Probabilités et Statistiques - M115

Michel Fournié

fournie@mip.ups-tlse.fr

Définitions ("intuitives")

- Population
  ensemble de référence (unités observées)
- Echantillon sous ensemble d'une population
- Individu ou unité statistique
  tout élément de la population
- Effectif total nombre d'individu observé, noté n
- Caractère : particularité de l'individu qui peut être
  qualitatif : observation non mesurée
  ordinal : observations ordonables sinon nominal
  quantitatif : mesurable
  discret : valeurs observées isolées
  continu s'il peut prendre toute valeur d'un intervalle réel
  (regroupement des valeurs discrètes dans des classes)
- serie[suite] statistique
  valeurs d'un seul caractère sur une population donnée

Représentation (collecte des données)

- Tableau des données ponctuelles

  1ère colonne : individus observés

  2ième colonne : valeurs correspondantes du caractère

Exemple : Mesure du diamètre de la tête de 4 boulons

Code du boulon	Diamètre (mm)
$B01$	$25,5$
$B02$	$25,4$
$B03$	$25,6$
$B04$	$25,5$

- Tableau de distribution des observations

  1ère colonne : valeurs distinctes du caractère

  2ième colonne : liste des individus correspondante

Exemple :

Code du boulon	Diamètre (mm)
$25,4$	$\{B02\}$
$25,5$	$\{B01, B03\}$
$25,6$	$\{B04\}$

Tableau de distribution d'effectifs/fréquences

- Tableau de distribution d'effectifs (et/ou de fréquences)

fréquence $f_i = \dfrac{\text{effectif } n_i}{\text{effectif total } n}$

féquence cumulée croissante $F_i = \sum_{j=1}^{i} f_j = f_1 + \cdots + f_i$

Valeurs observées	Effectifs	Fréquences	Fréq.c.c.
x_1	n_1	f_1	F_1
.	.	.	.
x_p	n_p	f_p	F_p

Exemple :

Diamètre (mm)	Effectifs	Fréq.	Féq.c.c
25, 4	1	0.25	0.25
25, 5	2	0.5	0.75
25, 6	1	0.25	1
Total	4	1	

Graphiques

- Diagramme en bâtons
- Histogramme
- Camembert

Caractéristiques de position (Moyenne)

- Objectif: Résumer les valeurs observées par 1 ou 2 val.
- Moyenne arithmétique $\bar{x}$:

Partant d'un tableau de données ponctuelles (i, x_i)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \cdots + x_n)$$

Partant d'un tableau d'effectifs (ou fréquences)
$(i, x_i, n_i$ ou $f_i)$ (p: nbre de valeurs distinctes, n: effectif total)

$$\bar{x} = \frac{\sum_{i=1}^{p} n_i x_i}{\sum_{i=1}^{p} n_i} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$$

Attention : Non robuste $\Longrightarrow$ Hypothèse de non existance de valeurs observées anormalement petites ou grandes

Caractéristiques de position (Médiane)

• **Médiane** Me :

Valeur partageant la série statistique en deux séries de même taille

Partant d'un tableau de données ponctuelles (i, x_i) ordonné

$$\begin{cases} \text{si } n \text{ impair } Me = x_{\frac{n+1}{2}} \\ \text{si } n \text{ pair } Me = \text{toute valeur de l'intervalle médian } \left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right] \\ \qquad \text{On pose parfois } Me = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) \end{cases}$$

Partant d'un tableau d'effectifs (ou fréquences)

On calcule les effectifs cumulés croissant $F(x_i)$ et décroissant $G(x_i)$

On résoud le système d'inéquations $F(x) \geq \frac{n}{2}$ et $G(x) \geq \frac{n}{2}$

$$\begin{cases} \text{S'il existe une unique valeur observée } x_i \text{ solution} : Me = x_i \\ \text{S'il existe 2 valeurs solutions } x_i \text{ et } x_{i+1} : Me = \frac{1}{2}(x_i + x_{i+1}) \end{cases}$$

Exemple : Médiane

Etudiant	Note
1	8
2	9
3	11
4	20

$$Me = 10, \quad (\bar{x} = 12)$$

Caractéristiques de dispersion

Exemple :

Notes de Pierre	7	8	11	12	13	13	13
Notes de Paul	4	7	9	12	13	13	19

Même médiane $Me = 12$, même moyenne 11
Or les notes de Paul sont plus dispersées ...

