

PROBABILITÉS ET STATISTIQUES

POLYCOPIÉ DE COURS

Pierre MOREAU
IUT de Bordeaux 1 - Département HSE
pierre.moreau@u-bordeaux1.fr

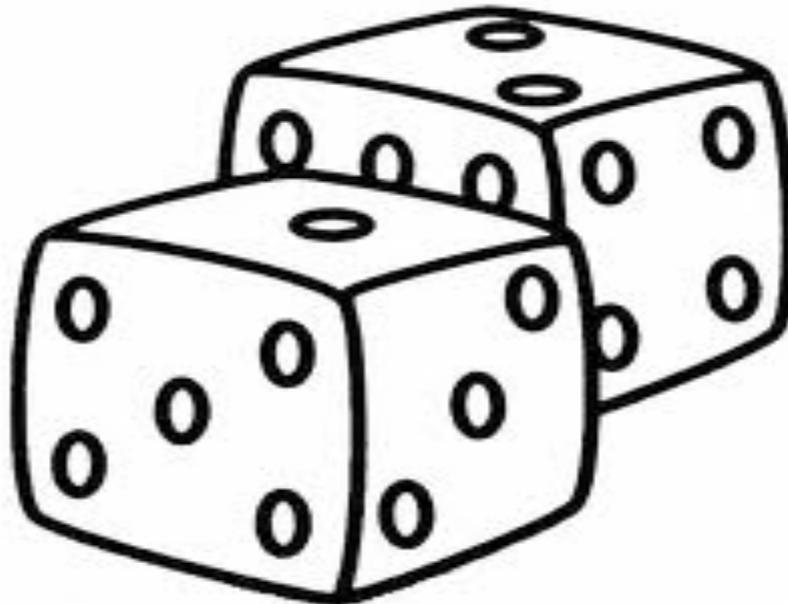


Table des matières

Table des matières	3
1 Introduction	5
2 Introduction aux probabilités	9
2.1 Les ensembles	9
2.1.1 Notion d'ensemble	9
2.1.2 Union	9
2.1.3 Intersection	10
2.1.4 Différence	10
2.1.5 Complémentarité	10
2.1.6 Produit cartésien	11
2.2 Dénombrement	11
2.2.1 Formules générales	11
2.2.2 Nombre de tirages	12
2.3 Probabilités	13
2.3.1 Définitions	13
2.3.2 Concept de Probabilité	13
2.3.3 Équiprobabilité	14
2.3.4 Probabilités conditionnelles	15
3 Variables aléatoires et applications	17
3.1 Variables aléatoires	17
3.1.1 Définition	17
3.1.2 Loi de probabilités	18
3.1.3 Fonction de répartition	19
3.1.4 Espérance	20
3.1.5 Variance et écart-type	21
3.2 Lois de probabilités classiques	24
3.2.1 Variables aléatoires discrètes	24
3.2.2 Variables aléatoires continues	26
4 Applications des probabilités aux statistiques	29
4.1 Échantillon et série statistique	29
4.1.1 Définition	29
4.1.2 Échantillonnage	29
4.1.3 Espérance et variance empirique	30
4.2 Théorèmes de convergence et applications	31
4.2.1 La loi forte des grand nombres	31

4.2.2	Le théorème central limite	32
4.3	Tests d'hypothèses	35
4.3.1	Décisions statistiques	35
4.3.2	Hypothèses statistiques	35
4.3.3	Tests d'hypothèses	36
4.3.4	Réalisation de tests à l'aide du T.C.L	37
4.4	Statistique bidimensionnelle	38
4.4.1	coefficient de corrélation	38
4.4.2	Test de linéarité et droite d'ajustement	39

Chapitre 1

Introduction

Avant toute chose je tiens à remercier vivement Marie Dusquene et Brigitte Carles pour m'avoir permis de m'inspirer de leurs notes de cours.

