

INTRODUCTION À LA MODÉLISATION STATISTIQUE

– L2–MIASHS 2015–2016 –

JUAN VIU-SOS

TABLE DES MATIÈRES

1. Notions de probabilité.....	1
1.1. Univers, événements et opérations.....	2
1.2. Probabilité : définition axiomatique.....	3
1.3. Probabilité : modèle finie et équiprobabilité.....	4
1.4. Notion de probabilité conditionnelle. Indépendance.....	5
2. Variables aléatoires.....	7
2.1. Variables aléatoires discrètes. Espérance. Variance.....	7
2.2. Épreuve de Bernoulli.....	9
2.3. Schéma de Bernoulli. Loi binomiale.....	10
2.4. Simulation de variables aléatoires discrètes.....	12
2.5. Une variable aléatoire continue : la loi normale.....	12
3. Estimation ponctuelle d'une proportion.....	17
3.1. Modèle statistique.....	17
3.2. Estimation du modèle.....	17
3.3. L'inégalité de Bienaymé-Tchebychev.....	18
4. Estimation d'une proportion à l'aide d'un intervalle de confiance.....	19
4.1. Théorème de Moivre-Laplace.....	20
4.2. Intervalle de fluctuation asymptotique.....	20
4.3. Intervalle de confiance asymptotique.....	21

1. NOTIONS DE PROBABILITÉ

La théorie des probabilités fournit des modèles mathématiques permettant l'étude des expériences dont le résultat est fruit du hasard :

- le lancer d'un dé ou de pièce de monnaie,
- faire tourner une roulette,
- durée de vie d'un appareil électronique,
- temps d'attente dans une file,
- mesurer la hauteur des Français(es),...

Définition 1.1. On appelle *expérience aléatoire* à toute épreuve dont :

- a) on ne peut pas prévoir le résultat de façon certaine.
- b) on peut indiquer l'ensemble des résultats possibles.

Tout résultat possible d'une expérience aléatoire est appelé *éventualité*.

1.1. Univers, événements et opérations.

Définition 1.2. On considère une expérience aléatoire.

- On appelle *univers*, au ensemble Ω qui représente toutes les éventualités.
- Soit $A \subset \Omega$, on dit que A est un *événement* si, une fois obtenu le résultat de l'expérience aléatoire, on peut affirmer si A est-il vérifié ou non. Un événement est-il appelé *élémentaire* s'il contient qu'une seule éventualité. On dénotera par \mathcal{A} l'ensemble de tous les événements de Ω .
- On appelle Ω et \emptyset les événements *certain* et *impossible*, respectivement.

Exemple 1.3. On considère l'expérience aléatoire qui consiste à lancer d'un dé de six faces non truqué et à regarder le nombre inscrit sur sa face supérieure.

- Les éventualités sont : “obtenir un 1”, “obtenir un 2”, ..., “obtenir un 6”, qui peuvent être représentées par les nombres $1, 2, \dots, 6$, respectivement.
- Dont, l'univers : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Les sous-ensembles $A = \{\text{“obtenir 2”}\} = \{2\}$, $B = \{\text{“obtenir pair”}\} = \{2, 4, 6\}$, $C = \{\text{“obtenir au moins 3”}\} = \{3, 4, 5, 6\}$ sont des événements.

Définition 1.4. Soient A et B des événement d'une expérience aléatoire avec univers associé Ω . On définit les opérations suivantes :

- *événement contraire de A* : $\bar{A} = \{\omega \in \Omega \mid \omega \notin A\}$.
- *événement “ A et B ”* : $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ et } \omega \in B\}$.
- *événement “ A ou B ”* : $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ ou } \omega \in B\}$.
- *événement “ A et non B ”* : $A \setminus B = \{\omega \in \Omega \mid \omega \in A \text{ et } \omega \notin B\}$.

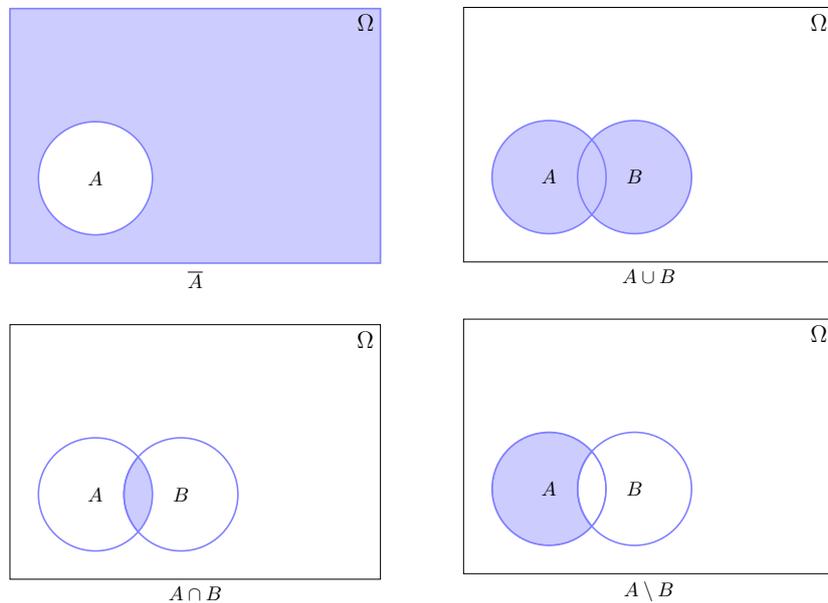


FIGURE 1. Diagrammes de Venn

Définition 1.5. Deux événements A et B sont appelés *incompatibles* si $A \cap B = \emptyset$.

Remarque 1.6.

- (1) ces opérations sont exactement les opérations classiques entre ensembles : complémentaire, intersection, réunion et différence (Figure 1).
- (2) On a assumé que la famille des événements \mathcal{A} est “suffisamment riche” pour que toutes les opérations précédentes soient bien définies. Pendant le cours, on n’aura besoin que des considérations précédentes, mais mathématiquement on a besoin d’assumer aussi que

$$\text{Si } \{A_i\}_{i=1}^{+\infty} \subset \mathcal{A} \implies \bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$$

(notion de σ -algèbre, *Introduction aux probabilités, S4*).

Rappel 1.7 (Lois de Morgan). Soient $A, B \subset \Omega$:

- (1) $\overline{A \cup B} = \overline{A} \cap \overline{B}$.
- (2) $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Exemple 1.8. Si on revient sur l'exemple précédent :

$$\begin{aligned} \overline{A} &= \{1, 3, 4, 5, 6\} & \overline{B} &= \{1, 3, 5\} = \{\text{“obtenir impair”}\} & \overline{C} &= \{1, 2\} = \{\text{“obtenir au plus 2”}\} \\ B \cap C &= \{4, 6\} & B \cup C &= \{2, 3, 4, 5, 6\} & A \cup B &= \{2, 4, 6\} = B \text{ (car } A \subset B) & \overline{B} \setminus C &= \{1, 3\} \\ \overline{B} \cap \overline{C} &= \{\text{“obtenir impair et au plus 2”}\} = \{1\} = \overline{B \cup C} \\ \overline{B} \cup \overline{C} &= \{\text{“obtenir impair ou au plus 2”}\} = \{1, 2, 3, 5\} = \overline{B \cap C} \end{aligned}$$

1.2. Probabilité : définition axiomatique.

On considère une expérience aléatoire d’univers Ω et famille d’événements \mathcal{A} .

Définition 1.9. Une fonction de *probabilité* est une fonction $P : \mathcal{A} \rightarrow \mathbb{R}$ telle que :

- (1) $0 \leq P(A) \leq 1$ pour tout $A \in \mathcal{A}$.
- (2) Si $\{A_i\}_{i=1}^{+\infty}$ est une famille d’événements deux à deux incompatibles (i.e. $A_i \cap A_j = \emptyset, \forall 1 \leq i \neq j$), alors

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

Le triplet (Ω, \mathcal{A}, P) s’appelle espace probabilisé.

Proposition 1.10. *Toute fonction de probabilité vérifie :*

- (1) $P(\Omega) = 1$ et $P(\emptyset) = 0$.
- (2) Pour tout $A, B \in \mathcal{A}$:
 - (a) $P(\overline{A}) = 1 - P(A)$.
 - (b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
En particulier, si A et B incompatibles : $P(A \cup B) = P(A) + P(B)$.
 - (c) $P(A \setminus B) = P(A) - P(A \cap B)$.

