STA401

Statistique et Calcul des Probabilités

Responsable *: Carole Durand-Desprez*

14 Séances de Cours Amphi

13 Séances de TD

13 séances de TP

Modalités d'examen

CC1 : Tp (20%)

CC2 : Td (20%)

Examen : Ecrit final (60%)

STA401

CHAPITRE 1 : Calcul de Probabilité et Statistique (révision + nouveautés).

1. Statistiques descriptives sur un échantillon.

2. Rappels de Probabilité (généralités et formules – Proba conditionnelle - Bayes)

3. Lois de probabilité discrètes (Uniforme, Bernoulli, Binomiale, Hypergéométrique, Poisson).

4. Inégalités de Markov et de Bienaymé Tchebychev.

CHAPITRE 2 : Lois continues ("Autour" de la loi Normale)

1. Lois de probabilité continues (Généralités. Uniforme, Exponentielle ...).

2. Loi Normale (Gauss) - Proriétés - Lecture de table.

3. Théorème Central Limite. Théorème de Moivre-Laplace.

4. Intervalles de fluctuation.

5. Quelques autres lois usuelles continues (liens avec la loi norrmale).

STA401

CHAPITRE 3 : Statistique décisionnelle - Estimation

1. Estimation ponctuelle (estimateur, maximum de vraisemblance, lois des estimateurs)

2. Estimation par intervalles de confiance.

CHAPITRE 4 : Tests Statistiques

1. Généralités sur les tests. Notion de risques et d'hypothèses.

2. Tests de conformité (paramétriques) : moyenne, variance, probabilité.

3. La Pvaleur d'un test.

4. La puissance d'un test.

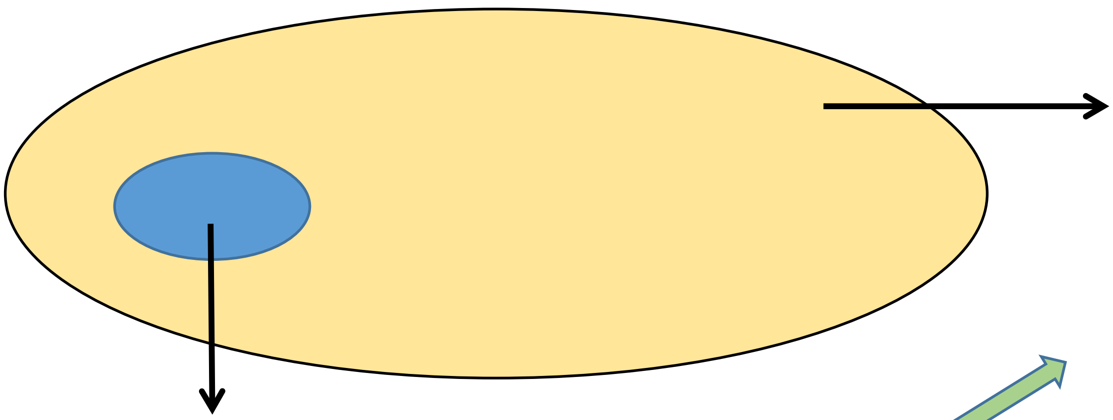
5. Un test paramétrique particulier : le test de données appariées.

6. Tests de comparaison de 2 éch. indépendants (moyennes, variances, probabilités)

7. Tests du Khi-deux (ajustement et indépendance).

CHAPITRE 1 : Remise à niveau – Révision

A quoi servent les statistiques ? Pour quoi faire ?



Population totale (N)

- Moyenne μ
- Variance σ^2
- Probabilité **P**

Echantillon (n)

- Moyenne empirique $\bar{X}$
- Variance empirique **S²**
- Fréquence empirique **F**

???