- Variance

$$V = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

$$V_{Pierre} = \frac{38}{7} \approx 5.42 \text{ et } V_{Paul} = \frac{142}{7} \approx 20.2$$

- Ecart type

$$\sigma = \sqrt{V}$$

$$\sigma_{Pierre} \approx 2.32 \text{ et } \sigma_{Paul} \approx 4.50$$

Remarque: Calculer $(x_1 - \bar{x}) + \cdots + (x_n - \bar{x})$

Caractéristiques de dispersion (suite)

- Partant d'un tableau de données ponctuelles (i, x_i) ordonné

$$V = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{x_1^2 + \cdots + x_n^2}{n} - \bar{x}^2$$

- Partant d'un tableau d'effectifs (ou fréquences)

$$V = \frac{n_1(x_1 - \bar{x})^2 + \cdots + n_p(x_p - \bar{x})^2}{n} = \frac{n_1 x_1 + \cdots + n_p x_p}{n} - \bar{x}^2$$

Series statistiques à 2 variables (Non développé)

- Tableau de données, Nuage de points

- Ajustement affine
    - Méthode des moindres carrés
    - Droites de régression
    - Covariance

- Coefficient de corélation linéaire

Probabilités - Vocabulaire

• Dans une expérience aléatoire, l'ensemble des résultats possibles est l'Univers Ω

• Evénement : partie de l'Univers Ω

• Evénement élémentaire : événement à 1 élément

• Les événements A et B sont disjoints ou incompatibles :

$$A \cap B = 0$$

• Evénement contraire de A : ensemble des éléments de Ω n'appartenant pas à A noté $\bar{A}$

Exemple : Cas d'un lancé de dé.

$$A = \{5, 6\} \text{ événement de } \Omega = \{1, 2, 3, 4, 5, 6\}$$

Probabilité

Définition :

Soit Ω un univers fini. Une probabilité sur Ω est une application P de l'ensemble des événements de Ω dans l'intervalle $[0,1]$ telle que :

- **Axiome 1**

  $P(\Omega) = 1$

- **Axiome 2**

  Pour tous événements A et B, si $A \cap B = 0$ alors

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

Propriétés :

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Equiprobabilité ("au hasard")

Définition :
On parle d'équiprobabilité quand tous les événements élémentaires ont la même probabilité. Dans ce cas

$$\boxed{P(A) = \frac{\text{nbre d'éléments de } A}{\text{nbre d'éléments de } \Omega} = \frac{\text{nbre de cas favorables}}{\text{nbre de cas possibles}}}$$

Exemple : Dans un lancé de dé.
Pour $A = \{2, 4, 6\}$ alors $P(A) = \frac{1}{2}$

Probabilité conditionnelle

M115

Michel Fournié

Définition :
Soit P une probabilité sur Ω et A un événement de probabilité non nulle.
La probabilité sachant que A (est réalisé) est l'application P_A qui à tout événement B associe le nombre

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

Notation : $P_A(B) = P(B|A)$

Exemple : Dans un lancé de dé.
Pour $A = \{2, 4, 6\}$ et $B = \{4, 5, 6\}$ alors
$P(A) = \frac{1}{2}$ et $P(A \cap B) = \frac{1}{3}$ d'où $P_A(B) = \frac{2}{3}$

Evénements indépendants

Définition :
Les événements A et B sont indépendants si et seulement si $P(A \cap B) = P(A)P(B)$

Rappel: $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Exemple : Tirage d'une carte dans un jeu de 32 cartes.

- A événement de "Tirer un As" : $P(A) = \frac{4}{32} = \frac{1}{8}$
- B événement de "Tirer un Coeur" : $P(B) = \frac{8}{32} = \frac{1}{4}$
- $A \cap B$ évt de "Tirer un As de Coeur": $P(A \cap B) = \frac{1}{32}$

$$P(A)P(B) = P(A \cap B) \Longrightarrow \text{événements indépendants}$$

- C événement "Tirer un As Rouge" alors $P(C) = \frac{2}{32} = \frac{1}{16}$
- $B \cap C$ événement de "Tirer l'As de Coeur": $P(B \cap C) = \frac{1}{32}$

$$P(B)P(C) \neq P(B \cap B) \Longrightarrow \text{événements dépendants}$$

Application d'un ensemble fini dans un ensemble fini

Définition :
Soit E_n un ensemble à n éléments et F_p un ensemble à p éléments.

$$\boxed{\text{Le nombre d'applications de } E_n \text{ dans } F_p \text{ est } p^n}$$

Rappel: Une application associe à tout élément de l'ensemble de départ un et un seul élément dans l'ensemble d'arrivé

Remarque: Notion d'arbre

Exemple : On a 3 usines A, B, C pouvant accepter 5 personnes. Comment répartir 4 personnes X, Y, Z, T dans ces 3 usines ? $\longleftrightarrow$ application

On doit associer à chaque personne une usine et une seule

$E_4 = X, Y, Z, T, F_3 = A, B, C \Longrightarrow 3^4 = 81$ répartitions

Bijections - Permutations

Définition :

- Soit E_n et F_n des ensembles à n éléments.

$$\boxed{\text{Le nombre de bijections de } E_n \text{ sur } F_n \text{ est } n!}$$

- Une bijection de E_n sur E_n est appelée permutation de E_n.