Remarque 1.11. \triangle On est en train de “mesurer” des ensembles !

1.3. Probabilité : modèle finie et équiprobabilité.

Un cas particulier d'expérience aléatoire est celle où l'univers est *finie* :

$$\Omega = \{e_1, e_2, \dots, e_n\}$$

Exemples : le lancé d'une pièce $\Omega = \{\text{"pile"}, \text{"face"}\}$, le lancé d'un dé $\Omega = \{1, 2, \dots, 6\}$, ...

Dans ce cas :

- $\mathcal{A} = \mathcal{P}(\Omega)$. (tout sous-ensemble de Ω)
- Pour définir une fonction de probabilité sur Ω , il suffit d'associer un réel $p_i \in [0, 1]$ à chaque éventualité $e_i \in \Omega$ tel que $P(\{e_i\}) = p_i$.

Exemple 1.12.

- Monnaie non-truquée : $P(\{\text{"pile"}\}) = P(\{\text{"face"}\}) = 1/2$.
- Monnaie truquée : $P(\{\text{"pile"}\}) = 3/4$. Quel est $P(\{\text{"face"}\})$?

Définition 1.13. On dit qu'un modèle finie est *équiprobable* (ou d'*équiprobabilité*) lorsque toutes les probabilités élémentaires sont égales, c'est-à-dire lorsque $P(\{e_i\}) = P(\{e_j\})$ quelque soit les éventualités $e_i, e_j \in \Omega$.

Théorème 1.14 (Règle de Laplace). Dans un modèle équiprobable :

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{nb des cas favorables}}{\text{nb total de cas}}, \quad \forall A \in \mathcal{A}.$$

Exemple 1.15. Le lancé d'un dé non-truqué, on a $\Omega = \{1, 2, \dots, 6\}$:

- $A = \{\text{"obtenir pair"}\} = \{2, 4, 6\} \implies P(A) = 3/6 = 1/2$.
- $B = \{\text{"obtenir impair"}\} = \{1, 3, 5\} \implies P(B) = 3/6 = 1/2 (= 1 - P(\overline{B}) = 1 - P(A))$.
- $C = \{\text{"obtenir multiple de 3"}\} = \{3, 6\} \implies P(C) = 2/6 = 1/3$
- $D = \{\text{"obtenir au moins 3"}\} = \{3, 4, 5, 6\} \implies P(D) = 4/6 = 2/3$

Exemple 1.16. Dans un collège, les 100 élèves de troisième sont répartis selon leur seconde langue vivante comme le montre le tableau suivant :

	allemand	espagnol	total
garçons	18	22	40
filles	33	27	60
total	51	49	100

Une expérience aléatoire consiste à prendre un élève au hasard. On modélise cette expérience de façon équiprobable sur l'ensemble Ω des 100 élèves.

Notons $A = \{\text{"l'élève étudie l'allemand"}\}$ et $F = \{\text{"l'élève est une fille"}\}$. Quel est la probabilité que l'élève pris au hasard

- a) soit une fille ? $P(F) = 60/100 = 0.6$
- b) soit une fille germaniste ? $P(F \cap A) = 33/100 = 0.33$
- c) soit un garçon ou fasse de l'allemand ?
 $P(\overline{F} \cup A) = P(\overline{F}) + P(A) - P(\overline{F} \cap A) = 1 - 60/100 + 51/100 - 18/100 = 73/100 = 0.73$
 En utilisant les lois de Morgan : $P(\overline{F} \cup A) = 1 - P(\overline{\overline{F} \cup A}) = 1 - P(F \cap \overline{A}) = 1 - 27/100 = 73/100 = 0.73$

1.4. Notion de probabilité conditionnelle. Indépendance.

Comment doit-on modifier la probabilité que l'on attribue à un événement lorsqu'on dispose d'une information supplémentaire ou on a déjà obtenu des résultats *a priori* sur l'expérience ?

Exemple 1.17. On reprend l'exemple précédent. Quelle est la probabilité qu'un élève *filles* pris au hasard étudie l'allemand ? On dispose ici d'une information supplémentaire : on sait que l'élève choisi est une fille. Notre univers se restreint aux 60 filles, dont 33 étudient l'allemand. D'après la Règle de Laplace, cette probabilité est de $33/60 = 11/20$. Comment peut-on définir mathématiquement cet concept ?

Définition 1.18. Soient $A, B \in \mathcal{A}$, avec $P(A) \neq 0$. La *probabilité de B sachant que A est réalisé* (on dit la *probabilité de B sachant A*), noté $P(B|A)$, est le nombre réel défini par :

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

Remarque 1.19. \triangle En général : $P(B|A) \neq P(A|B)$.

Exemple 1.20.

- Probabilité de qu'une fille prise au hasard étudie l'allemand : $P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{|A \cap F|}{|F|} = 11/20$.
- Probabilité de qu'un étudiant d'allemand pris au hasard soit une fille : $P(F|A) = \frac{P(A \cap F)}{P(A)} = \frac{|A \cap F|}{|A|} = 33/51$.

Remarque 1.21. \triangle En traduisant un énoncé, attention à ne pas confondre $P(A \cap B)$ avec $P(A|B)$.

Ce qui fait l'intérêt du concept de probabilité conditionnelle, c'est qu'il est souvent bien plus facile d'attribuer directement une valeur à $P(B|A)$ en tenant compte des conditions expérimentales (liées à l'information A) et d'en déduire ensuite la valeur de $P(A \cap B)$.

Proposition 1.22. Soient A et B deux événements de probabilité non nulle :

$$P(A \cap B) = P(A) \cdot P(B|A)$$

Remarque 1.23. Il est très pratique pour cela de remanier la définition ci-dessus et de représenter la situation par un arbre pondéré (Figure 2) :

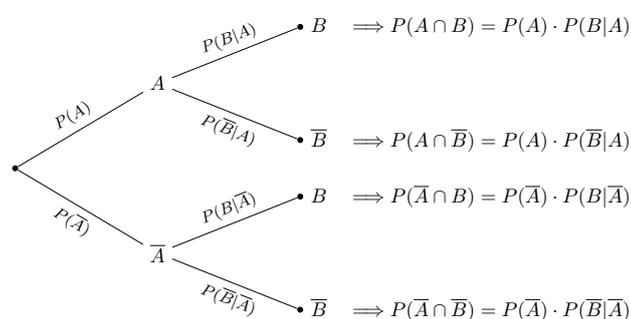


FIGURE 2

La probabilité de l'intersection de deux (ou plusieurs...) événements est égale au produit des probabilités des branches du chemin passant par ces événements.

Exemple 1.24. Une urne contient 2 boules rouges et 3 boules vertes indiscernables au toucher. On en tire au hasard deux l'une après l'autre, *sans remise*. Quelle est la probabilité d'obtenir deux rouges ?

On prends comme univers Ω l'ensemble des $20 = 5 \cdot 4$ possibles extractions des boules 5 boules avec la couleur :

$$\Omega = \{(i, j) \in \{R1, R2, V3, V4, V5\}^2 \mid i \neq j\} = \{(R1, R2), (R1, V3), \dots\}$$

(arrangement de 2 éléments de $\{R1, R2, V3, V4, V5\}$)

et dans ce cas on prend $\mathcal{A} = \mathcal{P}(\Omega)$ et P l'équiprobabilité. On définit les événements :

- $A = \{ \text{“ obtenir une rouge dans la premier extraction ”} \}$.
- $B = \{ \text{“ obtenir une rouge dans la seconde extraction ”} \}$.

On construit l'arbre pondéré avec les deux extractions et on obtient :

$$P(A) = 2/5, \quad P(B|A) = 1/4 \implies P(A \cap B) = 2/5 \cdot 1/4 = 1/10$$

Remarque 1.25. “ $B|A$ ” ne désigne pas un nouvel événement différent de A . Quand on écrit $P(B|A)$, ce que l'on a modifié, ce n'est pas l'événement B , mais la valeur numérique qui lui était attribuée par la fonction d'ensembles P !

