

Probabilités-Statistique pour la santé - 2022 – 2023

Clément Marteau - Thibault Espinasse - Mathieu Fauvernier

Ce résumé a vocation à être un aide mémoire, à compléter, et éventuellement à corriger. Si vous trouvez des coquilles, merci de contacter l'adresse mail mathieu.fauvernier@chu-lyon.fr pour les corriger. Il s'agit évidemment d'une version provisoire. Quelques références en suppléments

- Le livre “Cours et exercices corrigés de statistiques inférentielles” d'Olivier Marchal (rigoureux et en même temps assez facile d'accès car très bien écrit. Le présent cours s'en inspire largement)
- Le livre “Statistique mathématique, cours et exercices corrigés” de C. Vial et B. Cadre
- Le polycopié de M. Delarue de Statistique : <https://math.unice.fr/~delarue/CoursL2Stats.pdf>
- Le site internet Wikistat très riche : <http://wikistat.fr/>
- Le livre “Probabilités, analyse des données et statistique” de Gilbert Saporta
- Le livre “Intégration, Probabilités et Processus Aléatoires” de Jean-François Le Gall (présentation très rigoureuse du cadre probabiliste)

En outre, pour celles et ceux que ça intéresse, voici quelques chaînes youtube de vulgarisation qui peuvent donner un éclairage supplémentaire sur le cours.

1. StatQuest (un des meilleurs sur les stats, c'est en anglais certes mais c'est ultra pédagogique et avec illustration)
2. La statistique expliquée à mon chat
3. Le chat sceptique
4. Science4all
5. ScienceÉtonnante
6. Mickaël Launay
7. Hygiène Mentale
8. El Jj
9. Data Gueule
10. 3Blue1Brown
11. François Husson
12. ...

Sommaire

1	Introduction (L1)	3
1.1	Univers, événement, et probabilité	3
2	Indépendance et conditionnement d'événements (L1)	4
2.1	Conditionnement d'événements	4
2.2	Indépendance d'événements	5
2.3	Formule des probabilités totales et formule de Bayes	6
3	Variables aléatoires réelles (L1)	6
3.1	Variables aléatoires discrètes : introduction	6
3.1.1	Motivations	6
3.1.2	Variables aléatoires discrètes usuelles	6
3.2	Variables aléatoires continues : introduction	7
3.2.1	Motivations	7
3.2.2	Variables aléatoires à densité usuelles	8
3.3	Cas général	8
3.3.1	Définitions	8
4	Espérance et variance (L1)	10
5	Vers l'inférence statistique : théorèmes limites (L2)	13
6	Inférence statistique (L2)	18
6.1	Estimateur	18
6.2	Propriétés des estimateurs	19
6.2.1	Estimateurs sans biais	19
6.3	Méthode des moments	19
6.4	Méthode du maximum de vraisemblance	20
6.5	Intervalles de confiance	23
6.5.1	Intervalle de confiance d'une proportion	23
6.5.2	Intervalle de confiance d'une moyenne	23
6.6	Tests d'hypothèse	24
6.6.1	Comparaison d'une proportion à une valeur théorique	24
6.6.2	Comparaison d'une moyenne à une valeur théorique	26
6.6.3	Comparaison de deux proportions (échantillons indépendants)	27
6.6.4	Comparaison de deux moyennes (échantillons indépendants)	27
7	Exercices	28
7.1	Espace de probabilité	28
7.2	Conditionnement	28
7.3	Variables aléatoires à densité	28
7.4	Avancé	29
7.5	Maximum de vraisemblance	30
7.6	Intervalles de confiance et test d'hypothèses	31
8	Rappel sur quelques bases de mathématiques	31
8.1	Notations	31
8.2	Rappel sur les ensembles	31
8.3	Calculs d'intégrales	32
8.4	Séries	32

1 Introduction (L1)

De façon générale, lorsqu'on parle de probabilités, d'événement aléatoire, ou de hasard, on cherche à prendre en considération *l'imprévisibilité* d'une expérience (ex : lancer de pièce, guérison d'une maladie, durée de vie d'un téléphone portable, ordre des cartes dans un paquet mélangé). Que l'aléa soit réel ou modélise notre ignorance à propos du phénomène en question importe peu pour les calculs que l'on souhaite faire.

Pour pouvoir néanmoins apprendre d'expériences passées sur ce qui pourrait se passer sur les futures expériences, on parlera souvent de "réaliser de nombreuses fois l'expérience dans des conditions similaires". la probabilité pouvant être comprise comme la *fréquence* (i.e. proportion des expériences) limite d'un résultat, *lorsqu'on réalise l'expérience dans les mêmes conditions un très grand nombre de fois*. On reviendra sur cette notion lorsqu'on évoquera la **Loi des grands nombres** qui décrit mathématiquement cette propriété.¹

Pour formaliser la notion de probabilité, et de variable aléatoire, en mathématique, on va devoir décrire l'ensemble des résultats possibles d'une expérience, ainsi que la probabilité de chaque résultat. L'objet de la première partie de ce cours, est d'introduire toutes ces notions.

1.1 Univers, événement, et probabilité

Définition 1.1 On appelle **univers** (noté Ω) l'ensemble des résultats possibles d'une expérience aléatoire.

Remarque : On en reparlera plus tard, mais en réalité, le choix de l'univers est un *choix de modélisation*, et il n'y a pas qu'une seule façon de modéliser une expérience aléatoire. On y reviendra en parlant de variables aléatoires, mais reprenez déjà cette remarque.

- Exemple 1.1**
1. Pour un lancer d'une pièce, on pourra prendre $\Omega = \{Pile, Face\}$
 2. Pour le résultat d'un dé, on pourra prendre $\Omega = \{1, 2, 3, 4, 5, 6\}$
 3. Pour l'âge d'une personne prise au hasard, on pourra prendre $\Omega = \{0, 1, 2, \dots, 150\}$, ou $\Omega = \mathbb{N}$ ou encore $\Omega = \mathbb{R}^+$ selon si l'on considère la mesure bornée, entière, ou réelle.
 4. Pour un lancer de deux pièces, on pourra prendre $\Omega = \{(Pile, Pile), (Pile, Face), (Face, Pile), (Face, Face)\} = \{Pile, Face\}^2$
 5. Pour un nombre au hasard entre 0 et 1, on pourra prendre $\Omega = [0, 1]$

Remarque : Comme expliqué au dessus, on verra qu'en réalité, on aura rarement besoin de connaître Ω explicitement. De plus, quand on introduira la notion de *variable aléatoire*, on verra qu'une même variable aléatoire peut-être obtenue avec des univers différents. Cela est très intuitif : on peut "simuler" un pile ou face équilibré avec un dé équilibré par exemple. L'ensemble Ω est donc un **choix** de modélisation, et une même expérience pourra être étudiée avec des Ω très différents.

Chaque résultat de l'expérience $\omega \in \Omega$ est appelé **événement ponctuel** ou **événement élémentaire**. Un **événement aléatoire** est un ensemble d'événement ponctuel, et correspond donc à une partie $A \subset \Omega$ ($A \in \mathcal{P}(\Omega)$).

Par exemple, pour un lancer de dé à 6 faces, l'événement aléatoire "le dé est tombé sur une face avec un numéro pair" correspond au sous-ensemble $\{2, 4, 6\}$ de l'ensemble $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Pour un résultat d'expérience $\omega \in \Omega$, on dira qu'un événement A est **réalisé** si $\omega \in A$. (par exemple, si l'expérience donne $\omega = 4$, alors l'événement

$$A = \text{"le dé est tombé sur une face avec un numéro pair"} = \{2, 4, 6\},$$

est réalisé car $4 \in \{2, 4, 6\}$!)

- Définition 1.2**
1. L'événement \emptyset est appelé **événement impossible**.
 2. L'événement Ω est appelé **événement certain**.
 3. Pour un événement $A \subset \Omega$, l'événement A^C (le complémentaire de l'ensemble) est appelé **événement contraire**.
 4. Pour deux événements $A, B \subset \Omega$ l'événement $A \cap B$ sera appelé **A et B**
 5. Pour deux événements $A, B \subset \Omega$ l'événement $A \cup B$ sera appelé **A ou B**
 6. Deux événements $A, B \subset \Omega$ sont dits **incompatibles** si $A \cap B = \emptyset$

¹Cette vision (réaliser l'expérience un grand nombre de fois dans les mêmes conditions) sera souvent très utile pour aider à se représenter différentes notions. Par exemple dire que l'événement "le patient est guéri" a une probabilité de 1/4 pourra se comprendre plus facilement en se représentant 1000 patients similaires pour lesquels on a en moyenne 250 guéris.

Une fois l'univers Ω défini, il ne reste plus qu'à donner la probabilité de chaque événement. Mathématiquement, certaines propriétés d'une "probabilité" sont assez intuitives : la probabilité de l'événement impossible doit être de 0, la probabilité de l'événement certain doit être de 1, la probabilité de l'union de deux événements incompatibles doit être la somme des probabilités de chacun de ces deux événements.

Lorsque Ω est fini, donner la probabilité de chacun des événements ponctuels élémentaires suffit à obtenir les autres par sommation, mais quelques subtilités existent lorsque $\Omega = \mathbb{R}$.

Nota bene : Pour des raisons qui ne seront pas explicitées dans ce cours, il n'est pas possible de définir une probabilité pour tous les éléments de $\mathcal{P}(\Omega)$ quel que soit Ω (notamment quand $\Omega = \mathbb{R}$). En toute rigueur, au lieu de travailler sur $\mathcal{P}(\Omega)$, il est nécessaire de travailler sur $\mathcal{A} \subset \mathcal{P}(\Omega)$ où \mathcal{A} est appelée une tribu (ou σ -algèbre).

Définition 1.3 Une **probabilité** sur un univers Ω , muni d'une tribu $\mathcal{A} \subset \mathcal{P}(\Omega)$ (de l'ensemble des événements auxquels on peut associer une probabilité) est une application \mathbb{P} de \mathcal{A} dans $[0, 1]$ qui vérifie

1. $\mathbb{P}(\Omega) = 1$
2. Pour toute suite $(A_n)_{n \geq 1} \in \mathcal{A}^{\mathbb{N}}$ d'événements 2 à 2 incompatibles,

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

On peut déduire de la définition les propriétés suivantes :

Propriétés 1.4 1. $\mathbb{P}(\emptyset) = 0$

2. Pour toute suite $(A_n)_{n \geq 1} \in \mathcal{A}^{\mathbb{N}}$ d'événements,

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

3. $\forall A \in \mathcal{A}, \mathbb{P}(A^C) = 1 - \mathbb{P}(A)$
4. $\forall A, B \in \mathcal{A}$, si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$
5. $\forall A, B \in \mathcal{A}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
6. Si Ω est fini ou dénombrable, alors, $\forall A \in \mathcal{A}$,

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

Remarque : à cause de la dernière propriété, lorsque Ω est fini ou dénombrable, on peut se contenter de donner la probabilité de chacun des événements élémentaires seulement. Par exemple :

1. Pour un lancer d'une pièce équilibrée, on pourra prendre $\Omega = \{Pile, Face\}$; $\mathbb{P}(\{\omega\}) = 1/2$
2. Pour le résultat d'un dé équilibré, on pourra prendre $\Omega = \{1, 2, 3, 4, 5, 6\}$; $\mathbb{P}(\{\omega\}) = 1/6$
3. Pour l'âge d'une personne prise au hasard, on pourra prendre $\Omega = \{0, 1, 2, \dots, 150\}$, la probabilité pourrait être donnée par une pyramide des âges
4. Pour un lancer de deux pièces équilibrées, on pourra prendre $\Omega = \{(Pile, Pile), (Pile, Face), (Face, Pile), (Face, Face)\} = \{Pile, Face\}^2$; $\mathbb{P}(\{\omega\}) = 1/4$

2 Indépendance et conditionnement d'événements (L1)

Maintenant que l'on dispose de la notion d'expérience aléatoire, d'univers et de probabilité, on peut aborder les notions (fondamentales pour la suite) de probabilité conditionnelle et d'indépendance.