X : caractère
variable mesurée sur individus
→ Variabilité, incertitude

Données	→	Statistiques descriptives	→	Echantillon
Modèle	→	Probabilités	→	Population

1. Statistiques descriptives. Données sur un échantillon :

X *variable numérique (quantitative). Données : valeurs sur un échantillon :* $x_1, x_2, \dots, x_n$

$$Moyenne\ empirique : \bar{x} = \frac{1}{n}\sum x_i$$

$$Variance empirique : s^2 = \frac{1}{n}\sum (x_i)^2 - (\bar{x})^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$$

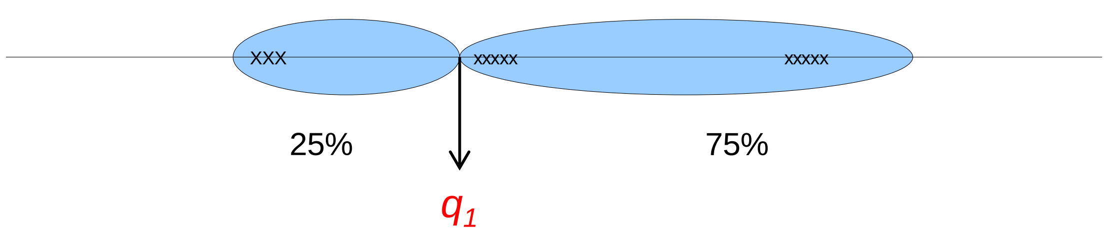
$$Ecart\ type\ empirique : s$$

$$Fréquence\ empirique : f = \frac{k}{n} \quad (k\ nombre\ de\ réalisations\ d'un\ évènement\ A)$$

<u>*La médiane*</u> m_d est la plus petite valeur prise par X telle qu'au moins la moitié des effectifs soit inférieur.

<u>*Le premier quartile*</u> q_1 est la plus petite valeur prise par X telle qu'au moins le quart des effectifs soit inférieur.

<u>*Le troisième quartile*</u> q_3 est la plus petite valeur prise par X telle qu'au moins les ¾ des effectifs soit inférieur.



Un échantillon est centré si sa moyenne est 0. Il est réduit si sa variance est 1

X variable → Qualitative (nominale) : ex. 'nom'
X variable → Quantitative (numérique) : ex. 'taille' → Discrète (ex: dé)
Quantitative (numérique) : ex. 'taille' → Continue (ex: poids)

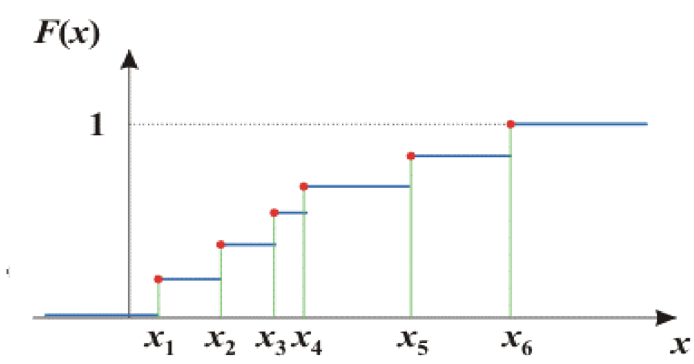
Cas de données regroupées en classes : $C_1, \ldots, C_k$ de centres $c_1, \ldots, c_k$
Soient $n_1, \ldots, n_k$ les effectifs de chaque classe, et n la taille de l'échantillon

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{n}\sum_{i=1}^{k} n_i\, c_i \qquad s^2 = \frac{1}{n}\sum (x_i - \bar{x})^{\,2} = \frac{1}{n}\sum (n_i c_i - \bar{x})^{\,2}$$

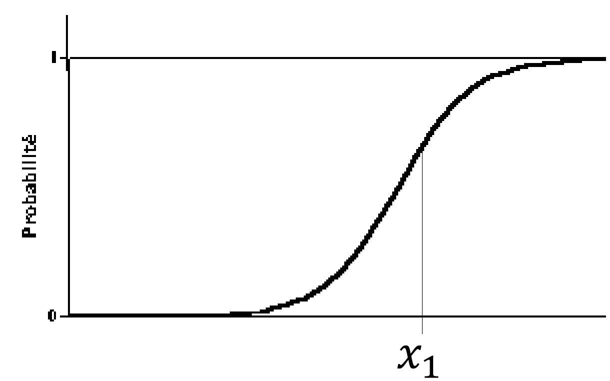
Graphiques : → Cas discret : diagramme en bâton
→ Cas continu : histogramme

Fonction de répartition : $F(x) = P(X \leq x)$

discret →



continu →



2. Probabilités (généralités et formules - révisions)

Soit Ω l'ensemble de toutes les éventualités d'une expérience. Soit E l'ensemble des évènements possibles. A un évènement et A son évènement complémentaire.