Proposition 1.26. Soit (Ω, \mathcal{A}, P) un espace probabilisé et $A \in \mathcal{A}$ fixe tel que $P(A) \neq 0$. Alors la fonction :

$$\begin{aligned} P(\cdot | A) : \mathcal{A} &\longrightarrow [0, 1] \\ B &\longmapsto P(B|A) = \frac{P(A \cap B)}{P(A)} \end{aligned}$$

est une fonction de probabilité et $(\Omega, \mathcal{A}, P(\cdot | A))$ est un espace probabilisé.

Remarque 1.27. $P(\cdot | A)$ vérifie donc toutes les propriétés de la fonction de probabilité, p.ex.

- $P(\overline{B}|A) = 1 - P(B|A)$,
- $P(B \cup C|A) = P(B|A) + P(C|A) - P(B \cap C|A)$,
- ...

Remarque 1.28. Il existent des résultats très utiles pour relier les probabilités *a priori* et *a posteriori* d'avoir reçu des nouvelles informations sur les expériences qu'on est en train d'étudier. (Formule des probabilités totales, Formule de Bayes, *Introduction aux probabilités, S4*).

Exemple 1.29. Si on revienne sur l'exemple précédent : supposons qu'on a obtenu comme résultat une boule verte. Quel est la probabilité de que la boule proviens de l'urne U_1 ?

Définition 1.30. Deux événements A et B sont dits *indépendants* lorsque $P(A \cap B) = P(A) \cdot P(B)$.

Proposition 1.31. Deux événements A et B de probabilité non nulle sont indépendants si et seulement si $P(A|B) = P(A)$. (siii $P(B|A) = P(B)$)

Proposition 1.32. Deux événements A et B sont indépendants si et seulement si A et \overline{B} sont indépendants.

Remarque 1.33. \triangle Ne pas confondre événements incompatibles et événements indépendants (ensembles vs probabilité).

Exemple 1.34. Si on considère deux lancers consécutifs d'une pièce équilibrée les événements :

- $A = \{ \text{“ obtention de pile dans la premier lancé ”} \}$.
- $B = \{ \text{“ obtention de pile dans la seconde lancé ”} \}$.

sont indépendantes, car $P(A \cap B) = 1/4 = 1/2 \cdot 1/2 = P(A) \cdot P(B)$.

Exemple 1.35. On classe 1200 personnes d'un village par rapport au genre et si elles fument ou non :

	femme	homme	total
fume	200	600	800
ne fume pas	100	300	400
total	300	900	1200

Une expérience aléatoire consiste à prendre une personne au hasard. On modélise cette expérience de façon équiprobable sur l'ensemble Ω des 1200 personnes.

Les événements $H = \{ \text{“ la personne est un homme ”} \}$ et $F = \{ \text{“ la personne fume ”} \}$ sont indépendantes, car

$$P(H) = \frac{900}{1200} = \frac{3}{4}, \quad P(F) = \frac{800}{1200} = \frac{2}{3}, \quad P(H \cap F) = \frac{600}{1200} = \frac{1}{2}$$

et $P(H \cap F) = 1/2 = 3/4 \cdot 2/3 = P(H) \cdot P(F)$.

2. VARIABLES ALÉATOIRES

Pendant tout le chapitre, on considère une expérience aléatoire avec espace probabilisé associé (Ω, \mathcal{A}, P) .

2.1. Variables aléatoires discrètes. Espérance. Variance.

Définition 2.1. On appelle *variable aléatoire discrète (v.a.d.)* toute application

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

telle que :

- (1) L'ensemble des images $X(\Omega)$, appelé l'*univers image* de X , est un sous-ensemble de \mathbb{N} .
- (2) Pour tout $x \in X(\Omega)$, l'ensemble $\{ \omega \in \Omega \mid X(\omega) = x \}$ (dénoté simplement par $\{X = x\}$) fait partie des événements de l'expérience aléatoire \mathcal{A} .

Une variable aléatoire discrète est caractérisée de manière unique par l'ensemble des probabilités $P(X = x) \in [0, 1]$, pour tout $x \in X(\Omega)$. Cet ensemble est appelé la *loi de X* .

Remarque 2.2. Une variable aléatoire discrète est simplement une codification numérique (par des entiers) de certaines informations intéressantes de l'univers Ω d'une expérience aléatoire.

Exemple 2.3. Si on considère deux lancers consécutifs d'une pièce non-truquée, on a l'univers associé

$$\Omega = \{ \text{“pile-pile”}, \text{“face-pile”}, \text{“pile-face”}, \text{“face-face”} \}$$

On peut donc définir la variable aléatoire discrète $X = \text{“nombre de piles obtenues dans deux lancers consécutifs d'une pièce”}$ qui a la forme

$$X : \begin{array}{l} \Omega \\ \text{“pile-pile”} \\ \text{“face-pile”} \\ \text{“pile-face”} \\ \text{“face-face”} \end{array} \begin{array}{l} \longrightarrow \mathbb{N} \\ \longmapsto 2 \\ \longmapsto 1 \\ \longmapsto 1 \\ \longmapsto 0 \end{array} \left. \vphantom{\begin{array}{l} \Omega \\ \text{“pile-pile”} \\ \text{“face-pile”} \\ \text{“pile-face”} \\ \text{“face-face”} \end{array}} \right\} X(\Omega)$$

donc l'univers image $X(\Omega) = \{0, 1, 2\}$. On se trouve dans situation d'équiprobabilité, donc en Utilisant la Règle de Laplace, la loi de probabilité de la v.a.d. X peut s'exprimer par le tableau

x_i	0	1	2	(*)
$P(X = x_i)$	1/4	1/2	1/4	

Cette loi est représenté par le graphe de la fonction discrète (Figure 3) :

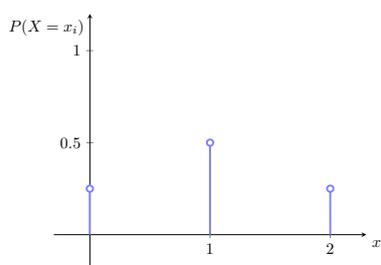


FIGURE 3

On peut exprimer l'événement {“ obtenir au plus une pile ”} = $\{X \leq 1\}$ comme l'union disjoint d'événements $\{X = 0\}$ et $\{X = 1\}$, donc :

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 1/4 + 1/2 = 3/4$$

On va définir deux paramètres qui nous résument des informations sur la loi de probabilité d'une v.a.d. :

Définition 2.4. Soit X une v.a.d. On appelle :

- *espérance* de X au nombre réel donné par la somme pondéré

$$E[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x)$$

Plus généralement, si ϕ est une fonction réelle définie sur $X(\Omega)$:

$$E[\phi(X)] = \sum_{x \in X(\Omega)} \phi(x) \cdot P(X = x)$$

- *variance* de X au nombre réel définie par

$$\text{Var}[X] = E[(X - E[X])^2] = \sum_{x \in X(\Omega)} (x - E[X])^2 \cdot P(X = x)$$

Remarque 2.5. Intuitivement, $E[X]$ et $\text{Var}[X]$ donnent la *moyenne théorique* et la *dispersion théorique* des valeurs prises par X qu'on l'obtiendrait sur un grand nombre d'expériences.

Propriété 2.6. $\text{Var}[X] = E[X^2] - E[X]^2$.

Exemple 2.7. Revenons sur le dernier exemple. On peut calculer l'espérance et la variance de la v.a.d. X à partir du tableau

$$\begin{aligned} E[X] &= \sum_{x \in X(\Omega)} x \cdot P(X = x) = \sum_{i=1}^3 x_i \cdot P(X = x_i) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) \\ &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1 \end{aligned}$$

$$\begin{aligned}\text{Var}[X] &= \sum_{x \in X(\Omega)} (x - E[X])^2 \cdot P(X = x) = \sum_{i=1}^3 (x_i - E[X])^2 \cdot P(X = x_i) \\ &= (0 - 1)^2 \cdot P(X = 0) + (1 - 1)^2 \cdot P(X = 1) + (2 - 1)^2 \cdot P(X = 2) = 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1/2 \\ (\text{plus simple}) &= E[X^2] - E[X]^2 = \sum_{i=1}^3 x_i^2 \cdot P(X = x_i) - 1 = 0 + 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} - 1 = 1/2\end{aligned}$$

Donc la moyenne théorique de X est 1 avec pas beaucoup de dispersion.