2.1 Conditionnement d'événements

Définition 2.1 Soient $A, B \in \mathcal{A}$ deux événements, tel que $\mathbb{P}(B) > 0$. On définit la probabilité conditionnelle de A sachant B par

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

Exemple 2.1 • On lance un dé équilibré à 6 faces ($\Omega = \{1, 2, 3, 4, 5, 6\}$, $\forall \omega \in \Omega, \mathbb{P}(\omega) = \frac{1}{6}$), on peut alors calculer la probabilité que “le résultat soit pair” ($A = \{2, 4, 6\}$) sachant qu’“il est strictement supérieur à 3” ($B = \{4, 5, 6\}$) :

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{4, 6\})}{\mathbb{P}(\{4, 5, 6\})} = \frac{2/6}{3/6} = \frac{2}{3}.$$

- On imagine qu’un virus circule fortement dans le monde. Heureusement il existe un vaccin. Une personne non vaccinée a aujourd’hui 0.05% de chances d’être hospitalisée pour le virus, c’est 10 fois moins pour une personne vaccinée². Si 80% de la population est vaccinée, quelle est la probabilité qu’une personne (tirée uniformément dans la population) soit vaccinée sachant qu’elle est hospitalisée ? En définissant l’événement H : “la personne tirée est hospitalisée” et V : “la personne tirée est vaccinée” on peut calculer :

$$\begin{aligned} \mathbb{P}(H | \bar{V}) &= 0.0005 \\ \mathbb{P}(H | V) &= 0.0005(1 - 0.90) = 0.00005 \\ \mathbb{P}(V) &= 0.8 \\ \mathbb{P}(V | H) &= \frac{\mathbb{P}(H \cap V)}{\mathbb{P}(H)} \\ &= \frac{\mathbb{P}(H | V)\mathbb{P}(V)}{\mathbb{P}(H | V)\mathbb{P}(V) + \mathbb{P}(H | \bar{V})\mathbb{P}(\bar{V})} \\ &= \frac{0.00005 * 0.8}{0.00005 * 0.8 + 0.0005 * 0.2} \\ &= \frac{2}{7} \approx 28\% \end{aligned}$$

- J’ai deux enfants, dont au moins un garçon, quelle est la probabilité que l’autre soit un garçon ?

2.2 Indépendance d’événements

Définition 2.2 Soient $A, B \in \mathcal{A}$ deux événements, on dit que A et B sont des événements **indépendants** si

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

Soient $n \geq 1, A_1, \dots, A_n \in \mathcal{A}$ des événements. On dit que les $(A_i)_{i=1, \dots, n}$ sont

- Indépendants deux à deux si

$$\forall 1 \leq i, j \leq n, \quad \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

- Indépendants dans leur ensemble si

$$\forall 2 \leq k \leq n, \forall 1 \leq i_1, \dots, i_k \leq n, \quad \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}).$$

Exemple 2.2 1. On tire un jour au hasard dans l’année passée. On note A l’événement “il a fait beau ce jour là” et B l’événement “il s’agit d’un dimanche”. On suppose que $\mathbb{P}(A) = 1/3$ et $\mathbb{P}(B) = 1/7$, et que A et B sont indépendants. On a alors $\mathbb{P}(A \cap B) = \frac{1}{21}$, ce qui est intuitif si on se représente Ω à partir du calendrier.

2. On considère un gène formé de 2 allèles A et a . Calculer la probabilité, pour chaque génotype des parents, de chaque génotype des enfants.
3. On considère deux gènes possédant chacun deux allèles A, a et B, b , l’un dominant et l’autre récessif, associés à des phénotypes $\phi_{AB}, \phi_{Ab}, \phi_{aB}, \phi_{ab}$. En supposant les gènes indépendants car sur deux chromosomes différents, donner la probabilité pour chaque phénotype d’un enfant issu de deux parents hétérozygotes pour chacun des deux gènes. On peut le faire facilement en considérant

$$\begin{aligned} \mathbb{P}(AA) &= \mathbb{P}(BB) = 1/4 \\ \mathbb{P}(Aa) &= \mathbb{P}(Bb) = 1/2 \\ \mathbb{P}(aa) &= \mathbb{P}(bb) = 1/4 \\ \mathbb{P}(\phi_A) &= \mathbb{P}(AA) + \mathbb{P}(Aa) = 3/4 = \mathbb{P}(\phi_B) \\ \mathbb{P}(\phi_a) &= \mathbb{P}(aa) = 1/4 = \mathbb{P}(\phi_b), \end{aligned}$$

²Le vaccin a donc une efficacité de 90% pour les cas graves.

puis en utilisant l'indépendance des caractères :

$$\begin{aligned}\mathbb{P}(\phi_{AB}) &= \mathbb{P}(\phi_A)\mathbb{P}(\phi_B) \text{ par indépendance} \\ &= 3/4 * 3/4 = \frac{9}{16} \\ \mathbb{P}(\phi_{Ab}) &= \mathbb{P}(\phi_A)\mathbb{P}(\phi_b) \text{ par indépendance} \\ &= 3/4 * 1/4 = \frac{3}{16} \\ \mathbb{P}(\phi_{aB}) &= \mathbb{P}(\phi_a)\mathbb{P}(\phi_B) \text{ par indépendance} \\ &= 1/4 * 3/4 = \frac{3}{16} \\ \mathbb{P}(\phi_{ab}) &= \mathbb{P}(\phi_a)\mathbb{P}(\phi_b) \text{ par indépendance} \\ &= 1/4 * 1/4 = \frac{1}{16}\end{aligned}$$

2.3 Formule des probabilités totales et formule de Bayes

Les deux formules ci-dessous sont très utilisées en pratique et sont à connaître absolument !

Propriétés 2.3 Formule des probabilités totales Soit $(A_i)_{i \in I}$ une famille d'événements formant une partition de Ω , c'est-à-dire avec $\bigcup (A_i) = \Omega$ et $A_i \cap A_j = \emptyset$ pour tout $i \neq j$. Supposons de plus que $P(A_i) \neq 0$ pour tout i .

Alors, pour tout événement B

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(B|A_i)P(A_i)$$

Propriétés 2.4 Formule de Bayes Soient A et B deux événements tels que $P(B) \neq 0$.

Alors,

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

3 Variables aléatoires réelles (L1)

3.1 Variables aléatoires discrètes : introduction

3.1.1 Motivations

On reprend l'exemple d'un lancer de 2 dés indépendants. Dans ce cas, on peut vouloir s'intéresser à 2 observations différentes, par exemple, la somme des résultats. Pour éviter de devoir changer l'ensemble Ω à chaque fois que l'on veut étudier une variable différente, il est nécessaire d'introduire la notion de **variable aléatoire**.

Une variable aléatoire est définie comme une fonction de Ω dans \mathbb{R} , par exemple :

$$\begin{aligned}\Omega &= \{1, \dots, 6\}^2 \\ S : \Omega &\rightarrow \mathbb{R} \\ (\omega_1, \omega_2) &\mapsto \omega_1 + \omega_2\end{aligned}$$

Lorsque la fonction considérée (par exemple ici S qui prend 2 résultats de dés et renvoie la somme) ne peut prendre qu'un nombre fini de valeurs (ou dénombrable), on parlera de variable aléatoire discrète.

Pour décrire le comportement, ou plus précisément **la loi** de S , il suffit de connaître :

- L'ensemble des valeurs possibles que peut prendre cette variable (ex : $E = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ pour S)
- La probabilité pour chaque résultat (ex : $\mathbb{P}(S = 2) = \frac{1}{36}, \mathbb{P}(S = 3) = \frac{2}{36} \dots, \mathbb{P}(S = 7) = \frac{6}{36}, \mathbb{P}(S = 8) = \frac{5}{36}, \dots, \mathbb{P}(S = 12) = \frac{1}{36}$)

3.1.2 Variables aléatoires discrètes usuelles

On peut enfin introduire les variables aléatoires discrètes usuelles, souvent rencontrées dans des problèmes réels :

1. Loi uniforme discrète (ex : lancer de dé). X est une variable aléatoire de **loi uniforme discrète** sur $\{1, \dots, n\}$ (noté $X \sim \mathcal{U}(\{1, \dots, n\})$) si X est à valeurs dans $\{1, \dots, n\}$, et si

$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}(X = k) = \frac{1}{n}.$$

2. Loi de Bernoulli (ex : lancer de pièce truquée). Cette loi est utilisée dès que l'on considère une variable ne pouvant prendre que 2 valeurs (ex : Pile ou Face, présence ou absence de maladie, caractère fumeur ou non.... En général, on considère que ces valeurs seront 0 et 1, et on dira que X est une variable aléatoire de **loi de Bernoulli** de paramètre $p \in [0, 1]$ (noté $X \sim \mathcal{B}(p)$ ou $X \sim \text{Ber}(p)$) si X peut prendre les valeurs 0 et 1, et que

$$\mathbb{P}(X = 1) = p ; \quad \mathbb{P}(X = 0) = 1 - p.$$

Remarque : Pour tout événement A , la variable aléatoire $X = \mathbb{1}_A$ (i.e. $X(\omega) = \mathbb{1}_{\omega \in A}$) ne prend que les valeurs 0 et 1 par définition de l'indicatrice. C'est donc une variable aléatoire de loi de Bernoulli de paramètre $\mathbb{P}(A)$. Cette propriété est à retenir.

3. Loi Binomiale. Cette loi est utilisée dès que l'on modélise le nombre de "succès" à plusieurs "épreuves de Bernoulli indépendantes", par exemple lorsqu'on compte le nombre d'individus présentant un certain caractère dans une population homogène (ex : nombre de cellules qui se sont divisées sur une certaine unité de temps, dans une population fixée et homogène). X est une variable aléatoire de **loi Binomiale** de paramètres N et p (noté $X \sim \mathcal{B}(N, p)$ ou $X \sim \text{Bin}(N, p)$) si X peut prendre les valeurs $\{0, \dots, N\}$, et

$$\forall k \in \{0, \dots, N\}, \quad \mathbb{P}(X = k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

Remarque : Une variable aléatoire de Bernoulli compte le nombre de succès à une seule épreuve de Bernoulli, donc une Bernoulli de paramètre p correspond à une Binomiale de paramètres 1 et p . Cela sera parfois utile de réécrire les probabilités pour une Bernoulli sous la forme

$$\forall k \in \{0, 1\}, \quad \mathbb{P}(X = k) = p^k (1-p)^{1-k}.$$

4. Loi de Poisson (ex : nombre de cellules affichant un caractère rare dans une population homogène). Cette loi apparaît naturellement comme limite de la loi Binomiale, quand le nombre d'épreuves tend vers l'infini, tout en conservant le nombre moyen de succès. Cela permet de modéliser le nombre d'événements ponctuels arrivant durant une certaine durée, ou dans une certaine zone spatiale, lorsqu'il y a stationnarité et indépendance des arrivées. X est une variable aléatoire de **loi de Poisson** de paramètre λ (noté $X \sim \mathcal{P}(\lambda)$) si elle est à valeurs dans \mathbb{N} et vérifie

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

5. Loi Géométrique (ex : Nombre de tentative avant de réussir une expérience, réalisée dans les mêmes conditions). Cette loi représente l'instant du premier succès à une suite d'épreuves de Bernoulli indépendantes de paramètre p . X est une variable aléatoire de **loi géométrique** de paramètre p (noté $X \sim \mathcal{G}(p)$) si elle est à valeurs dans \mathbb{N}^* , et vérifie

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = p(1-p)^{k-1}.$$

3.2 Variables aléatoires continues : introduction

3.2.1 Motivations

Le formalisme précédent ne permet pas de modéliser le choix d'un nombre au hasard dans $[0, 1]$. En effet, on ne peut pas donner seulement la probabilité de chaque résultat, puisque n'importe quel nombre a une probabilité nulle de sortir.

Il va falloir trouver une autre voie. Pour comprendre cette idée, on peut partir de l'intuition qu'un tel nombre "au hasard" (uniforme) U entre 0 et 1 doit vérifier

$$\forall a < b \in [0, 1], \quad \mathbb{P}(U \in [a, b]) = b - a.$$

ou de façon équivalente, que

$$\forall t \in [0, 1], \quad \mathbb{P}(U \leq t) = t.$$

Connaitre ceci suffit à avoir une idée de comment se comporte la variable U . Pour décrire **la loi** d'une variable X quelconque, il suffit en fait de donner la fonction $F_X : t \mapsto \mathbb{P}(X \leq t)$.

