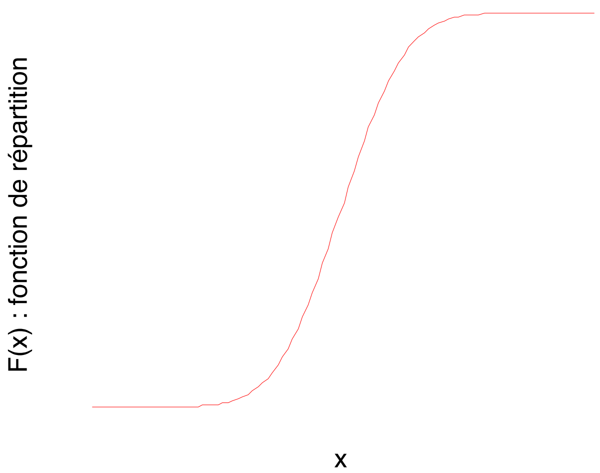
$\boxed{P(X>x)=1-F(x)}$

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

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

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

    Si φ bijective et strictement croissante

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

    Si φ bijective et strictement décroissante

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

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 :

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

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

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

E(X^2)=? E(XY)=?

Moments des variables aléatoires

- Remarques sur l'Espérance Mathématique :
  - L'espérance peut ne pas exister !!!
  - Valeur de l'espérance mathématique issue d'un calcul :
    - ponctuel pour f connue
    - théorique pour f inconnue !!!
- Propriétés :

$$
\boxed{\begin{aligned}
&E(a)=a\\
&E(aX)=aE(X)\\
&E(aX+bY)=aE(X)+bE(Y)\\
&E(XY)=E(X)E(Y)\, si\ X\ et\ Y\ v.a.\, indépendantes
\end{aligned}}
$$

Moments des variables aléatoires

- Variance d'une variable aléatoire : $var(X) = \sigma_X^2 = E([X - E(X)]^2)$

  où σ est l'écart type de la v. a. X $\quad var(X) = E(X^2) - (E(X))^2$

- Propriétés :

$$E((X-a)^2) = var(X) + (E(X)-a)^2 \, (Konig - Huyghens)$$

$$\boxed{\begin{array}{l} var(X+a) = var(X) \\ var(aX) = a^2 var(X) \\ var(X) = 0 \quad \Leftrightarrow X = a \, (presque\ sûrement) \end{array}}$$

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

$$\boxed{\begin{array}{l} var(X+Y) = var(X) + var(Y) + 2\,cov(X,Y) \\ cov(X,Y) = E(XY) - E(X)E(Y) = E((X-E(X))(Y-E(Y))) \\ var(X+Y) = var(X) + var(Y) \; si\ X\ et\ Y\ v.\,a.\ indépendantes \end{array}}$$

Moments des variables aléatoires

- Moment centré d'ordre k: $\boxed{\mu_k = E([X - E(X)]^k)}$
  - $\mu_1 = 0 \quad et \; \mu_2 = var(X)$
  - Pour une loi symétrique: $\mu_{2k+1} = 0 \; \forall k$
- Moment non-centré d'ordre k: $\boxed{m_k = E(X^k)}$
- 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 $\gamma_2 = \frac{\mu_4}{\sigma^4}$

Principales lois DISCRETES de probabilité

- Loi uniforme U(*n*)
  loi d'une v. a. *X* prenant les valeurs *1, 2,...,n* avec la même probabilité

$$\boxed{P(X=x)=\frac{1}{n}\ \forall\, x\in V=\{1,2,\ldots,n\}}$$

- Moments :

$$E(X)=\frac{n+1}{2} \qquad Var(X)=\frac{n^2-1}{12}$$

- Exemple :
  - Réalisation d'un nombre (entre 1 et 6) après avoir jeté un dé

Principales lois DISCRETES de probabilité

- Loi de Bernoulli $\mathcal{B}(1, p)$
  loi d'une v. a. X ne pouvant prendre que 2 valeurs *1* et *0* avec les probabilités p et $1-p$