Exemple 2.8. Imaginons qu'on en train de jouer avec un ami au jeu de *pile-pile* : on lance deux pièces équilibrées, si on obtient "pile-pile" notre ami doit nous donner 1 euro, si on obtient "face-face" c'est nous qui donnons l'euro à notre ami. Dans un autre cas, rien se passe. Quelle est le bénéfice attendu dans ce jeu ?

À partir de l'exemple précédente, on peut définir la variable aléatoire $Y =$ "Notre bénéfice dans le jeu de *pile-pile*" comme une fonction de X , c.-a.-d. $Y = \phi(X) = X - 1$. Si on calcule l'espérance associée à Y :

$$\begin{aligned}E[Y] &= E[X - 1] = \sum_{i=1}^3 (x_i - 1) \cdot P(X = x_i) = -1 \cdot P(X = 0) + 0 \cdot P(X = 1) + 1 \cdot P(X = 2) \\ &= -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0\end{aligned}$$

Donc, si on joue un nombre suffisamment grand de fois, les bénéfices et les pertes auront tendance à se compenser. C'est qu'on appelle un *jeu équitable*.

Remarque 2.9. Dans ce cours, on va se centrer sur des v.a.d. définies sur des univers finis : $\Omega = \{e_1, e_2, \dots, e_n\}$. L'univers image $X(\Omega)$ est donc fini, ainsi comme les sommes arithmétiques précédentes.

Définition 2.10. Soient X et Y deux v.a.d. définies sur l'univers Ω . On dit que X et Y sont indépendantes si et seulement si

$$P(X = k_x, Y = k_y) = P(X = k_x)P(Y = k_y), \quad \forall k_x \in X(\Omega), \forall k_y \in Y(\Omega).$$

2.2. Épreuve de Bernoulli.

Définition 2.11. On appelle *épreuve de Bernoulli* de probabilité $p \in [0, 1]$ toute expérience aléatoire admettant deux issues tel que

- la première issue appelée *succès* (notée S) se réalise avec une probabilité p .
- la seconde issue appelée *échec* (notée \bar{S} ou E) se réalise avec une probabilité $q = 1 - p$.

Exemple 2.12.

- Obtenir pile ou face en lançant une pièce équilibrée ($p = 1/2$) ou truquée ($p \neq 1/2$).
- Obtenir 1 avec un dé à 6 faces.
- Obtenir un as sur 32 cartes.
- Qu'une personne pris au hasard aille une certaine maladie.

Définition 2.13. On dira la variable aléatoire X suit la *loi de Bernoulli* de paramètre $p \in [0, 1]$ si elle ne prend que deux valeurs 0 et 1 avec :

$$P(X = 1) = p \quad \text{et} \quad P(X = 0) = q$$

On notera $X \sim \mathcal{B}(p)$.

2.3. Schéma de Bernoulli. Loi binomiale.

Définition 2.14. On appelle *schéma de Bernoulli comportant n épreuves* ($n \in \mathbb{N}^*$) toute expérience consistant à répéter n fois de façon indépendante la même épreuve de Bernoulli de paramètre p .

Exemple 2.15. On répète $n = 3$ fois le lancer d'une pièce et on observe l'événement "obtenir face". On peut illustrer un schéma de Bernoulli par un arbre pondéré (Figure 4) :

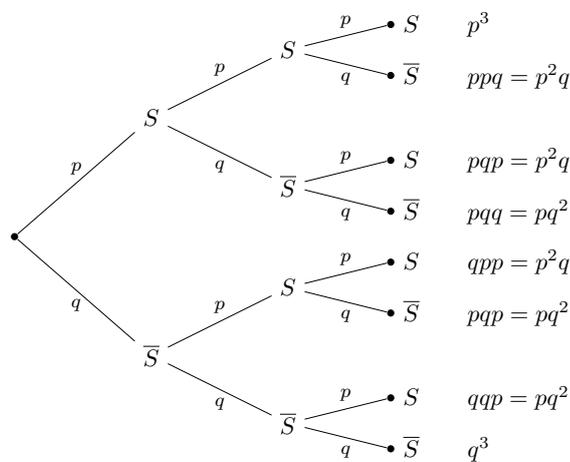


FIGURE 4

La probabilité d'observer une réalisation donnée est facile à calculer grâce à l'indépendance des épreuves !

$$P(\{\text{"obtenir 3 faces"}\}) = p^3 \quad \text{et} \quad P(\{\text{"obtenir 2 faces"}\}) = 3p^2q$$

Définition 2.16. On considère un schéma de Bernoulli avec n épreuves et de probabilité p d'observer un succès S à chacune des épreuves :

$$\Omega = \{(e_1, \dots, e_n) \mid e_i = S \text{ ou } \bar{S}\} = \{S, \bar{S}\}^n$$

Soit X la v.a.d. à valeurs dans $\{0, 1, \dots, n\}$ qui compte le nombre de succès d'une tel schéma de Bernoulli. On dira que la variable aléatoire X suit la *loi binomiale* de paramètres $n \in \mathbb{N}$ et $p \in [0, 1]$, noté $X \sim \mathcal{B}(n, p)$.

Remarque 2.17. Plus mathématiquement : si X_1, \dots, X_n sont des v.a.d. indépendantes tels que $X_i \sim \mathcal{B}(p)$, alors la v.a.d. définie par $Z = \sum_{i=1}^n X_i$ suit une binomiale de paramètres n et p .

En utilisant l'arbre pondéré décrivant le schéma de Bernoulli, on peut facilement obtenir la loi binomiale :

Propriété 2.18. Soit $X \sim \mathcal{B}(n, p)$, alors :

$$\forall k \in \{0, 1, \dots, n\} : \quad P(X = k) = C_n^k p^k q^{n-k}$$

où C_n^k = combinaisons sans répétition de n éléments pris k à k .

Remarque 2.19. $\sum_{k=0}^n P(X = k) = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p + q)^n = 1$ (Formule du binôme de Newton).

Rappel 2.20. Sur les combinaisons et la factorielle. Soient $k, n \in \mathbb{N}$ tels que $k \leq n$ et $n \geq 1$.

- a) $k! = k \cdot (k - 1) \cdot (k - 2) \cdots 2 \cdot 1$.
 $0! = 1$.
- b) $C_n^k = \frac{n!}{k!(n-k)!}$. En particulier, $C_n^0 = \frac{n!}{0!n!} = 1$ et $C_n^n = \frac{n!}{n!0!} = 1$.
- c) $C_n^k = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!k!} = C_n^{n-k}$.
- d) $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$.

Ces nombres peuvent se calculer facilement avec le Triangle de Pascal (Figure 5) :

$n \backslash k$	0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

FIGURE 5. Triangle de Pascal : $C_4^1 + C_4^2 = C_5^2$.

Propriété 2.21. Soit $X \sim \mathcal{B}(n, p)$ avec $n \in \mathbb{N}$ et $p \in [0, 1]$, alors

$$E[X] = np \quad \text{et} \quad \text{Var}[X] = npq$$

Démonstration. On sait que $P(X = k) = C_n^k p^k q^{n-k}$, pour tout $k \in \{0, 1, \dots, n\}$. Alors, par déf.

$$E[X] = \sum_{k=0}^n k \cdot P(X = k) = \sum_{k=1}^n k \cdot C_n^k p^k q^{n-k}$$

On remarque que :

$$\begin{aligned} k \cdot C_n^k &= \frac{kn!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \\ &= n \cdot C_{n-1}^{k-1} \end{aligned}$$

D'où :

$$\begin{aligned} E[X] &= n \sum_{k=1}^n k \cdot C_n^k p^k q^{n-k} = \left[j = k - 1; \begin{array}{l} k = 1 \rightarrow j = 0 \\ k = n \rightarrow j = n - 1 \end{array} \right] \\ &= n \sum_{j=0}^{n-1} C_{n-1}^j p^j q^{n-1-j} = np \sum_{j=0}^{n-1} C_{n-1}^j p^j q^{n-1-j} = np(p + q)^{n-1} \\ &= np \end{aligned}$$

De même, on peut prouver $\text{Var}[X] = npq$. (TD) □

2.4. Simulation de variables aléatoires discrètes.

Vouloir utiliser un ordinateur pour obtenir des nombres aléatoires apparaît paradoxal, sinon impossible : par définition, un nombre aléatoire n'est pas prévisible, tandis que l'ordinateur ne peut appliquer qu'une formule prédéfinie (un algorithme). Pour cela, on utilise ce qu'on appelle les *nombres pseudo-aléatoires*, générés à partir de certains données de l'ordinateur, comme p.ex, en fonction du nombre de millisecondes de l'horloge de l'ordinateur au moment où l'on commence la simulation.