Ainsi $\boxed{\text{le nombre de permutations de } n \text{ éléments est } n!}$

Rappel: Une bijection est une application tq tout élt de l'ens. d'arrivé est l'image d'un et d'un seul élt de l'ens. de départ

Factoriel n = $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$, $0! = 1$

Exemple : Comment répartir 4 lots entre 4 personnes ?

E_4 ens. des 4 lots et $F_4 = A, B, C, D$ ens. des 4 personnes

On doit associer à chaque lot une personne et une seule

$\longleftrightarrow$ application, mais en plus la personne ne doit avoir qu'un lot $\longleftrightarrow$ bijection d'où $4! = 24$ répartitions possibles

Arrangements

Définition : Soit F_n un ensemble à n éléments. Un arrangement (sans répétition) d'ordre p ($p \leq n$) est une liste ordonnée de p éléménts de F_n.
Le nombre d'arrangements d'ordre p de F_n est

$$\boxed{A_n^p = \frac{n!}{(n-p)!}}$$

Remarques :

- Lire A n p
- Il y a une notion d'ordre

Exemple : Combien existe-t-il de mots de passe à 4 lettres disctinctes minuscules ?
On doit choisir $p = 4$ lettres parmi $n = 26$
Notion d'**ordre** et **sans remise** d'où
$A_n^p = A_{26}^4 = 26 \times 25 \times 24 \times 23 = 358800$

Combinaisons

Définition : Soit F_n un ensemble à n éléments. Une combinaison (sans répétition) d'ordre p ($p \leq n$) est une partie de F_n à p éléments.
Le nombre de combinaisons d'ordre p de F_n est

$$\boxed{C_n^p = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}}$$

Remarques :

- Lire C n p
- Il n'y a pas de notion d'ordre

Exemple : Combien de binômes peut on former avec un groupe de 30 élèves ?
Il faut prendre $p = 2$ élèves parmi $n = 30$.
Pas de notion **d'ordre** ni de **remise** d'où $C_{30}^2 = 435$

Variables aléatoires - Présentation

- Soit une urne avec trois boules : Rouge, Verte, Bleue
- On tire au hasard 1 boule, on la repose et on re-tirer
- On peut définir Ω = ensemble de tous les couples tirés
- A chaque boule est associé un gain (+) ou une perte (-)

Boule	R	V	B	
Gain	$+6$	$+1$	-4	euros

- Soit X l'application de Ω dans $\boldsymbol{R}$ qui :

  à tout tirage de 2 boules ω associe un gain ou une perte

$$X : \begin{array}{ccc} \omega & \longrightarrow & X(\omega), \quad \text{gain obtenu avec } \omega \\ \Omega & \longrightarrow & \boldsymbol{R} \end{array}$$

$\boxed{X \text{ est une variable aléatoire (V.A.) à valeurs réelles}}$

On note $X(\Omega)$ l'ensemble des valeurs de X

Ici $X(\Omega)$ est l'ens. des gains possibles $\{-8, -3, 2, 7, 12\}$

Remarque : Le jeu a disparu !!!

Probabilité iamge définie par une varaible aléatoire (V.A.)

Première approche :

On peut se demander quelle est la probabilité de tirer une certaine boule ?

$\Longrightarrow$ définition d'une proba. P sur Ω

Deuxième approche :

On peut se demander quelle est la probabilité de gagner 12 euros ?

$\Longrightarrow$ définition d'une proba. P' sur $X(\Omega)$

Notation :

Pour tout nombre k de $X(\Omega)$ on note

$$\boxed{P(X = k) = P'(\{k\}) = P(\{\omega \in \Omega; X(\omega) = k\})}$$

Probabilité ou distribution de la variable aléatoire (V.A.)

