Statistique Appliquée

Probabilités

Luc Deneire – deneire@unice.fr

Xidian University, Polytech Nice Sophia

Février 2016

Introduction

Informations

- Cours sur : http ://jalon.unice.fr/public/pqg729/
- Questions par e-mail : deneire@unice.fr

La statistique pour modéliser *le monde*

Permet de répondre à des questions du type

- Quelle est l'efficacité d'une usine ?
- Quelle est l'influence d'un changement dans un process ?

EN

- Recueillant des données
- En en déduisant un modèle
- En appliquant les modifications à ce modèle

L'outil du modèle : les probabilités

Les probabilités sont un outil pour *apprivoiser* l'aléatoire

Notions de bases :

- Evénements
- Probabilités associées à un événement
- Espace des événements et calcul de probabilités
- Inférences

Modèle probabiliste

Pour modéliser :

- Un hasard “simple” (jet de pièces, naissances d’enfants, ...)
- Temps d’attente à un guichet
- Probabilité de réussir un examen
- Moyenne du nombre de pièces défectueuses ...

Plan du cours - Partie Probabilités

1. Définitions / Axiomes / Probabilités conditonnelles
2. Bayes - Indépendance - Calculs de probabilités
3. Variables aléatoires : définitions
4. Espérance mathématique
5. Principales lois discrètes et continues
6. Covariance et corrélation entre deux variables aléatoires
7. Fonctions de variables aléatoires

Expérience aléatoire

Une expérience aléatoire est une expérience dont l'issue est incertaine (on ne peut savoir avec certitude quelle sera le résultat de l'expérience).

Exemples : jet de dés – Météo

Issue – Issue élémentaire

Quand une expérience a été *effectuée*, elle a une *issue*. Dans un cas simple, on définit toutes les *issues élémentaires* (dés : $\{"1", "2", "3", "4", "5", "6"\}$)

Univers

L'univers est l'ensemble des issues de cette expérence aléatoire. On le note Ω (dés : $\Omega = \{"1", "2", "3", "4", "5", "6"\}$)

Événement

Un événement **A** est un ensemble d'issues élémentaires liées à l'expérience aléatoire et est donc un sous-ensemble de l'Univers Ω.

Exemple : **A** = "*Le dé est pair*" = $\{"2", "4", "6"\}$

Espace probabilisé

Espace d'événements – Tribu

L'espace d'événements est un ensemble qui contient “tous les événements d'intérêt”, c'est-à-dire toutes les compositions de sous-ensembles de l'univers. Cet ensemble doit avoir une structure d'espace algébrique (σ-algèbre)
On le note $\mathcal{A}$.

Probabilité

La probabilité est un nombre, compris entre 0 et 1, associé à chaque événement (à chaque élément de la tribu). On notera la probabilité de l'événement A : $\mathbb{P}(A)$

Espace Probabilisé

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé un espace probabilisé

Exemple d'espace probabilisé

Exemple : Lancer de deux dés

- Les issues élémentaires : $\omega_1 = (1,1)$, $\omega_2 = (3,4)$, $\omega_3 = (4,3)$, $\ldots$
- L'univers : $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$
- Un événement : $\mathbf{A} = \{$la somme est égale à 6$\}$
- Un événement :
  $\mathbf{B} = \{$le 1$^{\text{er}}$ est entre 3 et 5 ; le 2$^{\text{nd}}$ entre 2 et 4$\}$
- La tribu $\mathcal{A} = \{$ tous les sous-ensembles de $\Omega\}$.
- Les probabilités $\mathbb{P}$: si on a des dés non pipés et que les lancer sont indépendants, $\mathbb{P}$ est caractérisé par $\mathbb{P}(\omega_i) = 1/36$.

$\triangleleft$

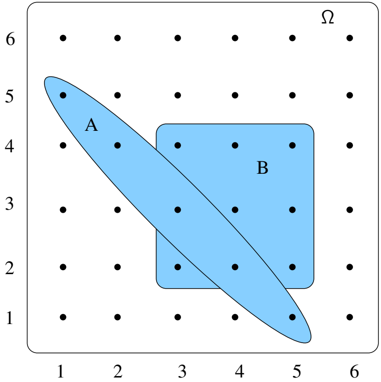
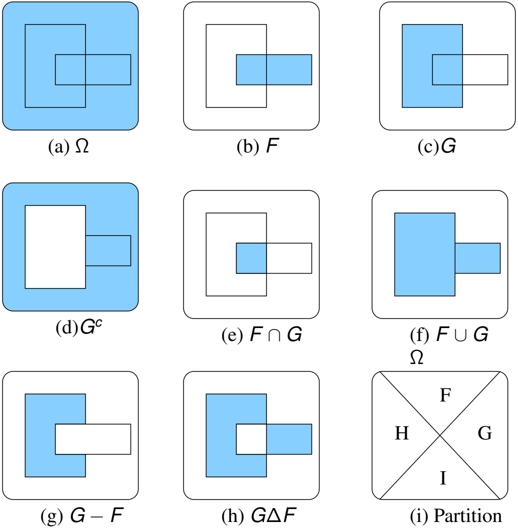


FIGURE: Exemple : lancer deux dés

exemples d'univers

- $\Omega = \emptyset$
- $\Omega = \{0, 1\}$; $\Omega = \{''\text{pile}'', ''\text{face}''\}$
- t-uples de longueur k ($k = 2$: $\Omega = \{\{0, 0\}; \{0, 1\}; \{1, 1\}; \{1, 0\}\}$
- : télécommunications : séquences infinies de "0" et de "1"
- nombres réels r entre $-V$ et V : $\Omega = \{r : -V \leq r \leq V\}$

Opérations de base des ensembles



a) L'univers Ω comprend toutes les issues possibles d'une expérience (si on tire un nombre réel au hasard, $\Omega = \mathcal{R} = \{\omega : -\infty \leq \omega \leq \infty\}$).
b) Un sous-ensemble de l'univers est un événement, p.ex. $F = \{\omega : -2 \leq \omega \leq 4\}$
c) Le complément à F, noté F^c est défini comme étant l'ensemble des éléments n'appartenant pas à F

$$F^c = \{\omega : \omega \notin F\}$$

d) L'intersection : $F \cap G = \{\omega : \omega \in F \text{ et } \omega \in G\}$
e) L'union : $F \cup G = \{\omega : \omega \in F \text{ ou } \omega \in G\}$
f) La différence : $G - F = \{\omega : \omega \in G \text{ et } \omega \notin F\} = G \cap F^c$
g) La différence symétrique :

$$G \Delta F = \{\omega : \omega \in G \text{ ou exclusif } \omega \in F\} = (F \cup G) - (F \cap G)$$

h) La partition :

$$\Omega = F \cup G \cup H \cup I \text{ et}$$

$$\forall X, Y \in \{F, G, H, I\} \text{ et } X \neq Y : X \cap Y = \emptyset$$

i) Première loi de De Morgan : $(F \cap G)^c = (F^c \cup G^c)$
j) Deuxième loi de De Morgan : $(F \cup G)^c = (F^c \cap G^c)$

Soit le triplet $(\Omega, \mathcal{A}, P)$,

1. Définir l'univers Ω.
2. Définir une tribu $\mathcal{A}$ (un exemple simple de tribu peut être $\mathcal{A}$ = tous les sous-ensembles de Ω).
3. Attribuer un nombre $\mathbb{P}(A) \in [0, 1]$ à un événement A.
   - Définition classique (Laplace)
   $$\mathbb{P}(A) = \frac{\text{nombre de cas équiprobables favorables}}{\text{nombre de cas équiprobables possibles}}$$
   - Définition intuitive (fréquence relative)
   $$\mathbb{P}(A) = \lim_{n \to \infty} \frac{N_n(A)}{n}$$
   - Définition axiomatique (Kolmogorov)

Axiome

$$\mathbb{P}(A) \geq 0 \quad \textit{pour chaque événement } A \in \mathcal{A}.$$

Axiome

$$\mathbb{P}(\Omega) = 1$$

Axiome

$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *pour A et B disjoints, qui se généralise à :*
soit les événements $A_i, i = 1, 2, ..., n$ disjoints, alors

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Une généralisation plus forte encore est : soit les événements $A_i, i = 1, 2, ...$ disjoints, alors

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
   dém. : $\mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) \overset{A \cap A^c = \emptyset}{=} \mathbb{P}(A) + \mathbb{P}(A^c) = 1$
2. $\mathbb{P}(\emptyset) = 0 = \mathbb{P}(\Omega^c)$
3. Toutes les probabilités sont comprises entre 0 et 1 (démonstration : utiliser les deuxième et troisième axiomes ainsi que la définition de complément).
4. **Partition** : si $\{A_i\}$ est une partition (finie ou infiniment dénombrable) de Ω, alors, $\mathbb{P}(B) = \sum_i \mathbb{P}(B \cap A_i)$, pour tout événement B.
5. Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$
6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
7. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
8. $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A^c \cap B^c \cap C)$

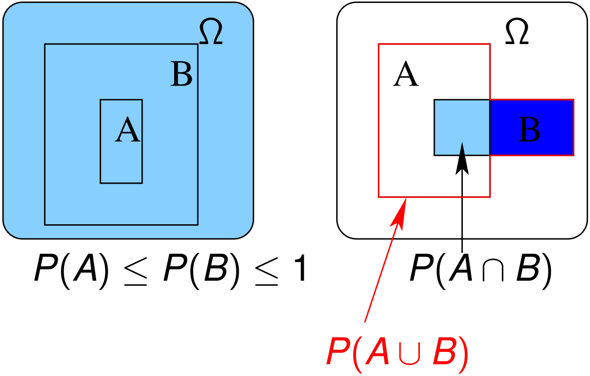


FIGURE: Exemples simples de la relation entre probabilités et ensembles.

Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$

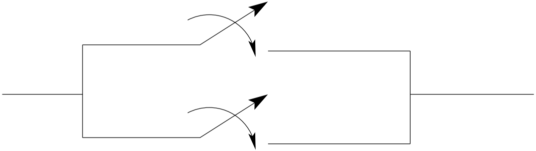
Exemple : Interrupteurs en série

Soit $\mathbb{P}$(Les deux interrupteurs sont fermés) $= 1/4$ et $\mathbb{P}$(Un interrupeur est fermé) $= 1/2$. On définit les événements : $A = \{$interrupteur 1 et 2 fermés$\}$ et $B = \{$interrupteur 1 fermé$\}$ (avec $A \subset B$) . La probabilité que le circuit fonctionne est alors $\mathbb{P}(A) = 1/4 < \mathbb{P}(B)$=1/2.

$\triangleleft$

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

Exemple : Interrupteurs en parallèle



Soit $A = \{\text{interrupteur 1 fermé}\}$ et $B = \{\text{interrupteur 2 fermé}\}$, avec $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$ et $\mathbb{P}(A \cap B) = 1/4$.
Alors, la probabilité que le circuit fonctionne est :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4} \end{aligned} \quad (1)$$

$\triangleleft$

La probabilité qu'il pleuve sachant que je suis à Xi'An est plus grande que si je suis dans le désert de Gobi ...

Probabilité conditionnelle

$\mathbb{P}(A|B) \in [0,1]$: associée à A, *sachant que* l'événement B ($\mathbb{P}(B) \neq 0$) a été réalisé.

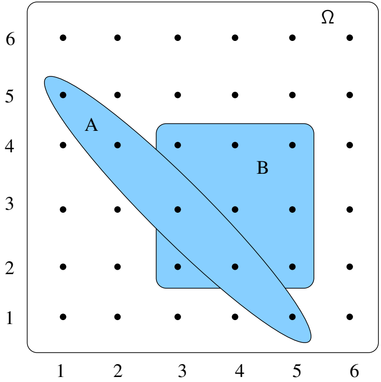


FIGURE: Exemple : lancer deux dés

Exemple : Lancer de deux dés

Dans l'exemple de la figure 3, en supposant des dés non pipés, les événements A et B ont les probabilités indiquées ci-dessous. Toutes les issues ω_i ($i = 1, \dots, 36$) sont équiprobables ($\mathbb{P}(\omega_i) = \frac{1}{36}$).

- $\mathbb{P}(A) = \frac{5}{36}$
- $\mathbb{P}(B) = \frac{9}{36}$
- $\mathbb{P}(A|B) = \frac{2}{9} = \frac{2/36}{9/36}$
- $$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

$\triangleleft$

Exemple : Le tabac et les jeunes

Soit l'exemple suivant :

	Fréquences		
	Fumeurs	Non fumeurs	Total
Hommes	340	314	654
Femmes	289	384	673
Total	629	698	1327

Probabilités (approche fréquentiste)

- $\mathbb{P}(\text{Hommes}) = 654/1327 = 0.49$
- $\mathbb{P}(\text{Femmes}) = 673/1327 = 0.51$
- $\mathbb{P}(\text{Fumeurs}) = 629/1327 = 0.47$
- $\mathbb{P}(\text{Non fumeurs}) = 698/1327 = 0.53$
- $\mathbb{P}(\text{Fumeurs} \cap \text{Hommes}) = 340/1327 = 0.26$
- $\mathbb{P}(\text{Fumeurs}|\text{Hommes}) = 340/654 = 0.53 = \boxed{0.26/0.49}$
- $\mathbb{P}(\text{Fumeurs}|\text{Femmes}) = 289/673 = 0.43 = \boxed{0.22/0.51}$

Exemple : Le tablac et les jeunes (suite)

Les fréquences relatives approximent les probabilités des événements “Hommes” (H), “Femmes” (nH), “Fumeurs” (F), et “non Fumeurs” (nF).

Fréquences relatives			
	Fumeurs	Non fumeurs	Total
Hommes	0.26	0.24	0.49
Femmes	0.22	0.29	0.51
Total	0.47	0.53	1

- **Espace Probabilisé**
  $\Omega = \{(H, F); (H, nF); (nH, F); (nH, nF)\}$, $\mathcal{A} = \Omega$,
  $\mathbb{P} = (0.26, 0.24, 0.22, 0.29)$.
- **Probabilité conjointe** (“Homme” et “Fumeur”, ...)
- **Probabilité marginale**(“Homme” ; “Fumeur”, ...)

◁

Les probabilités conditionnelles satisfont Kolmogorov

- La probabilité conditionnelle satisfait les trois axiomes :
  1. $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0$ pour chaque événement $A \subseteq \Omega$
  2. $\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)$ pour A_1 et A_2 disjoints
  3. $\mathbb{P}(B|B) = 1$
- Les propriétés générales restent valables , p.ex., $\mathbb{P}(A \cup C|B) \leq \mathbb{P}(A|B) + \mathbb{P}(C|B)$
- On peut remplacer 3. par 3'. $\mathbb{P}(B|B) = \frac{\mathbb{P}(B \cap B)}{\mathbb{P}(B)} = 1$
- $\mathbb{P}(A|B)$: *loi* de probabilité ;
- **Approche séquentielle (appelée théorème de la multiplication ou chain rule)** :
  - $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A|B)}$
  - $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}\left(A_n | \bigcap_{i=1}^{n-1} A_i\right)$

Fausse alarme

Soit un radar de détection d'un avion, on cherche la probabilité d'une fausse alarme

- $\mathbb{P}(\text{Avion présent}) = 0.05$
- $\mathbb{P}(\text{ Détection d'avion s'il est présent}) = 0.99$
- Fausse détection : $\mathbb{P}(\text{Détection d'avion si pas présent}) = 0.1$

Modélisation de l'univers "système radar"

- Avion : Présent / Absent
- Radar : Détection / Non détection
- En fonction des quatre issues possibles, l'univers vaut : $\Omega = \{(P, D), (A, D), (P, N), (A, N)\}$,

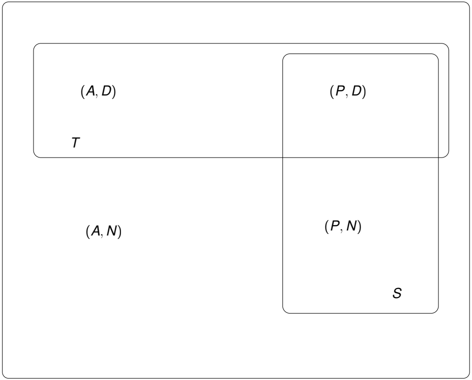


FIGURE: Détection de présence de l'avion : les 4 points de l'univers.

Les probabilités associées

- $S = \{\text{un avion est présent}\} = \{(P, D), (P, N)\}$
- $T = \{\text{le radar signale la présence d'un avion}\} = \{(P, D), (A, D)\}$
- $\mathbb{P}(S) = 0.05$ (présence d'un avion)
- $\mathbb{P}(T|S) = 0.99$ (détection si avion présent)
- $\mathbb{P}(T|S^c) = 0.10$ (fausse détection : ≪ détection ≫ si avion absent)

Les probabilités associées calculées grâce aux axiomes / propriétés

- Quelle est la probabilité d'une fausse alarme ?

  $\mathbb{P}((A, D)) = \mathbb{P}(S^c \cap T) =$
  $\mathbb{P}(S^c)\,\mathbb{P}(T|S^c) = [1 - \mathbb{P}(S)]\mathbb{P}(T|S^c) = 0.95 \cdot 0.10 = 0.095$
- Quelle est la probabilité qu'un avion ne soit pas détecté ?

  $\mathbb{P}((P, N)) = \mathbb{P}(S \cap T^c) =$
  $\mathbb{P}(S)\,\mathbb{P}(T^c|S) = \mathbb{P}(S)\,[1 - \mathbb{P}(T|S)] = 0.05 \cdot 0.01 = 0.0005$
- pour obtenir l'ensemble des probabilités, il faut calculer $\mathbb{P}((P, D)) = \mathbb{P}(S) - \mathbb{P}((P, N)) = 0.05 - 0.0005 = 0.0495$ et on en déduit directement $\mathbb{P}((A, N)) = 0.855$ en invoquant que la probabilité de l'univers vaut 1.

- soient $A_1, A_2, \ldots, A_n$ une partition de Ω
- soit un événement B, on peut écrire :
  $B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$, où $B \cap A_1, B \cap A_2, \ldots, B \cap A_n$ sont disjoints ;
- $\Rightarrow \mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \ldots + \mathbb{P}(B \cap A_n)$
- et donc $\mathbb{P}(B) = \mathbb{P}(A_1)\,\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\,\mathbb{P}(B|A_2) + \ldots + \mathbb{P}(A_n)\,\mathbb{P}(B|A_n)$.
- Ce résultat est connu comme étant le *théorème des probabilités totales* :

$$\boxed{\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{P}(B|A_i)}$$

Bayes : vers l'inférence

Soient :

- *la probabilité a priori* : $\mathbb{P}(A) = 0.001$; (exemple : $A = \{\text{j'ai un accident}\}$)
- *l'événement connu* : B (exemple = $B = \{\text{je roule trop vite}\}$)
- *la probabilité a posteriori* : $\mathbb{P}(A|B) = 0.01$.

Donc "J'ai un accident" *Parce que* "je roule trop vite" ... Schématiquement :

- ≪ Effet ≫ $A \longrightarrow$ ≪ Cause ≫ B, $\mathbb{P}(A|B)\,,\mathbb{P}(B) \neq 0$
- À partir de $\mathbb{P}(A|B)$, calculer $\mathbb{P}(B|A)$ (cause $\longrightarrow$ effet)
- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B|A) = \mathbb{P}(B)\,\mathbb{P}(A|B)$
- $$\boxed{\mathbb{P}(B|A) = \mathbb{P}(B)\,\frac{\mathbb{P}(A|B)}{\mathbb{P}(A)}}$$

S'il y a plusieurs causes possibles :

- $$\mathbb{P}(B_i|A) = \mathbb{P}(B_i)\,\frac{\mathbb{P}(A|B_i)}{\mathbb{P}(A)}$$
- $$\mathbb{P}(B_i)\,\mathbb{P}(A|B_i)$$

Exemple : Le tabac et les jeunes

	Fréquences		
	Fumeurs	Non fumeurs	Total
Hommes	340	314	654
Femmes	289	384	673
Total	629	698	1327

- $\mathbb{P}(\text{Hommes}) = 654/1327 = 0.49$
- $\mathbb{P}(\text{Fumeurs}) = 629/1327 = 0.47$
- $\mathbb{P}(\text{Fumeurs}|\text{Hommes}) = 340/654 = 0.53 = 0.26/0.49$
- $\mathbb{P}(\text{Hommes}|\text{Fumeurs}) = 340/629 = 0.54 = 0.26/0.47$
  $\mathbb{P}(\text{Hommes}|\text{Fumeurs}) = 0.49 \cdot 0.53/0.47$
- $\mathbb{P}(\text{Fumeurs}|\text{Hommes}) > \mathbb{P}(\text{Fumeurs})$
- $\mathbb{P}(\text{Hommes}|\text{Fumeurs}) > \mathbb{P}(\text{Hommes})$

$\triangleleft$

Indépendance

Deux événements A et B sont **indépendants** si et seulement si $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)}$.