Mais un cas particulier est vraiment intéressant, c'est quand F_X est dérivable et de dérivée continue, auquel cas, en posant $f_X = F'_X$, on peut écrire :

$$F_X(t) = \int_{-\infty}^t f_X(u) du.$$

Dans ce cas, plus $f_X(x)$ est élevé, plus la variable risque de tomber 'proche' de x . On dira donc qu'une variable aléatoire X est à densité, s'il existe une fonction f_X telle que

$$\forall a \leq b \in \mathbb{R}, \quad \mathbb{P}(X \in [a, b]) = \int_a^b f_X(x) dx.$$

³Certains peuvent la définir dans \mathbb{N} , en considérant $X - 1$

3.2.2 Variables aléatoires à densité usuelles

On peut aussi introduire les variables à densité usuelles :

6 Loi Uniforme Continue (ex: nombre “au hasard” entre 0 et 1). X est une variable aléatoire de loi Uniforme sur le segment $[a, b]$, $a < b$ (noté $X \sim \mathcal{U}([a, b])$) si elle est à valeurs dans $[a, b]$, et a pour densité la fonction définie par

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{x \in [a, b]}.$$

7 Loi Exponentielle (ex : temps entre 2 bursts d’un gène). Cette loi est l’analogie continue de la loi géométrique. Elle est utilisée pour modéliser des temps inter-arrivée d’événements ponctuels arrivant, lorsqu’il y a stationnarité et indépendance des arrivées. X est une variable aléatoire de **loi exponentielle** de paramètre $\lambda > 0$ (noté $X \sim \mathcal{E}(\lambda)$) si elle est à valeurs dans \mathbb{R}^+ , et a pour densité la fonction définie par

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \in \mathbb{R}^+}.$$

8 Loi Normale (ou Gaussienne) (ex : pression atmosphérique prise en une position et une date aléatoire). Elle peut modéliser de nombreuses variables continues “uni-modales” (avec un seul pic), et elle apparaît par exemple comme limite de la loi Binomiale lorsque l’on réalise un grand nombre d’épreuves de Bernoulli. Cette loi apparaît en particulier dans le Théorème Central Limite (que l’on abordera à la fin du cours), ce qui explique sa capacité pour modéliser de nombreux phénomènes réels. En biologie, de nombreuses quantités (titrages...) sont souvent modélisées comme des variables “**log-normales**”, ce qui signifie que le logarithme de ces quantités suit une loi normale. X est de **loi Normale** de paramètres m et σ^2 (noté $X \sim \mathcal{N}(m, \sigma^2)$) si elle est à valeurs dans \mathbb{R} , et a pour densité la fonction définie par

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

3.3 Cas général

3.3.1 Définitions

La notion de **variable aléatoire** est certainement la notion la plus importante de ce chapitre, et souvent la plus difficile à comprendre. La formalisation mathématique d’une variable aléatoire prend en compte, comme indiqué dans l’introduction, que plusieurs *expériences aléatoires* différentes peuvent donner la même *variable aléatoire*. Par exemple, pour jouer à Pile ou Face équilibré, on peut

- Tirer une pièce équilibrée
- Lancer un dé à 6 faces équilibré et regarder si le résultat est pair
- Tirer une carte dans un jeu uniformément, et regarder si la couleur est rouge
- Lancer 3 fois une pièce équilibrée et regarder le résultat du second lancer.
- Tirer un nombre uniforme entre 0 et 1, et le comparer à 0.5.

Pour chacun de ces cas, l’espace de probabilité serait différent :

- Tirer une pièce équilibré : $\Omega = \{Pile, Face\}$.
- Lancer un dé à 6 faces équilibré : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Tirer une carte dans un jeu uniformément : $\Omega = \{A\heartsuit, A\clubsuit, A\spadesuit, A\diamondsuit, 2\heartsuit, 2\clubsuit, 2\spadesuit, 2\diamondsuit, \dots, K\heartsuit, K\clubsuit, K\spadesuit, K\diamondsuit\}$.
- Lancer 3 fois une pièce équilibrée : $\Omega = \{P, F\}^3 = \{(P, P, P), (P, P, F), \dots, (F, F, F)\}$.
- Tirer un nombre uniforme entre 0 et 1 : $\Omega = [0, 1]$

Pour pouvoir introduire la notion de variable aléatoire sur chacun de ces espaces, on se donne la définition suivante :

Définition 3.1 Une variable aléatoire réelle X sur un univers Ω est une application de Ω dans \mathbb{R} :

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega \in \Omega &\mapsto X(\omega) \in \mathbb{R} \end{aligned}$$

Exemple 3.1 On reprend le cadre précédent pour définir la même variable aléatoire (Pile ou Face équilibré) qui vaut 1 si la pièce tombe sur Pile à partir d’espaces de probabilités différents :

- Tirer une pièce équilibré

$$\begin{aligned}\Omega &= \{Pile, Face\} \\ \mathbb{P}(\{\omega\}) &= \frac{1}{2} \\ X : \omega &\mapsto \mathbb{1}_{\omega=Pile}\end{aligned}$$

- Lancer un dé à 6 faces équilibré et regarder si le résultat est pair

$$\begin{aligned}\Omega &= \{1, 2, 3, 4, 5, 6\} \\ \mathbb{P}(\{\omega\}) &= \frac{1}{6} \\ X : \omega &\mapsto \mathbb{1}_{\omega \in \{2, 4, 6\}}\end{aligned}$$

- Tirer une carte dans un jeu uniformément, et regarder si la couleur est rouge

$$\begin{aligned}\Omega &= \{A\heartsuit, A\clubsuit, A\heartsuit, A\spadesuit, 2\heartsuit, 2\clubsuit, 2\heartsuit, 2\spadesuit, \dots, K\heartsuit, K\clubsuit, K\heartsuit, K\spadesuit\} \\ \mathbb{P}(\{\omega\}) &= \frac{1}{52} \\ X : \omega &\mapsto \mathbb{1}_{\omega \in \{A\heartsuit, A\heartsuit, 2\heartsuit, 2\heartsuit, \dots, K\heartsuit, K\heartsuit\}}\end{aligned}$$

- Lancer 3 fois une pièce équilibrée et regarder le résultat du second lancer.

$$\begin{aligned}\Omega &= \{Pile, Face\}^3 \\ \mathbb{P}(\{(\omega_1, \omega_2, \omega_3)\}) &= \frac{1}{8} \\ X : (\omega_1, \omega_2, \omega_3) &\mapsto \mathbb{1}_{\omega_2=Pile}\end{aligned}$$

- Tirer un nombre uniforme entre 0 et 1, et le comparer à 0.5.

$$\begin{aligned}\Omega &= [0, 1]; \\ \mathbb{P}([a, b]) &= b - a \\ X : \omega &\mapsto \mathbb{1}_{\omega < 1/2}\end{aligned}$$

Une variable aléatoire X définit une **loi de probabilité** \mathbb{P}_X sur (\mathbb{R}) donnée par⁴ :

$$\forall A \in \mathcal{P}(\mathbb{R}), \quad \mathbb{P}_X(A) = \mathbb{P}\left(\{\omega \in \Omega, X(\omega) \in A\}\right).$$

Remarque : à partir de maintenant, on notera tout simplement

$$\mathbb{P}(X \in A) := \mathbb{P}_X(A) \text{ et } \mathbb{P}(X = x) := \mathbb{P}_X(\{x\}).$$

La loi d'une variable aléatoire réelle est entièrement caractérisée par la fonction de répartition définie au dessous :

Définition 3.2 La fonction de répartition d'une variable aléatoire réelle X est la fonction

$$\begin{aligned}F_X : \mathbb{R} &\rightarrow [0, 1] \\ t &\mapsto \mathbb{P}(X \leq t)\end{aligned}$$

Propriétés 3.3 La fonction de répartition vérifie :

1. F_X est une fonction croissante
2. F_X tend vers 0 en $-\infty$ et vers 1 en $+\infty$
3. F_X est continue à droite.
4. $\forall a < b \in \mathbb{R}, \quad \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$

Dans ce cours, on s'intéressera à deux types de variables aléatoires :

⁴En réalité, on ne devrait pas travailler avec l'ensemble des parties de \mathbb{R} , mais c'est bien au delà du programme

1. Si X peut prendre un nombre fini ou dénombrable de valeurs $x_k, k \geq 0$, alors F_X est constante par morceaux, on dira que X est **une variable discrète**. Dans ce cas, il suffit de connaître $\mathbb{P}(X = x_k), \forall k \geq 0$ pour connaître la loi de X .
2. Si F_X est dérivable et de dérivée continue, alors, en posant $f_X = F'_X$, on a
 - $\forall x \in \mathbb{R}, \quad \mathbb{P}(X = x) = 0$
 - $\forall a < b \in \mathbb{R}, \quad \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \int_a^b f_X(x) dx.$

Dans ce cas là, on parle de **variable aléatoire continue** ou **variable aléatoire à densité**

Cela motive la définition suivante :

Définition 3.4 Soit X une variable aléatoire réelle. S'il existe une fonction f_X vérifiant

$$\forall a \leq b \in \mathbb{R}, \quad \mathbb{P}(a \leq X < b) = \int_a^b f_X(x) dx,$$

alors f_X est appelée **densité de probabilité** de la variable aléatoire X , et F_X est dérivable, de dérivée f_X .

Remarque : La probabilité que X tombe dans une région donnée correspond alors à l'aire sous la courbe de f_X dans cette région. En particulier, l'aire totale sous la courbe densité vaut 1 :

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Cette propriété est à retenir !

Définition 3.5 Soient X, Y deux variables aléatoires réelles. On dit que X et Y ont même loi si elles ont même fonction de répartition : $F_X = F_Y$. Pour des variables discrètes, il suffit que X et Y aient le même ensemble de valeurs $a_k, k \geq 0$ possibles, et que $\mathbb{P}(X = a_k) = \mathbb{P}(Y = a_k), \forall k \geq 0$. Pour des variables à densité, il suffit que $f_X = f_Y$. Si X_1, \dots, X_n ont même loi, on dira qu'elles sont **identiquement distribuées**.

Comme pour les événements, des variables aléatoires peuvent être indépendantes ou non (par exemple la taille et le poids d'un individu tiré au hasard ne sont pas indépendantes). Pour définir cette notion, nous avons besoin de parler de la loi jointe de plusieurs variables aléatoires (on parle alors de **vecteur aléatoire**, ou **couple de variable aléatoire** lorsqu'il y en a 2).

Définition 3.6 Soient (X_1, \dots, X_n) un vecteur aléatoire réel. La loi jointe de X_1, \dots, X_n est caractérisée par la fonction

$$F_{X_1, \dots, X_n} : (t_1, \dots, t_n) \mapsto \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n).$$

Définition 3.7 Soient (X_1, \dots, X_n) un vecteur aléatoire réel. On dit que X_1, \dots, X_n sont indépendantes si

$$\forall t_1, \dots, t_n \in \mathbb{R}, \quad \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \mathbb{P}(X_1 \leq t_1) \cdots \mathbb{P}(X_n \leq t_n).$$

i.e. si

$$\forall t_1, \dots, t_n \in \mathbb{R}, \quad F_{X_1, \dots, X_n}(t_1, \dots, t_n) = F_{X_1}(t_1) \cdots F_{X_n}(t_n).$$

Remarque : Lorsque les X_i sont discrètes, par exemple si X_i peut prendre les valeurs $a_k^{(i)}, k \geq 0$, il suffit de vérifier

$$\forall k_1, \dots, k_n \geq 0, \quad \mathbb{P}(X_1 = a_{k_1}^{(1)}, \dots, X_n = a_{k_n}^{(n)}) = \mathbb{P}(X_1 = a_{k_1}^{(1)}) \cdots \mathbb{P}(X_n = a_{k_n}^{(n)}).$$

Dans toute la suite, on aura souvent besoin de parler de suites de variables aléatoires réelles **indépendantes et identiquement distribuées**. On notera v.a.r. i.i.d..

4 Espérance et variance (L1)

La notion d'**espérance mathématique** d'une variable aléatoire correspond à la notion intuitive de "moyenne sur le hasard". Attention, il faut dès maintenant être très vigilant avec une confusion classique entre

- Moyenne d'un échantillon : on fait la moyenne **après** une expérience sur les résultats de l'expérience. Par exemple, je lance 4 fois un dé à 6 faces, et j'obtiens 3, 4, 2, 3, alors la moyenne de ces résultats est de $\frac{3+4+2+3}{4} = 3$
- Espérance d'une variable aléatoire : on peut la calculer **avant** une expérience. Par exemple, pour un dé à 6 faces, l'espérance correspond au résultat moyen que l'on peut espérer avec ce dé, en prenant en compte la probabilité d'obtenir chaque face :

$$E = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3.5$$

Notons aussi que si le dé est truqué, alors les probabilités de chaque face changent, et l'espérance change également.

Définition 4.1 Soit X une v.a.r. et h une application de \mathbb{R} dans \mathbb{R} . Par définition d'une variable aléatoire, $h(X)$ est aussi une v.a.r.. On définit l'**espérance** de $h(X)$, et on note $\mathbb{E}[h(X)]$, la quantité :

- Si X est discrète, et peut prendre les valeurs $a_k, k \geq 0$,

$$\mathbb{E}[h(X)] := \sum_{k \geq 0} h(a_k) \mathbb{P}(X = a_k),$$

(à condition que cette série soit convergente)

- Si X est continue, de densité f_X :

$$\mathbb{E}[h(X)] := \int_{-\infty}^{+\infty} h(x) f_X(x) dx.$$

(à condition que cette intégrale soit convergente)

Propriétés 4.2 L'espérance mathématique vérifie les propriétés suivantes :

1. $\forall a \in \mathbb{R}, \quad \mathbb{E}[a] = a.$
2. *Linéarité* : $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$
3. *Croissance* : Si $X \geq Y$ p.s., $\mathbb{E}[X] \geq \mathbb{E}[Y].$
4. $\forall A \in \mathcal{A}, \quad \mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$

L'espérance donne une indication sur la "localisation" de la variable : autour de quelle valeur elle tombe (donc autour de quelle valeur est/sont située(s) la ou les "bosses" sur la densité pour une v.a. continue). Par contre, cette quantité ne donne aucune indication sur la dispersion de la variable (tombe-t-elle "souvent" proche de l'espérance ?) ni donc sur "l'étalement" de la densité pour une v.a. continue.