Définition :

La loi de probabilité ou distribution de la variable aléatoire X est la fonction

$$\boxed{\begin{array}{ccc} k & \longrightarrow & P(X=k) \\ X(\Omega) & \longrightarrow & [0,1] \end{array}}$$

Exemple :

Tableau de valeurs

k	-8	-3	2	7	12
$P(X=k)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

Représentation (Diagramme en bâtons)

Fonction de répartition

Définition :

Pour tout $x \in \mathbb{R}$, on note

$$\boxed{P(X \leq x) = P(\{\omega \in \Omega ; X(\omega) \leq x\})}$$

Exemples :

- $P(X \leq -10) = 0$
- $P(X \leq +15) = 1$
- $P(X \leq 0) = P(X = -8) + P(X = -3) = \dfrac{1}{9} + \dfrac{2}{9} = \dfrac{1}{3}$

($X = -8$ et $X = -3$ incompatibles)

Définition :

La fonction de répartition de la variable aléatoire X est la fonction F

$$\boxed{\begin{array}{rcl} F : x & \longrightarrow & F(x) = P(X \leq x) \\ \mathbb{R} & \longrightarrow & [0, 1] \end{array}}$$

Fonction de répartition (suite)

Représentation graphique :

Remarque : La donnée de F suffit pour définir X

Exemple :

Partant de la fonction de répartition F retrouver $P(x = 7)$

On remarque que 7 est l'unique entier $\in [6.5, 7]$.

Or $F(7) = \dfrac{8}{9}$ et $F(6.5) = \dfrac{2}{3}$

D'où $P(X = 7) = F(7) - F(6.5) = \dfrac{2}{9}$

Variable aléatoire continue

• Au lieu de considèrer des valeurs isolées on va considèrer des variables aléatoires pouvant prendre n'importe quelle valeur de $\boldsymbol{R}$ ou d'un intervalle

Exemple : Variable aléatoire mesurant la durée de bon fonctionnement d'une machine (intervalle $[0, \infty[$)

Définition :

• Une variable aléatoire est continue si sa fonction de répartition F est continue, ce qui revient à dire qu'il existe une fonction f telle que

$$\text{Pour tout } x \in \boldsymbol{R},\ F(x) = \int_{-\infty}^{x} f(t)dt$$

• f est appelée densité de probabilité de la variable aléatoire X avec

$$f(x) \geq 0, \quad F'(x) = f(t)$$

Variable aléatoire continue (suite)

Théorème:

$$\left\{\begin{array}{l} f(x) \geq 0 \\ f \text{ continue par morceaux} \\ \int_{-\infty}^{+\infty} f(t)dt = 1 \end{array}\right. \Longrightarrow f \text{ est une densité de proba.}$$

Théorème : Pour tous a et b tels que $0 \leq a \leq b$, on a

$$\boxed{P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(t)dt} \longleftrightarrow \text{Aire}$$

Remarque : Probabilité par rapport à un point $P(X = a) = 0$

Exemple: La fonction $f(x) = \left\{\begin{array}{l} 0 \ \text{ pour } \ x \leq 0 \\ \frac{1}{2\sqrt{x}} \ \text{ pour } \ 0 < x \leq 1 \\ 0 \ \text{ pour } \ x > 1 \end{array}\right.$

La fonction f est-elle une densité de probabilité ?

Si oui quelle est sa fonction de répartition F ?

Espérance mathématique

Objectif : synthétiser l'information issue de la fct de répartition

Définition : L'espérance mathématique d'une variable aléatoire discrète prenant n valeurs x_i avec les probabilités $P(X = x_i) = p_i$, où $1 \leq i \leq n$ est

$$\boxed{E(X) = \sum_{i=1}^{n} p_i x_i}$$

Définition : L'espérance mathématique d'une variable aléatoire continue est

$$\boxed{E(X) = \int_{-\infty}^{+\infty} t f(t) dt}$$

Remarque :

Les propriétés de l'espérance sont comparables à celles de la moyenne

Exemple

Exemple :
Voici une distribution pour X déja étudiée
Gain lors du tirage des boules

k	-8	-3	2	7	12
$P(X=k)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

$E(X) = 2$

Interprétation du résultat : c'est le gain moyen qu'un joueur obtiendrait s'il jouait un très grand nombre de fois

Remarque : Voir la notion de moyenne arithmétique pour les séries statistiques

Variance - Ecart type

Définition :

La variance $V(X)$ d'une variable aléatoire X est, si elle existe

$$\boxed{V(X) = E((X - E(X))^2)}$$

Définition :

L'écart type de X est $\boxed{\sigma(X) = \sqrt{V(X)}}$

Remarques :

- $X \longrightarrow x_i$ alors $(X - E(X))^2 \longrightarrow (x_i - E(X))^2$
- Ces notions sont à rapprocher de celles vues pour les series statistiques (caractérisation de la dispersion par rapport à une tendance centrale)

Propriétés

Soit X une variable aléatoire

- $E(aX+b) = aE(X)+b$ pour a et b constantes réelles
- $(X-E(X))^2 = X^2 - (2E(X))X + (E(X))^2$
  $\Longrightarrow V(X) = E(X^2 - (2E(X))X + (E(X))^2)$
- $V(X) = E(X^2) - 2E(X)E(X) + (E(X))^2$
  $\Longrightarrow \boxed{V(X) = E(X^2) - (E(X))^2}$
- $V(aX+b) = a^2 V(X)$