L'intérêt est pourtant grand, car bien des applications utilisent des nombres aléatoires :

- sécurité informatique (génération automatique d'identifiants, de clés secrètes),
- méthodes d'optimisation dans des espaces de grande dimension (algorithmes génétiques),
- simulations numériques de systèmes complexes (physique, ingénierie, finance, assurance, météo. . .),
- jeux vidéos (paysages aléatoires, intelligence artificielle,. . .),

2.5. Une variable aléatoire continue : la loi normale.

Dans cette section, I désigne un intervalle $I \subset \mathbb{R}$ (borné ou non).

On a étudié que des expériences aléatoires avec un univers "discrète" muni d'une loi de probabilité P . Toute variable aléatoire ne prenait alors qu'un nombre fini de valeurs. P. ex., la "probabilité d'obtenir face ou pile dans le lancé d'une pièce".

Cependant, certaines expériences aléatoires conduisent à utiliser des variables aléatoires qui prennent toutes les valeurs d'un intervalle I de \mathbb{R} , p. ex. mesurer la taille ou le poids d'un homme de 19 ans prise au hasard dans une population.

Pendant le cours *Statistique Descriptive* (S2), on a fait la différence entre variables quantitatives "discrètes" et "continues". Pour définir la notion de fréquence dans les échantillons des "continues", on devait découper l'ensemble des modalités en intervalles disjoints d'une même amplitude, appelés *classes*.

Imaginons qu'on représente dans un histogramme l'échantillon donné par les tailles X des hommes de 19 ans dans une population. Au fur et au mesure qu'on augmente la taille de l'échantillon, on peut définir des classes avec des amplitudes plus petits : l'histogramme s'approchera vers une courbe $y = f(x)$ qui va modéliser la fréquence relative de X (Figure 6) :

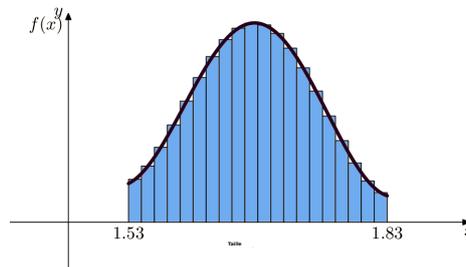


FIGURE 6. Taille.

Définition 2.22. On appelle *fonction de densité* sur un I toute fonction $f : I \rightarrow \mathbb{R}$ telle que :

- (1) f est positive et continue sur I (éventuellement, continue par morceaux).
- (2) $\int_I f(x)dx = 1$.

Remarque 2.23. Lorsque I est non-borné, p. ex. $I = [a, +\infty[$, on définit les intégrales par le passage au limite :

$$\int_I f(x)dx = \int_a^{+\infty} f(x)dx = \lim_{M \rightarrow +\infty} \int_a^M f(x)dx$$

(notion d'intégrale généralisée, Intégrales généralisées et multiples, S3)

Définition 2.24. On appelle *variable aléatoire continue (v.a.c.)* sur I toute application

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

telle que :

- (1) L'univers image $X(\Omega) = I$.
- (2) Pour tout J sous-intervalle de I , $\{\omega \in \Omega \mid X(\omega) \in J\} \in \mathcal{A}$ (dénnoté simplement par $\{X \in J\}$).
- (3) Il existe une fonction de densité f sur I telle que, pour tout sous-intervalle $J \subset I$:

$$P(X \in J) = \int_J f(x)dx.$$

On dira que X suit la loi de densité f sur I .

Remarque 2.25.

- (1) On utilisera (abusivement) les notations $\{a \leq X \leq b\}$, $\{a > X\}$, $\{X \leq b\}$ au lieu de $\{X \in [a, b]\}$, $\{X \in]a, +\infty[\}$, $\{X \in]-\infty, b]\}$, respectivement.
- (2) $P(X \in J) \geq 0$ car f est positive, et ce nombre correspond à l'aire de la région délimitée sous la courbe intégrale par J (Figure 7) :

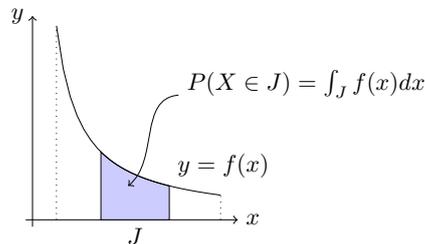


FIGURE 7

On a aussi que $P(X \in J) \leq 1$, car $\int_I f(x)dx = 1$, par définition.

- (3) Notez que, dans le cas des v.a. continues :
 - $P(X = a) = \int_a^a f(x)dx = 0, \forall a \in I$.
 - $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b), \forall a, b \in I$.

Dans le cas continu, on peut aussi définir la notion d'espérance et la variance.

Définition 2.26. Soit X une v.a.c. de loi de densité f sur I . On appelle *espérance* de X au nombre réel

$$E[X] = \int_I xf(x)dx$$

Plus généralement, si ϕ est une fonction réelle définie sur I :

$$E[\phi(X)] = \int_I \phi(x)f(x)dx$$

De la même façon, on appelle *variance* de X au nombre réel $\text{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$

Exemple 2.27 (Loi uniforme). On va définir la loi de densité équivalente à l'équiprobabilité dans le cas continu.

Une v.a.c. X suit une *loi uniforme* sur $I = [a, b]$, noté $X \sim \mathcal{U}([a, b])$, si la fonction de densité associé a X vient donnée par

$$f(x) = \frac{1}{b-a} = \text{cte}, \quad \forall x \in [a, b]$$

On voit bien que $\int_a^b \frac{dx}{b-a} = 1$. Pour tout $J = [c, d] \subset I = [a, b]$:

$$P(X \in J) = P(c \leq X \leq d) = \int_c^d \frac{dx}{b-a} = \frac{1}{b-a} \int_c^d dx = \frac{c-d}{b-a} = \frac{\text{longueur de } J}{\text{longueur de } I}$$

Exo : Calculer $E[X]$ et $\text{Var}[X]$.

On va définir une loi de densité fortement associé à la loi Binomiale : la loi Normale, une des plus adaptées pour modéliser des phénomènes naturels, p.ex. tailles anatomiques des animaux, chute d'objets, mesure d'erreurs dans des expériences physiques, distribution des notes dans un examen,...

Définition 2.28. Une v.a.c. X suit une *loi normale de paramètres* $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^*$, notée $X \sim \mathcal{N}(\mu, \sigma^2)$, si la fonction de densité associé a X vient donnée par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}.$$

En particulier, $X \sim \mathcal{N}(0, 1)$ est appelé la *loi normale centrée réduite*.

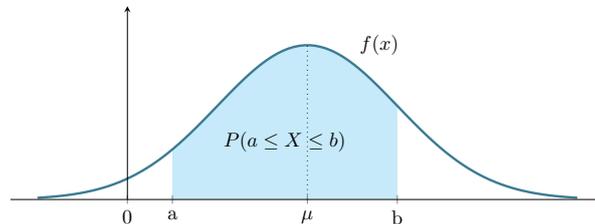


FIGURE 8

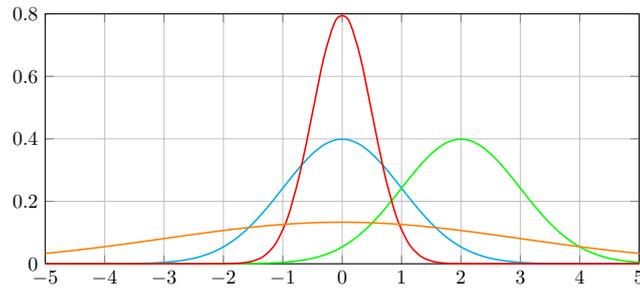
Propriété 2.29. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $E[X] = \mu$ et $\text{Var}[X] = \sigma^2$.

Remarque 2.30.

- 1) On assume que $f(x)$ correspond bien à une densité de probabilité (en particulier, $\int_{-\infty}^{+\infty} f(x)dx = 1$).
- 2) $f(x)$ est paire par rapport à $x = \mu$. Pour des différents valeurs de $\sigma \in \mathbb{R}^*$, on obtient un *allongement* ($\sigma < 1$) ou un *aplatissement* ($\sigma > 1$) de la graphique autour la droite défini par $x = \mu$ (Figure 9).