Pour contrôler l'écart à l'espérance, la quantité la plus usuelle considère la "moyenne (sur le hasard)" du carré de l'écart à la moyenne. Plus précisément, $(X - \mathbb{E}[X])^2$ donne une bonne indication de si la variable X est tombée proche ou loin de son espérance, mais cette quantité est aléatoire (car X l'est). On considère donc l'espérance de cette quantité, et cela définit la variance.

Définition 4.3 La **variance** d'une v.a.r. X est définie par

$$\text{Var}(X) := \mathbb{E} \left[(X - \mathbb{E}[X])^2 \right].$$

Propriétés 4.4 La variance vérifie :

1. $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
2. $\forall a \in \mathbb{R}, \quad \text{Var}(a) = 0$
3. $\forall a, b \in \mathbb{R}, \quad \text{Var}(aX + b) = a^2 \text{Var}(X)$

Définition 4.5 On appelle **écart-type** de la v.a.r. X la racine carré de la variance :

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Remarque : Si X est une v.a.r. telle que $\mathbb{E}[X] = m$ et $\text{Var}(X) = \sigma^2$, alors la variable $Y = \frac{X-m}{\sigma}$ est d'espérance nulle, et de variance 1. On dit que Y est centrée (d'espérance nulle) et réduite (de variance 1).

Définition 4.6 Soit (X, Y) un couple de v.a.r.

$$\text{Cov}(X, Y) := \mathbb{E} \left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) \right].$$

À comparer avec la définition de la variance...

Proposition 4.7 Si X, Y sont indépendantes, alors

$$\text{Cov}(X, Y) = 0.$$

La réciproque n'est pas vraie !!!!

Propriétés 4.8 Soient X, Y deux variables aléatoires⁵

1. $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
2. La fonction $(X, Y) \mapsto \text{Cov}(X, Y)$ est une forme bilinéaire symétrique :
 - $\forall a_1, a_2 \in \mathbb{R}, \quad \text{Cov}(a_1X_1 + a_2X_2, Y) = a_1 \text{Cov}(X_1, Y) + a_2 \text{Cov}(X_2, Y)$
 - $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. Conséquence : $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.
4. Lorsque X, Y sont indépendantes, l'égalité précédente devient : $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Proposition 4.9 (Stabilité des lois normales) Si X et Y sont deux variables aléatoires **indépendantes** de lois normales, alors toute combinaison affine de X, Y est aussi de loi normale :

$$\forall a, b, c \in \mathbb{R}, aX + bY + c \text{ est de loi normale.}$$

Attention, cela ne suffit pas à caractériser la loi, puisqu'il reste à calculer espérance et variance. En particulier, en posant $b = 0$, on voit que toute transformation affine d'une loi normale est aussi de loi normale⁶

Quelques calculs d'espérances et de variances (lois classiques):

Savoir faire ces calculs est important, car cela donne une bonne idée des outils de calculs qui peuvent être utilisés.

- Lois discrètes :

- Bernoulli $\mathcal{B}(p)$

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

et

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1 - p).$$

- Binomiale $\mathcal{Bin}(n, p)$

On utilise que $X \sim \sum_{i=1}^n Y_i$ où $Y_i \stackrel{i.i.d.}{\sim} \mathcal{B}(p)$, et on obtient

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n \mathbb{E}[Y_i] = np.$$

et, en utilisant que la variance d'une somme de v.a. indépendantes est la somme des variances, on obtient

$$\text{Var}(X) = np(1 - p).$$

- Géométrique $\mathcal{G}(p)$

On va utiliser $\forall |x| < 1, \sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}$, obtenue en dérivant terme à terme l'égalité $\sum x^k = \frac{1}{1-x}$.
On obtient

$$\mathbb{E}[X] = \sum_{k \geq 1} kp(1-p)^{k-1} = \frac{p}{(1 - (1-p))^2} = \frac{1}{p}.$$

Pour la variance, on dérive une nouvelle fois terme à terme, et on obtient $\forall |x| < 1, \sum_{k=1}^{+\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$. Ce qui donne

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \frac{2p(1-p)}{(1 - (1-p))^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{2 - 2p + p - 1}{p^2} = \frac{1-p}{p^2}.$$

- Uniforme $\mathcal{U}([a, b])$

$$\mathbb{E}[X] = \int_a^b \frac{dx}{b-a} = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}.$$

et

$$\text{Var}(X) = \mathbb{E}[X^2] - \frac{(a+b)^2}{4} = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{4(b^2 + ab + a^2) - 3(a^2 + b^2 + 2ab)}{12} = \frac{(b-a)^2}{12}.$$

⁵définies sur le même espace de probabilité, on risque de ne plus le préciser, mais c'est sous-entendu

⁶en réalité cette propriété cache des propriétés beaucoup plus générales, de ce qu'on appelle les "vecteurs gaussiens" définis comme des familles de loi normales (non nécessairement indépendantes) telle que toute combinaison affine reste de loi normale. Mais c'est en dehors du programme de ce cours.

5 Vers l'inférence statistique : théorèmes limites (L2)

Afin de passer de la probabilité à l'inférence statistiques, nous avons besoin de présenter les deux théorèmes sans doute les plus importants relatifs à ces deux disciplines.

Nota bene : dans la suite, on parlera brièvement de convergence de variables aléatoires sans en donner de définition rigoureuse et sans rentrer dans les détails des différents types de convergence possible.

Théorème 5.1 (Loi des grands nombres) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables indépendantes et de même loi (i.i.d.) telles que $\mathbb{E}[|X_n|] < +\infty$. Alors, avec probabilité 1,

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mathbb{E}[X_1].$$

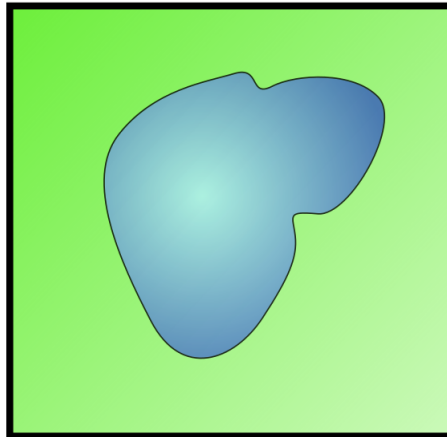
Ce théorème correspond à l'intuition que l'on se fait d'une probabilité : si on fait l'expérience un grand nombre de fois, alors la moyenne des résultats s'approche de l'espérance de la variable aléatoire. Un cas particulier intéressant est celui d'une variable aléatoire de Bernoulli de paramètre p pour laquelle $\mathbb{E}[X] = p$. Ce résultat dit alors que si l'on regarde la proportion de réussites $\frac{1}{n} \sum_{i=1}^n X_i$ sur un grand nombre d'expérience, alors cette proportion, s'approche de la probabilité de réussite. En lançant une pièce un grand nombre de fois, la proportion de pile s'approchera de la probabilité que la pièce tombe sur pile.

En réalité, cela correspond à l'intuition que l'on se fait de la notion de "probabilité" d'un événement : la proportion du temps où l'événement se produit, si on réalise l'expérience un grand nombre de fois.

Il reste une question : **à quel point on s'éloigne de cette espérance ?** C'est l'objet du théorème central limite énoncé ci dessous.

Application de la loi des grands nombres : méthodes de Monte-Carlo

Quelle est la superficie de ce lac ?



Si l'on connaît la superficie du terrain, on peut facilement avoir une approximation de la superficie du lac grâce à la loi des grands nombres.

Supposons que l'on tire n coups de canon, de manière aléatoire, sur le terrain. Si l'on note k le nombre de boulets tombés dans l'eau au final, on a :

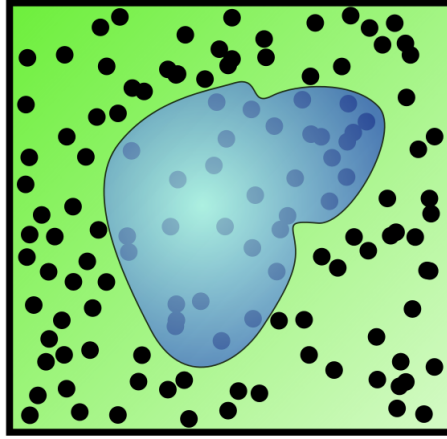
$$\frac{\text{superficie}_{\text{lac}}}{\text{superficie}_{\text{terrain}}} \approx \frac{k}{n}$$

Notons X_i la variable de Bernoulli $\mathcal{B}(p)$ telle que $X_i = 1$ si le i^e boulet atterrit dans l'eau et $X_i = 0$ sinon.

La méthode précédente revient alors à approximer p (la proportion réelle $\frac{\text{superficie}_{\text{lac}}}{\text{superficie}_{\text{terrain}}}$), en sommant les X_i

En effet, la loi des grands nombres nous assure que :

$$\frac{X_1 + \dots + X_n}{n} \rightarrow p$$



Théorème 5.2 (Théorème Central Limite) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables indépendantes et de même loi (i.i.d.) telles que $\text{Var}(X_n) < +\infty$. Alors, pour tout $a < b \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1]}{\sqrt{\text{Var}(X_1)}} \in [a, b]\right) \rightarrow \mathbb{P}(Z \in [a, b]),$$

où $Z \sim \mathcal{N}(0, 1)$ désigne une loi normale “centrée réduite” (d’espérance 0 et de variance 1).

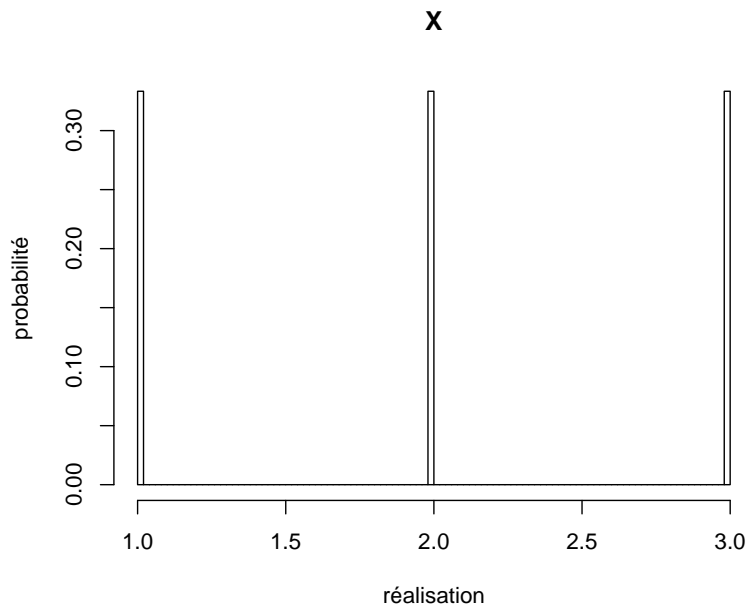
Ce résultat affirme essentiellement que l’écart à la moyenne se comporte (lorsqu’on a assez d’observations) comme une loi normale.