Remarque : $E(X^2) = \sum_{i=1}^{n} p_i x_i^2$

Exemple : Gain lors du tirage de 2 boules

$$\begin{cases} V(X) = \frac{100}{3} \\ \sigma(X) = \frac{10\sqrt{3}}{3} \approx 5.77 \end{cases}$$

Propriétés (suite)

Définitions :

- On dira qu'une variable aléatoire X est centrée si

$$\boxed{E(X) = 0}$$

- On dira qu'une variable aléatoire X est réduite si

$$\boxed{V(X) = 1}$$

Propriétés :

- La variable aléatoire $X - E(X)$ est centrée
- La variable aléatoire $Y = \frac{X - E(X)}{\sigma}$ est centrée réduite

Loi binomiale

Quand utiliser cette loi : (V.A. discrète)

- Dans une "épreuve aléatoire" élémentaire pouvant déboucher sur 2 résultats et 2 seulement ex. "succés" et "échec" de proba. respestives p et $q = 1 - p$
- On réalise n fois l'épreuve et on note X la variable aléatoire mesurant le nombre de "succés"
- Si les n épreuves aléatoires élémentaires sont indépendantes alors X suit la loi binomiale $B(n, p)$
- Tirage "avec remise"

Définition :

Une variable aléatoire X suit une loi binomiale $B(n, p)$ de paramètres $n \in \mathbb{N}$ et $p \in [0, 1] \subset \mathbb{R}$ lorsque sa loi de proba. est définie par : Pour tout $k \in \mathbb{N}$ tel que $0 \leq k \leq n$

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

Exemple - Loi binomiale

Exemple :

- Dans un fichier de fournisseurs (1 fiche par fournisseur), $p = \frac{1}{3}$ d'entre eux sont sur Toulouse les autres sont sur Paris (d'où $q = 1 - p = \frac{2}{3}$)
- On tire 5 fiches avec remise (tirages indépendants)
- Soit X la variable aléatoire qui à de tels tirages associe le nombre de fournisseurs sur Toulouse

- L'étude de la loi de probabilité de la variable aléatoire X qui suit donc la loi binomiale $B(5, \frac{1}{3})$ est donnée par

k	0	1	2	3	4	5
$P(X=k)$	$C_5^0 p^0 q^5$ 0.123	$C_5^1 p^1 q^4$ 0.329	$C_5^2 p^2 q^3$ 0.329	$C_5^3 p^3 q^2$ 0.165	$C_5^4 p^4 q$ 0.041	$C_5^5 p^5 q^0$ 0.004

Propriétés de la loi binomiale

Propriétés :

Soit X une variable aléatoire suivant la loi binomiale $B(n, p)$ alors

$$\boxed{E(X) = np, \;\; V(X) = np(1-p), \;\; \sigma(X) = \sqrt{np(1-p)}}$$

Démonstration :

A faire en exercice

Loi de Poisson

Quand utiliser cette loi : (V.A. discrète)

- Dans la modélisation de phénomènes aléatoires où le futur est indépendant du passé
- Lors du tirage d'un nombre entier "le plus au hasard possible"
- Situations : pannes de machines, sinistres, appels téléphoniques dans un standard, files d'attentes, mortalité, stocks, $\cdots$

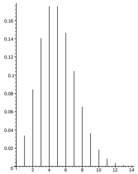
Définition :

Une variable aléatoire X suit une loi Poisson $P(\lambda)$ de paramètres λ positif lorsque sa loi de probabilité est définie par : Pour tout $k \in \boldsymbol{N}$

$$\boxed{P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}}$$

Exemple - Loi de Poisson

- Soit X la variable aléatoire mesurant le nombre de forets cassés en une journée suivant une loi $P(5)$
- Une table de formulaire donne pour $\lambda = 5$ les probabilités des événements $P(X = k)$ pour un certain nombre de valeurs de k
- Représentation graphique $P(5)$



- Calculer la proba. qu'en 1 jour 8 forets soient cassés

$$\begin{aligned} P(X \geq 8) &= 1 - P(X < 8) \\ &= 1 - (P(X = 0) + P(X = 1) + \cdots + P(X = 7)) \\ &= 1 - 0.867 = 0.133 \end{aligned}$$

Propriétés de la loi de Poisson

Propriétés :