On en déduit les propriétés suivantes :

- si $\mathbb{P}(B) \neq 0$, $\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)}$
- Soit deux événements indépendants A et B, conditionnés par C, $(\mathbb{P}(C) \neq 0)$:
  - $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\,\mathbb{P}(B|C)$
  - si $\mathbb{P}(B|C) \neq 0, \mathbb{P}(A|B \cap C) = \mathbb{P}(A|C)$
- Soit plusieurs événements indépendants $A_1, A_2, \ldots, A_n$:
  - $\mathbb{P}(\bigcap_{i \in S} A_i) = \prod_{i \in S} \mathbb{P}(A_i)$
    pour *chaque* S, sous-ensemble de $\{1, 2, \ldots, n\}$

- définir l'univers Ω.
- définir les probabilités associées.
  Par exemple considérer que les issues sont équiprobables :
  $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$ (Laplace).

DANGEREUX

- Utiliser – Indépendance – Bayes – Probabilités totales

Exemple : Chaîne de production

◁

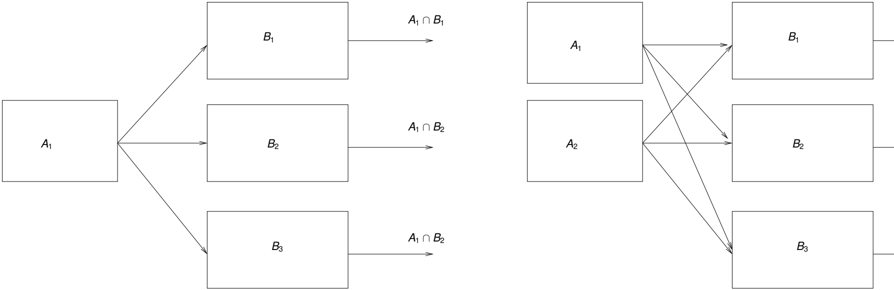


FIGURE: chaîne de production : risque de panne

$\mathbb{P}(A_i \text{ en panne}) = 0.2$ et $\mathbb{P}(B_i \text{ en panne}) = 0.4$

Question

Quelle est la probabilité de panne dans les deux cas ?

1.
   - Soit A_i l'événement " la chaîne A_i fonctionne correctement" et de même pour B_i. On a $\mathbb{P}(A_i) = 0.8$ et $\mathbb{P}(B_i) = 0.6$
   - Soit l'événement F : "la chaîne fonctionne". Selon le premier schéma, on a $F = (A_1 \cap B_1) \cup (A_1 \cap B_2) \cup (A_1 \cap B_3) = A_1 \cap (B_1 \cup B_2 \cup B_3)$.
   - A_1 est indépendant des B_i et donc $\mathbb{P}(F) = \mathbb{P}(A_1)\,.\mathbb{P}(B_1 \cup B_2 \cup B_3)$.
   - De plus, par les lois des ensembles, $\mathbb{P}(B_1 \cup B_2 \cup B_3) = 1 - \mathbb{P}(B_1 \cup B_2 \cup B_3)^c = 1 - \mathbb{P}\left(B_1^c \cap B_2^c \cap B_3^c\right)$, et par indépendance de B_i, on a indépendance des B_i^c et donc $\mathbb{P}\left(B_1^c \cap B_2^c \cap B_3^c\right) = \mathbb{P}\left(B_1^c\right).\mathbb{P}\left(B_2^c\right).\mathbb{P}\left(B_3^c\right)$.
   - Globalement on a alors $\mathbb{P}(F) = \mathbb{P}(A_1)\,.\mathbb{P}(B_1 \cup B_2 \cup B_3) = \mathbb{P}(A_1)\,.(1 - \mathbb{P}\left(B_1^c\right).\mathbb{P}\left(B_2^c\right).\mathbb{P}\left(B_3^c\right)) = 0.8(1 - (0.4)^3) = 0.75$, soit une probabilité de panne de 25 % .
2. Sur le deuxième schéma, $F = (A_1 \cup A_2) \cap \mathbb{P}(B_1 \cup B_2 \cup B_3)$, soit, toujours par indépendance,$\mathbb{P}(F) = \mathbb{P}(A_1 \cup A_2)\,.\mathbb{P}(B_1 \cup B_2 \cup B_3)$. Avec le même raisonnement que ci-dessus, on obtient $\mathbb{P}(A_1 \cup A_2) = 1 - \mathbb{P}(A_1^c)\,\mathbb{P}(A_2^c) = 0.96$ et donc $\mathbb{P}(F) = 0.96 \times 0.94 = 0.9$, soit une probabilité de panne de 10 %.

Associer un nombre à un événement

$\Omega = \{''\,faux'', ''\,vrai''\}$: compliqué à manipuler
$\Rightarrow$ Associer un nombre :

'' faux''	$\rightarrow 0$
'' vrai''	$\rightarrow 1$

Définition : Variable aléatoire

Une *Variable aléatoire* X est une fonction de Ω dans $\mathbb{R}$ telle qu'à tout ω correspond une valeur $X(\omega) = x$.

Définition : Domaine de variation

Le *domaine de variation* de X est l'ensemble $R_X \subset \mathbb{R}$ que peut prendre la variable aléatoire X.

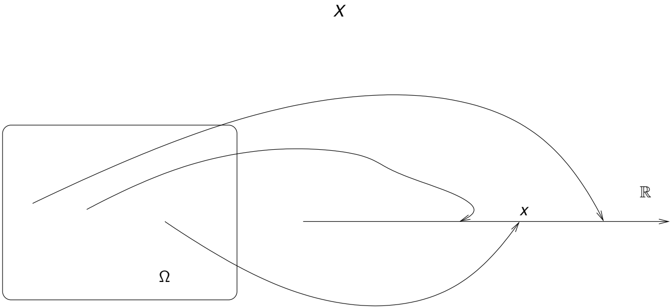


FIGURE: La variable aléatoire : une fonction de l'univers dans l'espace des réels

Définition : Réalisation

On appelle x une *réalisation* de la variable aléatoire liée à **l'événement** ω.

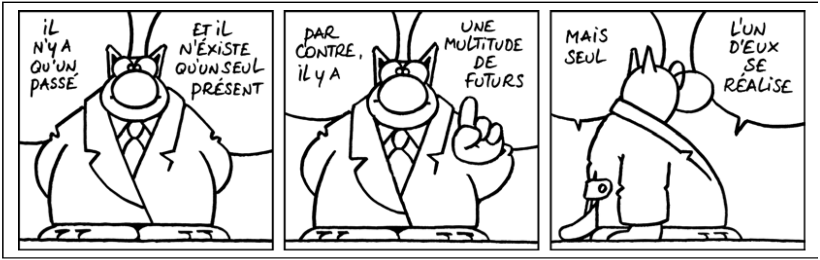


FIGURE: Définition d'une réalisation par le chat de Geluck

Exemple de variable aléatoire

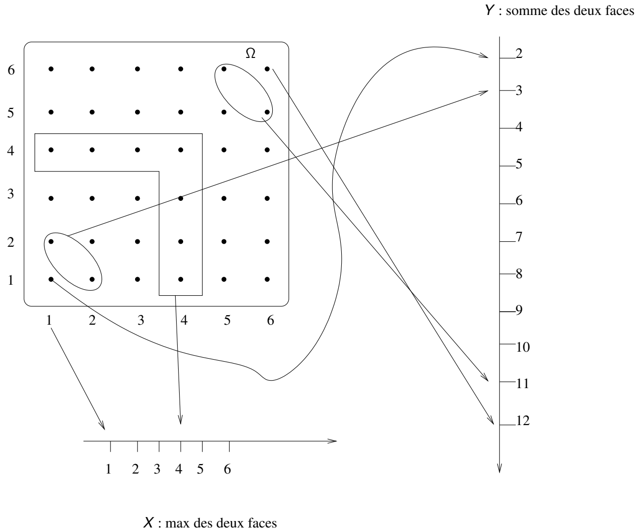


FIGURE: Exemple de deux variables aléatoires liées à la même expérience aléatoire

Variable aléatoire discrète

La *Variable aléatoire discrète* X prend ses valeurs dans un ensemble *fini* de valeurs *dénombrable* ou *non dénombrable*.

exemple Les domaines de variation de X et Y sont respectivement $R_X = \{1, 2, 3, 4, 5, 6\}$ et $R_Y = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

une v.a est définie par sa *fonction de probabilité* (*masse de probabilité* pour les v.a. discrètes).

Masse de probabilité

La fonction de probabilité est la fonction $p_X(x)$ définie par :

$$p_X(x) = \begin{cases} \mathbb{P}\left(\underbrace{\{X = x\}}_{\text{événement } \in \Omega}\right) \overset{\text{simpl.}}{=} \mathbb{P}(X = x) & \text{si } x \in R_X \\ 0 & \text{si } x \notin R_X \end{cases}$$

Propriétés

- $p_X(x) \geq 0 \qquad \forall x \in \mathbb{R}.$
- $\displaystyle\sum_{i=1}^{n} p_X(x_i) = 1 \qquad \text{si } R_X = \{x_1, x_2, \cdots x_n\}.$

Exemple : Détermination d'une masse de probabilité

Une classe de n élèves présente un examen. La masse de probabilité X du nombre d'élèves ayant la note x est une masse de probabilité triangulaire entre $x = 2$ et $x = 18$ sur 20, avec $p_X(2) = 0, p_X(10) = 8.a, p_X(18) = 0$.
On a donc

$$p_X(x) = \begin{cases} 0 & x < 3 \\ a * (x - 2) & 2 < x < 11 \\ a * (8 - (x - 10)) & 10 < x < 18 \\ 0 & x > 17 \end{cases}$$

Trouver a tel que $\sum p_X(x_i) = 1$.
$\Rightarrow \sum p_X(x_i) = 64.a$ et donc $a = 1/64$.

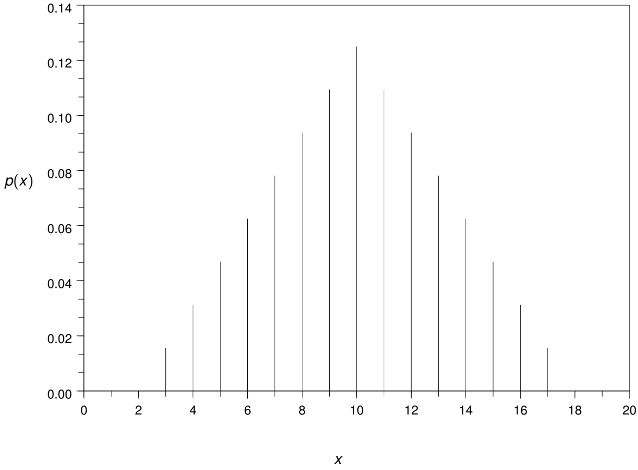


FIGURE: Fonction ou masse de probabilité des notes d'une classe

Definition : Fonction de répartition

$F(x) = F_X(x) \triangleq \mathbb{P}(\{X \leq x\}) = \sum_{x_i \leq x} p_X(x_i)$

- Si on classe les éléments de R_X par ordre : $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$,

$$F_X(x_{(k)}) = \mathbb{P}\big(\{X \leq x_{(k)}\}\big) = \sum_{i=1}^{k} p_X(x_{(i)})$$

- $0 \leq F(x) \leq 1, \qquad \forall x \in \mathbb{R}$
- $F(x_{(k)}) - F(x_{(k)}^{-}) = p_X\left(x_{(k)}\right)$
- La fonction $F(x)$ est monotone croissante (au sens large) :

$$\forall x_i < x_j \ , \ F(x_i) \leq F(x_j)$$

- La fonction $F(x)$ “démarre” en 0 et “termine” en 1
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to +\infty} F_X(x) = 1$
- $\forall x_i \in R_X, \qquad F(x_i) - F_X(x_{i-1}) = \mathbb{P}(\{x_{i-1} < X \leq x_i\})$.

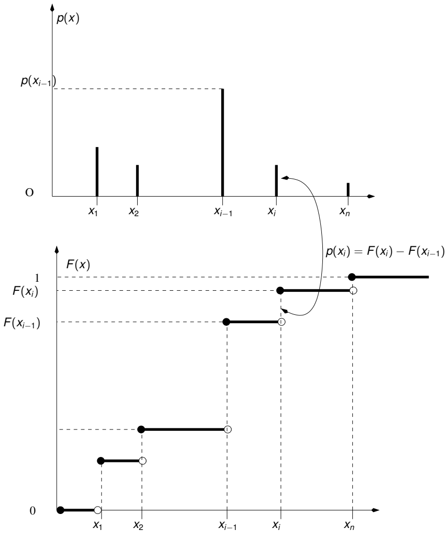


FIGURE: La fonction de répartition : lien avec la masse de probabilité.

Définition : Variable aléatoire continue

Une *variable aléatoire continue* est une variable aléatoire définie sur un domaine de variation continu.

Fonction de répartition

$F_X(x) = \mathbb{P}(X \leq x)$

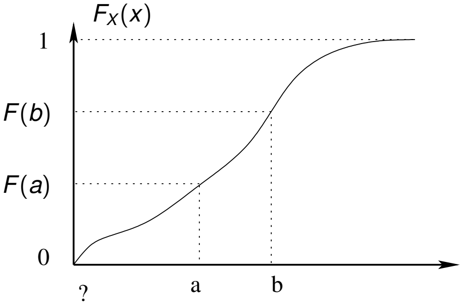


FIGURE: Fonction de répartition d'une v.a. continue.

Propriétés de la fonction de répartition

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$$

car $(a \leq X \leq b)$ = $(X \leq b)\backslash(X \leq a)$ et donc
$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$.

Probabilité et Fonction de répartition

Dans le cas continu, la probabilité $\mathbb{P}(X = x) = 0$ et donc
$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X < x)$.

Définition : Densité de probabilité

La *densité de probabilité* $f_X(x)$ est définie par :

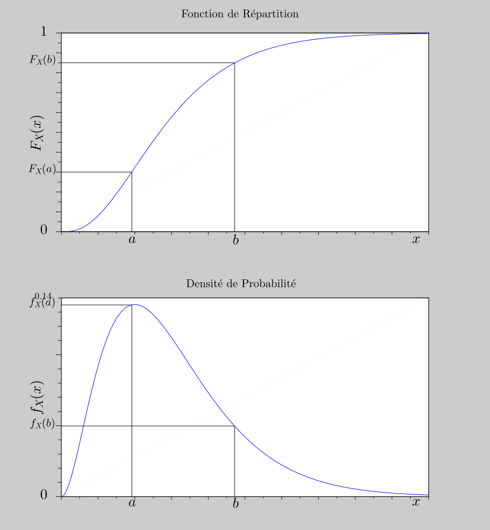
$$f_X(x) = \lim_{\epsilon \to 0} \frac{\mathbb{P}(x < X \leq x + \epsilon)}{\epsilon}$$

$\mathbb{P}(x < X \leq x + \epsilon) = F_X(x + \epsilon) - F_X(x)$ et donc, en supposant la fonction de répartition dérivable, on a la définition suivante :

Definition : Densité de probabilité

La *densité de probabilité* $f_X(x)$ est définie par :

$$f_X(x) = \frac{dF_X(x)}{dx}$$


Fonction de Répartition
1
$F_X(b)$
$F_X(x)$
$F_X(a)$
0
a
b
x
Densité de Probabilité
0.14
$f_X(a)$
$f_X(x)$
$f_X(b)$
0
a
b
x

Soit une v.a. X et un événement **A**,
que devient X si A ?

Definition : Fonction de répartition conditionnelle

Soit A et $\mathbb{P}(A) \neq 0$ la *fonction de répartition conditionnelle* $F_{X|\mathbf{A}}(x|\mathbf{A})$ est telle que :

$$F_{X|\mathbf{A}}(x|\mathbf{A}) = \mathbb{P}((X \leq x)|\mathbf{A}) = \frac{\mathbb{P}((X \leq x) \cap \mathbf{A})}{\mathbb{P}(\mathbf{A})}.$$

Definition : Masse de probabilité conditionnelle

Soit A et $\mathbb{P}(A) \neq 0$ la *masse de probabilité conditionnelle* $p_{X|\mathbf{A}}(x|\mathbf{A})$ est telle que :

$$p_{X|\mathbf{A}}(x|\mathbf{A}) = \mathbb{P}((X = x)|\mathbf{A}) = \frac{\mathbb{P}((X = x) \cap \mathbf{A})}{\mathbb{P}(\mathbf{A})}.$$

Definition : Densité de probabilité conditionnée sur un événement

Soit $\mathbf{A}$ et $\mathbb{P}(A) \neq 0$ la *densité de probabilité conditionnelle* $f_{X|\mathbf{A}}(x|\mathbf{A})$ telle que :

$$f_{X|\mathbf{A}}(x|\mathbf{A}) = \frac{dF_{X|\mathbf{A}}(x|\mathbf{A})}{dx}$$

Definition : masse de probabilité jointe

Soient deux v.a. discrètes X et Y, la **masse de probabilité jointe** est :

$$p_{XY}(x, y) = \mathbb{P}((X = x) \cap (Y = y))$$

En particulier, $p_{XY}(x, y) \geq 0$ et $\sum_x \sum_y p_{XY}(x, y) = 1$

Definition : masse de probabilité marginale

oient deux v.a. discrètes X et Y, les **masses de probabilité marginales** cont :

$$p_X(x) = \mathbb{P}((X = x)) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = \mathbb{P}((Y = y)) = \sum_x p_{XY}(x, y)$$

variable aléatoire conditionnelle discrète

Soient les v.a. X et Y,
Que devient X, si $Y = y_i$ est connu ?
De façon simple, il s'agit, $\forall x_i$ de connaitre $\mathbb{P}(X = x_i | Y = y_j)$. Il s'agit donc de la variable aléatoire conditionnée sur l'événement $(Y = y_j)$
ICI : $Y = y_i$ est un événement de proba non nulle

Definition : Fonction de répartition jointe

Soient deux v.a. continues X et Y, la **Fonction de répartition jointe** est :

$$F_{XY}(x,y) = \mathbb{P}((X \leq x) \cap (Y \leq y))$$

Definition : Densité de probabilité jointe

Soient deux v.a. continues X et Y, la **Densité de probabilité jointe** est :

$$f_{XY}(x,y) = \frac{dF_{XY}(x,y)}{dxdy}$$

En particulier, $f_{XY}(x,y) \geq 0$ et $\int_x \int_y f_{XY}(x,y) = 1$ et
Pour toute région R de l'espace engendré par x et y :

$$\mathbb{P}([X,Y] \in R) = \int\int_R f_{XY}(x,y)\, dxdy$$

Definition : Densité de probabilité marginale

Soient deux v.a. discrètes X et Y, les **masses de probabilité marginales** cont :

$$f_X(x) = \int_{y=-\infty}^{y=+\infty} f_{XY}(x, y)\, dy$$

$$f_Y(y) = \int_{x=-\infty}^{x=+\infty} f_{XY}(x, y)\, dx$$

variable aléatoire conditionnelle continue

Soient les v.a. X et Y,
Que devient X, si $Y = y$ est connu ?

On se rappelera que $\mathbb{P}(x < X \leq x + \Delta_x) = f_X(x)\,\Delta_x$, pour Δ_x petit.
On notera que $\mathbb{P}((x < X \leq x + \Delta_x)|\mathbf{A}) = f_{X|A}(x)\,\Delta_x$ car la dépendance sur **A** Donc

$$\mathbb{P}((x < X \leq x + \Delta_x)|(y < Y \leq x + \Delta_y)) = \frac{\mathbb{P}((x < X \leq x + \Delta_x), (y < Y \leq y + \Delta_y))}{\mathbb{P}(y < Y \leq y + \Delta_y)}$$

On en déduit directement que

$$f_{X|Y}(x|y)\Delta_x = \frac{f_{XY}(xy)\Delta_x\Delta_y}{f_Y(y)\Delta_y}$$

Definition : Densité de probabilité conditionnelle

Soient deux variable aléatoires continues X et Y, de densité conjointe $f_{XY}(xy)$ et de densité $f_Y(y)$ non nulle sur le support de Y, alors la densité conditionnelle, dite de X conditionnée sur Y est donnée par :

$$f_{X|Y}(x|y) = \frac{f_{XY}(xy)}{f_Y(y)}.$$

On exprime également la densité conditionnelle de la manière suivante, en explicitant la valeur de Y sur laquelle on conditionne :

$$f_{X|Y}(x|Y = y_o) = \frac{f_{XY}(xy)}{f_Y(y_o)}.$$

Le Mode d'une v.a. X est la valeur x_m : $x_m = \arg\max_x f_X(x)$. Valeur la plus *vraisemblable*.