TCL : Illustration 1

Soit la variable aléatoire suivante

$$X = \begin{cases} 1 & \text{avec probabilité } 1/3 \\ 2 & \text{avec probabilité } 1/3 \\ 3 & \text{avec probabilité } 1/3 \end{cases}$$

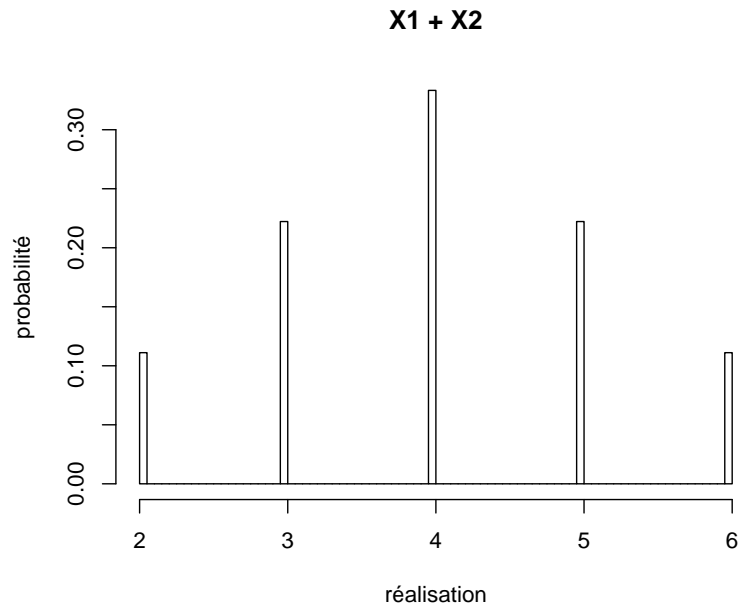
La loi de probabilité de X peut être illustrer par l’histogramme théorique suivant :



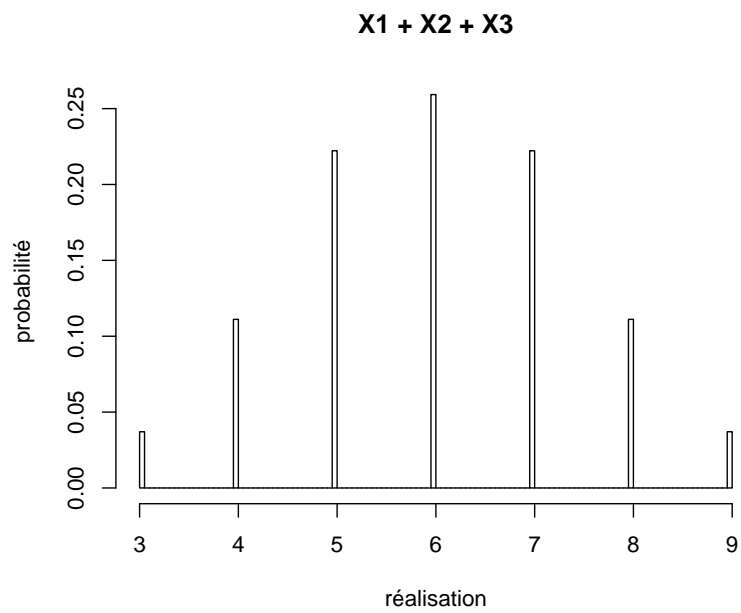
On considère maintenant la somme de deux copies indépendantes de X , X_1 et X_2 :

$$X_1 + X_2 = \begin{cases} 1 + 1 = 2 & \text{avec probabilité } 1/9 \\ 1 + 2 = 3 & \text{avec probabilité } 1/9 \\ 1 + 3 = 4 & \text{avec probabilité } 1/9 \\ 2 + 1 = 3 & \text{avec probabilité } 1/9 \\ 2 + 2 = 4 & \text{avec probabilité } 1/9 \\ 2 + 3 = 5 & \text{avec probabilité } 1/9 \\ 3 + 1 = 4 & \text{avec probabilité } 1/9 \\ 3 + 2 = 5 & \text{avec probabilité } 1/9 \\ 3 + 3 = 6 & \text{avec probabilité } 1/9 \end{cases} = \begin{cases} 2 & \text{avec probabilité } 1/9 \\ 3 & \text{avec probabilité } 2/9 \\ 4 & \text{avec probabilité } 3/9 \\ 5 & \text{avec probabilité } 2/9 \\ 6 & \text{avec probabilité } 1/9 \end{cases}$$

L'histogramme associé à $X_1 + X_2$ est

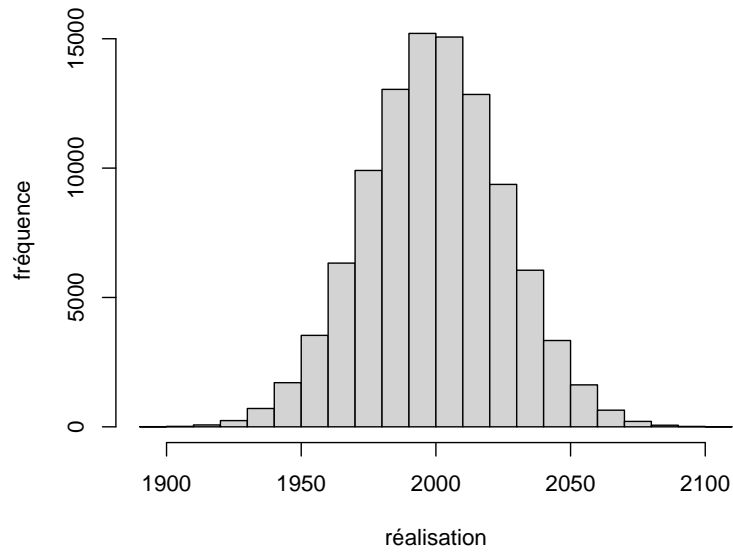


Si l'on répète le procédé avec trois variables au lieu de deux, il vient



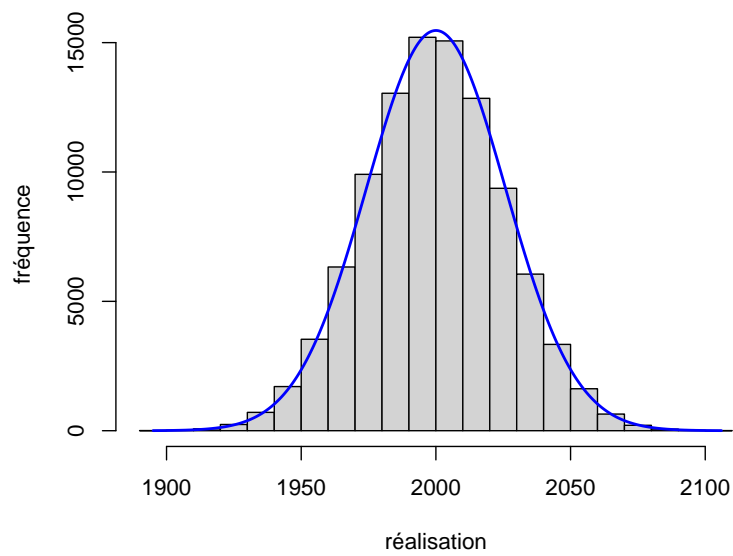
L'histogramme suivant donne les fréquences associées à la somme de 1000 copies indépendantes de X (100 000 sommes ont été simulées au total)

X1 + ... + X1000, Simulation de 100 000 sommes

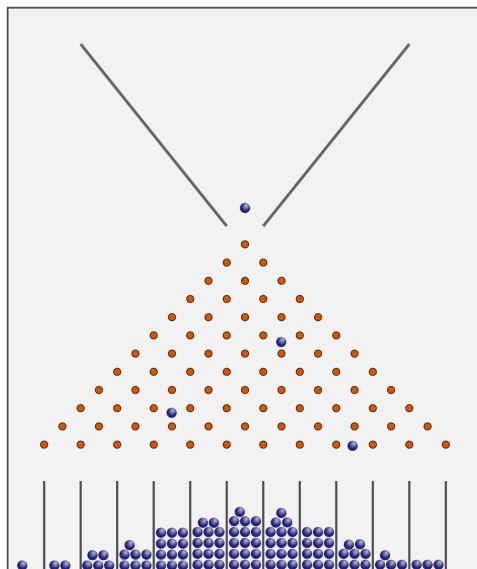


La figure suivante montre la densité de probabilité d'une loi normale de mêmes moyenne et variance que les données observées.

X1 + ... + X1000, Simulation de 100 000 sommes



TCL : Illustration 2, la Planche de Galton



La planche de Galton a été inventée par Francis Galton (1822 - 1911). Elle démontre la convergence de la loi binomiale $\mathcal{B}(n, p)$ vers la loi normale $\mathcal{N}(np, np(1-p))$ lorsque $n \rightarrow \infty$. Son principe est donné ci-après :

Sur la i^e ligne, il y a i clou(s).

Sur chaque ligne, la probabilité qu'une bille passe à gauche ou à droite du clou est la même et vaut $p = 0.5$. On considère qu'il y a n lignes et donc $n + 1$ boîtes en bas de la planche

Le nombre de chemins possibles amenant vers la k^e boîte (celle tout à gauche est la 0^e) est de C_n^k

Au final, la probabilité qu'une bille atterrisse dans la k^e boîte est

$$C_n^k p^k (1-p)^{n-k}$$

⇒ C'est la loi binomiale !

6 Inférence statistique (L2)

Les statistiques inférentielles permettent d'inférer, c'est à dire de déduire, des quantités d'intérêt (moyenne, variance, ...) d'une **population** d'étude à partir de données issues d'un **échantillon représentatif** de cette population.

Dans la suite on considère des données $(x_i)_{i \leq n} = (x_1, \dots, x_n)$ comme réalisation de variables aléatoires $(X_i)_{i \leq n} = (X_1, \dots, X_n)$.

Dans ce cours, les variables aléatoires $(X_i)_{i \leq n}$ seront considérées comme indépendantes et identiquement distribuées (i.i.d.) selon une loi P_θ qui dépend d'un vecteur de paramètres θ . Ainsi θ peut être de dimension quelconque même si en pratique dans ce cours il sera le plus souvent de dimension 1 ou 2.

L'ensemble des valeurs possibles pour θ sont contenues dans un ensemble \mathcal{I} (on note $\theta \in \mathcal{I}$). Notre objet d'étude est alors l'ensemble des lois définies par tous les θ possibles, noté $(P_\theta)_{\theta \in \mathcal{I}}$ et appelé **modèle paramétrique**.

Si \mathcal{I} est fini ou dénombrable, on parlera de modèle discret, et si \mathcal{I} est non-dénombrable on parlera de modèle continu.

6.1 Estimateur

Définition 6.1 Un *estimateur* de θ , noté $\hat{\theta}$ ou $\hat{\theta}_n$, est une fonction des variables $(X_i)_{i \leq n}$ de loi P_θ

$$\hat{\theta} = f(X_1, \dots, X_n)$$

Comme fonction de variables aléatoires, **un estimateur est une variable aléatoire**

Nota bene : Un estimateur peut dépendre des $(X_i)_{i \leq n}$, de la taille de l'échantillon n , **mais jamais de θ lui-même**

Exemple 6.1 1. $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur appelé **moyenne arithmétique empirique**

2. $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur appelé **variance empirique non-corrigée**

3. $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur appelé **variance empirique corrigée**

4. $\sigma_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$ est un estimateur appelé **écart-type empirique non-corrigé**

5. $s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$ est un estimateur appelé **écart-type empirique corrigé**

Il existe une infinité d'estimateurs possibles et dans la suite nous allons voir sur quels critères nous pouvons choisir les plus pertinents.

6.2 Propriétés des estimateurs

6.2.1 Estimateurs sans biais

Définition 6.2 Un estimateur sans biais de θ vérifie

$$E(\hat{\theta}) = \theta$$

Exemple 6.2 Supposons que les $(X_i)_{i < n}$ sont de moyenne μ ($E(X_i) = \mu$), alors la moyenne arithmétique empirique est un estimateur sans biais de μ

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

Définition 6.3 Un estimateur asymptotiquement sans biais de θ vérifie

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$$

La dépendance à n de $\hat{\theta}$ est ici implicite. Pour la rendre explicite, on trouve souvent l'écriture $\hat{\theta}_n$

Définition 6.4 Un estimateur convergent de θ vérifie, avec probabilité 1

$$\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{} \theta$$

Définition 6.5 Un estimateur sans biais est dit **efficace** s'il possède une variance minimale

De la même manière qu'un estimateur peut être asymptotiquement sans biais, il peut être asymptotiquement efficace.

6.3 Méthode des moments

Définition 6.6 Le **moment** d'ordre k associé à une variable aléatoire X est donné par

$$M_k = E(X^k)$$

La moyenne (espérance mathématique) est ainsi le moment d'ordre 1.

Définition 6.7 La **méthode des moments** consiste à faire correspondre les moments de la loi P_θ , i.e $(M_k(\theta))_{k \geq 1} = E(X_i^k)$ avec les moments empiriques $m_1 = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, $m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$, etc afin d'estimer θ via un système d'équations.

Exemple 6.3 **Méthode des moments dans le modèle gaussien**

On a $X_i \mathcal{N}(\mu, \sigma^2)$ et donc $\theta = (\mu, \sigma^2)$

Les moments associés au modèle sont

$$M_1(\mu, \sigma^2) = E(X_i) = \mu \quad M_2(\mu, \sigma^2) = E(X_i^2) = \mu^2 + \sigma^2$$

Les moments empiriques sont

$$m_1 = \bar{X}_n \quad m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

L'association donne les estimateurs de μ et σ^2

$$\hat{\mu} = \bar{X}_n \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 = \sigma_n^2$$

Avec σ_n^2 la variance empirique non-corrigée

Avantages et limites de la méthode des moments

La méthode des moments est extrêmement simple à mettre en place et conduit à des estimateurs asymptotiquement sans biais et convergent.

Elle est toutefois confrontée à trois problèmes majeurs :

- Tous les moments ne sont pas toujours définis (si la dimension du modèle est plus grande que le nombre de moments disponibles c'est peine perdue)
- La loi des estimateurs est très difficile à étudier

6.4 Méthode du maximum de vraisemblance

Définition 6.8 Soient $(X_i)_{i \leq n}$ des variables modélisant des données $(x_i)_{i \leq n}$ appartenant à un modèle paramétrique $(P_\theta)_{\theta \in \mathcal{I}}$. La vraisemblance, notée \mathcal{L} , associée à une valeur $\theta_0 \in \mathcal{I}$ du paramètre correspond à la **probabilité d'avoir observé les données** $(x_i)_{i \leq n}$ si le paramètre du modèle était θ_0 .