Soit X une variable aléatoire suivant la loi de Poisson de paramètre λ ($P(\lambda)$) alors

$$\boxed{E(X) = \lambda, \; V(X) = \lambda, \; \sigma(X) = \sqrt{\lambda}}$$

Démonstration:

Admise

Remarque : (Approximation de la loi binomiale)

Si n est "grand", p "voisin" de 0 et np pas "trop grand" alors la loi $B(n, p)$ est très proche de la loi $P(\lambda)$ où $\lambda = np$

(ex: $n \geq 30, p \leq 0.1, np < 15$ ou $n \geq 50, p \leq 0.1, np \leq 10$)

Loi normale

Quand utiliser cette loi : (V.A. continue)

- Dans la modélisation de phénomènes aléatoires posssédant de nombreuses causes indépendantes dont les effets s'ajoutent sans que l'un d'eux soit dominant
- Situations : qualité, économie, sociologie $\cdots$

Définition :

Une variable aléatoire X suit une loi normale $N(m, \sigma)$ de paramètres m et σ lorsque sa densité de probabilité est la fonction f définie sur $\boldsymbol{R}$ par

$$\boxed{f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2}}$$

Propriétés de la loi normale

Propriétés :

Soit X une variable aléatoire suivant la loi normale $N(m, \sigma)$ alors

$$\boxed{E(X) = m,\ \ V(X) = \sigma^2,\ \ \sigma(X) = \sigma}$$

Démonstartion :

Admise

Remarque :

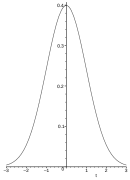
La loi normale $N(0, 1)$ a pour espérance 0
et pour écart type 1
On dira que $N(0, 1)$ est une loi normale centrée réduite

Loi normale centrée réduite $N(0,1)$

Théorème (important) :
Si une variable aléatoire suit une loi normale $N(m,\sigma)$ alors la variable aléatoire $T = \frac{X-m}{\sigma}$ suit la loi normale centrée réduite $N(0,1)$

Conséquence : Toute loi normale se ramène à $N(0,1)$

Représentation graphique de la densité de probabilité ("courbe en cloche")



Calcul de la probabilité d'un événement concernant la variable aléatoire T suivant la loi $N(0,1)$
$\longrightarrow$ utilisation de tables

Propriétés de la Loi normale

Propriétés :

Soit $\Pi(t_0)$ l'aire délimitée par $\begin{cases} \text{la courbe en cloche} \\ \text{l'axe des } t \\ \text{la droite } t = t_0 \end{cases}$

$$\boxed{P(t_1 \leq T \leq t_2) = \Pi(t_2) - \Pi(t_1)}$$

$$\boxed{P(-t \leq T \leq t) = 2\Pi(t) - 1}$$

Savoir utiliser les tables

Pour calculer une probabilité d'un événement concernant une variable aléatoire qui suit une loi normale $N(0,1)$ on tient compte des propriétés suivantes

- la courbe repésentative de la densité de probabilité est symétrique par rapport à l'axe des y

- l'aire totale comprise entre la courbe et l'axe des x est égale à 1

- ne pas oublier qu'une probabilité d'un événement peut s'exprimer à l'aide de la probabilité de l'événement contraire

Table Loi normale $N(0, 1)$

Table de la loi normale centrée réduite *N* (0 , 1)

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6654	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	08186	0,8212	0,8238	0,8254	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8797	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8988	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table pour les grandes valeurs de t

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
P(T<t)	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

Exemple sur la Loi normale

Exemple :

• Contexte : en contrôle de qualité sur une chaîne de production

• La variable aléatoire X mesure la longueur d'une pièce et suit la loi normale

de moyenne $m = 20cm$

d'écart type $\sigma = 2cm$

$$\bullet P(X \leq -4) = P\left(T = \frac{X-20}{2} < -12\right)$$
$$= \Pi(-12) = 1 - \Pi(12) \approx 0$$

(Ce résultat est bien proche de la réalité)

Loi normale et loi binomiale

Approximation de la loi binomiale par la loi normale

On admet que si n est "grand" et p ni "trop voisin" de 0 ni "trop voisin" de 1 alors la loi $B(n, p)$ est très proche de la loi $N(m, \sigma)$ où $m = np$ et $\sigma = \sqrt{np(1-p)}$

Comme par exemple quand np et $n(1-p) > 15$

Idée : Recherche d'informations sur une population d'effectif relativement important à partir de l'étude d'un échantillon

- Estimation ponctuelle (moyenne, pourcentage, écart type)
- Intervalle de confiance

Exemple :

Dans un atelier de réparation, on veut étudier la durée d'une intervention

On tire au hasard avec remise 50 fiches de suivis, on obtient les résultats

Durée	$[0, 20[$	$[20, 40[$	$[40, 60[$	$[60, 80[$	$[80, 100[$	$[100, 120[$
Nb. Int.	1	6	11	15	10	7

Estimation ponctuelle - Moyenne

Moyenne

Une estimation de la moyenne inconnue m d'une population est donnée par la moyenne de l'échantillon.