Quelles sont les différences entre probabilités et statistiques? Il est très fréquent, même pour des mathématiciens confirmés, de faire la confusion entre les deux et ce pour deux raisons. Premièrement les outils mathématiques sont les mêmes et un problème concret ne relève jamais exclusivement d'un domaine ou de l'autre. Pour le traiter, on doit jongler en permanence entre probabilités et statistiques. Deuxièmement, si les deux domaines ont leur propre vocabulaire afin justement de permettre la distinction, on utilisera dans la pratique indifféremment un terme probabiliste ou statistique. On le fait par commodité car ceci permet d'alléger grandement le discours, et c'est le contexte qui déterminera s'il s'agit de probabilité ou de statistique.

La différence entre ces deux branches des mathématiques s'explique en terme d'objectif. Pour le comprendre étudions le petit jeu suivant : un joueur parie sur le résultat du lancé d'un dé à 6 faces. S'il annonce le bon résultat, il gagne la valeur de la face en euros, s'il perd il ne gagne rien du tout. Quelle stratégie le joueur doit-il adopter pour maximiser ses gains?

Il s'agit de faire la remarque suivante : quelle que soit la face qu'il annonce, il a toujours une chance sur six de gagner. Par contre, les gains en cas de bonne réponse dépendent de la face : s'il annonce 1 il a une chance sur six de gagner 1 euros, s'il annonce 2 il a une chance sur six de gagner 2 euros etc...On devine ainsi que le joueur à tout intérêt à parier sur le 6 : il gagnera aussi souvent qu'avec une autre face mais gagnera plus. Mathématiquement cela s'écrit en terme **de gain moyen**, c'est-à-dire les chances de gagner multipliées par le gain, cela représente ce que le joueur gagne en moyenne par partie. On voit alors que le gain moyen en pariant sur le 6 est de $1/6 * 6 = 1$ euros tandis que les gains moyens sur les autres faces sont plus faibles : on vient de prouver que jouer le 6 est la meilleure stratégie. Le joueur vient de faire des **probabilités** .

Le raisonnement se tient mais on part tout de même d'une hypothèse : toutes les faces ont la même probabilité de sortir, à savoir $1/6$, mais est-ce bien le cas? Pour le vérifier, le joueur lance avant de commencer à parier 100 fois le dé et remarque que le 6 ne sort pas une seule fois et que le 4 sort une fois sur deux... il se dit donc naturellement que le dé n'est pas équilibré et que le 4 a une chance sur deux de sortir

tandis que le 6 n'a aucune chance de sortir, il vient de faire des **statistiques**. Il en déduit donc qu'il vaut mieux jouer le 4 (gain moyen de $1/2 * 4 = 2$ euros) que le 6 (gain moyen de $0 * 6 = 0$ euros), il vient de refaire des probabilités.

La connaissance des chances de sortie d'une face s'appelle la **loi de probabilité** du dé. Faire des statistiques c'est essayer de déterminer cette loi ; faire des probabilités, c'est utiliser cette loi pour faire des calculs de chance, déterminer des stratégies etc...

Dans la pratique, on fait sans cesse des allers-retours entre statistiques et probabilités : on commence par observer un phénomène et en déduire quelques informations sur la loi (stat), on en déduit quelques propriétés (proba), on vérifie que ces propriétés se réalisent bien (stat) etc...

Nous verrons dans le cours trois types de méthodes d'obtenir des informations sur la loi de probabilité :