La Médiane d'une v.a. X est la valeur $x_{\frac{1}{2}}$: $\mathbb{P}\left(X \leq x_{\frac{1}{2}}\right) = 0.5$.

Les Quantiles d'une v.a. X, plus précisément le *p-quantile* est d'une v.a. X est la valeur x_p : $\mathbb{P}(X \leq x_p) = p$.

L'espérance mathématique d'une v.a. X est la moyenne de cette v.a., pondérée par sa densité de probabilité. Intuitivement, c'est la valeur qu'on s'attend à observer en moyenne (que l'on "espère" observer).

La variance d'une v.a. X est une mesure (du carré) de la variation qu'on peut observer autour de la moyenne. L'espérance et la variance donnent une bonne idée du domaine de variation de la variable aléatoire X.

Les moments d'une v.a. X sont l'espérance des puissances de la v.a. (donc la moyenne est le moment d'ordre 1, puisque c'est l'espérance de la X à la puissance 1, le moment d'ordre 2 est lié à la variance, etc.).

Definition : Mode

Le mode d'une variable aléatoire X est la valeur x_m telle que :

$$\forall x \in R_X, x \neq x_m, \qquad p_X(x_m) > p_X(x) \text{ pour une v.a. discrète ;}$$

$$\forall x \in R_X, x \neq x_m, \qquad f_X(x_m) > f_X(x) \text{ pour une v.a. continue.}$$

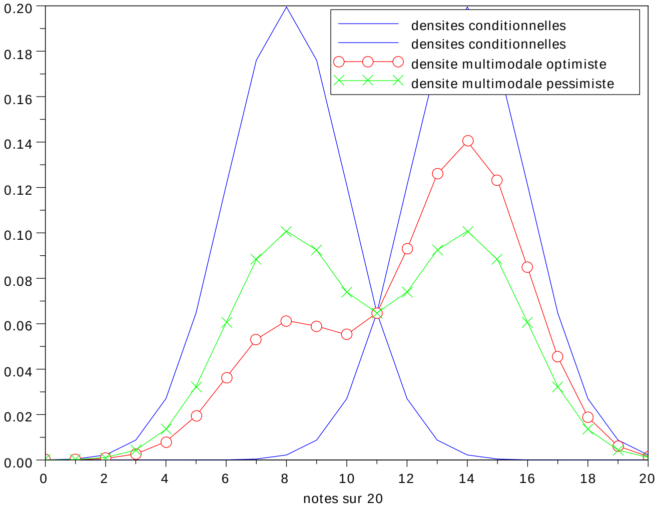
On impose une inégalité stricte ($f_X(x_m) > f_X(x)$), donc :

- le mode n'est pas toujours défini (exemple de l'uniforme) ;
- à strictement parler, il n'y a qu'un seul mode. Cependant, on parle souvent de *v.a. multimodale* s'il y a plusieurs maxima locaux ; dans le cas contraire, on parle de *v.a. unimodale*.

v.a. multimodale : notes d'un D.S.

Un bon examen a souvent un résultat bi-modal :

- Ceux qui ont travaillé : notes autour de 14
- Ceux qui n'on pas travaillé : notes autour de 8
- soit N la v.a. représentant les notes et l'événement **P** : **P** $\equiv$ "l'étudiant a préparé".
  1. $N|P \sim N(14, 4)$
  2. $N|P^c \sim N(8, 4)$


0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
10
12
14
16
18
20
notes sur 20
densites conditionnelles
densites conditionnelles
densite multimodale optimiste
densite multimodale pessimiste

Definition : Médiane

La médiane d'une variable aléatoire X est la valeur $x_{\frac{1}{2}}$ telle que :

$$\mathbb{P}\left(X \leq x_{\frac{1}{2}}\right) \triangleq F(x_{\frac{1}{2}}) = \frac{1}{2}$$

N'existe pas toujours pour une v.a. discrète.

Definition : $p-$quantile

Le $p-$quantile d'une variable aléatoire X est la valeur x_p telle que :

$$\mathbb{P}(X \leq x_p) \triangleq F(x_p) = p, \qquad p \in [0, 1].$$

On distingue en particulier

La médiane : pour $p = 1/2$, qui “divise” en deux le domaine de variation de la v.a.

Les quartiles : pour $p = 1/4$ (le premier quartile), $p = 1/2$ et $p = 3/4$ (le troisième quartile. Les quartiles “divisent” le domaine de variation de la v.a. en quatre parties “égales” (c'est-à-dire dont la surface sous la densité de probabilité est divisée en quatre parties égales). On a donc que la probabilité de se trouver entre deux quartiles successifs vaut 1/4.

Les déciles : pour $p = k/10$, $k = 1, 2, ..., 9$. $x_{0.1}$ est le premier décile, etc. On a donc que la probabilité de se trouver entre deux déciles successifs vaut 1/10.

Les centiles : pour $p = k/100, k = 1, 2, , 99$. L'utilité est plutôt pour les grands et petits centiles, par exemple, a probabilité d'obtenir une valeur supérieur au 99$^{\text{ème}}$ centile est de 1 pourcent.

Definition : Intervalle de confiance

Un intervalle de confiance (bilatéral) $[a, b]$ au niveau p est tel que

$$\mathbb{P}(a \leq X \leq b) = p.$$

Un intervalle de confiance (unilatéral) $[-\infty, b]$ au niveau p est tel que :

$$\mathbb{P}(X \leq b) = p.$$

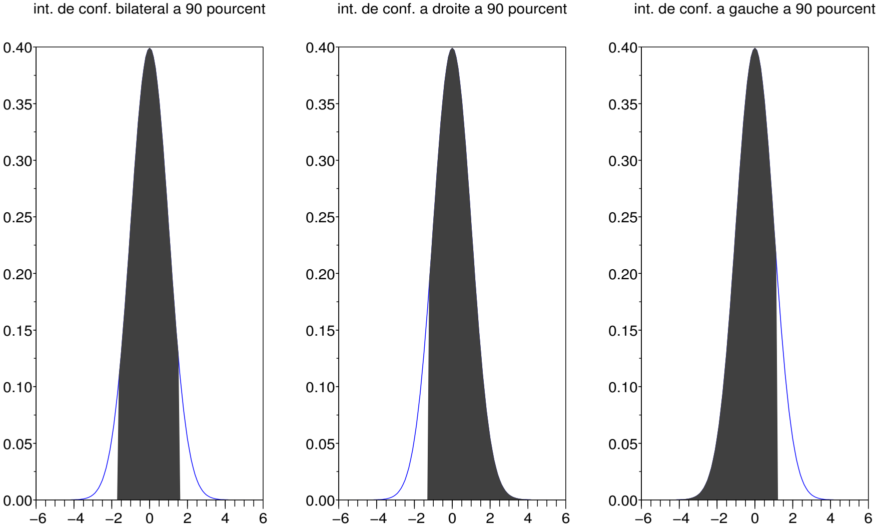
Un intervalle de confiance (unilatéral) $[a, \infty]$ au niveau p est tel que :

$$\mathbb{P}(a \leq X) = p.$$

Dans le cas bilatéral $\mathbb{P}(X < a) = \mathbb{P}(X > b) = (1-p)/2 = \alpha/2$, où $\alpha = 1 - p$.

Confiance et erreur

α représente alors la probabilité qu'on a de se tromper si on fait l'hypothèse que la réalisation x de la v.a. X est dans l'intervalle $[a, b]$, et on a que $a = x_{\alpha/2}$ et $b = x_{1-\alpha/2}$.


int. de conf. bilateral a 90 pourcent
int. de conf. a droite a 90 pourcent
int. de conf. a gauche a 90 pourcent
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
–6 –4 –2 0 2 4 6

Definition : Espérance mathématique

L'espérance mathématique μ ou $\mathrm{E}[X]$ est définie par :

$$\mu = \mathrm{E}[X] = \begin{cases} \displaystyle\sum_{i=1}^{n} x_i p_X(x_i) & \text{cas discret} \\ \displaystyle\int_{-\infty}^{\infty} x f_X(x)\, dx & \text{cas continu} \end{cases}$$

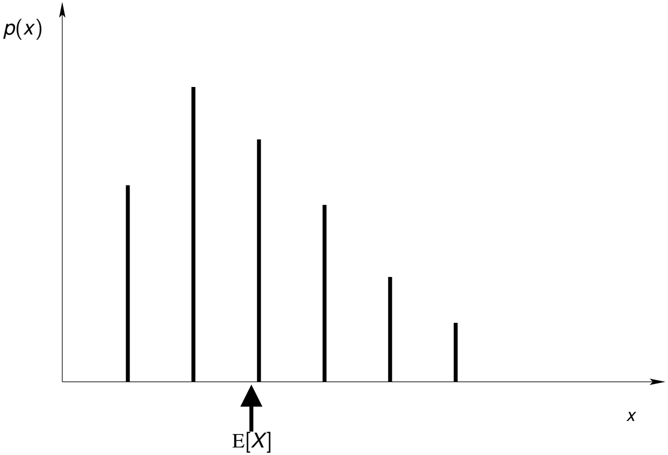


FIGURE: Interprétation d'un espérance comme étant un centre de gravité

L'espérance d'une fonction d'une variable aléatoire, contrairement à la loi de probabilité de cette fonction, s'obtient très facilement :

Definition : Espérance d'une fonction d'une v.a.

Si $Y = g(X)$, où $g(.)$ est une fonction, alors :

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \begin{cases} \displaystyle\sum_{i=1}^{n} g(x_i)p_X(x_i) & \text{pour une v.a. discrète} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f_X(x)\,dx & \text{pour une v.a. continue} \end{cases}$$

Si $Y = aX + b$, alors, l'application directe de la linéarité de la somme et de l'intégrale donne la propriété suivante :

Propriété : Linearité de l'espérance

Si $Y = aX + b$ alors :

$$\mathrm{E}[Y] = a\mathrm{E}[X] + b$$

Definition : v.a. de Bernoulli

x	$p_X(x)$
1	p
0	1-p
$x \neq \{0,1\}$	0

On dira que $X \sim Be(p)$: la variable aléatoire X est distribuée selon la loi de Bernoulli

Definition : Variable aléatoire binomiale

Soit $Y_1 \sim \cdots \sim Y_n \sim Be(p)$ avec les Y_i mutuellement indépendantes, alors :

$$X = Y_1 + \cdots + Y_n \sim Bi(n,p)$$

est une variable aléatoire binomiale.

Exemple : v.a. de Bernoulli et Binômiale

Une variable aléatoire de Bernoulli aura un espérance égale à (pour $X \sim Be(p)$) :

$$\mathrm{E}[X] = 0.(1-p) + 1.p = p.$$

Une variable aléatoire Binômiale $X \sim Bi(n,p)$ aura une espérance :

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i}^{n} X_i\right] = \sum_{i}^{n} \mathrm{E}[X_i] = np.$$

$\triangleleft$

Definition : Variance

La variance d'une variable aléatoire X est donnée par :

$$\mathrm{var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \int_{R_X} (X - \mathrm{E}[X])^2 f_X(x)\,dx.$$

Bernoulli

Si $X \sim Be(p)$, alors

$$\begin{aligned} \mathrm{E}[X] &= \textstyle\sum_{k=0}^{1} k p_X(k) \\ &= 0(1-p) + 1.p \\ &= p. \end{aligned} \tag{2}$$

$$\begin{aligned} \mathrm{var}[X] &= \textstyle\sum_i (x_i - \mathrm{E}[X])^2 p_X(x_i) \\ &= \textstyle\sum_{k=0}^{1} (k-p)^2 P_X(k) \\ &= (0-p)^2(1-p) + (1-p)^2 p \\ &= p(1-p) \end{aligned} \tag{3}$$

Definition : Variable aléatoire binomiale

Une variable aléatoire est binomiale $X \sim Bi(n, p)$ si et seulement si :

$$p_X(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \cdots, n \text{ avec } C_n^x = \dfrac{n!}{x!(n-x)!} \\ 0 & \text{sinon} \end{cases}$$

Binomiale

si $X \sim Bi(n, p)$, alors

$$\begin{aligned} \mathrm{E}[X] &= \textstyle\sum_{k=0}^n k p_X(k) \\ &= \textstyle\sum_{k=0}^n = C_n^k p^k (1-p)^{n-k} \\ &= np. \end{aligned} \tag{4}$$

Espérance d'une fonction de v.a. multiples

$$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dxdy$$

Cas particulier :

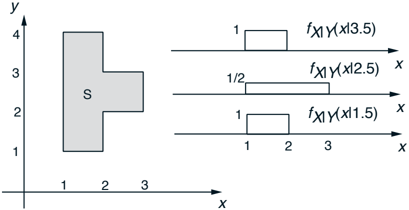
$$\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$$

Rappel de densité conditionnelle

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

et

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1$$


y
4
3
2
1
S
1
2
3
x
1
$f_{X|Y}(x|3.5)$
x
1/2
$f_{X|Y}(x|2.5)$
x
1
$f_{X|Y}(x|1.5)$
1
2
3
x

$$
f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of the circle}} & \text{if } (x,y) \text{ is in the circle,} \\ 0 & \text{otherwise,} \end{cases}
$$

$$
= \begin{cases} \dfrac{1}{\pi r^2} & \text{if } x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases}
$$

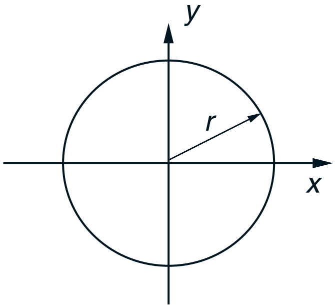


Figure 3.19: Circular target for Example 3.16.

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx \\
&= \frac{1}{\pi r^2} \int_{x^2+y^2} \leq r^2 dx \\
&= \frac{1}{\pi r^2} \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx \\
&= \frac{2}{\pi r^2} \sqrt{r^2-y^2}
\end{aligned}
$$

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{XY}(x,y)}{f_Y(y)} \\
&= \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2}\sqrt{r^2-y^2}} \\
&= \frac{1}{2\sqrt{r^2-y^2}}
\end{aligned}
$$

$$\mathrm{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$$

$$\begin{aligned}
\mathrm{E}[X|Y=y] &= \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{x}{2\sqrt{r^2-y^2}}\,dx \\
&= \left.\frac{x^2}{4\sqrt{r^2-y^2}}\right]_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \\
&= 0
\end{aligned}$$

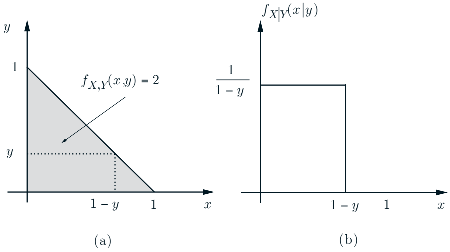


Figure 4.6: (a) The joint PDF in Example 4.15. (b) The conditional density of X.

We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{1-y} 2\,dx = 2(1-y), \qquad 0 \le y \le 1,$$

and

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{1-y}, \qquad 0 \le x \le 1-y.$$

On a : $\mathrm{E}[X|Y=y] = \frac{1-y}{2}$
Si on considère que y varie, on peut le considérer comme une v.a. (Y)
On écrit alors $\mathrm{E}[X|Y] = \frac{1-Y}{2}$ (! ! ! ! ! c'est une v.a. ! ! ! ! !)
On peut donc en prendre l'espérance :

$$\begin{aligned} \mathrm{E}_Y(\mathrm{E}_X(X|Y)) = \mathrm{E}[EX|Y] = & \int_{-\infty}^{\infty} \mathrm{E}[X|Y=y]\, f_Y(y)\, dy \\ & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, f_Y(y)\, dxdy \\ & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\, dxdy \\ & = \mathrm{E}[X] \end{aligned}$$

Espérances itérées

$$\mathrm{E}_Y(\mathrm{E}_X(X|Y)) = \mathrm{E}[\mathrm{E}[X|Y]] = \mathrm{E}[X]$$

Exemple

- $\mathrm{E}[X|Y] = \frac{1-Y}{2}$
- $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]] = \frac{1-\mathrm{E}[Y]}{2}$
- par symétrie $\mathrm{E}[X] = \mathrm{E}[Y]$
- Donc $\mathrm{E}[X] = 1/3$

Definition : Variable aléatoire de Bernoulli

X est une *variable de Bernoulli*

x	$p_X(x)$
1	p
0	1-p
$x \neq \{0,1\}$	0

On dira que $X \sim Be(p)$:

Exemple : Encore la famille et garçon/fille

Soit $\mathbf{A}$ = “le cadet est un garçon” et $\mathbb{P}(\mathbf{A}) = 1/2$. On a donc $X \sim Be(1/2)$.
Soit $\mathbf{A}$ = “les trois premiers enfants sont des garçons”, $\mathbb{P}(\mathbf{A}) = 1/8$ et donc $X = Be(1/8)$. $\triangleleft$

Definition : Variable aléatoire binomiale

Soit $Y_1 \sim \cdots \sim Y_n \sim Be(p)$ avec les Y_i mutuellement indépendantes, alors :

$$X = Y_1 + \cdots + Y_n \sim Bi(n, p)$$

est une variable aléatoire binomiale.

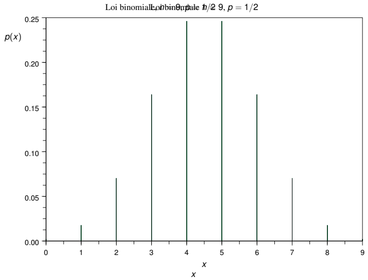
Exemple : Jet de n pièces de monnaie

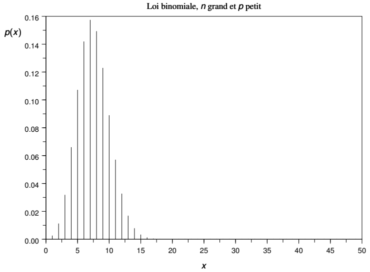
Supposons que l'on jette une pièce n fois.
$\mathbb{P}(\text{pile}) = p$ indépendamment d'un jet à l'autre.

Le succès est ici “pile” et donc $Y_i = 1 \Rightarrow X = \sum_{i=1}^{n} Y_i$.

$$p_X(x) = \mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} \text{ si } x = 0, 1, 2, \cdots, n \text{ avec } C_n^x = \frac{n!}{x!(n-x)!}$$

$\triangleleft$


Loi binomiale, $n = 9$, $p = 1/2$
$p(x)$
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
6
7
8
9
x


Loi binomiale, n grand et p petit
$p(x)$
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25
30
35
40
45
50
x

Definition : Variable aléatoire binomiale

Une variable aléatoire est binomiale $X \sim Bi(n, p)$ si et seulement si :

$$p_X(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \cdots, n \text{ avec } C_n^x = \dfrac{n!}{x!(n-x)!} \\ 0 & \text{sinon} \end{cases}$$

Definition : Variable aléatoire géométrique

Une variable aléatoire suit une loi géométrique de paramètre p (on notera $X \sim Ge(p)$) si et seulement si :

$$p_X(x) = \begin{cases} p(1-p)^{x-1} & \text{si } x = 1, 2, \ldots \\ 0 & \text{sinon} \end{cases}$$

et

$$F(x) = 1 - (1-p)^x \qquad x = 1, 2, \ldots$$

Soit une v.a. géométrique $X \sim Ge(p)$, son espérance vaut :

$$\mathrm{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}.p = \frac{1}{p} \tag{5}$$

Sa variance vaut :

$$\mathrm{var}[X] = \frac{1-p}{p^2} \tag{6}$$

v.a. Pascal

Réussir une épreuve k fois.