Il n'existe pas de primitive s'exprimant avec des fonctions élémentaires pour $f(x)$. Le calcul de l'aire sous la courbe demande des méthodes numériques sur $f(x)$. Cependant, nous pouvons nous ramener toujours a l'étude de la loi normale centrée réduite :

Propriété 2.31. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, alors la v.a.c. $Z = \frac{X-\mu}{\sigma}$ vérifie $Z \sim \mathcal{N}(0, 1)$.

FIGURE 9. Les lois $\mathcal{N}(0, 1)$, $\mathcal{N}(2, 1)$, $\mathcal{N}(0, 1/4)$ et $\mathcal{N}(0, 9)$.

Pour calculer des probabilités sur une v.a.c. $Z \sim \mathcal{N}(0, 1)$, on utilisera des tables qui nous donnent des bons approximations de la fonction

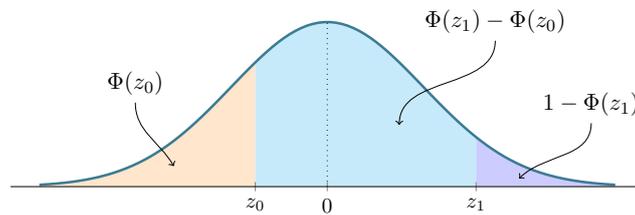
$$\begin{aligned} \Phi: \mathbb{R} &\longrightarrow [0, 1] \\ z &\longmapsto \Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx \end{aligned}$$

appelé *fonction de répartition* de la loi normale centrée réduite.

Propriété 2.32. Soient $z_0, z_1 \in \mathbb{R}$ tels que $z_0 < z_1$, alors :

- (1) $P(Z \geq z_0) = 1 - \Phi(z_0) = \Phi(-z_0)$.
- (2) $P(z_0 \leq Z \leq z_1) = \Phi(z_1) - \Phi(z_0)$.

On apprêtera en TD à calculer des probabilités $P(Z \in J)$ en utilisant une table d'approximations de $\Phi(z)$ et les propriétés précédentes (Figure 10).

FIGURE 10. Calcul d'aires dans une $\mathcal{N}(0, 1)$ en utilisant $\Phi(z)$.

Exemple 2.33. Soit $X \sim \mathcal{N}(5, 4)$:

- (1) Déterminer les probabilités $P(X \leq 8)$, $P(X \leq 2)$:

On a vu dans le cours que la v.a. $Z = \frac{X-5}{2}$ suit une loi normal centrée réduite $\mathcal{N}(0, 1)$.
Donc

$$P(X \leq 8) = P\left(\frac{X-5}{2} \leq \frac{8-5}{2}\right) = P(Z < 1.5) = \Phi(1.5) = 0.9332$$

De même,

$$\begin{aligned} P(X \leq 2) &= P\left(\frac{X-5}{2} \leq \frac{2-5}{2}\right) = P(Z < -1.5) \\ &= \Phi(-1.5) = 1 - \Phi(1.5) = 0.0668 \end{aligned}$$

- (2) En déduire la valeur de $P(2 \leq X \leq 8)$:

On a

$$\begin{aligned} P(2 < X < 8) &= P\left(\frac{2-5}{2} < \frac{X-5}{2} < \frac{8-5}{2}\right) = P(-1.5 < Z < 1.5) \\ &= \Phi(1.5) - \Phi(-1.5) = 0.9332 - 0.0668 = 0.8664 \end{aligned}$$

Du à la symétrie de la loi normal $X \sim \mathcal{N}(\mu, \sigma^2)$ autour $E[X] = \mu$ et au fait que $\text{Var}[X] = \sigma^2$, les paramètres μ et σ nous donnent des intervalles dépendant de μ et σ qui nous donnent des valeurs fixes de probabilité. Classiquement, en prenant des multiples entières (Figure 11) :

Proposition 2.34.

- (1) $P(X \in [\mu - \sigma, \mu + \sigma]) \simeq 0.683$.
- (2) $P(X \in [\mu - 2\sigma, \mu + 2\sigma]) \simeq 0.954$.
- (3) $P(X \in [\mu - 3\sigma, \mu + 3\sigma]) \simeq 0.997$.

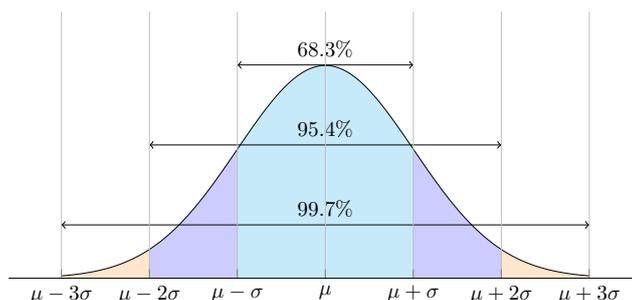


FIGURE 11

On peut faire le chemin inverse, en fixant une probabilité d'erreur $\alpha \in]0, 1[$, on peut trouver un intervalle symétrique autour la moyenne correspondant à cette proba.

Proposition 2.35. Soit $Z \sim \mathcal{N}(0, 1)$. Pour tout $\alpha \in]0, 1[$, il existe un réel $z_{\alpha/2} > 0$ t.q.

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

On appelle $1 - \alpha$ le seuil et $z_{\alpha/2}$ la valeur critique correspondante.

Remarque 2.36. Les concepts précédents seront très utiles à l'heure de construire des intervalles centrés dans la moyenne pour lesquels les résultats des observations de v.a. "tombent" dedans avec une certaine proba $1 - \alpha$ (Figure).

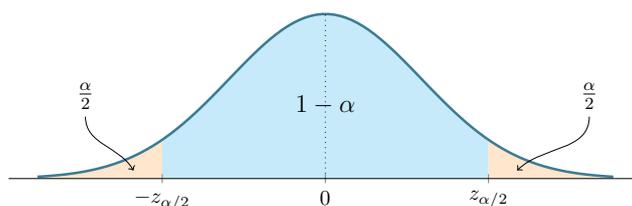


FIGURE 12

Remarque 2.37. Notez que $P(Z \leq -z_{\alpha/2}) = P(Z \geq z_{\alpha/2}) = \alpha/2$.

Classiquement en statistique, on prend des seuils du 95% ou 99% :

Propriété 2.38. $z_{0.025} \simeq 1.96$ et $z_{0.005} \simeq 2.58$.

Exemple 2.39. Calculer approximativement la valeur critique $z_{\alpha/2}$ pour $\alpha = 0.1$:

On veut trouver $z_{0.05} > 0$ tel que $P(Z > z_{0.05}) = 0.05$. On peut exprimer $P(Z > z_{0.05}) = 1 - P(Z \leq$

$z_{0.05}) = 1 - \Phi(z_{0.05})$. On doit donc chercher dans la table de la normal une valeur approx. qui vérifie

$$\Phi(z_{0.05}) = 1 - 0.05 = 0.95$$

On remarque que $\Phi(1.64) \simeq 0.9495$ et $\Phi(1.65) \simeq 0.9505$, donc on prendra la moyenne comme valeur approx. $\Phi(1.645) \simeq 0.95$. D'où $z_{0.05} \simeq 1.645$ et

$$P(-1.645 < Z < 1.645) \simeq 0.9$$

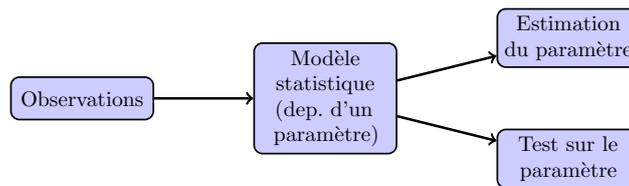
3. ESTIMATION PONCTUELLE D'UNE PROPORTION

3.1. Modèle statistique.

On sait qu'on observe un succès avec proba p dans une épreuve de Bernoulli. Mais on ne connaît pas la valeur de p et on souhaite l'estimer à partir des observations ou bien on souhaite vérifier la valeur de p à partir des observations. On est en face à deux problèmes :

- (1) Estimation statistique de p .
- (2) Test statistique sur p .

En résumé et de manière générale :



3.2. Estimation du modèle.

On suppose que les n observations sont les réalisations de n variables aléatoires *indépendantes* de la même loi de Bernoulli de paramètre p .