1. **Les approches par valeurs moyennes.** Le nom est assez explicite, il s'agit d'approcher la loi par la moyenne des observations. Dans l'exemple précédent le joueur observe 50 fois la face 4 sur 100 lancers, il en déduit que la probabilité d'obtenir 4 est de $50/100 = 1/2$. Cette approche naturelle se justifie mathématiquement par **la loi des grands nombres**. Cette approche très simple a un défaut majeur : son imprécision. En effet rien n'indique combien de lancers il faut faire pour obtenir une bonne approximation, et pire, rien n'indique quelle marge d'erreur on risque de commettre en se contentant de 100 lancers. Il est tout à fait possible, bien qu'improbable, que le dé soit équilibré bien que le 4 soit sorti une fois sur deux !
2. **Les approches par intervalles de confiance.** Cette approche plus fine se base sur le **théorème central limite**, il ne s'agit plus de trouver une valeur exacte mais de trouver un encadrement assez fin de la quantité recherchée. Si l'on applique cette méthode au même problème que ci-dessus, on aura une conclusion de la forme : "il y a 99% de chances que la probabilité d'obtenir un 4 soit entre $0,9/2$ et $1,1/2$ ". L'estimation obtenue est moins fine mais on sait quels sont les risques d'erreurs que l'on prend.
3. **Les approches par tests d'hypothèses.** Dans cette approche, un peu plus générale que la précédente, il s'agit de rejeter une hypothèse en sachant avec quelle chance on risque de se tromper. Le principe est un peu le même que celui du "qui est-ce ?" :
mon "adversaire" choisit un nom d'acteur que je dois deviner uniquement en posant des questions dont les réponses sont oui ou non. Je commence en demandant s'il s'agit d'un homme, il me répond que non. A ce stade, je peux rejeter avec certitude l'hypothèse Clint Eastwood, par contre je ne peux pas affirmer avec certitude qu'il s'agit d'Uma Thurman. C'est exactement pareil en statistique : on peut écarter mais pas affirmer. Du moins presque exactement pareil, car la réponse de l'adversaire lorsque l'on fait des stats n'est jamais "non" mais plutôt "non à 99%"...
Reprenons l'exemple du dé et faisons l'hypothèse "le dé est équilibré". Sous ces

hypothèses, la probabilité d'obtenir la face 4 cinquante fois ou plus est d'environ $1,7 * 10^{-14}$. Pour comprendre ce que cela représente, on peut se dire que 10^{14} est l'ordre de grandeur du nombre de mètre carré de la Terre, on a donc autant de chances de faire plus de cinquante fois 4 que de trouver la bonne parcelle d'un mètres carrés sur Terre. Ainsi, on peut rejeter l'hypothèse "le dé est équilibré" avec $1,7 * 10^{-14}$ chances de se tromper.

Dans le deuxième chapitre, nous présenterons les concepts de base des probabilités et le vocabulaire associé. Dans le troisième, nous nous intéresserons à des concepts probabilistes plus avancés, les variables aléatoires et les distributions de probabilité. Dans le quatrième chapitre enfin, nous verrons comment utiliser les résultats du second chapitre pour faire des statistiques.

Chapitre 2

Introduction aux probabilités

2.1 Les ensembles

2.1.1 Notion d'ensemble

La notion d'ensemble est la base des probabilités et de la statistique. Un ensemble peut être considéré comme une collection, ou une classe, d'objets appels éléments. Il est équivalent de dire “ x est un élément de l'ensemble E ” et “ x appartient à E ”, mathématiquement cela se note $x \in E$. Quand c'est possible, on décrit un ensemble par la liste de ses éléments entre des accolades. Par exemple si E est l'ensemble de tous les résultats possibles d'un lancer de dé, $E = \{1, 2, 3, 4, 5, 6\}$.

Si E et F sont deux ensembles, on dit que F est une partie de E , ou que F est inclus dans E , ou que F est un sous-ensemble de E , si est seulement si tous les éléments de F sont aussi des éléments de E ; mathématiquement cela se note $F \subset E$. Par exemple si F est l'ensemble de tous les résultats pairs possibles d'un lancer de dé, alors $F = \{2, 4, 6\}$ et $F \subset E$.

Enfin on note \emptyset l'ensemble vide, c'est-à-dire l'ensemble qui ne contient aucun élément.

Dans les sous-sections suivantes, E , E_1 et E_2 seront trois sous-ensembles d'un ensemble Ω .

2.1.2 Union

L'ensemble de tous les éléments qui appartiennent à E_1 ou E_2 ou aux deux est appelé **union** de E_1 et E_2 , noté $E_1 \cup E_2$.

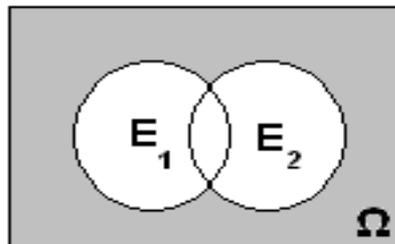


FIGURE 2.1 – Union

2.1.3 Intersection

L'ensemble de tous les éléments qui appartiennent à la fois à E_1 et E_2 est appelé **intersection** de E_1 et E_2 , noté $E_1 \cap E_2$.