→ Une probabilité P est une application de E dans [0;1].

$$P(\Omega) = 1, \qquad P(\emptyset) = 0, \qquad 0 \leq P(A) \leq 1,$$

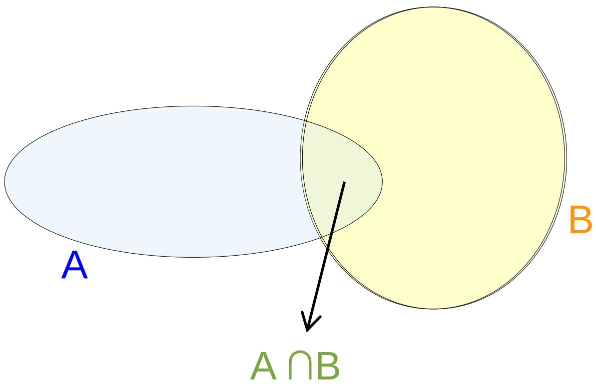
$$P(\bar{A}) = 1 - P(A)$$

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

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A) = P(A \mid B)P(B) + P(A \mid \bar{B})P(\bar{B})$$



A et B sont *disjoints (incompatibles)* si :

$$P(A \cap B) = 0 \quad (ou\ bien :\ A \cap B = \emptyset\ )$$

A et B sont *indépendants* si :

$$P(A \cap B) = P(A) * P(B) \quad ou\ \ P(A \mid B) = P(A)$$

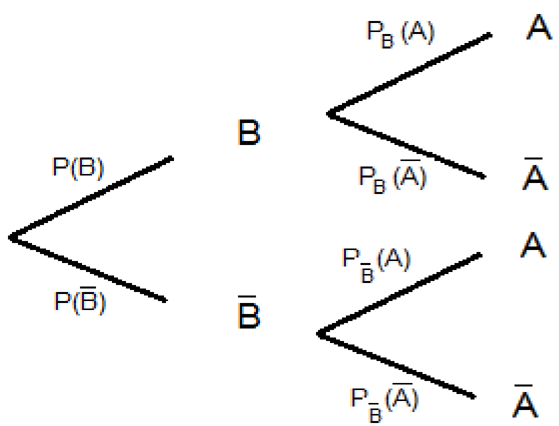
Autre notation : $\boxed{P(A \mid B) = P_B(A)}$

Formule de Bayes :

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})}$$

$Cas\ général\ (Si\ \ A = A_1 \cup \ldots \cup A_n, et\ disjoints)$:

$$\boxed{P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{\sum_{i=1}^{n} P\,(B \mid A_i)P(A_i)}}$$



Var1 \ Var2	A	$\bar{A}$	Total
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
$\bar{B}$	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
Total	$P(A)$	$P(\bar{A})$	1

Arbre pondéré

Tableau croisé

Sur le graphique …

$$P(A \cap B) = P(B) * P(A \mid B)$$

Dans le tableau …

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

3. Lois de probabilités usuelles discrètes :

→ Une variable variable X est une application sur Ω.

Exemple : On lance un dé (expérience). $\Omega = \{f_1 , \ldots, f_6 \}$.
Evènement A : “tomber sur un nombre pair“ ; B : “tomber sur un nombre inférieur à 5“
X la variable : $X(f_i) =$ 'gain de i euros' … Les valeurs possibles de X : {1,2,3,4,5,6}
Y variable : $Y(f_i) =$ 'gain de i € si i pair, sinon 0' … Les valeurs possibles de Y : {0,2,4,6}

→ Loi de probabilité P d'une variable X discrète (finie), valeurs possibles $\{x_1, \ldots, x_n\}$:

$$P(X = x_i) = p_i \quad \forall i, \qquad \sum_{i=1}^{n} p_i = 1$$

$$E[X] = \mu = \sum_{i=1}^{n} x_i\, p_i$$

→ Espérance

$$V[X] = \sigma^2 = E\left[X - E[X]\right]^2$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 p_i = \sum_{i=1}^{n} (x_i)^2 p_i - \mu^2$$

→ Variance

→ *Propriété* : $E[aX + bY] = aE[X] + bE[Y]$ Linéarité de l'espérance évidente.