Definition : Variable aléatoire de Pascal

Une variable aléatoire est dite de Pascal $X \sim Pa(k, p)$ si et seulement si :

$$p_X(x) = \begin{cases} C_{x-1}^{k-1} p^k (1-p)^{x-k} & \text{si } x = k, k+1, \cdots \\ 0 & \text{sinon} \end{cases}$$

Exemple : Poker

$\mathbb{P}$ une paire, un brelan et un carré ? (nombre infini de cartes : $\mathbb{P}(''As'') = 1/13$)
Dans ce cas, on obtient :

- Pour une paire : $p_X(x) = C^1_{x-1} p^2 (1-p)^{x-2}$ Soit :

x	2	3	4	5	6	$\cdots$
$p_X(x)$	0.0059	0.0109	0.0151	0.0186	0.0215	$\cdots$

  Et donc la probabilité d'avoir une paire sur 5 cartes vaut $13.0.0186 = 0.242$

- Pour un brelan : $p_X(x) = C^2_{x-1} p^3 (1-p)^{x-3}$ Soit :

x	2	3	4	5	6	$\cdots$
$p_X(x)$	0.0	0.00045	0.00126	0.00232	0.00358	$\cdots$

  Et donc la probabilité d'avoir un brelan sur 5 cartes vaut $13.0.00232 = 0.06$

$\triangleleft$

Loi binômiale : comportement particulier pour p petit et n grand.
Nombre d'occurences de A si $\mathbb{P}(A) = p$ pour p petite ?
: X tend vers la *loi de Poisson* de paramètre μ

$$X \sim Po(\mu)$$

.
La fonction de probabilité $p_X(x)$ de la v.a. de Poisson vaudra alors :

Definition : Variable aléatoire de Poisson

Une v.a. discrète est dite de Poisson ($X \sim Po(\mu)$) si et seulement si :

$$p_X(x) = \begin{cases} \dfrac{e^{-\mu}\mu^x}{x!} & \text{si } x = 0, 1, 2, \cdots \qquad \mu > 0 \\ 0 & \text{sinon} \end{cases}$$

Soit une v.a. de Poisson $X \sim Po(\lambda)$, son espérance vaut :

$$\begin{aligned} \mathrm{E}[X] &= \sum_{i=0}^{\infty} i \frac{e^{-\lambda}\lambda^i}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda \end{aligned} \tag{7}$$

Sa variance vaut également λ.

Soit une loi binomiale $X \sim Bi(n,p)$ de valeurs n grande et p petite et de produit np fini. Dans ce cas, on a un événement de faible probabilité, mais qui après un grand nombre d'essais, se produira np fois en moyenne. Cette loi binomiale peut être approximé par une loi de Poisson W de paramètre $w = np$. On a donc, pour une variable aléatoire binomiale :

$$\boxed{X \overset{n\to\infty}{\sim} Po(np)}$$

En effet, la masse de probabilité $p_X(x)$ de la binomiale vaut :

$$\begin{aligned}
p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!(n-x)!}\frac{w^x}{n^x}\left(1-\frac{w}{n}\right)^{n-x} \\
&= \frac{n(n-1)\dots(n-x+1)}{x!}\frac{w^x}{n^x}\left(1-\frac{w}{n}\right)^{n-x} \\
&= \frac{n}{n}\frac{n-1}{n}\dots\frac{n-x+1}{n}\frac{w^x}{x!}\left(1-\frac{w}{n}\right)^{n}\left(1-\frac{w}{n}\right)^{-x} \\
&\underset{n\to\infty}{\longrightarrow} 1\cdot 1\cdot\ldots\cdot 1\frac{w^x}{x!}\mathrm{e}^{-w}\cdot 1 = \mathrm{e}^{-w}\frac{w^x}{x!} = p_W(x)
\end{aligned}$$

Definition : Variable aléatoire uniforme

Une variable aléatoire uniforme que l'on note

$$X \sim Un(a, b)$$

est définie par sa densité de probabilité :

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

Elle est définie de façon équivalente par sa fonction de répartition :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{sinon} \end{cases}$$

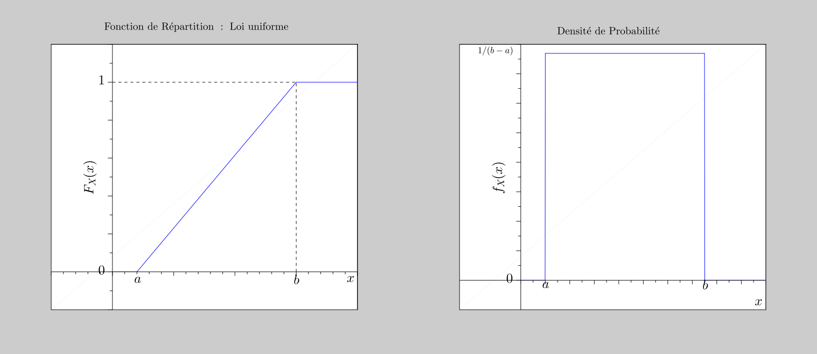


FIGURE: Loi de probabilité uniforme

On obtient aisément que si $X \sim Un(a, b), \mathrm{E}[X] = (a + b)/2$, et que variance vaut $\mathrm{var}[X] = (b - a)^2/12$.

Definition : Variable aléatoire exponentielle

Une variable aléatoire est dite exponentielle de paramètre λ, notée $X \sim Exp(\lambda)$ si et seulement si :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \; ; F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

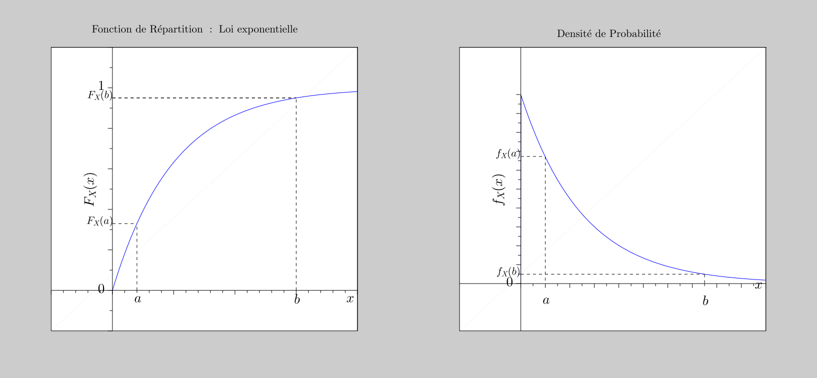


FIGURE: Loi de probabilité exponentielle

Exemple : Durée de vie d'un composant

Les composants suivent une loi exponentielle de paramètre $\lambda = \frac{1}{20}\text{ans}^{-1}$. On demande de calculer :

- La probabilité que le composant fonctionne plus de (10 ; 15 ; 20 - ;25) ans.
- La demi-vie, i.e. le temps x tel que la probabilité que la durée de vie excède x soit égale à 0.5.

Solution

- On a que $\mathbb{P}(X > x) = 1 - \mathbb{P}(X < x) = 1 - F_X(x)$, ce qui donne des probabilités de (0.61 ; 0.47 ; 0.37 ; 0.29). On notera que si l'espérance de vie est de 20 ans, la probabilité d'atteindre 20 ans n'est que de 37 % !.
- On cherche x tel que $1 - F_X(x) = 0.5$, donc $F_x = e^{-x/\lambda} = 0.5$: $x = -20 * \ln 0.5 \simeq 13.8$, et la demi-vie est de presque 14 ans.

$\triangleleft$

Si $X \sim \exp(\lambda)$, alors

$$\mathrm{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right] \Big|_0^\infty = \frac{1}{\lambda}. \qquad (8)$$

De même, $\mathrm{var}[X] = 1/\lambda^2$.

La plus aléatoire

Definition : Variable aléatoire Normale

Une variable aléatoire normale (encore appelée Gaussienne) de paramètres μ (fini) et σ^2 (positif), notée

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

est définie par sa densité de probabilité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in \mathbb{R}$$

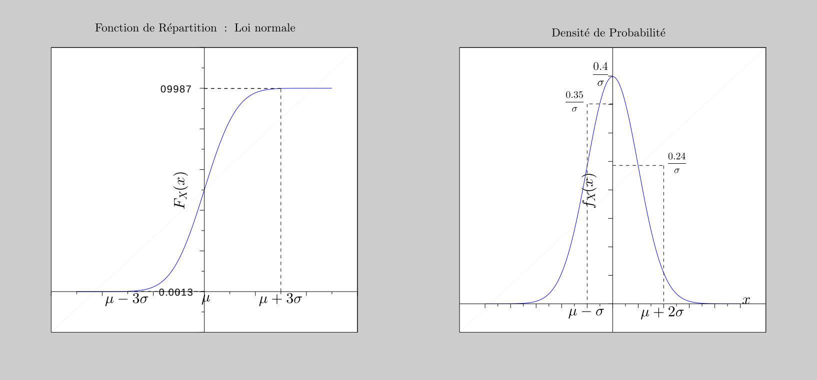


FIGURE: Loi de probabilité normale

Pas d'expression analytique de la primitive de $f_X(x)$ et que la fonction de répartition $F_X(x)$ est donc définie sous forme intégrale.

La variable aléatoire Laplacienne est similaire à la normale, sauf qu'elle décroit plus lentement (on dit que c'est une variable aléatoire à queue lourde).

Definition : Variable Aléatoire Laplacienne

Une variable aléatoire continue X est dite Laplacienne si et seulement si sa densité de probabilité est donnée par :

$$f_X(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}|x-\mu|\right) \qquad x \in \mathbb{R}. \tag{9}$$

Sa fonction de répartition est alors donnée par :

$$F_X(x) = 0.5*(1+\text{sign}(x)*\exp\left(-\sqrt{\sigma^2/2}|x-\mu|\right) \qquad x \in \mathbb{R}. \tag{10}$$

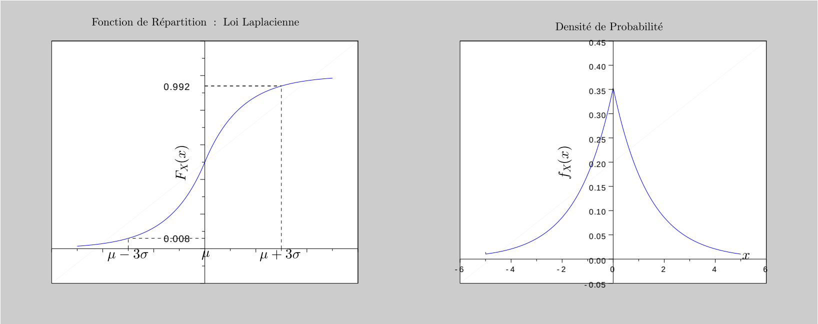


FIGURE: Loi de probabilité laplacienne

La variable aléatoire de Cauchy est similaire à la normale, sauf qu'elle décroit plus lentement (on dit que c'est une variable aléatoire à queue lourde.

Definition : Variable Aléatoire de Cauchy

Une variable aléatoire continue X est dite de Cauchy si et seulement si sa densité de probabilité est donnée par :

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}. \tag{11}$$

Sa fonction de répartition est alors donnée par :

$$F_X(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2} \qquad x \in \mathbb{R}. \tag{12}$$

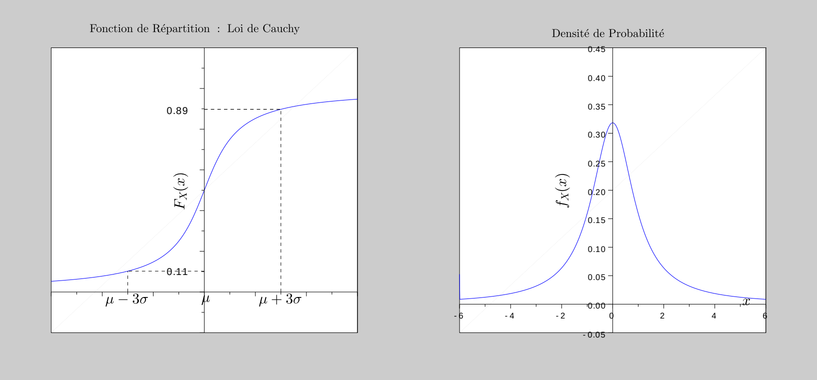


FIGURE: Loi de probabilité de Cauchy

Definition : v.a. de Rayleigh

Une v.a. de Rayleigh est définie par sa densité de probabilité :

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & x \geq 0 \\ 0 & x \leq 0 \end{cases} \tag{13}$$

Sa fonction de répartition est donnée par :

$$F_X(x) = \begin{cases} 1 - \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{14}$$

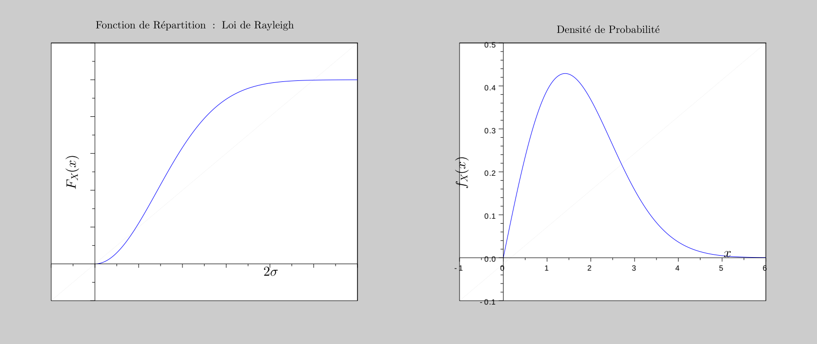


FIGURE: Loi de probabilité de Rayleigh (variance = 4)

La variance donne une idée de la variabilité des issues possibles autour de la moyenne.
Dans ce cadre, on peut se poser la question suivante :
“Quelle est la probabilité que la réalisation d’une variable aléatoire quelconque soit écartée de la moyenne de plus d’une quantité donnée ?”. En termes mathématiques :

$$\mathbb{P}(|X - \mathrm{E}[X]| > \gamma) \leq p.$$

En d’autres termes, si $\mu = 0$ et $\sigma^2 = 1$, quand je dis que la réalisation de la variable aléatoire sera comprise entre -3 et 3, quelle est la probabilité de se tromper ? (ici p).

On peut trouver une borne supérieure de p de la manière suivante :

$$\begin{aligned}
\mathrm{var}[X] &= \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f_X(x)\,dx \\
&= \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} (x - \mathrm{E}[X])^2 f_X(x)\,dx + \underbrace{\int_{\{x:|x-\mathrm{E}[X]|\leq\gamma\}} (x - \mathrm{E}[X])^2 f_X(x)\,dx}\\
&\geq \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} (x - \mathrm{E}[X])^2 f_X(x)\,dx \\
&\geq \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} \gamma^2 f_X(x)\,dx \qquad \text{car dans ce domaine, } |x - \mathrm{E}[X]\,| > \gamma \\
&= \gamma^2 \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} f_X(x)\,dx \\
&= \gamma^2 \mathbb{P}(|X - \mathrm{E}[X]\,| > \gamma)\,,
\end{aligned} \tag{15}$$

Inégalité de Tchebychev

$$\mathbb{P}(|X - \mathrm{E}[X]\,| > \gamma) \leq \frac{\mathrm{var}[X]}{\gamma^2} \tag{16}$$

Si $\gamma = 3$, on obtient $\mathbb{P}(|X - \mathrm{E}[X]\,| > \gamma) \leq \frac{\sigma^2}{9} \simeq 0.11\sigma^2$.

Definition : Fonction de répartition jointe

Soient deux v.a. continues X et Y, la **Fonction de répartition jointe** est :

$$F_{XY}(x, y) = \mathbb{P}((X \leq x) \cap (Y \leq y))$$

Definition : Densité de probabilité jointe

Soient deux v.a. continues X et Y, la **Densité de probabilité jointe** est :

$$f_{XY}(x, y) = \frac{dF_{XY}(x, y)}{dxdy}$$

En particulier, $f_{XY}(x, y) \geq 0$ et $\int_x \int_y f_{XY}(x, y) = 1$ et
Pour toute région R de l'espace engendré par x et y :

$$\mathbb{P}([X, Y] \in R) = \int \int_R f_{XY}(x, y)\, dxdy$$

Definition : Densité de probabilité marginale

Soient deux v.a. discrètes X et Y, les **masses de probabilité marginales** cont :

$$f_X(x) = \int_{y=-\infty}^{y=+\infty} f_{XY}(x, y)\, dy$$

$$f_Y(y) = \int_{x=-\infty}^{x=+\infty} f_{XY}(x, y)\, dx$$

Definition : Covariance

La *Covariance* de deux v.a. X et Y, notée $\mathrm{cov}[X, Y]$ vaut :

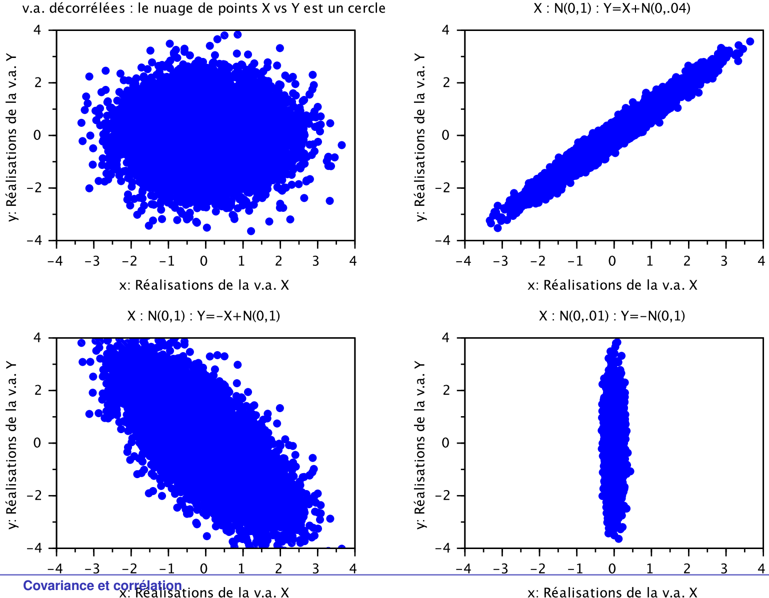
$$\mathrm{cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Si $\mathrm{cov}[X, Y] = 0$, X et Y sont **décorrélées**.

Relation avec l'indépendance

X et Y indépendants $\Rightarrow \mathrm{cov}[X, Y] = 0$

Mais pas l'inverse !


v.a. décorrélées : le nuage de points X vs Y est un cercle
X : N(0,1) : Y=X+N(0,.04)
X : N(0,1) : Y=−X+N(0,1)
X : N(0,.01) : Y=−N(0,1)
y: Réalisations de la v.a. Y
x: Réalisations de la v.a. X
−4
−3
−2
−1
0
1
2
3
4

1. $X \simeq \mathcal{N}(0,1)$, $Y \simeq \mathcal{N}(0,1)$, X et Y Indépendants :
$$\begin{aligned}\operatorname{cov}[X,Y] = \mathrm{E}[XY] &= \int_x \int_y x.yf_{XY}(x,y)\,dxdy \\ &= \int_x xf_X(x)\,dx \int_y yf_Y(y)\,dy \\ &= \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= 0\end{aligned}$$
2. $X \simeq \mathcal{N}(0,1)$, $Y = X + \mathcal{N}(0,.04)$ notons $Z \simeq \mathcal{N}(0,.04)$:
$\operatorname{cov}[X,Y] = \mathrm{E}[X.(X+Z)] = \mathrm{E}\left[X^2\right] + \mathrm{E}[X]\,\mathrm{E}[Z] = 1$
3. $X \simeq \mathcal{N}(0,1)$, $Y = -X + \mathcal{N}(0,1)$ notons $Z \simeq \mathcal{N}(0,1)$:
$\operatorname{cov}[X,Y] = \mathrm{E}[X.(-X+Z)] = -\mathrm{E}\left[X^2\right] + \mathrm{E}[X]\,\mathrm{E}[Z] = -1$
4. $X \simeq \mathcal{N}(0,.01)$, $Y = -\mathcal{N}(0,1)$:
$\operatorname{cov}[X,Y] = \mathrm{E}[X.Y] = \mathrm{E}[X]\,\mathrm{E}[Y] = 0$

La covariance ne dit pas tout

Definition : Coefficient de corrélation

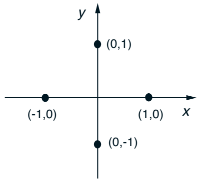
Le *Coefficient de corrélation* ρ de deux v.a. X et Y de variance non nulle est défini par :

$$\rho \triangleq \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]}}$$

1. $\rho = 0$
2. $\rho = 1/\sqrt{1.(1+0.04)} = 0.98$
3. $\rho = -1/\sqrt{1.(1+1)} = -0.71$
4. $\rho = 0$

Exemple : Covariance nulle n'implique pas indépendance

Soient les v.a. X et Y pouvant prendre 4 valeurs équiprobables comme indiqué ci-dessous :



$\mathrm{E}[X] = \mathrm{E}[Y] = 0$
$\Rightarrow \mathrm{cov}[X, Y] = \mathrm{E}[XY] = 1/4.(0 * 1 - 1 * 0 + 0 * (-1) + 1 * 0) = 0$
Or X et Y sont clairement dépendants (si X=1, alors Y=0)
$\lhd$

Exemple : pièces truquées

Soient n jets de pièces truquées, les jets sont indépendants.
Soit X le nombre de “pile” et Y le nombre de “face”.
On a toujours $x + y = n$ et donc (par linéarité) $\mathrm{E}[X] + \mathrm{E}[Y] = n$

$$\Rightarrow x - \mathrm{E}[X] = -(y - \mathrm{E}[Y]), \qquad \forall (x, y)$$

$$\Rightarrow \mathrm{cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = -\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = -\mathrm{var}[X]$$

$$\Rightarrow \rho(X, Y) = \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]}} = \frac{-\mathrm{var}[X]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[X]}} = -1$$

$\triangleleft$

Exemple : Somme de v.a. non indépendantes

Soient n v.a. $X_1, X_2, ..., X_n$:

$$\mathrm{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{var}[X_i] + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathrm{cov}[X_i, X_j]$$

Exemple : sinusoïdes en phase, ou avec des phases aléatoires.