$$\boxed{P(X=x)=p^x(1-p)^{1-x}}$$

- Moments :

$$E(X)=p \qquad Var(X)=p(1-p)$$

- Exemple :
  - Réalisation de pile (ou face) après avoir jeté une pièce

Principales lois DISCRETES de probabilité

- Loi binomiale $\mathcal{B}(n, p)$
  Répétition de l'expérience de Bernouilli *n* fois,
  X est la somme des résultats des expériences $X \in \{0 \ldots n\}$

$$\boxed{P(X=x)=C_n^x p^x (1-p)^{n-x}}$$

- Moments :

$$E(X)=np \qquad Var(X)=np(1-p)$$

- Propriétés :
  - Si *n* grand (*n*>50) alors $\begin{cases} \mathcal{B}(n,p) \to \mathcal{P}(\lambda=np) & \text{si } p \text{ petit } (p<0.1) \\ \mathcal{B}(n,p) \to \mathcal{N}(\mu,\sigma) & \text{sinon} \end{cases}$
  - $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$ v.a. indépendantes $\Rightarrow X_1+X_2 \sim \mathcal{B}(n_1+n_2, p)$
- Exemple : nombre de réalisations de pile après *n* essais

Principales lois DISCRETES de probabilité

- Loi de Poisson $P(\lambda)$

$$P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}\ \forall x\in\mathbb{N}$$

- Moments :

$$E(X)=\lambda \qquad Var(X)=\lambda$$

- Propriétés :
  - Si λ grand alors $P(\lambda) \to N(\mu, \sigma)$Texte
  - $X_1 \sim P(\lambda_1)$ et $X_2 \sim P(\lambda_2)$ v.a. indépendantes $\Rightarrow X_1+X_2 \sim P(\lambda_1+\lambda_2)$

- Exemples :
  - nombre de personnes à la queue du bus après un intervalle de temps
  - Nombre d'appels téléphoniques pendant un intervalle de temps

Distributions CONTINUES usuelles

- Loi uniforme $U_{[0, a]}$ sur [0,a] :

Densité :

$$\boxed{f(x)=\begin{cases}\frac{1}{a} \; sur\,[0,a] \\ 0 \;\; ailleurs\end{cases}}$$

Fonction de répartition :

$$F(x)=\begin{cases}\frac{x}{a} \; sur\,[0,a] \\ 0 \;\; sur\,(-\infty,0)\end{cases}$$

1 si x>a

- Moments :

$$E(X)=\frac{a}{2} \qquad Var(X)=\frac{a^2}{12}$$

- Remarque :
  - La somme de 2 lois uniformes n'est pas une loi uniforme !!!

Distributions CONTINUES usuelles

- Loi exponentielle E(λ) :

Densité :

$$\boxed{f(x)=\lambda e^{-\lambda x} \, si \, x>0}$$

Fonction de répartition :

$$F(x)=1-e^{-\lambda x}$$

- Moments :

$$E(X)=\frac{1}{\lambda} \qquad Var(X)=\frac{1}{\lambda^2}$$

- Exemples :
  - Temps d'attente à la queue du bus
  - Durée de vie d'un composant électrique

Distributions CONTINUES usuelles

- Loi normale $N(\mu, \sigma^2)$

  Densité :
$$\boxed{f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad \forall x\in\mathbb{R}}$$

- Moments : $E(X)=\mu$ $\qquad V(X)=\sigma^2$

- Propriété :
  - $X_1 \sim N(\mu_1, \sigma_1^2)$ et $X_2 \sim N(\mu_2, \sigma_2^2)$ v. a. Indépendantes

    $\Rightarrow X_1 \mp X_2 \sim N(\mu_1 \mp \mu_2, \sigma_1^2+\sigma_2^2)$

- Exemple :
  - Variation du diamètre d'une pièce ;
  - Répartition des erreurs de mesure autour de la « vraie valeur »