→ *Propriétés* :
$$V[X] = E[X - E[X]]^2 = E[X^2] - (E[X])^2$$
$$V[aX + b] = a^2 V[X]$$

Démo : Dans le cas discret, $V(X) = E[X-E(X)]^2 = E[X-\mu]^2 = \Sigma (x_i - \mu)^2 p_i = \Sigma (x_i^2 - 2\mu x_i + \mu^2) p_i$
donc, $V(X) = \Sigma (x_i^2) p_i - 2\mu \ \Sigma x_i p_i + \mu^2 \Sigma p_i = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - [E(X)]^2$
$V(aX+b) = E[(aX+b)^2] - [E(aX+b)]^2 = E(a^2X^2+2abX+b^2) - [aE(X)+b]^2 = a^2 V(X)$

→ **Loi uniforme (équiprobable) :** $\forall i, \quad p_i = 1/n$

Exemple : Un dé non pipé, $\Omega = \{1, \ldots, 6\}$, variable $X(i) = i$, $P(X= i) = p_i = 1/6$
$E[X] = 1*1/6 + 2*1/6 + \ldots + 6*1/6 = 3,5$ $V[X]= (1^2+\ldots+6^2)/6 - 3,5^2 = 2,916\ldots$

→ **Loi Bernoulli (p) :** *A un évènement qui se réalise avec une probabilité p.*

$$X = \begin{cases} 1 \ \ si \ \ A \ \ se\, réalise \ \ avec \ \ proba \ (p) \\ 0 \ \ si \ \ A \ \ non\, réalisé \ \ proba \ (1-p) \end{cases}$$

X suit la loi Bernoulli (p)

⇒ $P(X = 1) = p \ \ et \ \ P(X = 0) = 1 - p \ ; \ E[X] = p \ ; \ V[X] = p(1-p)$

→ Loi Binomiale (n,p) : *A est un évènement qui se réalise avec une probabilité p.*

On refait n expériences indépendantes. X_i variables de Bernoulli (p) indépendantes

$$X = \sum_i X_i$$ est le nombre de réalisations de A parmi les n possibles.

X suit la loi *Binomiale (n,p)* *Valeurs possibles de X : {0,1, …, n}*

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \;\; ; \;\; E[X] = np \;\; ; \;\; V[X] = np(1-p)$$

Rappel : Combinatoire

On prend *simultanément* k objets parmi un ensemble en contenant n ($n \geq k$).

Le nombre de combinaisons possibles est : $\binom{n}{k} = \dfrac{n!}{(n-k)!\,k!}$ $autre\ \ notation : C_n^k$

Exemple : A : "être un étudiant de la filière INF dans l'amphi", probabilité p=90%.
Dans un amphi N=60, on prend n=10 individus (tirages indépendants, avec remise).
X : nombre d'étudiants du groupe INF parmi les n=10

Indépendance (avec remise) ⇒ *X suit B(10 ; 0,9)*

$P(X=0)=(0{,}1)^{10} = 10^{-10}$ ... voir calculatrice aussi ...
E(X) = 10*0,9 = 9
V(X) = 10*0,9*0,1 = 0,9

→ Loi Hypergéométrique (N,m,n) :

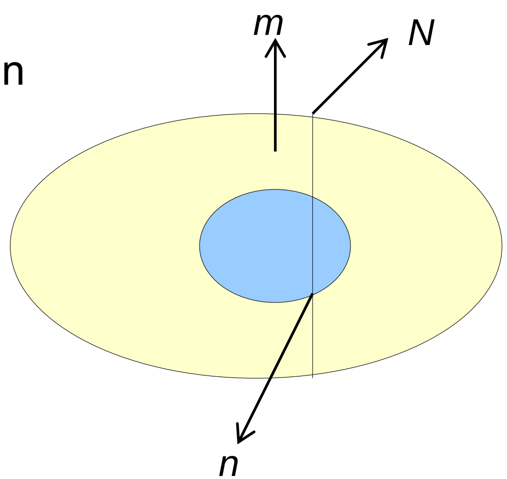
Soit une population de N individus dont m individus réalisent un évènement A.
On prend un échantillon de n individus (sans remise).