$\triangleleft$

Fonction d'une variable

Par exemple : Puissance $W = \text{cste}\, V^2 = V.I$

Soit $Y = g(X)$:

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

Comment trouver $f_Y(y)$

Méthode générale : par la fonction de répartition

$$f_Y(x) = \frac{dF_Y(y)}{dy}(y)$$

Exemple : Calcul de la tension à partir de la puissance

Soit $W \sim \mathrm{Un}[1:10] \Rightarrow V$?
On a donc $V = \sqrt{W}$, avec $f_W(w) = 1/9, 1 \leq w \leq 10$.

$$F_V(v) = \mathbb{P}(V \leq v) = \mathbb{P}\left(\sqrt{W} < v\right) = \mathbb{P}\left(W < v^2\right) = v^2/9, \quad 1 \leq v \leq \sqrt{10}$$

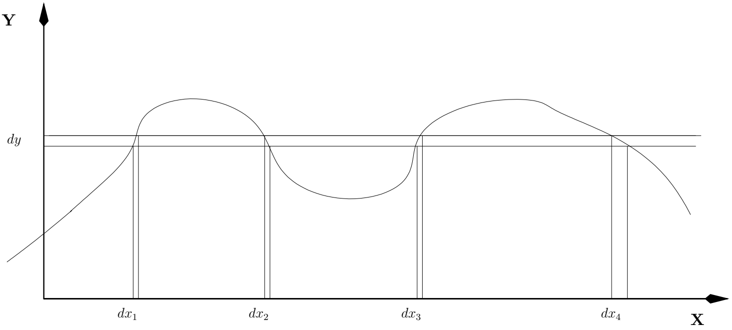
$$f_V(v) = \frac{dF_V(dv)}{dv}(v) = \frac{dv^2/9}{dv} = 2\frac{v}{9}, \quad 1 \leq v \leq \sqrt{10}$$

On vérifie que $\displaystyle\int_{-\infty}^{\infty} f_V(v)\, dv = 1$

$\triangleleft$

Cas unidimensionnel

Soit une v.a. $\mathbf{Y}$, telle que $\mathbf{Y} = g(\mathbf{X})$. Fonction a priori non bijective



Tronçons i où la fonction est bijective ($\mathbf{Y} = g_i(\mathbf{X}), i = 1, ..n$) : on peut écrire $\mathbf{X} = g_i^{-1}(\mathbf{Y})$ où $g_i^{-1}(.)$ est la fonction inverse de $g_i(.)$.
On a $\mathbb{P}(y < Y \leq y + dy) = f_Y(y)dy = f_Y(y)|dy| = \sum_{x_i | y = g(x_i)} f_X(x_i)|dx_i|$: :

$$f_Y(y) = f_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + \cdots + f_X(g_n^{-1}(y)) \left| \frac{dg_n^{-1}(y)}{dy} \right| .$$

Exemple : Changement de variable $\mathbf{Y} = \mathbf{X}^2$

Soit $\mathbf{Y} = \mathbf{X}^2$ sur [-1,1]. Dans ce cas, on a deux tronçons :

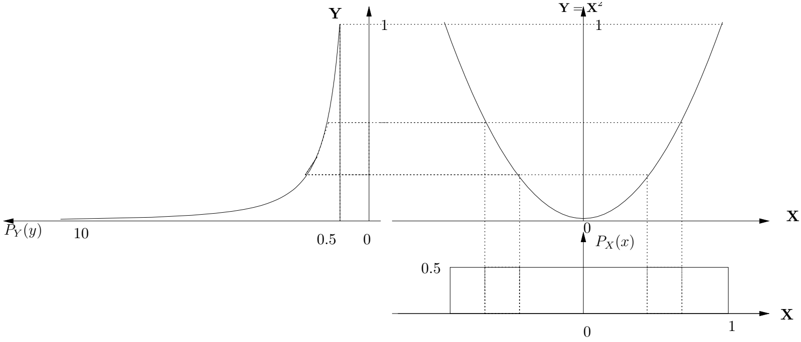
1. sur $x \in [-1, 0]$: $\mathbf{X} = -\sqrt{\mathbf{Y}}$ $(g_1^{-1}(.) = -\sqrt{(.)})$
2. sur $x \in [0, 1]$: $\mathbf{X} = \sqrt{\mathbf{Y}}$ $(g_2^{-1}(.) = \sqrt{(.)})$

On obtient alors

$$\begin{aligned} f_Y(y) &= f_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + f_X(g_2^{-1}(y)) \left| \frac{dg_2^{-1}(y)}{dy} \right| \\ &= f_X(\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} \\ &= \frac{1}{2} \frac{1}{\sqrt{y}}. \end{aligned}$$

$\triangleleft$

La figure ci-dessous illustre ce résultat.



$$\begin{aligned}
\mathbf{Y}_1 &= g_1(\mathbf{X}_1, \ldots, \mathbf{X}_n) \\
&\vdots \\
\mathbf{Y}_n &= g_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)
\end{aligned}$$

Soient $g_i(.)$ continues et différentiables

$$f_Y(y_1, \ldots, y_n) = f_X(x_1, \ldots, x_n) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|,$$

on notera que dans ce cas-ci, il est plus compliqué d'écrire l'expression en fonction de g_i^{-1}, mais on aurait plutôt des fonctions de type $\mathbf{X}_i = h_i(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$.

Précision de fabrication en micro-électronique

Dans le processus de fabrication de circuits intégrés, une des parties cruciales est la précision de la lithographie. On peut quantifier cette précision comme étant la déviation en coordonnées horizontales et verticales (x et y) par rapport à l'endroit à graver.
Dans le cas de technologies “70 nm”, on peut considérer que les déviations en x et y sont des variables aléatoires indépendantes qui suivent des lois gaussiennes de moyenne nulle et de variance $\sigma^2 = 0.2\text{nm}^2$. La densité de probabilité conjointe des déviations $(\mathbf{X}, \mathbf{Y})$ est donnée par :

$$f_{XY}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

La caractérisation de la précision en x et en y ne répond pas à la question suivante : “quelle est la loi de probabilité de la distance entre le point désiré et le point obtenu par la lithographie ?” Pour obtenir cette loi, on passe des coordonnées cartésiennes aux coordonnées polaires selon la transformation :

$$(\mathbf{X}, \mathbf{Y}) = (\mathbf{R}\cos(\Theta), \mathbf{R}\sin(\Theta)).$$

Le jacobien de la transformation vaut r :

$$\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| = \begin{vmatrix} \cos(\theta) & \sin(\theta) \\ -r\sin(\theta) & r\cos(\theta) \end{vmatrix} = r,$$

et donc

$$\begin{aligned} f_{\mathbf{R},\Theta}(r,\theta) &= \frac{r}{2\pi\sigma^2} e^{-[(r\cos(\theta))^2+(r\sin(\theta))^2]/2\sigma^2} \\ &= \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} \end{aligned}$$

Pour trouver les marginales, il faut intégrer sur l'autre variable aléatoire, soit :

$$\begin{aligned} f_\Theta(\theta) &= \int_{r=0}^{\infty} f_{\mathbf{R},\Theta}\, dr \\ &= \int_{r=0}^{\infty} \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}\, dr \\ &= \frac{1}{2\pi} \end{aligned}$$

L'angle est donc uniformément distribué sur $[0, 2\pi]$. D'autre part, on en déduit que

$$f_\mathbf{R}(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}$$

On en déduit également que les variables aléatoires **R** et Θ sont indépendantes (la loi conjointe est donnée par le produit des lois marginales). La densité de probabilité suivie par la distance **R** est celle d'une variable dite de Rayleigh.

Somme de variables aléatoires indépendantes

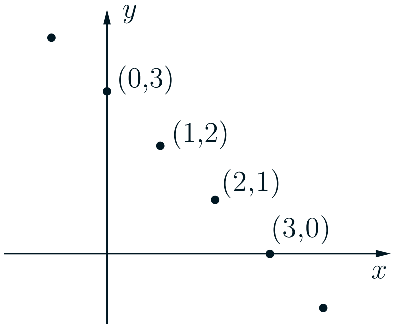
Cas Discret

Soient X et Y, deux v.a. indépendantes de masse $p_X(x)$ et $p_Y(y)$, à valeurs entières.
Soit $W = X + Y$, alors, pour tout entier w :

$$\begin{aligned} p_W(w) &= \mathbb{P}(X + Y = w) \\ &= \sum_{(x,y):x+y=w} \mathbb{P}(X = x \text{ et } Y = y) \\ &= \sum_x \mathbb{P}(X = x \text{ et } Y = w - x) \\ &= \sum_x p_X(x)\, p_Y(w - x) \\ &= p_X(x) * p_Y(y) \end{aligned}$$

où $p_X(x) * p_Y(y)$ est la convolution de $p_X(x)$ et de $p_Y(y)$

La masse $p_W(3)$ est la probabilité que $X + Y = 3$: c'est la probabilité de toutes les paires (x, y) telles que $x + y = 3$



et $p_{X,Y}(x, 3 - x) = p_X(x)\, p_Y(3 - x)$

Trois Tireurs de Penalty

Somme de v.a. continues

Soient deux v.a. X et Y continues et $W = X + Y$. Pour déterminer $f_W(w)$, nous allons dériver $F_W(w)$.

$$\begin{aligned}
F_W(w) &= \mathbb{P}(W \leq w) \\
&= \mathbb{P}(X + Y \leq w) \\
&= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{w-x} f_X(x) f_Y(y)\, dydx \\
&= \int_{x=-\infty}^{\infty} f_X(x) \left[\int_{y=-\infty}^{w-x} f_Y(y)\, dy \right] dx \\
&= \int_{x=-\infty}^{\infty} f_X(x) F_Y(w-x)\, dx
\end{aligned}$$

On en déduit :

$$
\begin{aligned}
f_W(w) &= \frac{F_W(w)}{dw} \\
&= \frac{d}{dw}\int_{x=-\infty}^{\infty} f_X(x)\,F_Y(w-x)\,dx \\
&= \int_{x=-\infty}^{\infty} f_X(x)\,\frac{dF_Y(w-x)}{dw}dx \\
&= \int_{x=-\infty}^{\infty} f_X(x)\,f_Y(w-x)\,dx
\end{aligned}
$$

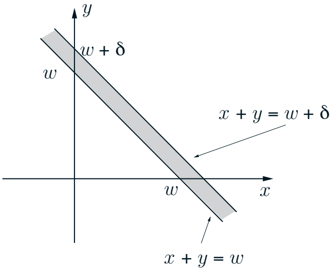


Illustration de la convolution

$$\mathbb{P}(w \leq X + Y \leq w + \delta) = f_W(w)\,.\delta$$

$$\begin{aligned} f_W(w)\,.\delta &= w \int_{x=-\infty}^{\infty} f_X(x) \int_{y=-w-x}^{w-x+\delta} f_Y(y)\, dydx \\ &\simeq \int_{x=-\infty}^{\infty} f_X(x)\, f_Y(w-x)\, \delta dx \end{aligned}$$

Convolution avec l'applet Java

http ://pages.jh.edu/ signals/convolve/

Rappels sur la covariance et la correlation

Slides 107-114

Estimateur aux moindres carrés

Soit Y, une v.a. qui est une “mesure” de X.
On cherche “c” qui minimise l'erreur quadratique $(X - c)^2$

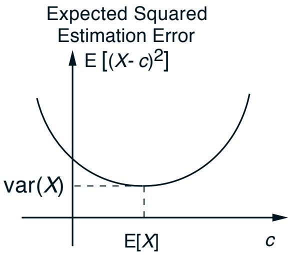
En l'absence d'Y, l'espérance de $\mathrm{E}[X]$ est le meilleur estimateur

Soit $\mu_X = \mathrm{E}[X]$

$$\begin{aligned}
\mathrm{E}\left[(X-c)^2\right] &= \mathrm{E}\left[X - \mu_X + \mu_X - c)^2\right] \\
&= \mathrm{E}\left[(X-\mu_X)^2\right] + 2\mathrm{E}[(X-\mu_X)(\mu_X - c)] + \mathrm{E}\left[(\mu_X - c)^2\right] \\
&= \mathrm{E}\left[(X-\mu_X)^2\right] + 2\mathrm{E}[(X-\mu_X)]\,(\mu_X - c) + (\mu_X - c)^2 \\
&= \mathrm{E}\left[(X-\mu_X)^2\right] + (\mu_X - c)^2
\end{aligned}$$

D'où $c = \mu_X$.

L'erreur quadratique moyenne $\mathrm{E}\left[(X-c)^2\right]$ en fonction de c est minimale en $c = \mathrm{E}[X]$.



Ò

Estimateur “Least Squares” sur base de mesure

Soit Y une v.a. de “mesure” de X, et y une réalisation de cette mesure (une vraie mesure).
Le nouvel univers est “conditionné” sur $Y = y$, donc :

$$c = \mathrm{E}[X|Y = y]$$

Exemple : Mesure bruitée : énoncé

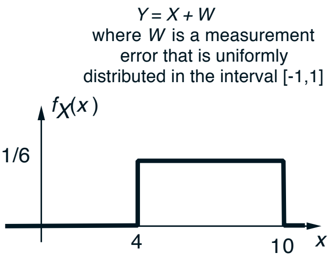
- X uniformément répartie sur [4,10] : $f_X(x) = \frac{1}{6}\mathbf{1}_{[4,10]}(x)$
- Bruit W uniformément réparti sur [-1,1] : $f_W(w) = \frac{1}{2}\mathbf{1}_{[-1,1]}(w)$
- Mesure $Y = X + W$: $f_{Y|X=x}(Y|x) = \frac{1}{2}\mathbf{1}_{[x-1,x+1]}(y)$

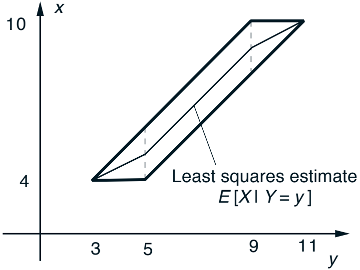
$$f_{X,Y}(x,y) = f_X(x)\, f_{Y|X=x}(Y|x) = \frac{1}{2}\cdot\frac{1}{6}\mathbf{1}_{[4,10],y\in[x-1,x+1]}(x,y)$$

- $\mathrm{E}[X|Y = y]$

◁

Exemple d'un estimateur aux moindres carrés (non-lineaire)





Estimateur aux moindres carrés

Soient deux v.a. X et sa "mesure" Y, l'estimateur qui minimise $\mathrm{E}\left[(X - g(Y))^2\right]$ est $g(Y) = \mathrm{E}[X|Y]$.

Propriétés de l'estimateur aux moindres carrés (non-lineaire)

Soit l'estimateur $\hat{X} = \mathrm{E}[X|Y]$ et l'erreur d'estimation $\tilde{X} = X - \hat{X}$

Par les espérances itérées : $\mathrm{E}\left[\tilde{X}\right] = \mathrm{E}[X - EX|Y] = \mathrm{E}[X] - \mathrm{E}[X] = 0$

$\mathrm{E}\left[\tilde{X}\right] = 0$ est toujours valide, même conditionné sur Y, car

$\mathrm{E}\left[\tilde{X}|Y\right] = \mathrm{E}\left[X - \hat{X}|Y\right] = \mathrm{E}[X|Y] - \mathrm{E}\left[\hat{X}|Y\right] = \hat{X} - \hat{X} = 0$

De même : $\mathrm{E}\left[(\hat{X} - \mathrm{E}[X])\tilde{X}|Y\right] = (\hat{X} - \mathrm{E}[X])\mathrm{E}\left[\tilde{X}|Y\right] = 0$

Par les espérances itérées :

$$\mathrm{E}\left[(\hat{X} - \mathrm{E}[X])\tilde{X}\right] = 0$$

Variances de la v.a, de son estimée et de son erreur

Soit X, on a $X - \mathrm{E}[X] = \hat{X} - \mathrm{E}[X] + \tilde{X}$:

$$\begin{aligned}
\mathrm{var}[X] &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[(\hat{X} - \mathrm{E}[X] + \tilde{X})^2\right] \\
&= \mathrm{E}\left[(\hat{X} - \mathrm{E}[X])^2\right] + \mathrm{E}\left[\tilde{X}^2\right] + 2\,\underbrace{\mathrm{E}\left[(\hat{X} - \mathrm{E}[X])\tilde{X}\right]}_{=0} \\
&= \mathrm{var}\left[\hat{X}\right] + \mathrm{var}\left[\tilde{X}\right]
\end{aligned}$$

La variance de X est la somme des variances de l'estimée et de l'erreur

$$\mathrm{var}[X] = \mathrm{var}\left[\hat{X}\right] + \mathrm{var}\left[\tilde{X}\right]$$

Estimateur basé sur plusieurs mesures

Si on a plusieurs mesures $Y_1, Y_2, \ldots, Y_n$ alors l'estimateur est

$$\mathrm{E}[X | Y_1, \ldots, Y_n]$$

Compliqué : nécessite $p_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$

Estimateur linéaire

$$\hat{X} = a_1 Y_1 + \cdots + a_n Y_n + b$$

Si une seule donnée :

$$\hat{X} = aY + b$$

Estimateur linéaire simple ($\hat{X} = aY + b$)

Solution :

- b : estimateur de $X - aY$. Donc $b = \mathrm{E}[X - aY] = \mathrm{E}[X] - a\mathrm{E}[Y]$
- a : minimiser $\mathrm{E}\left[(X - aY - \mathrm{E}[X] + a\mathrm{E}[Y])^2\right]$
  $\mathrm{E}\left[(X - aY - \mathrm{E}[X] + a\mathrm{E}[Y])^2\right] =$

$$= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] + a^2\mathrm{E}\left[(Y - \mathrm{E}[Y])^2\right] - 2a\mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$
$$= \sigma_X^2 + a^2\sigma_Y^2 - 2.a\mathrm{cov}[X, Y]$$

  Minimisé pour (dérivée = 0) :

$$a = \frac{\mathrm{cov}[X, Y]}{\sigma_Y^2} = \frac{\rho\sigma_X\sigma_Y}{\sigma_Y^2} = \rho\frac{\sigma_X}{\sigma_Y}$$

  where $\rho = \frac{\mathrm{cov}[X,Y]}{\sigma_X\sigma_Y}$
- Erreur :

$$\sigma_X^2 + a^2\sigma_Y^2 - 2.a\mathrm{cov}[X, Y] \quad = \sigma_X^2 + \rho^2\frac{\sigma_X^2}{\sigma_Y^2}\sigma_Y^2 - 2\rho\frac{\sigma_X}{\sigma_Y}\rho\sigma_X\sigma_Y$$
$$= (1 - \rho^2)\sigma_X^2$$

Estimateur linéaire aux moindres carrés

Résultats principaux

L'estimateur linéaire aux moindres carrés de X basé sur Y vaut :

$$\hat{X} = \mathrm{E}[X] + \frac{\mathrm{cov}[X, Y]}{\sigma_Y^2}(Y - \mathrm{E}[Y])$$

L'erreur quadratique moyenne (minimale) vaut :

$$(1 - \rho^2)\mathrm{var}[X]$$

Qu'est-ce qu'une statistique

Definition : Une *statistique*

est une quantité calculée à partir d'un certain nombre d'observations.

exemples : la moyenne, la médiane, les quantiles, ...

Signification des termes dépendante du contexte

- En statistique : moyenne basée sur des observations
- En probabilité : moyenne basée sur un modèle probabiliste

Definition : *Un individu*

est l'unité statistique de base.

exemple du sondage : un "sondé"

Definition : *Une population*

est l'ensemble des individus que l'on souhaite étudier. Cette population peut être infinie ou finie.

exemple de l'élection présidentielle : les "citoyens"

Definition : *Un échantillon*

(d'une population) est un sous-ensemble de la population.

exemple du sondage : 1034 "sondé"

Definition : *Un caractère*

(à étudier) est une variable statistique que l'on souhaite étudier.

exemple qualitatif "politique" : gauche / droite
exemple quantitiatif : Puissance moteur des voitures en Chine

Definition : ***Les fréquences***

liées à un caractère d'un échantillon sont le nombre d'individus présentant le caractère étudié (on parlera de *fréquences absolues* ou *d'effectifs*). On parlera également de *fréquences relatives* ou *proportions* si on s'intéresse à la proportion d'individus de l'échantillon qui présentent le caractère étudié.