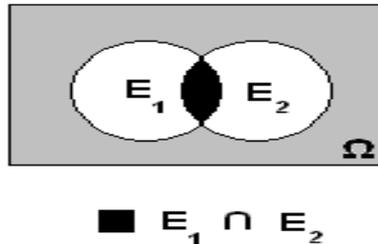


FIGURE 2.2 – Intersection

2.1.4 Différence

L'ensemble de tous les éléments de E_1 qui n'appartiennent pas à E_2 est appelé **différence** de E_1 et E_2 , noté $E_1 - E_2$.

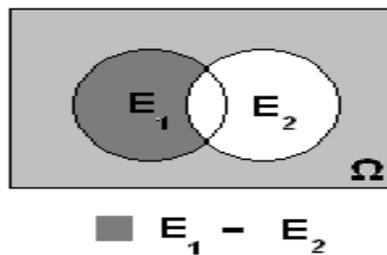


FIGURE 2.3 – Différence

2.1.5 Complémentarité

L'ensemble de tous les éléments de Ω qui n'appartiennent pas à E est le **complémentaire** de E dans Ω noté \bar{E} , alors $E \cap \bar{E} = \emptyset$ et $E \cup \bar{E} = \Omega$. Il faut faire attention

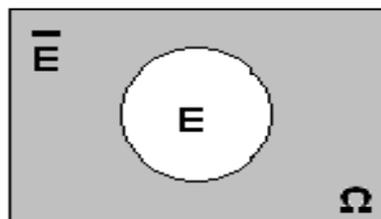


FIGURE 2.4 – Complémentarité

car le complémentaire dépend de l'ensemble de départ. Ainsi le complémentaire de $\{1, 2\}$ dans $\{1, 2, 3\}$ est $\{3\}$, mais son complémentaire dans $\{1, 2, 3, 4, 5\}$ est $\{3, 4, 5\}$.

2.1.6 Produit cartésien

L'ensemble de tous les couples d'un élément de E_1 et d'un élément de E_2 est appelé produit cartésien de E_1 et E_2 , noté $E_1 \times E_2$. Attention car $E_1 \times E_2 \neq E_2 \times E_1$ comme on peut le voir dans l'exemple suivant :

$$\{1, 2\} \times \{3, 4\} = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$$

$$\{3, 4\} \times \{1, 2\} = \{(3, 1), (3, 2), (4, 1), (4, 2)\}$$

2.2 Dénombrement

Le nombre d'éléments d'un ensemble E est appelé cardinal de E , noté $Card(E)$ ou $|E|$; dénombrer un ensemble c'est trouver son cardinal. Lorsque l'on fait ceci, on est fréquemment en présence de nombres très grands qui ne sont pas très parlant. Avant de présenter les techniques de dénombrement, nous allons donc donner quelques ordres de grandeur qui permettent de se faire une idée de ce que représentent ces nombres.

- 10^{12} : une heure-lumière, soit le nombre de mètres parcourus par la lumière en une heure.
- 10^{16} : une année-lumière, également le nombre de cellules dans le corps humain.
- $4 * 10^{26}$: nombre de mm^3 d'eau sur terre.
- 10^{28} : nombre d'atomes dans le corps humain.
- 10^{50} : nombre d'atomes de la Terre.
- 10^{80} : nombre d'atomes de l'Univers.

Ces nombres semblent gigantesques mais sont pourtant faciles à atteindre. Cherchons par exemple le nombre de mélanges possibles pour un jeu de 32 cartes. Il y a 32 possibilités pour la première carte, fois 31 possibilités pour la seconde...on a donc $32! \simeq 10^{35}$ mélanges possibles soit dix millions de fois plus qu'il n'y a d'atomes dans un corps humain ! Avec un jeu de 52 cartes on arrive à environ 10^{67} mélanges possibles soit 1000000000000000000 de fois plus qu'il n'y a d'atomes sur Terre...Citons également le nombre de Shannon, 10^{120} , correspondant au nombre **minimum** de parties d'échec possibles.