X est le nombre d'individus qui réalisent A parmi les n possibles.

X suit la loi Hypergéométrique *(N,m,n)* *Valeurs possibles de X : {0,1, …, min(n;m)}*

$$P(X=k)=\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \quad ; \quad E[X]=\frac{nm}{N} \quad ; \quad V[X]=n\left(\frac{m}{N}\right)\left(\frac{N-m}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Exemple : Population : “amphi de sta401“ avec N=60 étudiants, A : “être un étudiant de la filière INF“, m=50 étudiants sont des INF.
On prend un échantillon de n=8 individus (tirages sans remise).
X : nombre d'étudiants du groupe INF parmi les n=8



Indépendance (sans remise) ⇨ *X suit H(60;50;8)*

P(X=0) = 1*45 / 2558620845 = $1{,}7588.10^{-8}$ ... voir calculatrice aussi ...
E(X) = 8*50/60 = 6,66666...

→ Loi de Poisson (λ) :

Soit un évènement A qui arrive 'rarement' dans le temps.
On observe pendant un certain laps de temps cet évènement.
X est le nombre d'évènements qui se réalisent.
X suit la loi de Poisson (λ) *Valeurs possibles de X : {0,1, … }*

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad ; \quad E(X)=\lambda \quad ; \quad V(X)=\lambda$$

Exemple : En sta401, on a constaté qu'en moyenne 2 étudiants posaient une question pendant l'examen.
X : nombre d'étudiants qui posent des questions cette année ⇒ *X suit P(2)*
$P(X=4) = 2^4 e^{-2} / 4! = 0{,}0902...$

Propriété 1 :

Soient X suit $P(\lambda_1)$ et Y suit $P(\lambda_2)$ indépendantes, alors : $X+Y$ *suit* $P(\lambda_1+\lambda_2)$

Propriété 2 : Approximation d'une loi Binomiale par une loi de Poisson

Lorsque n est grand (n>50) et p est petit (p<0,1) alors : $B(n;p) \approx P(\lambda) \quad avec \quad \lambda = np$

Dém : Voir les 2 démonstrations en TD à titre d'exercices.

4. Inégalités à connaître (également applicables au cas continu du chapitre suivant) :

Inégalité de Markov

→ Soit X v.a. positive, pour tout a > 0 : $\boxed{P(X \geq a) \leq \frac{E(X)}{a}}$

Dém : $Soient\ (x_i)_i\ \ les\ valeurs\ possibles\ prises\ par\ X, alors:$

$$E(X) = \sum_{x_i} P(X = x_i)x_i = \sum_{x_i<a} P(X = x_i)x_i + \sum_{x_i \geq a} P(X = x_i)x_i \geq 0 + \sum_{x_i \geq a} P(X = x_i)x_i$$

$$\geq a \sum_{x_i \geq a} P(X = x_i) \geq aP(X \geq a).\quad Donc, \quad P(X \geq a) \leq \frac{E(X)}{a}$$

Inégalité de Bienaymé Tchebychev

→ Soit X v.a. positive, pour tout a > 0 : $\boxed{P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}}$

Dém : $Dans\ \ l'inégalité\ de\ Markov\ \ on\ remplace\ a\ par\ a^2\ \ et\ \ X\ par\ \left(X - E(X)\right)^2:$

$$Donc, \quad P\left(\left(X - E(X)\right)^2 \geq a^2\right) \leq \frac{E\left(X - E(X)\right)^2}{a^2}.\quad On\ déduit\ \ aisément:$$

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

Application : Trouver le plus petit h tel que (interv. centré sur moyenne): $P[E(X) - h \leq X \leq E(X) + h] \geq 0{,}95$

$$\longrightarrow \quad P[|X - E(X)| \leq h] \geq 0{,}95 \quad \Rightarrow \quad P[|X - E(X)| \geq h] \leq 0{,}05 \quad IBT \ :\ h = \sqrt{V(X)/0{,}05}$$