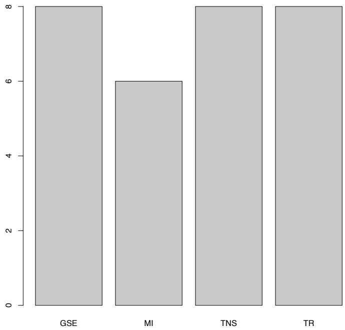
Paramètres statistiques d'un échantillon

- Mesures de tendance centrale (position)
  - Moyenne : $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Médiane : partage les valeurs en deux parties
  - Quantiles : partagent les valeurs en k parties
  - Quartiles ($k = 4$) : Q_1, Q_2 (médiane), Q_3
  - Mode(s) : la (les) valeur(s) avec la plus grande fréquence
- Mesures de dispersion
  - Étendue : $x_{(n)} - x_{(1)}$
  - Intervalle interquartile (IQR) : $Q_3 - Q_1$
  - Variance de l'échantillon :
    $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{n\sum_{i=1}^{n}(x_i)^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$
  - Écart-type de l'échantillon : s
  - Écart absolu médian par rapport à la médiane
  - Coefficient de variation : $s/\overline{x}$

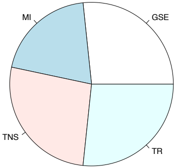
L'exemple suivant étudie la population des étudiants XXX. On a un échantillon de 30 étudiants et on s'intéresse aux caractères suivants :

- option (qualitatif)
- moyenne tp (quantitatif)
- contrôle final (quantitatif)

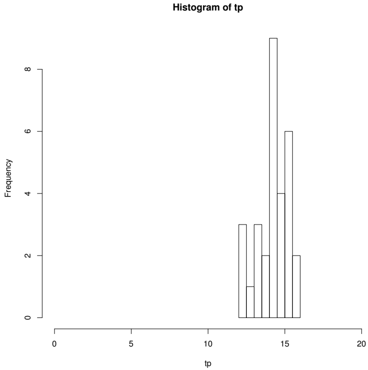
Caractère : option

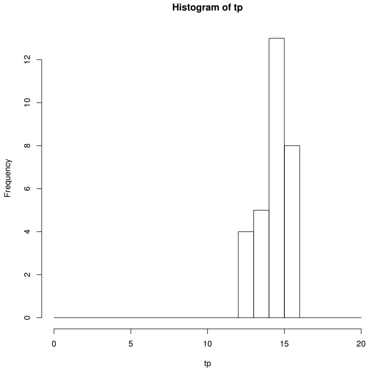


Caractère : option

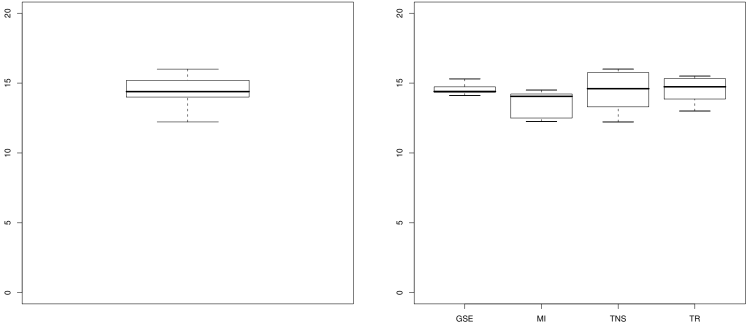


Caractère : moyenne tp (classes différentes)

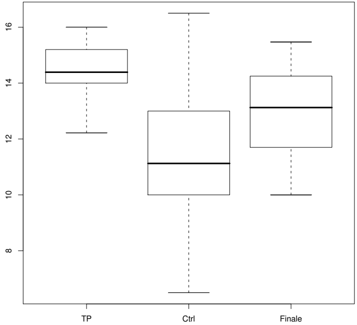




Caractère : moyenne tp

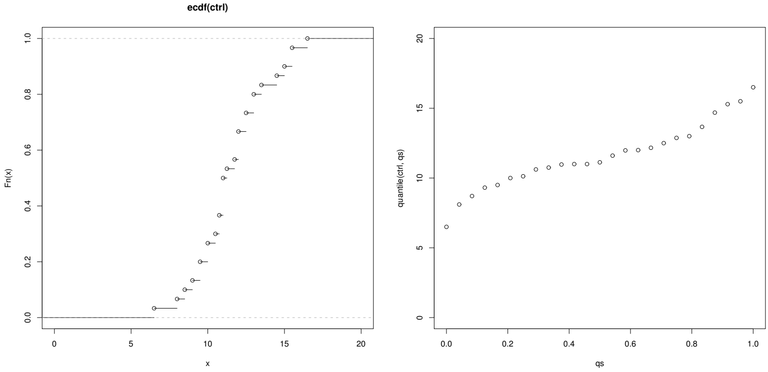


Caractère : moyenne tp / contrôle final / note finale (50-50)



Fréquence relative cumulée / quantiles

Caractère : contrôle final



$$x \xrightarrow{F_X} F_X(x) = P(X \leq x) = p \qquad p \xrightarrow{Q_X = F_X^{-1}} Q_X(p) = x$$

Objectif de la Statistique Inférentielle

Definition : Objectif principal :

obtenir, à partir de mesures sur une *partie* de la population (échantillon), des informations (de caractère *probabiliste*) sur la *totalité* de celle-ci.

Inférence :

$\boxed{\text{échantillon}} \quad \Rightarrow \quad \boxed{\text{population}}$

On prélève un **échantillon** de la population, on en *déduit* (ou encore on *infère*) des caractéristiques de la **population**.

L'échantillonnage : une expérience aléatoire

Definition : em Échantillonnage

choisir *au hasard* n individus de la population
Il y a deux types d'échantillonnage :

1. *avec* remplacement de l'individu choisi, ce qui mène à un traitement théorique plus simple - population infinie ;
2. *sans* remplacement : échantillonnage exhaustif, ce qui est une procédure naturelle ou obligatoire (contrôle destructif).

L'échantillonnage : une expérience aléatoire

L'échantillonnage est une **expérience aléatoire**, : choisir *au hasard* un individu (ou un "petit nombre" d'individus) de la population.
Chaque individu doit avoir la même probabilité d'être choisi.

$$\begin{array}{ccccc} \text{population} & \xrightarrow{\text{éch.}} & \text{individu} & \xrightarrow{\text{caract.}} & \text{valeur} \\ \Omega & \xrightarrow{\text{éch.}} & \omega & \xrightarrow{\text{caract.}} & x \end{array}$$

Enfin, à partir de l'échantillon, on étudie la variable aléatoire X associée au caractère étudié et on peut, dans le meilleur des cas, déterminer la densité de probabilité $f_X(x)$ (ou sa masse de probabilité $p_X(x)$ s'il s'agit d'une variable aléatoire discrète).

Relation entre la statistique et la loi de probabilité

Soit X une v.a. sur la population $(f_X(x))$
l'échantillonnage correspond à la répétition de n expériences aléatoires identiques, : n v.a. indépendantes X_i $(i = 1, \ldots, n)$ ayant la (même) densité de probabilité $f_X(x)$.

l'échantillonnage correspond à la répétition de *n* expériences aléatoires identiques

X_i *i.i.d.* (indépendantes et identiquement distribuées).

$$f_{X_1}(x) = f_{X_2}(x) = \ldots = f_{X_n}(x) = f_X(x) \text{ et}$$

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = f_X(x_1) f_X(x_2) \ldots f_X(x_n).$$

Echantillonnage *avec* remplacement et *même probabilité* de choisir chaque individu.

Definition : Une statistique est

une fonction des variables aléatoires X_i $(i = 1, \ldots, n)$ obtenue à partir d'un échantillon.

Théorie d'échantillonnage : Population $\longrightarrow$ Échantillon
Statistique inférentielle : Échantillon $\longrightarrow$ Population

Échantillon		Population $p_X(x)$
v.a.	valeur	paramètre
une population		
$\overline{X}$ S^2	$m = \overline{x}$ s^2	$\mu_X = \mathrm{E}[X]$ $\sigma_X^2 = \mathrm{var}[X]$
$\hat{P}$	$\hat{p}$	π
deux populations		
$\overline{X}_2 - \overline{X}_1$ S_2^2/S_1^2	$m_2 - m_1 = \overline{x}_2 - \overline{x}_1$ $(s_2/s_1)^2$	$\mu_2 - \mu_1$ $(\sigma_2/\sigma_1)^2$
$\hat{P}_2 - \hat{P}_1$	$\hat{p}_2 - \hat{p}_1$	$\pi_2 - \pi_1$

- Estimer les paramètres de la population
- Calculer des intervalles de confiance
- Formuler des hypothèses et les tester

Théorème limite central

Soit n v.a. (i.i.d) :

- $X_1, X_2, \ldots, X_n$: série de v.a. indépendantes
- $f_{X_1}(x) = \ldots = f_{X_n}(x) = f_X(x)$ (même distribution)
- $\mathrm{E}[X_1] = \ldots = \mathrm{E}[X_n] = \mu_X\ ,\ \sigma_{X_1} = \ldots = \sigma_{X_n} = \sigma_X$
- $$S_n = X_1 + X_2 + \ldots + X_n\ ,\ \mathrm{E}[S_n] = n\mu_X\ ,\ \sigma^2_{S_n} \overset{\text{ind}}{=} n\sigma^2_X$$

$$\boxed{Z_n = \frac{S_n - \mu_{S_n}}{\sigma_{S_n}}} = \frac{X_1 + X_2 + \ldots + X_n - n\mu_X}{\sqrt{n}\sigma_X}\ ,\ \boxed{\mathrm{E}[Z_n] = 0}\ ,\ \boxed{\sigma^2_{Z_n} = 1}$$

Théorème Limite Central (2)

Théorème Limite Central

$$\lim_{n \longrightarrow \infty} P(\{Z_n \leq z\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u$$

Théorème Limite Central

$$\boxed{n \to \infty \ : \ Z_n \to N(0,1)} \ , \ S_n \to N(n\mu_X, n\sigma_X^2) \ , \ \frac{S_n}{n} \to N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

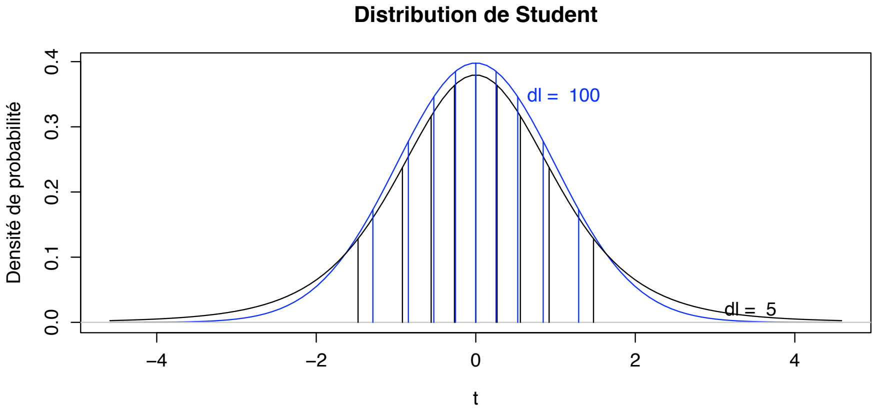
http ://onlinestatbook.com/simulations/CLT/clt.html

- Échantillon aléatoire de taille n ; moyenne $\overline{X}$
- Population normale $N(\mu, \sigma^2)$
  - $\overline{X}$: normale (combinaison linéaire de v.a. normales)
  - $\mu_{\overline{X}} = \mu$
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ (σ connu)
- Population non normale (σ connu)
  - $n > 30$: $\overline{X} = N\left(\mu, \frac{\sigma^2}{n}\right)$ (tlc)
  - $n < 30$: $\overline{X} = N\left(\mu, \frac{\sigma^2}{n}\right)$ si $p_X(x)$ presque normale
- Presque toujours : $\boxed{\overline{X} = N(\mu, \sigma/\sqrt{n})}$
  - $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$
  - $\mathbb{P}(Z > z_\alpha) = \alpha$ (définition de z_α « valeur critique »)
  - $\mathbb{P}(Z < -z_\alpha) = \alpha$ (symétrie de la normale)

- $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$
- $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{(\overline{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{V/(n-1)}} = \frac{Z}{\sqrt{V/\nu}}$
- $V = \frac{(n-1)S^2}{\sigma^2}$: loi du χ^2 à $\nu = n-1$ d.l.
- Condition : population normale
- Z, V indépendantes
- $\boxed{T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \text{ : loi de Student à } \nu = n-1 \text{ d.l.}}$
- $\mathrm{E}[T] = 0$
- $\sigma_T^2 = \frac{\nu}{\nu-2} > 1$ (non définie pour $\nu \leq 2$)
- $P(T > t_\alpha) = \alpha$ (définition de t_α, valeur critique)
- $P(T < -t_\alpha) = \alpha$ (symétrie de la loi t)
- $n \geq 30$: $s \rightarrow \sigma$ donc $T \rightarrow Z$
- "Student" : W.S. Gosset, 1908

La distribution de Student

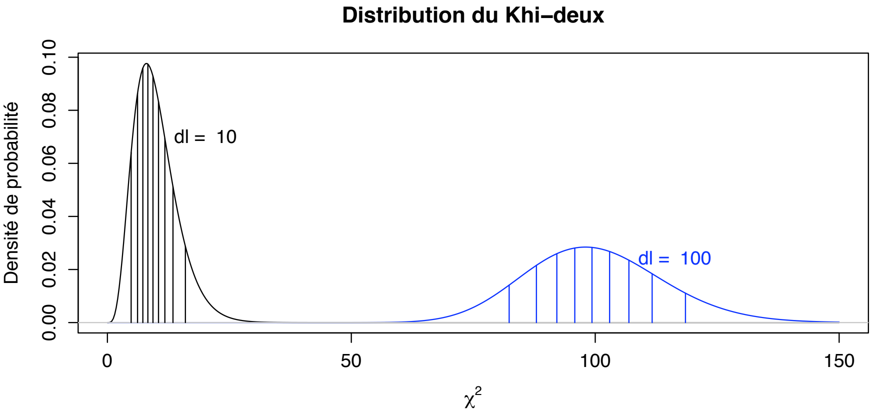
La distribution de Student



$$\mathrm{E}[T] = 0 \quad , \quad \sigma_T^2 = \frac{\nu}{\nu-2} > 1 \text{ (non définie pour } \nu \leq 2)$$

- Échantillon aléatoire de taille n ; variance S^2
  - Condition : population normale $N(\mu, \sigma^2)$
  - $\boxed{X^2 = \dfrac{(n-1)S^2}{\sigma^2}} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$
  - $\boxed{X^2 \text{ : v.a. loi du } \chi^2 \text{ à } \nu = n-1 \text{ degrés de liberté (d.l.)}}$
  - $X^2 > 0$
  - $\mathrm{E}\left[X^2\right] = n - 1 \longrightarrow \boxed{\mathrm{E}\left[S^2\right] = \sigma^2}$
  - $\sigma^2_{X^2} = 2(n-1) \longrightarrow \sigma^2_{S^2} = 2\sigma^4/(n-1)$
  - $P(X^2 > \chi^2_\alpha(\nu)) = \alpha$ (définition de $\chi^2_\alpha(\nu)$, valeur critique)

Distribution du χ^2



$$\mathrm{E}\left[X^2\right] = n - 1 \quad , \quad \sigma^2_{X^2} = 2(n-1)$$

1

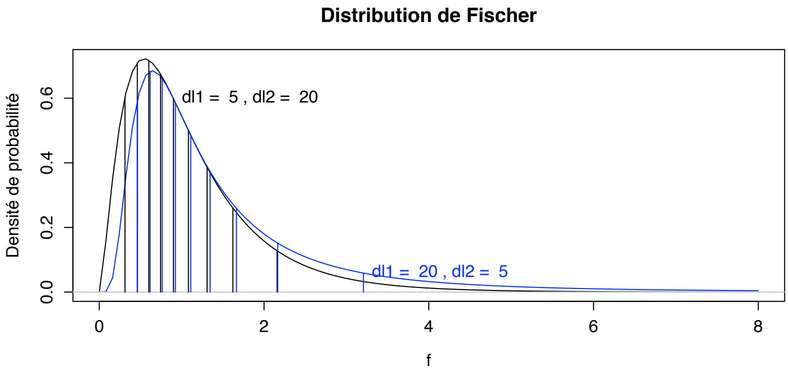
1. La moyenne est égale à $n-1$. À gauche elle se trouve à 9. Comme la courbe n'est pas

- Population
  - π : proportion d'individus possédant un caractère qualitatif ($\pi \neq 3.14$!)
- Échantillon aléatoire de taille n
  - n v.a. X_i ; $x_i \in \{0, 1\}$: Bernoulli indépendantes, de paramètre π
  - $\sum_{i=1}^{n} X_i$: nombre d'individus possédant le caractère (fréquence)
  - $\boxed{\hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ : proportion d'individus (fréquence relative)}}$
- Conditions :
  - $n > 30$ (grand échantillon : théorème limite central)
  - $n\hat{p} \geq 5$ (fréquence de présence du caractère)
  - $n(1 - \hat{p}) = n - n\hat{p} \geq 5$ (fréquence d'absence du caractère)
  - ni $\hat{p} \approx 0$, ni $\hat{p} \approx 1$
- Distribution :
  - $\mu_{\hat{P}} = (n\mu_X)/n = \mu_X = \pi \quad , \quad \sigma^2_{\hat{P}} \overset{\text{ind}}{=} (n\sigma^2_X)/n^2 = \pi(1-\pi)/n$
  - $\boxed{\hat{P} \text{ : normale } N\left(\pi, \frac{\pi(1-\pi)}{n}\right)} \to Z$: normale $N(0, 1)$

- Conditions : σ_1, σ_2 connus et
  - populations normales $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ ou
  - $n_1 > 30$ **et** $n_2 > 30$, ou
  - populations ≪ presque ≫ normales
- Échantillons aléatoires et indépendants de tailles n_1, n_2 ; moyennes $\overline{X}_1, \overline{X}_2$
  - $\boxed{\overline{X}_1 - \overline{X}_2}$: normale
  - $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2$
  - $\sigma^2_{\overline{X}_1 - \overline{X}_2} \overset{\text{ind}}{=} \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
- D'autres cas à examiner ultérieurement...

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Provenant de populations normales de variances σ_1^2, σ_2^2
- Variances des échantillons : S_1^2, S_2^2
- $\boxed{F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}} = \frac{V_1/\nu_1}{V_2/\nu_2}$
- $V_i = \frac{(n_i-1)S_i^2}{\sigma_i^2}$: v.a. indépendantes, loi du χ^2 à $\nu_i = n_i - 1$ d.l.
- $\boxed{F \text{ : loi de Fisher (1924) - Snedecor (1934) avec } \nu_1 \text{ et } \nu_2 \text{ d.l.}}$
- $F \geq 0$
- $\mathrm{E}[F] = \frac{\nu_2}{\nu_2-2}$ $(\nu_2 > 2)$
- $\sigma_F^2 = \frac{\nu_2^2(2\nu_1+2\nu_2-4)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ $(\nu_2 > 4)$
- $P(F > f_\alpha(\nu_1, \nu_2)) = \alpha$ (définition de $f_\alpha(\nu_1, \nu_2)$, v.c.)
- $\boxed{f_\alpha(\nu_1, \nu_2) = \frac{1}{f_{1-\alpha}(\nu_2, \nu_1)}}$ (propriété de la loi F)

Distribution de Fisher



$$f_\alpha(\nu_1, \nu_2) = 1/f_{1-\alpha}(\nu_2, \nu_1)$$

Definition : **Un estimateur ponctue**

est une statistique qui donne une valeur (unique) estimée de la grandeur recherchée.

- Paramètre à estimer : θ
- Estimateur : v.a. $\hat{\Theta}$
- Estimateur non biaisé : $\mathrm{E}[\hat{\Theta}] = \theta$
- Biais = $\mathrm{E}[\hat{\Theta}] - \theta$
- Estimateur efficace : sans biais ; de faible variance
- Estimateur efficace : minimise l'erreur quadratique moyenne $\mathrm{E}\left[(\hat{\Theta} - \theta)^2\right] = \sigma^2_{\hat{\Theta}} + (\text{biais})^2$
- Estimateur convergent : $n \to \infty \quad : \quad \mathrm{E}[\hat{\Theta}] = \theta$ et $\mathrm{var}[\hat{\Theta}] = 0$

Definition : **Estimateur par intervalle de confiance**

: une statistique qui donne un intervalle dans lequel la grandeur recherchée se trouve, avec un indice de confiance. Cet indice de confiance donne le niveau de confiance avec lequel on peut “croire” que la grandeur recherchée se trouve à l’intérieur de cet intervalle.