Ecart type

Un estimation de la variance d'une population est donnée par $\frac{n}{n-1}\sigma'^2$ où n est l'effectif de l'échantillon et σ'^2 est la variance de l'échantillon

Remarques :

- La variance d'un échantillon est corrigée pour donner un estimation de la variance de la population
- $\lim_{n \longrightarrow \infty} \frac{n}{n-1} = 1$

Exemple :

- Moyenne $\bar{x} = 69.2$
- $\sigma'^2 \approx 655.36$ d'où $\sigma^2 = \frac{n}{n-1}\sigma'^2 \approx 668.73$

($\sigma' \approx 25.6$ et $\sigma \approx 25.86$)

Estimation par intervlle de confiance

Constat : les résultats dépendent de l'échantillon
$\Longrightarrow$ Théorie de l'échantillonnage qui permet de controler l'influence d'un échantillon (utilisation des probabilités)

Moyenne

L'intervalle $[\bar{x} - t\frac{\sigma}{\sqrt{n}}, \bar{x} + t\frac{\sigma}{\sqrt{n}}]$ est l'intervalle de confiance de la moyenne m de la population (inconnue) avec le coefficient de confiance $2\Pi(t) - 1$ ayant pour centre la moyenne $\bar{x}$ de l'échantillon considéré

Remarque :

- Ca se justifie par le <u>théo de la limite centrée</u> appliqué à la V.A. $\bar{X}$ qui à tout échantillon associe sa moyenne
- Théorème : Soit $X_1, \cdots, X_n$ des variables aléatoires indépendantes suivant la même loi, admettant une moyenne μ et une variance σ^2 (non nulle). Pour n suffisament grand, la variable aléatoire $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ suit approximativement la loi normale $N(\mu, \frac{\sigma}{\sqrt{n}})$

Exemple

- On part de $n = 50$, $\sigma = 25.86$, $\bar{x} = 69.2$ et m inconnue
- Pour un coefficient de confiance de 95 %, on a $2\Pi(t) - 1 = 0.95 \Longleftrightarrow \Pi(t) = 0.975$ ce qui correspond à $t = 1.96$ (voir Table loi normale)
- l'intervalle $[\bar{x} - t\frac{\sigma}{\sqrt{n}}, \bar{x} + t\frac{\sigma}{\sqrt{n}}] = [62.03, 76.37]$ est l'intervalle de confiance de la moyenne m de la population

Interprétation :

- Si on prélève un très grand nombre d'échantillons, environ 95 % d'entre eux contindraient la moyenne inconnue m de la population
- En fait ici, on n'en prélève qu'un seul et on ne peut pas savoir si celui-ci contient ou non m, mais la méthode mise en œuvre permet d'obtenir un "bon" intervalle dans 95 cas sur 100

Démonstration (intervalle de confiance)

- Soit $\bar{X}$ la V.A. qui à tout échantillon associe sa moyenne
- On utilise le théorème de la limite centrée : la V.A. $\bar{X}$ suit la loi normale $N(m, \frac{\sigma}{\sqrt{n}}) \Longrightarrow T = \frac{\sqrt{n}}{\sigma}(\bar{X} - m)$ suit la loi $N(0, 1)$
- Pour tout $t \geq 0, \quad P(-t \leq T \leq -t) = 2\Pi(t) - 1$
- Par exemple pour $2\Pi(t) - 1 = 0.95$ alors $t = 1.96$ (voir ex.)

$$P(-1.96 \leq \frac{\sqrt{n}}{\sigma}(\bar{X} - m) \leq 1.96) = 0.95$$

- On peut donc écrire que :
- $P(m - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq m + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$

Avant de prélever un echantillon, il y a 95 chances sur 100 pour que $\bar{X}$ prenne une valeur sur $[m - 1.96\frac{\sigma}{\sqrt{n}}, m + 1.96\frac{\sigma}{\sqrt{n}}]$

- $P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$

Avant de prélever un échantillon, il y a 95 chances sur 100 pour que la V.A. $\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}$ prenne une valeur $\leq m$ et que $\bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$ prenne une valeur $\geq m$

Démonstration (intervalle de confiance)

Après le prélévement d'un échantillon: il est Vrai ou Faux que la moyenne de la population m se situe dans l'intervalle

$$[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} - 1.96\frac{\sigma}{\sqrt{n}}] = [62.03, 76.37]$$