Observations : $x_1, \dots, x_n \in \{0, 1\}^n$ ($x_i = 1$ corresp. à un succès)

Vars. aléatoires : $X_1, \dots, X_n \sim \mathcal{B}(p)$

On construit une *statistique* $S_n = f(X_1, \dots, X_n)$ qui soit informative par rapport à l'estimation de p ou bien par rapport à un test sur p .

Remarque 3.1. S_n est aussi une variable aléatoire : si x_1, \dots, x_n sont des réalisations de X_1, \dots, X_n (observations), on dénotera par $s_n = f(x_1, \dots, x_n)$ la réalisation correspondante à S_n .

On cherche de donner une estimation ponctuelle de p . Précédemment, on a vu que si

$$S_n = \sum_{i=1}^n X_i$$

alors $S_n \sim \mathcal{B}(n, p)$ (une binomiale!). D'où, on peut calculer facilement la "valeur attendu" de S_n par l'espérance :

$$E[S_n] = np \quad \text{et} \quad \text{Var}[S_n] = np(1 - p)$$

Donc, un estimateur naturel de p vient donné par

$$\hat{p}_n = \frac{S_n}{n}$$

Question. Comment juger la qualité d'un estimateur ? Qu'est-ce qu'est un bon estimateur ?

Proposition 3.2. \hat{p}_n est un estimateur sans biais, i.e. $E[\hat{p}_n] = p$.

Démonstration. On a $E[\hat{p}_n] = E\left[\frac{S_n}{n}\right] = \frac{1}{n} E[S_n] = \frac{1}{n} np = p$. □

Proposition 3.3. $\text{Var}[\hat{p}_n] = \frac{p(1-p)}{n} \xrightarrow[n \rightarrow \infty]{} 0$.

Démonstration. (Rappel $\text{Var}[aX + b] = a^2 \text{Var}[X]$)
On a $\text{Var}[\hat{p}_n] = \text{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2} \text{Var}[S_n] = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$. □

Remarque 3.4. Interprétation : La variance est une *mesure de la dispersion* d'une v.a. autour son espérance, la *valeur moyenne attendue*. Plus la variance est petite, plus la loi de la variable aléatoire est concentrée autour l'espérance.

Ici, plus l'échantillon est de taille grande, plus \hat{p}_n est concentrée sur p .

3.3. L'inégalité de Bienaymé-Tchebychev.

Dans un cadre général, qu'est-ce qu'on peut considérer comme "convergence" d'une suite de v.a. ? On devra avoir en compte le caractère aléatoire de la suite. Comment est-ce qu'on peut mesurer ce convergence en utilisant des mesures de dispersion (p.e. la variance) ?

Proposition 3.5 (Inégalité de Bienaymé-Tchebychev). Soit X une v.a.d. et soit $k \in \mathbb{N}^*$ t.q. $E[|X|^k] < \infty$. Alors $\forall \varepsilon > 0$:

$$(0 \leq) P(|X| > \varepsilon) \leq \frac{E[|X|^k]}{\varepsilon^k}$$

Corollaire 3.6. Soit X une v.a.d. t.q. $\text{Var}[X] < \infty$. Alors

$$\forall \varepsilon > 0, \quad P(|X - E[X]| > \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

Démonstration. On applique l'inég. de B-T à la v.a.d. $X - E[X]$ et on prends $k = 2$. □

Démonstration de l'inég. de B-T. Par déf.

$$\begin{aligned} E[|X|^k] &= \sum_{x \in X(\Omega)} |x|^k P(X = x) = \underbrace{\sum_{\substack{x \in X(\Omega) \\ |x| < \varepsilon}} |x|^k P(X = x)}_{\geq 0} + \sum_{\substack{x \in X(\Omega) \\ |x| \geq \varepsilon}} |x|^k P(X = x) \\ &\geq \sum_{\substack{x \in X(\Omega) \\ |x| \geq \varepsilon}} \underbrace{|x|^k}_{\geq \varepsilon^k} P(X = x) \geq \varepsilon^k \sum_{\substack{x \in X(\Omega) \\ |x| \geq \varepsilon}} P(X = x) = \varepsilon^k P(X \geq \varepsilon) \end{aligned}$$

D'où, car $\varepsilon > 0$:

$$P(|X| > \varepsilon) \leq \frac{E[|X|^k]}{\varepsilon^k} \quad \square$$

L'inég. de B-T donne un moyen d'évaluer la distance entre les valeurs prises par une v.a. et son espérance, en donnant une majoration de la probabilité que l'écart soit grand par la variance. On peut donc donner une notion de convergence pour des v.a. (ce n'est pas la seule notion qui existe ! : *Intégration et calcul des probabilités*, S5).

Définition 3.7. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. représentant la répétition une \hat{m} . exp. al. On dit que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la v.a. X , noté $X_n \xrightarrow{pr.} X$ si,

$$\forall \varepsilon > 0, P(|X_n - X| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Est-ce que notre estimateur \hat{p}_n vérifie cette convergence sur le param. p ?

Proposition 3.8. $\hat{p}_n \xrightarrow{pr.} p$.

Démonstration. On applique le corollaire sur \hat{p}_n et :

$$\forall \varepsilon > 0, P(|\hat{p}_n - E[\hat{p}_n]| > \varepsilon) \leq \frac{\text{Var}[\hat{p}_n]}{\varepsilon^2}$$

D'où, car $E[\hat{p}_n] = p$ et $\text{Var}[\hat{p}_n] = \frac{p(1-p)}{n}$,

$$0 \leq P(|\hat{p}_n - p| > \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0$$

En utilisant le Thm. des gendarmes, on a donc $P(|\hat{p}_n - p| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0$. □

Remarque 3.9. Dans la dém. précédente, on a utilisé le fait que

$$P(|\hat{p}_n - p| > \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$

On remarque que $\{|\hat{p}_n - p| > \varepsilon\} = \{p \notin [\hat{p}_n - \varepsilon, \hat{p}_n + \varepsilon]\}$, d'où

$$P(p \in [\hat{p}_n - \varepsilon, \hat{p}_n + \varepsilon]) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

En prenant la proba $\alpha = \frac{p(1-p)}{n\varepsilon^2}$, on peut déterminer des extrêmes de l'intervalle précédente en fonction de α :

$$\varepsilon^2 = \frac{p(1-p)}{n\alpha} \iff_{\varepsilon > 0} \varepsilon = \sqrt{\frac{p(1-p)}{n\alpha}}$$

Alors, on obtiens :

$$P\left(p \in \left[\hat{p}_n - \sqrt{\frac{p(1-p)}{n\alpha}}, \hat{p}_n + \sqrt{\frac{p(1-p)}{n\alpha}}\right]\right) \geq 1 - \alpha$$

On a donc construit à partir de \hat{p}_n un intervalle (aléatoire!) “qui contient le paramètre p avec une probabilité sup. à $1 - \alpha$ ”. Peut-on donner des constructions similaires plus précises, à partir de l'estimateur, pour un seuil $1 - \alpha$ donné ?

4. ESTIMATION D'UNE PROPORTION À L'AIDE D'UN INTERVALLE DE CONFIANCE

On considère encore n réalisations (observations) de n variables aléatoires *indépendantes* de la même loi de Bernoulli de paramètre p , et l'estimateur \hat{p}_n . En utilisant une approximation des v.a. binomiales par la loi normal, on va construire des intervalles à partir de l'estimateur qui vont nous aider à tester notre modèle statistique : ou bien à estimer la valeur de p , ou bien à tester une certaine hypothèse sur p , dans tout les deux avec un certain seuil mesuré par la proba.

4.1. Théorème de Moivre-Laplace.

Pour un nombre n suffisamment grand, on va démontrer qu'on peut approcher une v.a. binomiale $\mathcal{B}(n, p)$ pour une normale, en utilisant une autre notion de convergence basée sur la convergence mathématique des fonctions de répartition,

Théorème 4.1 (de Moivre-Laplace). Soit X_n une v.a. suivant une loi binomiale $\mathcal{B}(n, p)$. On pose $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$, alors pour tout intervalle $J = [a, b]$ de \mathbb{R} :

$$P(Z_n \in J) = P(a \leq Z_n \leq b) \xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(b) - \Phi(a)$$

où $\Phi(z)$ est la fonction de répartition d'une v.a. normale centrée réduite.