Dans le cas discret, on a la probabilité jointe

$$\mathcal{L}(\theta) = P(X_1 = x_1, \dots, X_n = x_n)$$

Dans le cas continu, on a la densité jointe

$$\mathcal{L}(\theta) = f(x_1, \dots, x_n)$$

Dans notre cas, les $(X_i)_{i \leq n}$ **seront toujours indépendantes**, donc

Dans le cas discret, on obtient le produit des probabilités

$$\mathcal{L}(\theta) = P(X_1 = x_1) \times \dots \times P(X_n = x_n)$$

Dans le cas continu, on obtient le produit des densités

$$\mathcal{L}(\theta) = f(x_1) \times \dots \times f(x_n)$$

Définition 6.9 La méthode du **maximum de vraisemblance** consiste à retenir comme estimation du paramètre celle qui maximise la vraisemblance.

Exemple 6.4 Illustration de la vraisemblance

On effectue 10 lancers de pièces consécutifs et indépendants et l'on obtient les résultats suivants

$$\text{Res} = (\text{Pile}, \text{Pile}, \text{Pile}, \text{Pile}, \text{Face}, \text{Pile}, \text{Pile}, \text{Pile}, \text{Pile}, \text{Face})$$

Dans la suite, le résultat d'un lancer est noté X et on associe la valeur 1 à Pile et 0 à Face. Le lancer de pièce est soumis à la loi de probabilité dite de Bernoulli

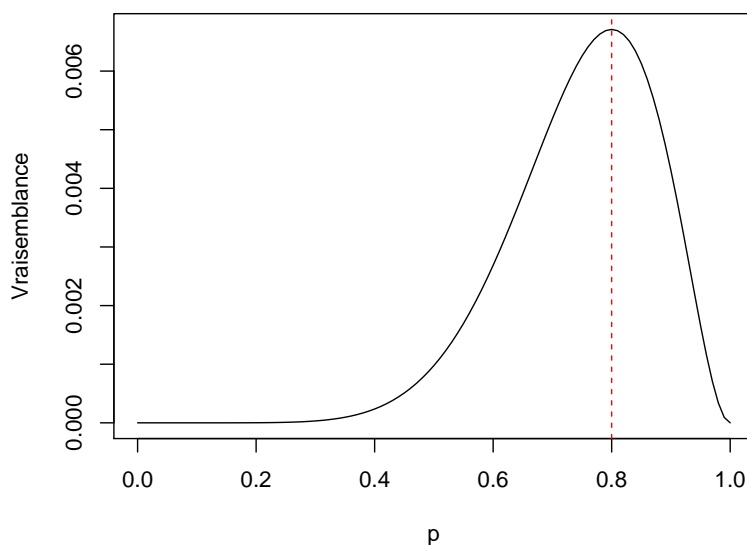
$$P(X = 1) = p \quad \text{et} \quad P(X = 0) = 1 - p$$

où p est le paramètre de la loi. C'est lui qu'on cherche généralement à estimer.

Si l'on suppose que la pièce est standard, on peut supposer que $p = 0.5$. Quelle est alors la probabilité d'observer les données obtenues lors de notre série de 10 lancers ?

$$P(\text{Res}) = P(X = 1)^8 \times P(X = 0)^2 = p^8 \times (1 - p)^2 = 0.5^8 \times 0.5^2 \approx 0.1\%$$

Cette valeur de 0.1%, assez faible, n'a que peu d'intérêt en soi. La question soulevée par la méthode du maximum de vraisemblance est la suivante : peut-on trouver une valeur de p pour laquelle cette probabilité serait supérieure à 0.1% ? Et si oui, quelle est la valeur maximale possible et pour quelle valeur de p ce maximum est-il atteint ?



La figure ci-dessus montre la vraisemblance obtenue en fonction de la valeur de p . On remarque qu'un maximum est atteint pour $p = 0.8$. Il s'agit de la valeur la plus compatible avec les données, dans le sens où elle maximise la probabilité de les observer.

Exemple 6.5 Exemple de détermination d'un estimateur du maximum de vraisemblance

Considérons une loi de Poisson $\mathcal{P}(\lambda)$ pour notre modèle paramétrique :

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

Alors la vraisemblance de données $(x_i)_{i \leq n}$ s'écrit

$$\begin{aligned} \mathcal{L}(\lambda) &= P(X_1 = x_1) \times \cdots \times P(X_n = x_n) \\ &= \prod_{i=1}^n \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) \\ &= e^{-n\lambda} \prod_{i=1}^n \left(\frac{\lambda^{x_i}}{x_i!} \right) \end{aligned}$$

Comment maximiser cette quantité en λ ?????

Il s'agit d'une fonction de λ , donc pour en trouver le maximum on peut chercher à annuler sa dérivée. Problème \Rightarrow la fonction est très compliquée à dériver !

Pour nous aider, nous allons définir une quantité appelée log-vraisemblance, qui correspond au **logarithme népérien** (même s'il est souvent noté **log en statistiques, désolé pour ça**) de la vraisemblance

$$\begin{aligned} l(\lambda) &= \log(\mathcal{L}(\lambda)) \\ &= -n\lambda + \sum_{i=1}^n \log\left(\frac{\lambda^{x_i}}{x_i!}\right) \\ &= -n\lambda + \sum_{i=1}^n (x_i \log(\lambda) - \log(x_i!)) \\ &= -n\lambda + \log(\lambda) \sum_{i=1}^n x_i - \sum_{i=1}^n \log(x_i!) \end{aligned}$$

Nous obtenons une somme qui sera beaucoup plus facile à dériver !

$$l'(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i$$

Et maintenant trouvons la valeur $\hat{\lambda}$ qui annule la dérivée

$$l'(\lambda) = 0 \quad \Leftrightarrow \quad \frac{1}{\lambda} \sum_{i=1}^n x_i = n$$

Et au final

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$$

Nota bene : $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$ est l'**estimation** de λ par la méthode du maximum de vraisemblance. En remplaçant les x_i par les variables aléatoires X_i correspondantes on obtient $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$ l'**estimateur** du maximum de vraisemblance.

Un estimateur du maximum de vraisemblance possède les propriétés remarquables suivantes

Propriétés 6.10 1. Il est asymptotiquement sans biais

2. Il est asymptotiquement distribué selon une loi normale

3. Il est asymptotiquement efficace (variance minimale pour un estimateur sans biais) et sa variance est donnée par l'inverse de l'information de Fisher, notée I_n et telle que

$$I_n(\theta) = E(-l''(\theta))$$

Exemple 6.6 Calcul de la variance de l'estimateur

Reprenons le modèle de Poisson précédent. La dérivée de la log-vraisemblance était

$$l'(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i$$

La dérivée seconde est donc

$$l''(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n x_i$$

Alors, en prenant l'espérance (on remplace donc les x_i par les X_i car on veut la variance de l'estimateur)

$$\begin{aligned} I_n(\lambda) &= E(-l''(\lambda)) \\ &= E\left(\frac{1}{\lambda^2} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{\lambda^2} E\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{\lambda^2} n\lambda \\ &= \frac{n}{\lambda} \end{aligned}$$

Et l'estimateur $\hat{\lambda}$ est asymptotiquement distribué selon la loi normale suivante

$$\hat{\lambda} \underset{n \rightarrow \infty}{\sim} \mathcal{N}\left(\lambda, \frac{\lambda}{n}\right)$$

où λ est la vraie valeur du paramètre. En pratique, étant donné que λ est inconnu, on le remplacera par l'estimation ponctuelle $\hat{\lambda}$ obtenue sur l'échantillon afin de calculer un intervalle de confiance.

6.5 Intervalles de confiance

6.5.1 Intervalle de confiance d'une proportion

On cherche à estimer la proportion p de personnes présentant une maladie dans une population donnée.

A partir d'un échantillon représentatif i.i.d de taille n , on construit l'estimateur suivant

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Avec $X_i = 1$ si l'individu i a la maladie et 0 sinon. X_i suit une loi de Bernoulli de paramètre p ($X_i \sim \mathcal{B}(p)$), calculons alors les espérance et variance de \hat{p}_n .

$$E(\hat{p}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} np = p$$
$$Var(\hat{p}_n) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{p(1-p)}{n}$$

D'après le théorème central limite, quand n est suffisamment grand ($n \geq 30$), la loi de \hat{p}_n peut être approchée par une loi normale. Ainsi, on a

$$\hat{p}_n \underset{n \rightarrow \infty}{\sim} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Si l'on centre et réduit, il vient

$$\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

Or, on sait qu'une loi normale centrée réduite a 95% de ses valeurs comprises entre -1.96 et 1.96

$$P\left(-1.96 < \frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} < 1.96\right) = 95\%$$
$$P\left(\hat{p}_n - 1.96\sqrt{\frac{p(1-p)}{n}} < p < \hat{p}_n + 1.96\sqrt{\frac{p(1-p)}{n}}\right) = 95\%$$

Pour construire un intervalle de confiance pour p , on remplace p par son estimation ponctuelle \hat{p}_n dans l'expression de la variance.

Supposons que $n = 300$ et $\hat{p}_n = 0.15$, l'intervalle de confiance à 95% pour p est

$$IC_{95\%}(p) = \left[\hat{p}_n - 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}; \hat{p}_n + 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right] = [0.11; 0.19]$$

6.5.2 Intervalle de confiance d'une moyenne

On cherche à estimer la moyenne μ des individus vivants en France.

A partir d'un échantillon représentatif i.i.d de taille n , on construit l'estimateur suivant

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

avec X_i la taille de l'individu i

On donne $E(X_i) = \mu$ et $Var(X_i) = \sigma^2$. Calculons alors les espérance et variance de $\hat{\mu}_n$.

$$E(\hat{\mu}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$
$$Var(\hat{\mu}_n) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}$$

D'après le théorème central limite, quand n est suffisamment grand ($n \geq 30$), la loi de $\hat{\mu}_n$ peut être approchée par une loi normale. Ainsi, on a

$$\hat{\mu}_n \underset{n \rightarrow \infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Si l'on centre et réduit, il vient

$$\frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

Or, on sait qu'une loi normale centrée réduite a 95% de ses valeurs comprises entre -1.96 et 1.96

$$P\left(-1.96 < \frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} < 1.96\right) = 95\%$$

$$P\left(\hat{\mu}_n - 1.96\sqrt{\frac{\sigma^2}{n}} < \mu < \hat{\mu}_n + 1.96\sqrt{\frac{\sigma^2}{n}}\right) = 95\%$$

$$P\left(\hat{\mu}_n - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

Pour construire un intervalle de confiance pour μ , on remplace σ^2 par son estimation ponctuelle $\hat{\sigma}_n^2$ dans l'expression de la variance (ou bien σ par $\hat{\sigma}_n$)

Supposons que $n = 50$, avec une moyenne $\hat{\mu}_n = 175\text{cm}$ et un écart-type $\hat{\sigma}_n = 50\text{cm}$, l'intervalle de confiance à 95% pour μ est

$$IC_{95\%}(\mu) = [161; 189]$$

6.6 Tests d'hypothèse

6.6.1 Comparaison d'une proportion à une valeur théorique

Reprenons l'exemple avec $n = 300$ et $\hat{p}_n = 0.15$. Ainsi sur notre échantillon on obtient 15% de malades.

Question : peut-on en déduire que la véritable proportion p (donc dans la population) est différente de 10% ?

Évidemment, la véritable proportion p est inconnue et il est impossible d'apporter une réponse certaine à cette question.

Démarche:

- on fait l'hypothèse que la véritable proportion est égale à 10%. Cette hypothèse est appelée l'hypothèse nulle et est notée $\mathcal{H}_0 : p = 0.1$
- on établit une hypothèse alternative notée \mathcal{H}_1 et qui est ici $\mathcal{H}_1 : p \neq 0.1$
- on définit enfin un niveau d'erreur ou risque α qui correspond à la probabilité de rejeter à tort \mathcal{H}_0 . En général, on prend $\alpha = 0.05$

On a vu précédemment que, grâce au TCL,

$$\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

Ainsi, si \mathcal{H}_0 est vraie (on dit qu'on est sous l'hypothèse nulle), on a

$$Z = \frac{\hat{p}_n - 0.1}{\sqrt{\frac{0.1(1-0.1)}{n}}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

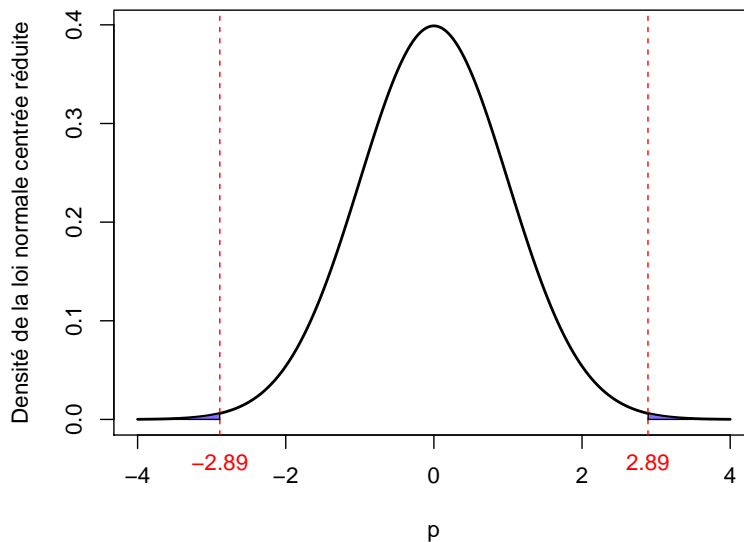
La quantité Z est appelée la statistique de test.