- v.a. $\hat{\Theta}_L$, $\hat{\Theta}_H$: estimateurs ponctuels
- $P(\hat{\Theta}_L < \theta < \hat{\Theta}_H) = 1 - \alpha$
- $\boxed{\hat{\theta}_L < \theta < \hat{\theta}_H}$: intervalle de confiance
- $1 - \alpha$: niveau de confiance

Propriétés et intervalle de confiance : Moyenne

- Variance σ^2 connue
- $\overline{X}$: normale $N(\mu, \sigma^2/n)$
- $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$: normale $N(0,1)$
- $\overline{X}$ estimateur non biaisé et convergent de μ
- $P(Z > z_{\alpha/2}) = \alpha/2$ (définition de $z_{\alpha/2}$)
- $P(Z < -z_{\alpha/2}) = \alpha/2$ (symétrie de la normale)
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- $P(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$
- $P(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
- $P(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
- $\hat{\Theta}_L = \overline{X} - z_{\alpha/2}\sigma_{\overline{X}}$, $\hat{\Theta}_H = \overline{X} + z_{\alpha/2}\sigma_{\overline{X}}$
- $1 - \alpha = 0.95$, $z_{\alpha/2} = 1.96$
- $1 - \alpha = 0.99$, $z_{\alpha/2} = 2.56$

Taille de l'échantillon

- $P(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
- $P(|\overline{X} - \mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
- $e = |\overline{X} - \mu|$: erreur
- $e_{\max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$: marge d'erreur à $1 - \alpha$
- $n_{\min} = \left(\frac{z_{\alpha/2}\sigma}{e_{\max}}\right)^2$: taille d'échantillon minimale
- $\overline{X} - e_{\max} < \mu < \overline{X} + e_{\max}$ à $1 - \alpha$
- Cas particulier : échantillonnage d'une population finie, sans remplacement
  - Population de taille N
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \overset{N \gg 1}{\approx} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}$
  - $n_{\min} = \frac{Nz^2_{\alpha/2}\sigma^2}{Ne^2_{\max} + z^2_{\alpha/2}\sigma^2}$: taille d'échantillon minimale

Cas où la variance est inconnue

- Variance σ^2 inconnue
- Population normale
- $T = (\overline{X} - \mu)/(S/\sqrt{n})$: Student à $n-1$ d.l.
- $P(T > t_{\alpha/2}) = \alpha/2$ (définition de $t_{\alpha/2}$)
- $P(T < -t_{\alpha/2}) = \alpha/2$ (symétrie de la loi t)
- $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$
- $P(-t_{\alpha/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2}) = 1 - \alpha$
- $P(-t_{\alpha/2}\frac{S}{\sqrt{n}} < \overline{X} - \mu < t_{\alpha/2}\frac{S}{\sqrt{n}}) = 1 - \alpha$
- $P(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}) = 1 - \alpha$
- $\hat{\Theta}_L = \overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}$, $\hat{\Theta}_H = \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$
- $1 - \alpha = 0.95$, $t_{\alpha/2} = 2.05$
- $1 - \alpha = 0.99$, $t_{\alpha/2} = 2.76$
- Rappel : $n \geq 30$, $T \to Z$
- T : petits échantillons !

Estimation de la variance (un échantillon)

- Condition : population normale $N(\mu, \sigma^2)$
- $X^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$
- X^2 : v.a. loi du χ^2 à $\nu = n - 1$ degrés de liberté (d.l.)
- $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$
- $P\left(\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}\right) = 1 - \alpha$
- $P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right) = 1 - \alpha$
- $P\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}}\right) = 1 - \alpha$
- Intervalle de confiance :
  $\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$ à un niveau de confiance de $(1 - \alpha)100\%$

Estimation de la proportion (= moyenne)

- Caractère quantitatif (rappel)
  - Moyenne : $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - $n > 30$, σ connu
  - $\overline{X} = N\left(\mu, \frac{\sigma^2}{n}\right)$
- Caractère qualitatif
  - Proportion : $\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - $n > 30$, $n\hat{p} \geq 5$, $n(1-\hat{p}) \geq 5$, ni $\hat{p} \approx 0$, ni $\hat{p} \approx 1$
  - $\hat{P} = N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$
- Les proportions (fréquences relatives) sont des moyennes !
- $\overline{X} \longrightarrow \hat{P}$: remplacer
  - $\mu \longrightarrow \pi$
  - $\sigma \longrightarrow \sqrt{\pi(1-\pi)}$

- Caractère quantitatif (rappel)
  - $P(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
  - Intervalle de confiance à un niveau de confiance de $(1-\alpha)100\%$:
    $\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
  - $n_{\text{min}} = \left(\frac{z_{\alpha/2}\sigma}{e_{\text{max}}}\right)^2$: taille d'échantillon minimale
- Caractère qualitatif
  - $P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} < \pi < \hat{P} + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1 - \alpha$
  - Intervalle de confiance à un niveau de confiance de $(1-\alpha)100\%$:
    $\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{\mathbf{p}}(1-\hat{\mathbf{p}})}{n}} < \pi < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{\mathbf{p}}(1-\hat{\mathbf{p}})}{n}}$
  - $n_{\text{min}} = \left(\frac{z_{\alpha/2}}{e_{\text{max}}}\right)^2 \hat{p}(1-\hat{p})$: taille d'échantillon minimale
    estimer $\hat{p}$ (1er échantillonage, $n \geq 30$) ou prendre $\hat{p} = 0.5$ (pire scénario)

Estimation du rapport des variances (deux échantillons)

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Provenant de populations normales de variances σ_1^2, σ_2^2
- Variances des échantillons : S_1^2, S_2^2
- $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{V_1/\nu_1}{V_2/\nu_2}$
- $V_i = \frac{(n_i-1)S_i^2}{\sigma_i^2}$: v.a. indépendantes, loi du χ^2 à $\nu_i = n_i - 1$ d.l.
- F : loi de Fisher - Snedecor avec ν_1 et ν_2 d.l.
- $P(f_{1-\alpha/2}(\nu_1, \nu_2) < F < f_{\alpha/2}(\nu_1, \nu_2)) = 1 - \alpha$
- $P\left(f_{1-\alpha/2}(\nu_1, \nu_2) < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2}(\nu_1, \nu_2)\right) = 1 - \alpha$
- $P\left(\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)}\right) = 1 - \alpha$
- $P\left(\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(\nu_2, \nu_1)\right) = 1 - \alpha$

Tests d'hypothèses

- Hypothèse : énoncé concernant les caractéristiques d'une population
- Hypothèse nulle : fixer un paramètre θ à une valeur particulière θ_0
  - $H_0 : \theta = \theta_0$
- Hypothèse alternative (trois choix possibles)
  - $H_1 : \theta \neq \theta_0$ (test bilatéral)
  - $H_1 : \theta < \theta_0$ (test unilatéral)
  - $H_1 : \theta > \theta_0$ (test unilatéral)
- Test : procédure suivie afin d'accepter/rejeter H_0
- Rejet ¿ Acceptation (non-rejet)
- En pratique : formuler H_0 comme l'opposé de ce qu'on veut démontrer !

Types et probabilités d'erreur

- Types d'erreur

décision \ *état du monde*	H_0 vraie	H_1 vraie
non-rejet de H_0	OK	Type II
rejet de H_0	Type I	OK

- $P(\text{Type I}) = P(\text{rejet de } H_0 | H_0 \text{ vraie}) = \alpha$
- $P(\text{Type II}) = P(\text{non-rejet de } H_0 | H_1 \text{ vraie}) = \beta$
- Probabilités d'erreur

décision \ *état du monde*	H_0 vraie	H_1 vraie
non-rejet de H_0	$1-\alpha$	β
rejet de H_0	α	$1-\beta$

- α : seuil de signification (calculé dans l'univers de H_0 , ok)
- $1-\beta$: puissance du test (calculée dans l'univers de H_1 , ? ? ?)
  - Préciser H_1 , ensuite calculer une valeur de β liée à cette H_1

Tests : la procédure à suivre

1. Formuler les hypothèses H_0 et H_1
2. Choisir le seuil de signification α (typiquement 1% ou 5%)
3. Déterminer la statistique utilisée ainsi que sa distribution
4. Définir la région critique (région de rejet de H_0)
5. Adopter une règle de décision (à partir des valeurs critiques)
6. Prélever un échantillon et faire les calculs
7. Décider

Test sur une moyenne

1. $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ (test bilatéral)
2. α à définir
3. Statistique à utiliser : $\overline{X}$; distribution :
   $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ si on connaît σ ou n grand (cas présenté dans la suite)
   $T = (\overline{X} - \mu)/(S/\sqrt{n})$ si on ne connaît pas σ et n petit (population normale)
4. $P(\text{non-rejet de } H_0 | H_0 \text{ vraie}) = 1 - \alpha$
   $P(\text{non-rejet de } H_0 | \mu = \mu_0) = 1 - \alpha$
   $P(z_{1-\alpha/2} < Z < z_{\alpha/2} | \mu = \mu_0) = 1 - \alpha$
   $P(-z_{\alpha/2} < Z < z_{\alpha/2} | \mu = \mu_0) = 1 - \alpha$
   $P(-z_{\alpha/2} < (\overline{X} - \mu)/(\sigma/\sqrt{n}) < z_{\alpha/2} | \mu = \mu_0) = 1 - \alpha$
   $P(-z_{\alpha/2} < (\overline{X} - \mu_0)/(\sigma/\sqrt{n}) < z_{\alpha/2}) = 1 - \alpha$
   région critique : $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n}) < -z_{\alpha/2}$ et
   $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n}) > z_{\alpha/2}$
5. Règle de décision :
   rejeter H_0 si $\overline{x} < \overline{x}_{c1} = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ ou $\overline{x} > \overline{x}_{c2} = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Test Unilatéral (moyenne)

1. $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$ (test unilatéral)
2. α à définir
3. Statistique à utiliser : $\overline{X}$; distribution :
   $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ si on connaît σ ou n grand (cas présenté dans la suite)
   $T = (\overline{X} - \mu)/(S/\sqrt{n})$ si on ne connaît pas σ et n petit (population normale)
4. $P(\text{non-rejet de } H_0 | H_0 \text{ vraie}) = 1 - \alpha$
   $P(\text{non-rejet de } H_0 | \mu = \mu_0) = 1 - \alpha$
   $P(Z < z_\alpha | \mu = \mu_0) = 1 - \alpha$
   $P((\overline{X} - \mu)/(\sigma/\sqrt{n}) < z_\alpha | \mu = \mu_0) = 1 - \alpha$
   $P((\overline{X} - \mu_0)/(\sigma/\sqrt{n}) < z_\alpha) = 1 - \alpha$
   région critique : $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n}) > z_\alpha$
5. Règle de décision :
   rejeter H_0 si $\overline{x} > \overline{x}_c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$

Taille de l'échantillon

- $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$ (test unilatéral)
- $\alpha = P(\text{rejet de } H_0 | H_0 \text{ vraie}) = P(\text{rejet de } H_0 | \mu = \mu_0)$
  $= P(Z > z_\alpha | \mu = \mu_0)$
  $= P((\overline{X} - \mu)/(\sigma/\sqrt{n}) > z_\alpha | \mu = \mu_0)$
  $= P((\overline{X} - \mu_0)/(\sigma/\sqrt{n}) > z_\alpha)$
- Règle de décision : rejeter H_0 si $\overline{x} > \overline{x}_c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$
- $\beta = P(\text{rejet de } H_1 | H_1 \text{ vraie}) = P(\text{non-rejet de } H_0 | H_1 \text{ vraie})$
  $= P(\overline{X} < \overline{x}_c | H_1 \text{ vraie})$
- Préciser $H_1 : \mu = \mu_0 + \delta$
- $\beta = P(\overline{X} < \overline{x}_c | \mu = \mu_0 + \delta) = P(Z < (\overline{x}_c - \mu)/(\sigma/\sqrt{n}) | \mu = \mu_0 + \delta)$
- $= P(Z < \frac{\overline{x}_c - \mu_0}{\sigma/\sqrt{n}} - \frac{\delta}{\sigma/\sqrt{n}})$
- $= P(Z < z_\alpha - \frac{\delta}{\sigma/\sqrt{n}})$
- $-z_\beta = z_\alpha - \frac{\delta}{\sigma/\sqrt{n}}$
- $n = (z_\alpha + z_\beta)^2 \frac{\sigma^2}{\delta^2}$

Test bilatéral (Variance)

1. $H_0 : \sigma = \sigma_0$, $H_1 : \sigma \neq \sigma_0$ (test bilatéral)
2. α à définir
3. Statistique à utiliser : S ; distribution :
   $X^2 = \frac{(n-1)S^2}{\sigma^2}$, v.a. loi du χ^2 à $\nu = n - 1$ degrés de liberté (population normale)
4. $P(\text{non-rejet de } H_0 | H_0 \text{ vraie}) = 1 - \alpha$
   $P(\text{non-rejet de } H_0 | \sigma = \sigma_0) = 1 - \alpha$
   $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2} | \sigma = \sigma_0) = 1 - \alpha$
   $P\left(\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{\alpha/2}\right) = 1 - \alpha$
   $P\left(\frac{\chi^2_{1-\alpha/2}\sigma_0^2}{(n-1)} < S^2 < \frac{\chi^2_{\alpha/2}\sigma_0^2}{(n-1)}\right) = 1 - \alpha$
   région critique : $X^2 < \chi^2_{1-\alpha/2}$ et $X^2 > \chi^2_{\alpha/2}$
5. Règle de décision :
   rejeter H_0 si $s^2 < s^2_{c1} = \chi^2_{1-\alpha/2}\sigma_0^2/(n-1)$ ou $s^2 > s^2_{c2} = \chi^2_{\alpha/2}\sigma_0^2/(n-1)$

Test unilatéral (variance)

1. $H_0 : \sigma = \sigma_0$, $H_1 : \sigma < \sigma_0$ (test unilatéral)
2. α à définir
3. Statistique à utiliser : S ; distribution :
   $X^2 = \frac{(n-1)S^2}{\sigma^2}$, v.a. loi du χ^2 à $\nu = n - 1$ degrés de liberté (population normale)
4. $P(\text{non-rejet de } H_0 | H_0 \text{ vraie}) = 1 - \alpha$
   $P(\text{non-rejet de } H_0 | \sigma = \sigma_0) = 1 - \alpha$
   $P(\chi^2_{1-\alpha} < X^2 | \sigma = \sigma_0) = 1 - \alpha$
   $P\left(\chi^2_{1-\alpha} < \frac{(n-1)S^2}{\sigma_0^2}\right) = 1 - \alpha$
   $P\left(\frac{\chi^2_{1-\alpha}\sigma_0^2}{(n-1)} < S^2\right) = 1 - \alpha$
   région critique : $X^2 < \chi^2_{1-\alpha}$
5. Règle de décision :
   rejeter H_0 si $s^2 < s_c^2 = \chi^2_{1-\alpha}\sigma_0^2/(n-1)$

Test sur une proportion

1. $H_0 : \pi = \pi_0$, $H_1 : \pi \neq \pi_0$ (test bilatéral)
2. α à définir
3. Statistique à utiliser : $\hat{P}$; distribution : $Z = (\hat{P} - \pi)/(\sqrt{\pi(1-\pi)}/\sqrt{n})$
4. $P(\text{non-rejet de } H_0 | H_0 \text{ vraie}) = 1 - \alpha$
   $P(\text{non-rejet de } H_0 | \pi = \pi_0) = 1 - \alpha$
   $P(-z_{\alpha/2} < (\hat{P} - \pi_0)/(\sqrt{\pi_0(1-\pi_0)}/\sqrt{n}) < z_{\alpha/2}) = 1 - \alpha$
   région critique : $Z < -z_{\alpha/2}$ et $Z > z_{\alpha/2}$
5. Règle de décision :
   rejeter H_0 si $\hat{p} < \hat{p}_{c1} = \pi_0 - z_{\alpha/2}\frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}}$ ou
   $\hat{p} > \hat{p}_{c1} = \pi_0 + z_{\alpha/2}\frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}}$

1. $H_0 : \pi = \pi_0$, $H_1 : \pi > \pi_0$ (test unilatéral)

   ...

5. Règle de décision : rejeter H_0 si $z > z_\alpha$
   c.à.d. $\hat{p} > \hat{p}_c = \pi_0 + z_\alpha \frac{\sqrt{\pi_0(1-\pi_0)}}{\sqrt{n}}$

Paramètre θ	μ			
Population Écart-type σ	$\approx$ normale connu	— connu	$\approx$ normale inconnu	
Échantillon	—	$n > 30$	$n > 30$	$n < 30$
Statistique $\hat{\Theta}$	$\overline{X}$			
St. normalisée	$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$		$Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$	$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$
Distribution	$N(0,1)$			Student (ν)
D.L.	—			$n-1$
Mesure $\hat{\theta}$	$\overline{x}$			

Paramètre θ	π	σ^2
Population Écart-type σ	— —	$\approx$ normale —
Échantillon	$n > 30$ [2]	—
Statistique $\hat{\Theta}$	$\hat{P}$	S^2
St. normalisée	$Z = \frac{\hat{P}-\pi}{\sqrt{\pi(1-\pi)/n}}$	$X^2 = \frac{(n-1)S^2}{\sigma^2}$
Distribution	$N(0,1)$	khi-deux (ν)
D.L.	—	$n-1$
Mesure $\hat{\theta}$	$\hat{p}$	s^2

2. En plus : $n\hat{p} \geq 5$, $n(1-\hat{p}) \geq 5$, ni $\hat{p} \approx 0$, ni $\hat{p} \approx 1$.

Stat. norm.	Intervalle de confiance	Test d'hypothèse $H_0 : \theta = \theta_0$		
		$H_1 : \theta \neq \theta_0$	$H_1 : \theta < \theta_0$	$H_1 : \theta > \theta_0$
Z	$-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$	$z < -z_{\frac{\alpha}{2}}$ ou $> z_{\frac{\alpha}{2}}$	$z < -z_\alpha$	$z > z_\alpha$
T	$-t_{\frac{\alpha}{2}} < t < t_{\frac{\alpha}{2}}$	$t < -t_{\frac{\alpha}{2}}$ ou $> t_{\frac{\alpha}{2}}$	$t < -t_\alpha$	$t > t_\alpha$
X^2	$\chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}$	$\chi^2 < \chi^2_{1-\frac{\alpha}{2}}$ ou $> \chi^2_{\frac{\alpha}{2}}$	$\chi^2 < \chi^2_{1-\alpha}$	$\chi^2 > \chi^2_\alpha$
	mettre sous la forme : $\theta_L < \theta < \theta_H$	« entrer dans le monde de H_0 » : $\theta = \theta_0$, calculer z, t, χ^2 à partir des mesures ; décisions de *rejet* de H_0		

- Intervalle de confiance : niveau de confiance $1 - \alpha$
- Tests d'hypothèse : seuil de signification α
- Voir tableaux unifiés en annexe.