On va utiliser le Thm. de Moivre-Laplace pour construire des intervalles soit *de fluctuation* soit *de confiance* sur le paramètre p d'une épreuve de Bernoulli, à partir des observations indépendantes. Rappelons que si X_1, \dots, X_n des v.a. indép. de loi $\mathcal{B}(p)$, alors $\sum_{i=1}^n X_i \sim \mathcal{B}(n, p)$.

4.2. Intervalle de fluctuation asymptotique.

En modélisant une épreuve de Bernoulli de paramètre p . Supposons qu'on est dans des cas où :

- a) on connaît le paramètre p ,
- b) on a formulé une hypothèse sur sa valeur,

et on veut, ou bien vérifier la valeur du paramètre, ou bien construire des échantillons bien distribués d'une population. (TEST SUR LE PARAMÈTRE).

En utilisant le Thm. de Moivre-Laplace, on va déterminer quels sont les "variations dues au hasard" qu'on obtient sur \hat{p}_n pour des échantillons de grande taille n , avec un certaine probabilité $1 - \alpha$.

Théorème 4.2. Soient X_1, \dots, X_n des v.a. indép. de loi $\mathcal{B}(p)$, et $\alpha \in]0, 1[$ fixé. Posons $\hat{p}_n = \frac{X_1 + \dots + X_n}{n}$. Alors

$$P\left(\hat{p}_n \in \left[p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}\right]\right) \xrightarrow{n \rightarrow \infty} 1 - \alpha$$

où $z_{\alpha/2} > 0$ est l'unique valeur qui vérifie.

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \quad Z \sim \mathcal{N}(0, 1).$$

Démonstration. On sait que $Y_n = \sum_{i=1}^n X_i$ suit une loi $\mathcal{B}(n, p)$. On pose $Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}$ et on applique le Thm. de Moivre-Laplace pour l'intervalle $[-z_{\alpha/2}, z_{\alpha/2}]$:

$$P(Z_n \in [-z_{\alpha/2}, z_{\alpha/2}]) \xrightarrow{n \rightarrow \infty} \int_{-z_{\alpha/2}}^{z_{\alpha/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha.$$

Or

$$\begin{aligned}
 Z_n \in [-z_{\alpha/2}, z_{\alpha/2}] &\iff -z_{\alpha/2} \leq \frac{Y_n - np}{\sqrt{np(1-p)}} \leq z_{\alpha/2} \\
 &\iff -z_{\alpha/2} \sqrt{np(1-p)} \leq Y_n - np \leq z_{\alpha/2} \sqrt{np(1-p)} \\
 &\iff np - z_{\alpha/2} \sqrt{np(1-p)} \leq Y_n \leq np + z_{\alpha/2} \sqrt{np(1-p)} \\
 &\iff p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq \frac{Y_n}{n} \leq p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \\
 &\iff \frac{Y_n}{n} \in \left[p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right],
 \end{aligned}$$

et par déf. $\hat{p}_n = \frac{Y_n}{n} = \frac{X_1 + \dots + X_n}{n}$, d'où le résultat. \square

Définition 4.3. On appelle $I_n = \left[p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$ l'intervalle de fluctuation asymptotique de \hat{p}_n au seuil de $1 - \alpha$.

Remarque 4.4. On admet que, sous certaines conditions, on peut approcher

$$P(\hat{p}_n \in I_n) \simeq 1 - \alpha$$

En pratique, on permet l'approximation si

$$\boxed{n \geq 30, \quad np \geq 5, \quad \text{et} \quad n(1-p) \geq 5}$$

Classiquement, on utilise ces au seuil de 95% et 99%, i.e. avec $\begin{cases} \alpha = 0.05 \rightarrow z_{0.025} \simeq 1.96 \\ \alpha = 0.01 \rightarrow z_{0.005} \simeq 2.58 \end{cases}$.

Exemple 4.5. Expérience "pile ou face" avec une pièce équilibré : on suppose $p = 1/2$. On va lancer la pièce $n = 200$ fois. Tout d'abord, on va tester si on vérifie les conditions pour prendre l'approximation :

$$n = 200 \geq 30, \quad np = n(1-p) = 200 \cdot 0.5 = 100 \geq 5$$

Pour $n = 200$, on obtient l'intervalle de fluctuation asymptotique de \hat{p}_{200} au

$$\begin{aligned}
 - \text{seuil du 95\%} : I_{200} &= \left[\frac{1}{2} - 1.96 \sqrt{\frac{1/2(1-1/2)}{200}}, \frac{1}{2} + 1.96 \sqrt{\frac{1/2(1-1/2)}{200}} \right] = [0.43, 0.57]. \\
 - \text{seuil du 99\%} : I_{200} &= \left[\frac{1}{2} - 2.58 \sqrt{\frac{1/2(1-1/2)}{200}}, \frac{1}{2} + 2.58 \sqrt{\frac{1/2(1-1/2)}{200}} \right] = [0.41, 0.59].
 \end{aligned}$$

Donc, au bout de 200 observations, la fréquence observé de \hat{p}_{200} "tombera" dans $[0.43, 0.57]$ et dans $[0.41, 0.59]$ un 95% et 99% des fois, respectivement.

Remarque 4.6. Le fait d'obtenir une valeur en dehors de cet intervalle s'interprète alors en mettant en cause la représentativité de l'échantillon ou la valeur de p .

ATTENTION : À l'inverse, le fait que la moyenne soit comprise dans l'intervalle n'est pas une garantie de la validité de l'échantillon ou du modèle.

4.3. Intervalle de confiance asymptotique.

On se place dans le cas où on a aucune information sur p et rien nous permet de faire une hypothèse : on veut donc faire une ESTIMATION SUR LE PARAMÈTRE, avec un certain niveau de confiance.

Dans ce cas, on veut construire un intervalle à partir des observations tel que la vraie valeur de

p soit contenu dedans avec une grande probabilité. En général, ce problème est compliqué car l'expression de l'intervalle pour \hat{p}_n :

$$z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

dépend de p . Une approche classique (Terminale) est de considérer, pour un seuil de 95% :

$$p(1-p) \leq \frac{1}{4} \implies \sqrt{p(1-p)} \leq \frac{1}{2} \implies 1.96 \sqrt{p(1-p)} \leq 1$$

donc,

$$P\left(p \in \left[\hat{p}_n - \frac{1}{\sqrt{n}}, \hat{p}_n + \frac{1}{\sqrt{n}}\right]\right) \geq 0.95$$

mais c'est en générale une mauvaise approximation, l'intervalle est trop grand !

Définition 4.7. On appelle

$$I_n = \left[\hat{p}_n - 1.96 \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + 1.96 \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right]$$

l'intervalle de confiance asymptotique de p au seuil de 0.95.

À partir du Thm. de Moivre-Laplace et en utilisant des utiles d'approximation des lois normales, on peut montrer qu'on peut approcher

$$P(p \in I_n) \simeq 0.95$$

au seuil de 95%, en vérifiant les conditions :

$$\boxed{n \geq 30, \quad n\hat{p}_n \geq 5, \quad \text{et} \quad n(1-\hat{p}_n) \geq 5}$$

Exemple 4.8. On dispose d'une urne avec des boules rouges et vertes. On réalise un tirage de 100 boules et on obtiens 59 rouges et 41 vertes. Alors, la fréquence observé d'apparition du caractère "boule rouge" est de $\hat{p}_{100} = 0.59$. Les conditions d'approximation $100 \geq 30$, $100 \cdot 0.59 = 59 \geq 5$, $100 \cdot 0.41 = 41 \geq 5$ sont vérifiées donc on peut construire l'intervalle de confiance au seuil de 95% pour la proportion des boules rouges p dans l'urne :

$$I_{100} = \left[0.59 - 1.96 \sqrt{\frac{0.59(1-0.59)}{100}}, 0.59 + 1.96 \sqrt{\frac{0.59(1-0.59)}{100}} \right] = [0.493, 0.685]$$

LABORATOIRE DE MATHÉMATIQUES ET DE LEURS APPLICATIONS UMR CNRS 5142 BÂTIMENT IPRA - UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR AVENUE DE L'UNIVERSITÉ - BP 1155 64013 PAU CEDEX
E-mail address : juan.viusos@univ-pau.fr