Nota bene : Z est une variable aléatoire puisque \hat{p}_n est une variable aléatoire. Si l'on répète l'expérience avec plusieurs échantillons, alors chaque échantillon donnera une estimation \hat{p}_n différente et donc une réalisation de Z différente. La réalisation de Z sur un échantillon donné est notée z .

La réalisation de Z sur notre échantillon est

$$z = \frac{0.15 - 0.1}{\sqrt{\frac{0.1(1-0.1)}{300}}} \approx 2.89$$

Question : la valeur observée sur notre échantillon est-elle compatible avec notre hypothèse nulle de départ ?



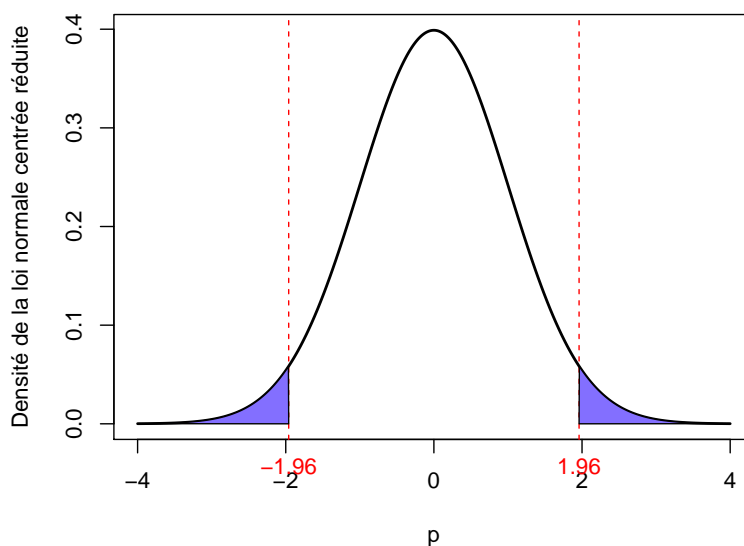
Si \mathcal{H}_0 était vraie, alors la probabilité que Z soit supérieur à 2.89 (en valeur absolue) serait de $P(|Z| > 2.89) = P(Z > 2.89) + P(Z < -2.89) = 0.004$

Cette probabilité, appelée **p-value**, correspond à l'aire en bleu sur le graphique précédent.

Puisque la p-value est inférieure au risque α choisi au départ, on va rejeter l'hypothèse nulle \mathcal{H}_0 et considérer que p est différent de 0.1.

Précédemment, l'intervalle de confiance calculé était de $[0.11; 0.19]$. Cet intervalle aurait déjà permis de répondre à la question posée par le test. En effet, l'intervalle exclut la valeur 0.1 et ainsi on pouvait déjà dire que la vraie proportion est statistiquement différente de 0.10 au seuil de $\alpha = 5\%$.

Important : étant donné que Z suit une loi normale centrée réduite, la valeur seuil de rejet/non rejet de \mathcal{H}_0 (pour un risque α de 5%) est le fameux 1.96 que vous connaissez sans doute déjà. En effet, $P(|Z| > 1.96) = P(Z > 1.96) + P(Z < -1.96) = 0.05$



Définition 6.11 Un test d'hypothèse repose sur une statistique de test Z qui est une variable aléatoire. On note z sa réalisation sur l'échantillon dont on dispose.

On définit une hypothèse nulle \mathcal{H}_0 , une hypothèse alternative \mathcal{H}_1 , et une probabilité de acceptable de rejeter à tort \mathcal{H}_0 appelé le risque α (= 5% dans ce cours)

Sous l'hypothèse nulle, on connaît la loi de Z .

Dans le cas où cette loi est une loi normale centrée réduite (ce qui sera toujours le cas dans ce cours), le processus de décision est le suivant

- si $|z| \geq 1.96$, alors on rejette \mathcal{H}_0
- si $|z| < 1.96$, alors on ne rejette pas \mathcal{H}_0

Définition 6.12 La p -value est la probabilité, sachant l'hypothèse nulle, que Z soit supérieur ou égal à z en valeur absolue.

$$p = P(|Z| > |z| \mid \mathcal{H}_0)$$

En termes de règles de décision d'un test, on peut donc également considérer

- si $p \leq \alpha$, alors on rejette \mathcal{H}_0
- si $p > \alpha$, alors on ne rejette pas \mathcal{H}_0

6.6.2 Comparaison d'une moyenne à une valeur théorique

Reprenons l'exemple avec $n = 50$, une moyenne $\hat{\mu}_n = 175cm$ et un écart-type $\hat{\sigma}_n = 50cm$.

Question : peut-on en déduire que la véritable moyenne μ (donc dans la population) est différente de $165cm$?

Démarche:

- on fait l'hypothèse que la véritable moyenne est égale à $165cm$. Cette hypothèse est appelée l'hypothèse nulle et est notée $\mathcal{H}_0 : \mu = 165cm$
- on établit une hypothèse alternative notée \mathcal{H}_1 et qui est ici $\mathcal{H}_1 : \mu \neq 165cm$
- on définit enfin un niveau d'erreur ou risque α qui correspond à la probabilité de rejeter à tort \mathcal{H}_0 . En général, on prend $\alpha = 0.05$

On a vu précédemment que, grâce au TCL,

$$\frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

Ainsi, si \mathcal{H}_0 est vraie (on dit qu'on est sous l'hypothèse nulle), on a

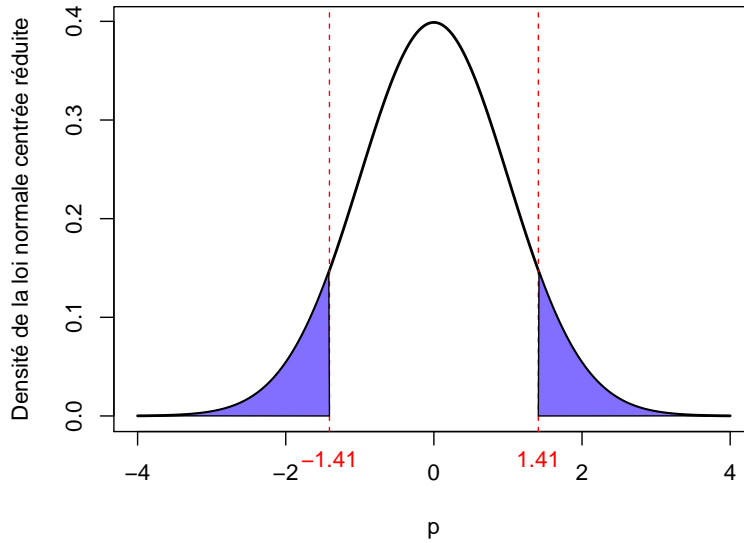
$$Z = \frac{\hat{\mu}_n - 165}{\sqrt{\frac{\sigma^2}{n}}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

La réalisation de Z sur notre échantillon est

$$z = \frac{175 - 165}{\sqrt{\frac{50^2}{50}}} = \frac{175 - 165}{\frac{50}{\sqrt{50}}} \approx 1.41$$

Ainsi, $z < 1.96$ donc on ne rejette pas \mathcal{H}_0 . La p -value associée au test est

$$p = P(|Z| > 1.41 \mid \mathcal{H}_0) \approx 15.9\%$$



6.6.3 Comparaison de deux proportions (échantillons indépendants)

Pour savoir si deux proportions p_1 et p_2 sont statistiquement différentes à partir de leurs estimations respectives \hat{p}_1 et \hat{p}_2 dans des échantillons indépendants de tailles respectives n_1 et n_2 , c'est à dire

- $\mathcal{H}_0 : p_1 = p_2$
- $\mathcal{H}_1 : p_1 \neq p_2$

On utilise la statistique de test suivante

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

avec

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

Sous \mathcal{H}_0 les proportions sont les mêmes donc on construit la variance à partir de la proportion globale

6.6.4 Comparaison de deux moyennes (échantillons indépendants)

Pour savoir si deux moyennes μ_1 et μ_2 sont statistiquement différentes à partir de leurs estimations respectives $\hat{\mu}_1$ et $\hat{\mu}_2$ dans des échantillons indépendants de tailles respectives n_1 et n_2 , c'est à dire

- $\mathcal{H}_0 : \mu_1 = \mu_2$
- $\mathcal{H}_1 : \mu_1 \neq \mu_2$

On utilise la statistique de test suivante

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1)$$

avec

$$\hat{\sigma}^2 = \frac{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2}{n_1 + n_2}$$

Nota bene : On fera ici l'hypothèse que les vraies variances σ_1^2 et σ_2^2 sont égales. Il existe de nombreuses variantes de ce test, en fonction des hypothèses faites (égalité de tailles d'échantillon, égalité des variance) ou des estimations retenues (variance empirique corrigée ou non corrigée). L'idée est ici de présenter l'un des versions les plus simples afin de faire comprendre le principe et également pour que les statistiques de tests utilisées pour comparer deux proportions ou deux moyennes soient les mêmes !

7 Exercices

7.1 Espace de probabilité

Exercice 1 :

On lance 1 dé à 6 faces.

1. Proposer un ensemble Ω , et définir \mathbb{P} sur $\mathcal{P}(\Omega)$ pour décrire l'expérience.
2. Expliciter l'événement A : "le résultat du dé est pair" et B : "le résultat est strictement supérieur à quatre", et donner les probabilités de ces événements.
3. Expliciter $A \cap B$, $A \cup B$, A^C et B^C , et donner les probabilités de ces événements.

Exercice 2 :

On lance (indépendamment) 2 dé à 6 faces.

1. Proposer un ensemble Ω , et définir \mathbb{P} sur $\mathcal{P}(\Omega)$ pour décrire l'expérience.
2. Expliciter l'événement A : "la somme des résultats des dés vaut six" et B : "le résultat de chaque dé est impair", et donner les probabilités de ces événements.
3. Expliciter $A \cap B$, A^C et B^C , et $(A^C \cup B^C)^C$, et donner les probabilités de ces événements.

Exercice 3 :

On lance (indépendamment) 2 dé à 6 faces. On pose $\Omega = \{1, 2, \dots, 6\}^2$ et $\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{36}$.

1. Pour tout k entre 2 et 12, expliciter l'événement S_k : "la somme des résultats des dés vaut k ", et donner la probabilité de ces événements.
2. Expliciter l'événement D_1 : "le résultat du premier dé est pair", D_2 : "les deux dés ont même parité, et $D_1 \cap D_2$ et donner les probabilités de ces événements.
3. Expliciter les événements suivants et calculer les probabilités associées :
 - A : "la somme des résultats des dés est supérieure ou égale à dix"
 - B : "Au moins un des résultats des deux dés est égal à six"
 - $A \cap B$

7.2 Conditionnement

Exercice 1 : (CC 2020)

Lors d'un QCM, on interroge des étudiants sur la signification du sigle ADN. A chaque personne, on propose trois réponses différentes: 1, 2 ou 3, la réponse correcte étant la réponse 1. Tout étudiant connaissant la réponse correcte la donne, sinon il choisit au hasard une des trois réponses proposées. On suppose que la probabilité qu'un étudiant connaisse la bonne réponse est égale à $3/4$. Soit R l'événement "l'étudiant connaît la bonne réponse" et B l'événement "l'étudiant donne la bonne réponse".

1. Dédire de l'énoncé que $\mathbb{P}(B/R) = 1$ et que $\mathbb{P}(B/R^c) = 1/3$.
2. Calculer la valeur de $\mathbb{P}(R/B)$.

Exercice 2 :

De nombreux logiciels de triche existent pour les jeux vidéos. A tel point qu'une proportion conséquente des joueurs peuvent se mettre à en utiliser. On suppose que pour un certain jeu, 20% des joueurs utilisent un logiciel de triche. De plus, la probabilité de gagner une partie en utilisant ce logiciel, est de 50%, alors qu'elle est seulement de 5% sans ce logiciel.

Quelle est la probabilité qu'un gagnant soit en réalité un tricheur ?