Cet intervalle est appelé intervalle de confiance de la moyenne de la population avec le coefficient de confiance de 95 % (ou avec le risque de 5 %)

Tests de validité d'hypothèse

• Nature du problème : à partir de l'étude d'un ou plusieurs échantillons, on cherche à préciser comment prendre des décisions concernant l'ensemble de la population

• Nous allons présenter cette notion sur l'exemple déja utilisé pour les estimations
$n = 50$, $\sigma = 25,86mn$, [62.03,76.37] intervalle de confiance de m avec un coefficient de confiance de 95 %

Problèmes les plus fréquement rencontrés
1- Comparaison de la moyenne d'une population à un nombre fixé
2- Comparaison des moyennes de deux populations

Comparaison de la moyenne d'une population à un nombre fixé

Problématique

On se demande si l'on peut considèrer qu'en moyenne la durée d'intervention est de 60 minutes ?

$\Longrightarrow$ notion de "proximité" (moyenne estimée de 69.2)

Démarche à suivre

On suppose que la moyenne de la population est $m = 60$.

C'est l'hypothèse nulle, notée H_0: $m = 60$

Remarque : On suppose que la différence entre la moyenne souhaitée et la moyenne effective est nulle

Hypothèse nulle (suite)

• On reprend la démonstration présentée pour les intervalles de confiance et qui repose sur le fait que la V.A. $\bar{X}$ qui à un echantillon associe sa moyenne suit approximativement la loi normale $N(m, \frac{\sigma}{\sqrt{n}})$
Pour un coefficient de confiance de 95 %

$$P(m - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq m + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

• Pour notre exemple $P(52.83 \leq \bar{X} \leq 67.17) = 0.95$

• Ainsi, on sait avant de prélever un échantillon que sa moyenne appartient à [52.83,67.17]
Si H_0 est vraie, il n'y a que 5 % de chances de préléver un échantillon dont la moyenne soit inférieure à 52.83 ou supèrieure à 67.17

Erreur de première espèce

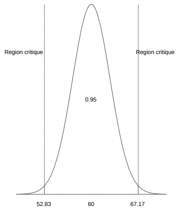
- On fixe la règle de décision:

On prélève un échantillon

On calcule sa moyenne $\bar{x}$

Si $\bar{x} \in [52.83, 67.17]$

on accepte H_0



Sinon on rejette H_0

- Si H_0 est vraie, on prend le risque de se tromper dans 5 % des cas en rejettant H_0
- On définit ainsi une région critique au seuil $\alpha = 5$ %

Le seuil α est la probabilité de rejeter H_0 alors que H_0 est vraie. Il correspond à l'erreur de première espèce

- En géneral α est fixé à priori. Ici α=0.05

Erreur de seconde espèce

• Si on choisit 1 % comme seuil pour diminuer le risque de rejeter H_0 alors que H_0 est vraie
On obtiendrait $P(50.56 \leq \bar{X} \leq 69.44) = 0.99$

• Dans notre exemple, au seuil de 1 % on accepte donc H_0

• En acceptant H_0 on court un second risque, celui d'accepter H_0 alors que H_0 est fausse c'est l'erreur de seconde espèce dont la probabilité est notée β
β est la probabilité d'accepter H_0 alors que H_0 est fausse

Remarques:
En général quand α diminue, β augmente et inversement
On essaye de limiter le plus grave !

Hypothèse alternative

• On cherche à définir précisément le cas où H_0 est fausse

• Dans notre exemple, on a choisi $m \neq 60$ comme hypothèse alternative H_1

• Le test est alors bilatéral car la région critique est située des deux cotés de la région où H_0 est accepté

• On parlera de test unilateral quand la région critique est entièrement située d'un coté (ex: $m < 60$)

Tests de validité - Résumé

Construction d'un test

- Choix de l'hypothèse nulle H_0
  Choix de l'hypothèse alternative H_1
- Détermination de la région critique à un seuil α donné
- Enoncé de la règle de décision :
  - si un paramètre de l'échantillon est dans la région critique, on rejette H_0
  - sinon on accèpte H_0

Utilisation du test

- Calcul du paramètre de l'échantillon mentionné dans la règle de décision
- Application de la règle de décision

Tests de validité - Résumé

Tableau des erreurs

Réalité / Décision	H_0 vraie	H_0 fausse
Accepter H_0	Bonne décision	erreur de 2ième espèce
Rejeter H_0	Erreur de 1ère espèce	Bonne décision

Tableau des risques

Réalité / Décision	H_0 vraie	H_0 fausse
Accepter H_0	$1-\alpha$	β
Rejeter H_0	α	$1-\beta$