Différence de moyennes : variances connues, populations presque normales

- Conditions : σ_1, σ_2 connus et
  - populations normales $N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$ ou
  - $n_1 > 30$ **et** $n_2 > 30$, ou
  - populations $\ll$ presque $\gg$ normales
- Échantillons aléatoires et indépendants de tailles n_1, n_2 ; moyennes $\overline{X}_1, \overline{X}_2$
  - $\boxed{\overline{X}_1 - \overline{X}_2}$: normale
  - $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2$
  - $\sigma^2_{\overline{X}_1 - \overline{X}_2} \overset{\text{ind}}{=} \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
- $Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \to N(0, 1)$
- Intervalle de confiance :

$$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Test d'hypothèse :
  1. $H_0 : \mu_1 - \mu_2 = d_0$, $H_1 : \mu_1 - \mu_2 \neq d_0$ (test bilatéral)

  5. Règle de décision : rejeter H_0 si $z < -z_{\alpha/2}$ ou $z > z_{\alpha/2}$

$(\overline{x}_1 - \overline{x}_2) < (\overline{x}_1 - \overline{x}_2)_{c} = d_0 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ ou

variances inconnues, populations normales et grands échantillons

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Populations normales **et** grands échantillons ($n_1 > 30$, $n_2 > 30$)
- σ_1, σ_2 : inconnus
- $Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \to \approx N(0, 1)$
- *Équivalent de $T \to Z$ pour grands échantillons*
- Intervalle de confiance :
  $$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
- Test d'hypothèse :
  1. $H_0 : \mu_1 - \mu_2 = d_0$, $H_1 : \mu_1 - \mu_2 > d_0$ (test unilatéral)
  5. Règle de décision : rejeter H_0 si $z > z_\alpha$
     $$(\overline{x}_1 - \overline{x}_2) > (\overline{x}_1 - \overline{x}_2)_c = d_0 + z_\alpha\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Variances inconnues mais égales, petits échantillons

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Populations normales **et** petits échantillons ($n_1 < 30$ ou $n_2 < 30$)
- σ_1, σ_2 : inconnus mais $\sigma_1 = \sigma_2$ (à tester)
- $T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_c^2}{n_1} + \frac{S_c^2}{n_2}}} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_c\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow$ Student
- Variance commune : $S_c^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2 + \sum_{i=1}^{n_2}(X_{2i} - \overline{X}_2)^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$
- T : Student à $(n_1 + n_2 - 2)$ d.l.
- Intervalle de confiance :
  $(\overline{x}_1 - \overline{x}_2) - t_{\alpha/2} s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2} s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Test d'hypothèse : ...
- À propos des conditions :
  - $\sigma_1 \approx \sigma_2$ ou populations $\approx$ normales : OK
  - $\sigma_1 \neq \sigma_2$ et normales : OK si $n_1 = n_2$

Variances inconnues et différentes - petits échantilons

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Populations normales **et** petits échantillons ($n_1 < 30$ ou $n_2 < 30$)
- σ_1, σ_2 : inconnus et $\sigma_1 \neq \sigma_2$ (à tester)
- $T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \to$ Student à ν d.l. ; $\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$
- Arrondir ν au nombre entier *inférieur*.
- Intervalle de confiance :
  $$(\overline{x}_1 - \overline{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
- Test d'hypothèse :
  1. $H_0 : \mu_1 - \mu_2 = d_0$, $H_1 : \mu_1 - \mu_2 < d_0$ (test unilatéral)
  5. Règle de décision : rejeter H_0 si $t < t_\alpha$
  $$(\overline{x}_1 - \overline{x}_2) < (\overline{x}_1 - \overline{x}_2)_c = d_0 - t_\alpha\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Echantillons appariés

- Échantillons aléatoires et **appariés** de tailles $n_1 = n_2 = n$
- Appariés : « avant / après »
- Population : nouvelle v.a. $D = X_1 - X_2$ (μ_D, σ_D)
- Échantillon : calculer $d_i = x_{1i} - x_{2i}$; oublier X_1, X_2 !
- Population normale ou grands échantillons ($n > 30$), σ_D connu : $Z = \frac{\overline{D}-\mu_D}{\sigma_D/\sqrt{n}} \rightarrow N(0, 1)$
- Population normale et petits échantillons ($n < 30$), σ_D inconnu : $T = \frac{\overline{D}-\mu_D}{s_D/\sqrt{n}}$ à $(n - 1)$ d.l.
- Intervalle de confiance : $\overline{d} - t_{\alpha/2}\frac{s_D}{\sqrt{n}} < \mu_D < \overline{d} + t_{\alpha/2}\frac{s_D}{\sqrt{n}}$
- Test d'hypothèse : ...
- Échantillons appariés : un seul nouvel échantillon !

Distribution de la différence des proportions

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Grands échantillons ($n_1 > 30$, $n_2 > 30$)
- Proportions : $\hat{P}_i = N(\pi_i, \sqrt{\pi_i(1-\pi_i)}/\sqrt{n_i})$
- $Z = \frac{(\hat{P}_1 - \hat{P}_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} \to N(0,1)$
- Intervalle de confiance : $(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}} <$
  $\pi_1 - \pi_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$; remplacer
  $\pi_i(1-\pi_i) \to \hat{\mathbf{p}}_i(1-\hat{\mathbf{p}}_i)$
- Test d'hypothèse :
  1. $H_0 : \pi_1 - \pi_2 = d_0$ $(\pi_1 = \pi_2 + d_0)$, $H_1 : \pi_1 - \pi_2 > d_0$ (test unilatéral)
  5. Règle de décision : rejeter H_0 si $z > z_\alpha$
     $(\hat{p}_1 - \hat{p}_2) > (\hat{p}_1 - \hat{p}_2)_c = d_0 + z_\alpha\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$
     Si $d_0 = 0$, $\pi_1 = \pi_2$: remplacer $\pi_j \to \hat{\mathbf{p}} = \frac{\sum_{i=1}^{n_1} x_{1i} + \sum_{i=1}^{n_2} x_{2i}}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$
     Si $d_0 \neq 0$: remplacer $\pi_j \to \hat{\mathbf{p}}_j$

- Échantillons aléatoires et indépendants de tailles n_1, n_2
- Provenant de populations normales de variances σ_1^2, σ_2^2
- Variances des échantillons : S_1^2, S_2^2
- $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{V_1/\nu_1}{V_2/\nu_2}$
- $V_i = \frac{(n_i-1)S_i^2}{\sigma_i^2}$: v.a. indépendantes, loi du χ^2 à $\nu_i = n_i - 1$ d.l.
- F : loi de Fisher (1924) - Snedecor (1934) avec ν_1 et ν_2 d.l.
- $F \geq 0$
- $P(F > f_\alpha(\nu_1, \nu_2)) = \alpha$ (définition de $f_\alpha(\nu_1, \nu_2)$)
- $f_\alpha(\nu_1, \nu_2) = \frac{1}{f_{1-\alpha}(\nu_2, \nu_1)}$ (propriété de la loi F)
- $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2}$
- Intervalle de confiance (niveau de confiance $1-\alpha$) :
  - $f_{1-\alpha/2}(\nu_1, \nu_2) < F < f_{\alpha/2}(\nu_1, \nu_2)$
  - $\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)}$
- Test d'hypothèse $H_0 : \sigma_1 = \sigma_2$
- Règle de décision : rejeter H_0 si
  - $H_1 : \sigma_1 \neq \sigma_2$
    $f < f_{1-\alpha/2}$ ou $f > f_{\alpha/2}$ c-à-d $s_1^2/s_2^2 < f_{1-\alpha/2}$ ou $s_1^2/s_2^2 > f_{\alpha/2}$
  - $H_1 : \sigma_1 > \sigma_2$
    $f > f_\alpha$ c-à-d $s_1^2/s_2^2 > f_\alpha$
  - $H_1 : \sigma_1 < \sigma_2$
    $f < f_{1-\alpha}$ c-à-d $s_1^2/s_2^2 < f_{1-\alpha/2}$

Paramètre θ	$\mu_2 - \mu_1$		
Populations	$\approx$ normales	—	$\approx$ normales
Écart-types σ_1, σ_2	connus	connus	inconnus
Échantillons	—	$n_1 > 30$ et $n_2 > 30$	$n_1 > 30$ et $n_2 > 30$
Statistique $\hat{\Theta}$	$\overline{X}_2 - \overline{X}_1$		
St. normalisée	$Z = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$		$Z = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
Distribution	$N(0, 1)$		
Degrés de liberté	—		
Mesure $\hat{\theta}$	$\overline{x}_2 - \overline{x}_1$		

Paramètre θ	$\mu_2 - \mu_1$	
Populations	$\approx$ normales	
Écart-types σ_1, σ_2	inc., $\sigma_1 = \sigma_2$ ou $n_1 = n_2$	inc., $\sigma_1 \neq \sigma_2$ et $n_1 \neq n_2$
Échantillons	$n_1 < 30$ ou $n_2 < 30$	
Statistique $\hat{\Theta}$	$\overline{X}_2 - \overline{X}_1$	
St. normalisée	$T = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{S_c\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$T = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
Distribution	Student (ν)	
Degrés de liberté	$n_1 + n_2 - 2$	ν^*
Mesure $\hat{\theta}$	$\overline{x}_2 - \overline{x}_1$	
Rappels	S_c : diapo #191	ν^* : diapo #192

Paramètre θ	$\pi_2 - \pi_1$	σ_1^2/σ_2^2
Populations	—	$\approx$ normales
Écart-types σ_1, σ_2	—	—
Échantillons	$n_1 > 30$ et $n_2 > 30$ [3]	—
Statistique $\hat{\Theta}$	$\hat{P}_2 - \hat{P}_1$	F
St. normalisée	$Z = \frac{(\hat{P}_2-\hat{P}_1)-(\pi_2-\pi_1)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}}}$	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$
Distribution	$N(0,1)$	Fischer (ν_1, ν_2)
Degrés de liberté	—	$n_1 - 1, n_2 - 1$
Mesure $\hat{\theta}$	$\hat{p}_2 - \hat{p}_1$	s_1^2/s_2^2

3. En plus : $n_i\hat{p}_i \geq 5$, $n_i(1-\hat{p}_i) \geq 5$, ni $\hat{p}_i \approx 0$, ni $\hat{p}_i \approx 1$ ($i = 1, 2$).

Stat. norm.	Intervalle de confiance	Test d'hypothèse $H_0 : \theta = \theta_0$		
		$H_1 : \theta \neq \theta_0$	$H_1 : \theta < \theta_0$	$H_1 : \theta > \theta_0$
Z	$-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$	$z < -z_{\frac{\alpha}{2}}$ ou $> z_{\frac{\alpha}{2}}$	$z < -z_\alpha$	$z > z_\alpha$
T	$-t_{\frac{\alpha}{2}} < t < t_{\frac{\alpha}{2}}$	$t < -t_{\frac{\alpha}{2}}$ ou $> t_{\frac{\alpha}{2}}$	$t < -t_\alpha$	$t > t_\alpha$
F	$f_{1-\frac{\alpha}{2}} < f < f_{\frac{\alpha}{2}}$	$f < f_{1-\frac{\alpha}{2}}$ ou $f > f_{\frac{\alpha}{2}}$	$f < f_{1-\alpha}$	$f > f_\alpha$
	mettre sous la forme : $\theta_L < \theta < \theta_H$	$\ll$ entrer dans le monde de H_0 $\gg$: $\theta = \theta_0$, calculer z, t, χ^2 à partir des mesures ; décisions de *rejet* de H_0		

- Intervalle de confiance : niveau de confiance $1 - \alpha$
- Tests d'hypothèse : seuil de signification α
- Voir tableaux unifiés en annexe

- Test statistique : « 2. Choisir le seuil de signification α »
- « Typiquement 1% ou 5% »
- Comment choisir ?
- Comment décider ?
- *Pourquoi* choisir α ?
- Tests classiques :
  - Mesurer $\hat{\theta}$; comparer $\hat{\theta}$ aux valeurs critiques $\hat{\theta}_c$
  - Valeurs critiques dépendent de α
- Alternative
  - Calculer α_p (p-value) telle que $\hat{\theta} = \hat{\theta}_c$
  - α_p : rejeter H_0 de façon marginale
- P-value (seuil descriptif) : la plus petite valeur de $\alpha = P(\text{rejeter} H_0 | H_0 \text{ vraie})$ qui conduirait au rejet de H_0
- La probabilité de se retrouver « au moins aussi loin » de la H_0 – dans le sens de la H_1 – que l'échantillon examiné, si H_0 est vraie.

- Test sur la moyenne, petit échantillon, population normale, σ inconnu

1. $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ (test bilatéral)
2. α à définir
3. Statistique à utiliser : $\overline{X}$; distribution :
   $T = (\overline{X} - \mu)/(S/\sqrt{n})$
4. Région critique : $T < -t_{\alpha/2}$ et $T > t_{\alpha/2}$
5. Règle de décision :
   rejeter H_0 si $t < -t_{\alpha/2}$ ou $> t_{\alpha/2}$
6. Prélever un échantillon et faire les calculs
7. Décider
8. Prélever un échantillon et faire les calculs
   Population $N(0.5, 1)$, $n = 5$

```
--> x = 0.5+rand(1,5,'normal')
x = 0.4303745 -1.2195277 -0.3570756 2.2734783
-0.5112132
--> mean(x)
ans = 0.1232073
--> stdev(x)
ans = 1.337359
```

$\mu_0 = 0$, calculer t :

```
--> t = ( mean(x) - 0 ) / ( stdev(x) / sqrt(5) )
t = 0.2060029
```

Comparer, à l'issue d'une expérience aléatoire, des fréquences expérimentales aux fréquences prévues par la théorie (Pearson, 1900).

- k : nombre de fréquences à comparer (nombre de classes)
- o_i : fréquences Observées (obtenues expérimentalement)
- e_i : fréquences « Espérées » (théoriques, à calculer)
- $$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}}$$
- Loi du χ^2 à ν degrés de liberté ; si $o_i = e_i$, $\chi^2 = 0$, sinon $\chi^2 > 0$
- Calculer χ^2 à partir de o_i, e_i ; obtenir $\alpha = P(X^2 > \chi^2)$, la p-value
- $$\boxed{\nu = k - 1 - (\text{nombre de paramètres estimés utilisés dans le calcul de } e_i)}$$
- Condition : $e_i \geq 5$ au moins pour 80% des classes ; $e_i > 0$ pour les autres
- Applications : test d'adéquation, d'indépendance, d'homogénéité, de proportions

H_0 : les données expérimentales ont été obtenues à partir d'une population suivant la loi $p_X(x)$ (p.ex., normale, uniforme, etc).

- Exemple : données sur plusieurs lancers d'un dé (données simulées...)
-

Face	1	2	3	4	5	6	Total N
Fréquence (o_i)	1037	937	1055	1034	929	1008	6000

```
O = [ 1037 937 1055 1034 929 1008]
```

- H_0 : le dé est bien équilibré ; $p_i = 1/6$, $\boxed{e_i = p_i N} = 1000$

```
e=ones(1,6)*1000
```

- Conditions : OK (sinon grouper des classes voisines)
- Calculer $\chi^2 = 14.624$ (`sum((O-e).^2)/1000`
- $\nu = 6 - 1 - 0 = 5$
- p-value : $P(X^2 > 14.624) =$

```
[P Q]=cdfchi(PQ,sum((O-e).^2)/1000,5)
Q= 0.0120957 P=0.9879047
```

- On peut rejeter H_0 au seuil de signification 5%

On mesure, sur chaque individu d'un échantillon aléatoire de taille n, deux caractères X et Y, à l et c modalités, respectivement.
H_0 : les deux caractères X et Y sont indépendants.

- Example : le tabac et les jeunes, INPES, baromètre santé 2000 (tr. #**??**)

Sexe \ Fumeur	Oui	Non	Total
Homme	340 (**310**)	314 (**344**)	654
Femme	289 (**319**)	384 (**354**)	673
Total	629	698	**1327**

- H_0 : X et Y sont indépendants ; $\pi_{ij} = \pi_i \pi_j$ $(i = 1, \ldots, l\,; j = 1, \ldots, c)$
- On *estime* π_i et π_j à partir des fréquences marginales de l'échantillon
- $\pi_{ij} = \pi_i \pi_j \rightarrow \frac{\mathbf{e_{ij}}}{n} = \frac{\sum_{j=1}^{c} o_{ij}}{n} \frac{\sum_{i=1}^{l} o_{ij}}{n} \rightarrow \boxed{\mathbf{e_{ij}} = \frac{1}{n} \sum_{j=1}^{c} o_{ij} \sum_{i=1}^{l} o_{ij}}$
- Degrés de liberté $\nu = (lc - 1) - 1 - [(l-1) + (c-1)] = \boxed{(l-1)(c-1)}$
- Conditions : OK (sinon ? augmenter la taille de l'échantillon !)

- Si $\nu = 1$ (tableau 2×2) utiliser :

$$\boxed{\chi^2 = \sum_{i,k} \frac{(|o_{ij} - e_{ij}| - 0.5)^2}{e_{ij}}}$$

- Calculer $\chi^2 = 10.5256$
- $\nu = (2-1)(2-1) = 1$
- p-value : $P(X^2 > 10.5256) =$
```
[P Q]=cdfchi(``PQ'', 10.5256, 1)
Q=0.0011773 P = 0.998227
```
- On peut rejeter H_0 au seuil de signification 1%

À partir de c populations, on obtient c échantillons aléatoires et indépendants, de taille n_j ($j = 1, \ldots, c$). On mesure sur chaque individu le même caractère X, à I modalités.
H_0 : la proportion d'individus appartenant à la i-ème modalité ($i = 1, \ldots, I$), reste la même pour toutes les populations (les populations sont homogènes par rapport au caractère étudié).

- Example : notes (fictives) échantillonnées dans trois parcours
-

Note \ Parcours	I	II	III	Total
$0 \leq x < 6$	32	15	8	55
$6 \leq x < 12$	123	60	43	226
$12 \leq x \leq 20$	145	125	149	419
Total (n_j)	**300**	**200**	**200**	**700**

- H_0 : proportion de chaque modalité constante ; $\pi_{i1} = \pi_{i2} = \ldots = \pi_{ic} = \pi_i$ ($i = 1, \ldots, I$)
- On *estime* π_i à partir des fréquences marginales de l'échantillon
-

Note \ Parcours	I	II	III	Total
$0 \leq x < 6$	32 (**23.57**)	15 (**15.71**)	8 (**15.71**)	55
$6 \leq x < 12$	123 (**96.86**)	60 (**64.57**)	43 (**64.57**)	226
$12 \leq x \leq 20$	145 (**179.57**)	125 (**119.71**)	149 (**119.71**)	419

À partir de c populations, on obtient c échantillons aléatoires et indépendants, de taille n_j ($j = 1, \ldots, c$). On mesure sur chaque individu le même caractère X, à 2 modalités (≪ oui ≫ / ≪ non ≫).
H_0 : la proportion de ≪ oui ≫ reste la même pour toutes les populations (cas spécial du test d'homogénéité, $I = 2$).

- Exemple : nombre de pièces défectueuses et moment de production
-

Pièces\ Créneau	Matin	Après-midi	Nuit	Tota
Défectueuses (≪ O ≫)	45 (**56.97**)	55 (**56.67**)	70 (**56.37**)	170
Normales (≪ N ≫)	905 (**893.03**)	890 (**888.33**)	870 (**883.63**)	266
Total (n_j)	**950**	**945**	**940**	**283**

- $H_0 : \pi_1 = \pi_2 = \ldots = \pi_c = \pi$
- On *estime* π à partir des fréquences marginales de l'échantillon
- ≪ Oui ≫ : $\pi_j = \pi \rightarrow \frac{\mathbf{e_{1j}}}{n_j} = \frac{\sum_{j=1}^{c} o_{1j}}{n}$
- ≪ Non ≫ : $1 - \pi_j = 1 - \pi \rightarrow \frac{\mathbf{e_{2j}}}{n_j} = \frac{\sum_{j=1}^{c} o_{2j}}{n}$

- $\mathbf{e_{ij}} = \frac{n_j}{n} \sum_{j=1}^{c} o_{ij} \rightarrow \boxed{\mathbf{e_{ij}} = \frac{1}{n} \sum_{j=1}^{c} o_{ij} \sum_{i=1}^{l} o_{ij}}$
- Même formule que le test d'indépendance / d'homogénéité !
- Degrés de liberté $\boxed{\nu = (2-1)(c-1) = c-1}$
- Conditions : OK (sinon ? augmenter les tailles des échantillons !)
- Calculer $\chi^2 = 6.2339$
- $\nu = (3-1) = 2$
- p-value : $P(X^2 > 6.2339) =$
```
[P Q]=cdfchi(``PQ'', 6.2339, 2)
Q=0.04429
```
- On peut rejeter H_0 au seuil de signification 5%

Même contexte qu'avant : c populations, c échantillons, caractère X à deux modalités.
H_0 : les proportions de « oui », $\pi_1, \ldots, \pi_c$, sont égales à $p_1, \ldots, p_c$ (pas d'estimation de paramètres).

- « Oui » : $\pi_j = p_j \rightarrow \frac{\mathbf{e_{1j}}}{n_j} = p_j$
- « Non » : $1 - \pi_j = 1 - p_j \rightarrow \frac{\mathbf{e_{2j}}}{n_j} = 1 - p_j$
- $\boxed{\nu = c}$: on ne perd aucun degré de liberté
- Exemple précédent avec :
  $p_1 = 0.05$, $p_2 = 0.06$, $p_3 = 0.08$ ($\neq 170/2835 \approx 0.06$)
- Calculer $\chi^2 = 0.5836$
- $\nu = 3$
- p-value : $P(X^2 > 0.5836) = 0.9002$
- On ne peut pas rejeter H_0

H_0 : les données expérimentales (échantillon de taille n) ont été obtenues à partir d'une population **normale***.*

- Procédure « classique » : test du χ^2 (cf. TD 6)
  1. Répartir les données en classes (histogramme)
  2. Estimer μ et σ avec `cdfnor`

  3a. Calculer les probabilités théoriques p_j des classes
  Calculer les fréquences théoriques $e_j = p_j n$
  Vérifier les conditions sinon regrouper les classes

  3b. Ou répartir en $(M+1)$ classes équiprobables : $e_j = n/(M+1)$

  4. Calculer χ^2 (on perd deux d.l. avec l'estimation de μ et σ !)
- Une grande p-value permet de ne pas rejeter l'hypothèse de normalité