Exercice 3 :

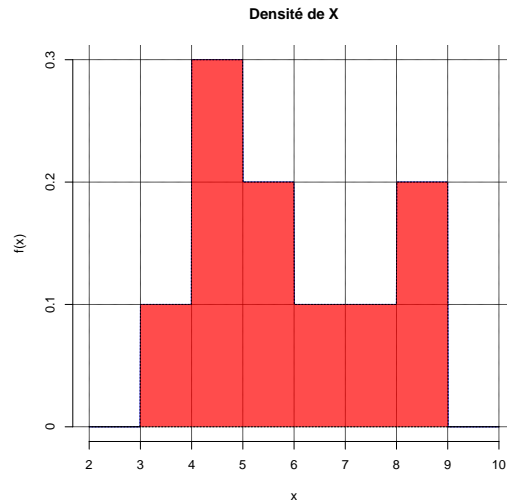
Une maladie rare touche 1% de la population. Il existe un test pour la détecter, qui détecte avec 99% de chances la maladie chez un patient malade, mais se trompe 5% du temps pour un patient non porteur de la maladie (le test est un faux positif).

1. Donner la probabilité d'être malade sachant qu'on a reçu un test positif.
2. Que devient la probabilité si 2 tests successifs sortent positifs ? (en supposant l'indépendance entre les deux résultats sachant le statut de maladie de la personne).

7.3 Variables aléatoires à densité

Exercice 1

Soit X une variable aléatoire de densité f représentée ci-dessous. Le quadrillage est représenté seulement pour aider à la lecture de la courbe.



1. Quelle est l'aire totale sous la courbe ?
2. Que vaut $\mathbb{P}(X = 5.5)$?
3. Donner, en lisant le graphique, $\mathbb{P}(X \leq 4)$.
4. Donner, en lisant le graphique, $\mathbb{P}(X \in [4, 7])$.
5. Donner, en lisant le graphique, $\mathbb{P}(X \leq 5 \mid X \leq 8)$.

Exercice 2

Soit X une variable aléatoire de loi uniforme sur $[2, 4]$. On rappelle que X a alors pour densité la fonction $f_X : x \mapsto f(x) = \frac{1}{2} \mathbb{1}_{x \in [2, 4]}$.

1. Calculer $\mathbb{P}(X \leq 3)$
2. Calculer $\mathbb{E}[X]$.
3. Calculer $\mathbb{E}[X^2]$.
4. En déduire $\text{Var}(X)$.

Exercice 3

Soit X une variable aléatoire de densité donnée par $f_X : x \mapsto f(x) = 2x \mathbb{1}_{x \in [0, 1]}$.

1. Pour $t \in [0, 1]$, calculer $\mathbb{P}(X \leq t)$ en fonction de t .
2. Calculer $\mathbb{E}[X]$.
3. Calculer $\mathbb{E}[X^2]$.
4. En déduire $\text{Var}(X)$.

7.4 Avancé

Exercice 1 :

On considère un gène formé de 2 allèles A et a . On tire deux personnes au hasard dans la population (on supposera donc l'indépendance de leur génotypes), et on définit les événements $AA_1, AA_2, aa_1, aa_2, Aa_1, Aa_2$ par " AA_1 : la personne 1 possède 2 allèles A "... (de même pour les autres).

Si ces deux personnes ont un enfant, alors on considère que cet enfant hérite aléatoirement (et uniformément) d'un allèle de chacun de ces parents. On note AA_3, Aa_3, aa_3 les événements : " AA_3 : l'enfant possède 2 allèles A "... (de même pour les autres).

On suppose que

$$\begin{aligned} \mathbb{P}(AA_1) &= \mathbb{P}(AA_2) = p_0 \\ \mathbb{P}(Aa_1) &= \mathbb{P}(Aa_2) = r_0 \\ \mathbb{P}(aa_1) &= \mathbb{P}(aa_2) = q_0 = 1 - p_0 - r_0 \end{aligned}$$

1. Donner $p_1 = \mathbb{P}(AA_3)$ en fonction de p_0, r_0 et q_0 .
2. Donner $r_1 = \mathbb{P}(Aa_3)$ en fonction de p_0, r_0 et q_0 .
3. Donner $q_1 = \mathbb{P}(aa_3)$ en fonction de p_0, r_0 et q_0 .
4. On suppose qu'une population a une répartition à la génération 0 de p_0, r_0, q_0 entre les 3 phénotypes AA, Aa, aa . En supposant que les couples se forment indépendamment des phénotypes, quelle sera la répartition à la génération 1 ? à la génération 2 ? Commenter.

Exercice 2 : On donne les probabilités conditionnelles de guérison de deux médicaments agissant sur les calculs rénaux, en fonction de la taille des calculs. On note P, L les événements “petits calculs” et “larges calculs”, A, B les événements “on a administré le médicament A ou B ”, et G, G^C les événements “le/la patient-e a “guéri” ou “pas guéri”.

$$\mathbb{P}(G | A, P) = 0.8$$

$$\mathbb{P}(G | B, P) = 0.7$$

$$\mathbb{P}(G | A, L) = 0.6$$

$$\mathbb{P}(G | B, L) = 0.5$$

1. Pour un premier plan de traitement, les médecins choisissent l’administration suivante des médicaments A et B :

$$\mathbb{P}(A | P) = 0.2 = 1 - \mathbb{P}(B | P)$$

$$\mathbb{P}(A | L) = 0.8 = 1 - \mathbb{P}(B | L)$$

De plus, on choisit la définition de “petits calculs” et “gros calculs” en prenant les 50% les plus petits ou plus larges, de sorte qu’on a

$$\mathbb{P}(P) = 0.5 = \mathbb{P}(L)$$

Calculer $\mathbb{P}(G | A)$ et $\mathbb{P}(G | B)$. Quel médicament semblerait le meilleur au vue de ce seul résultat ? Commenter.

2. Pour une étude **randomisée**, on choisira aléatoirement le médicament qu’on administrera à chaque patient, de sorte d’avoir

$$\mathbb{P}(A | P) = \mathbb{P}(A | L) = 1 - \mathbb{P}(B | P) = 1 - \mathbb{P}(B | L)$$

De plus, on fixe arbitrairement cette probabilité à $1/2$. Calculer de nouveau $\mathbb{P}(G | A)$ et $\mathbb{P}(G | B)$, et commenter.

Exercice 3 :

On modélise le temps d’efficacité d’une anesthésie locale (en heures) par une variable aléatoire T de densité $f_T(x) = x\mathbb{1}_{x \in [0,1]} + (2-x)\mathbb{1}_{x \in [1,2]}$.

1. Quelle est la probabilité que l’anesthésie soit efficace moins de 30 min ?
2. Quelle est la probabilité que l’anesthésie soit efficace plus de 1h30 ?
3. Que vaut le temps moyen d’efficacité de l’anesthésie ?

7.5 Maximum de vraisemblance

Exercice 1 :

Soient $(X_i)_{i <= n}$ des variables modélisant des données $(x_i)_{i <= n}$ appartenant à un modèle paramétrique exponentiel $\mathcal{E}(\lambda)$ de densité

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

Donner l’estimateur du maximum de vraisemblance de λ

Exercice 2 :

Soit $p \in]0; 1[$ et X une var suivant une loi géométrique $\mathcal{G}(p)$

$$P(X = k) = p(1-p)^{k-1} \quad k \in \mathbb{N}^*$$

1. Donner l’estimateur de p par la méthode du maximum de vraisemblance
2. Des observations de X sur 12 individus choisis au hasard sont les suivantes

$$1; 2; 1; 4; 2; 1; 1; 5; 3; 2; 10; 1$$

Donner une estimation ponctuelle pour p .

Exercice 3 :

Soient $\mu \in \mathbb{R}$ et X une var de densité

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

Donner l’estimateur du maximum de vraisemblance de μ

7.6 Intervalles de confiance et test d'hypothèses

Exercice 1 :

La durée moyenne de séjour de référence des patients hospitalisés pour un infarctus du myocarde (IDM) et sortis vivants de l'hôpital est de 16 jours. La durée de séjour moyenne de séjour de 408 patients de 80 ans et plus, hospitalisés pour un infarctus du myocarde dans 3 départements de la région Rhône-Alpes et sortis vivants de l'hôpital, a été estimée à 19 jours avec un écart type de 13,23 jours.

La durée moyenne de séjour de ces patients est-elle significativement différente de la durée de séjour de référence au risque alpha de 5% ?

Exercice 2 :

La durée moyenne de séjour de référence des patients hospitalisés pour un infarctus du myocarde (IDM) et sortis vivants de l'hôpital est de 16 jours. La durée de séjour moyenne de séjour de 60 patients de 80 ans et plus, hospitalisés pour un infarctus du myocarde dans 3 départements de la région Rhône-Alpes et sortis vivants de l'hôpital, a été estimée à 19 jours avec un écart type de 13,23 jours.

La durée moyenne de séjour de ces patients est-elle significativement différente de la durée de séjour de référence au risque alpha de 5% ?

Exercice 3 :

Dans une population, la prévalence d'une maladie est de 30%

Un échantillon de 64 patients indique une proportion de malades de 46%

1. La proportion estimée est-elle différente de 30% au risque alpha de 5% ?
2. Donner un intervalle de confiance à 95% pour la prévalence des patients

Exercice 4 :

On souhaite comparer le montant des salaires mensuels moyens des chefs de rayon dans les grandes surfaces pour les régions A et B. Pour cela, on effectue des sondages auprès de quelques salariés choisis au hasard. Dans la région A, on dispose d'un échantillon de 74 individus dont le salaire moyen est de 1820 euros. Dans la région B, on dispose d'un échantillon de 66 individus dont le salaire moyen est de 1968 euros.

L'écart-type commun est de 150.

Les salaires moyens sont-ils différents au risque alpha de 5% ?

Exercice 5 :

Un essai clinique compare la proportions de malades dans un groupe traité et un groupe placebo La proportion de malades dans le groupe traité est de 7% sur 30 individus La proportion de malades dans le groupe placebo est de 25% sur 60 individus

Les proportions sont-ils différentes au risque alpha de 5% ?

8 Rappel sur quelques bases de mathématiques

8.1 Notations

Lors d'une définition, il est possible de rencontrer le symbole $:=$ qui signifie seulement "égal par définition à".

On rappelle les quantificateurs utilisés dans ce poly : \forall signifie "Quel que soit" ou "pour tout", et \exists signifie "il existe". On pourra réfléchir à l'importance de l'ordre de ces quantificateurs, en méditant sur la différence entre les phrases

- "Il existe une solution à tous les problèmes" (*le chocolat ?*)
- "Pour tout problème, il existe une solution" (*adaptée au problème*)

Dans tout ce polycopié, nous utiliserons abondamment la notation $\mathbb{1}$ pour désigner la fonction indicatrice :

$$\mathbb{1}_{x \in A} := \begin{cases} 0 & \text{si } x \notin A \\ 1 & \text{si } x \in A \end{cases}$$

8.2 Rappel sur les ensembles

Un ensemble E est dit au plus dénombrable si l'on peut numéroter ses éléments, c'est à dire s'il existe une **injection** de E dans \mathbb{N} . Les ensembles finis, \mathbb{N} , \mathbb{Z} et \mathbb{Q} par exemple sont "au plus dénombrables", mais \mathbb{R} n'est pas dénombrable.

On note \cup l'union de deux ensembles, et \cap leur intersection. On rappelle

$$x \in A \cup B \Leftrightarrow x \in A \text{ ou } x \in B$$

$$x \in A \cap B \Leftrightarrow x \in A \text{ et } x \in B$$

8.3 Calculs d'intégrales

On ne rappelle pas ici les fondements mathématique de l'intégrale, mais seulement quelques propriétés :

- Intégrale d'une constante sur un segment : $\forall a, b, C \in \mathbb{R}, \int_a^b C dx = C(b - a)$.
- Intégrale d'un monôme sur un segment : $\forall a, b \in \mathbb{R}, \forall n \geq 2, \int_a^b x^n dx = \frac{b^{n+1} - a^{n+1}}{n+1}$.
- Intégrale de la fonction exponentielle sur un segment : $\forall a, b, \alpha \in \mathbb{R}, \int_a^b e^{\alpha x} dx = \frac{e^{\alpha b} - e^{\alpha a}}{\alpha}$
- Intégrale de la fonction inverse sur un segment : $\forall a, b, \alpha \in \mathbb{R}_+^*, \int_a^b \frac{1}{\alpha x} dx = \frac{\log(\alpha b) - \log(\alpha a)}{\alpha}$
- Pour intégrer une fonction positive sur un intervalle infini, on peut prendre la limite en "poussant" la borne à l'infini :

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

- Pour intégrer une fonction f quelconque sur un intervalle infini, on peut séparer la partie positive f^+ et négative $-f^-$ (si les intégrales existent) :

$$\int_a^{+\infty} f(x) dx = \int_a^{+\infty} f^+(x) dx - \int_a^{+\infty} f^-(x) dx$$

8.4 Séries

On rencontrera dans ce cours des sommes infinies.

- Lorsque tous les termes sont positifs, $\sum_{i=0}^{+\infty} a_i = \lim_{M \rightarrow \infty} \sum_{i=0}^M a_i$.