2.2.1 Formules générales

Si E_1 et E_2 sont deux ensembles finis alors,

$$Card(E_1 \cup E_2) = Card(E_1) + Card(E_2) - Card(E_1 \cap E_2)$$

$$Card(E_1 \times E_2) = Card(E_1) * Card(E_2)$$

Nous allons à présent étudier le dénombrement de listes d'éléments, en distinguant les cas avec ou sans ordre.

2.2.2 Nombre de tirages

Considérons le modèle suivant : on tire p boules dans une urne contenant n boules, combien y-a-t'il de tirages possibles ? Avant de répondre à cette question, remarquons les faits suivants :

- On peut avec ce modèle presque tous les phénomènes aléatoires. Par exemple lancer un dé est équivalent à tirer une boule dans une urne à 6 boules, piocher une carte d'un jeu de 32 cartes est équivalent à tirer une boule dans une urne à 32 boules etc...
- Il y a deux questions à se poser sur le tirage : est-il **avec ou sans ordre**, est-il **avec ou sans remise**. Suivant les réponses à ces deux questions, le nombre de tirages possibles est différent.
- Pour les tirages sans remise, on a nécessairement $p \leq n$.

Avec ordre, avec remise

Le nombre de tirages avec ordre et avec remise de p boules dans une urne à n boules (on parle alors de p -listes) est

$$n^p$$

Exemple :

Si $E = \{a, b, c, d, e\}$, il y a $5^3 = 125$ mots de 3 lettres.

Avec ordre, sans remise

Le nombre de tirages avec ordre et sans remise de p boules dans une urne à n boules (on parle alors d'arrangements) est noté A_n^p , il vérifie la formule :

$$A_n^p = n * (n - 1) * \dots * (n - p + 1) = \frac{n!}{(n - p)!}$$

On rappelle que par convention, $0! = 1$.

Exemple :

Avec les lettres A,B et C, combien y-a-t-il de façons de former des paires de lettres distinctes ?

Sans ordre, sans remise

Le nombre de tirages sans ordre et sans remise de p boules dans une urne à n boules (on parle alors de combinaisons) est noté C_n^p , il vérifie la formule :

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

On a alors les formules suivantes :

- $C_n^p = C_n^{n-p}$
- Pour $1 \leq p \leq n - 1$, $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$

Sans ordre, avec remise

Il n'y a pas de formule dans ce cas précis ! Cela pourrait être gênant à priori mais nous verrons dans la section 2.3.4 que ce cas de figure ne correspond pas à un cas d'équiprobabilité, et qu'il est alors inutile de dénombrer pour trouver des probabilités.

2.3 Probabilités

2.3.1 Définitions

1. On appelle **expérience aléatoire**, ou une **épreuve**, une expérience donnant un résultat que l'on ne peut prévoir à l'avance. Par exemple lancer un dé.
2. L'ensemble de tous les résultats possibles est l'**ensemble fondamental** Ω . Ici $\Omega = \{1, 2, 3, 4, 5, 6\}$.
3. Les sous ensembles de Ω contenant un seul résultat possible de cette expérience sont les **événements élémentaires**. Ici les événements élémentaires sont donc $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$ et $\{6\}$.
4. Plus généralement un sous-ensemble de Ω , c'est-à-dire un ensemble de résultats possibles, est un événement. Si par exemple E est l'événement "obtenir un résultat pair", alors $E = 2, 4, 6$.
5. Ω est l'événement certain, \emptyset est l'événement impossible, et si E est un événement, son complémentaire \bar{E} est l'événement contraire. Ici $\bar{E} = \{1, 3, 5\}$ c'est-à-dire "obtenir un résultat impair".
6. Deux événements E_1 et E_2 sont dits **incompatibles** ou **disjoints** si leur réalisation simultanée est impossible, c'est-à-dire si $E_1 \cap E_2 = \emptyset$. Par exemple "obtenir un 4 ou un 6" et "obtenir un résultat impair" sont deux événements incompatibles.

2.3.2 Concept de Probabilité

Une probabilité P sur un ensemble fondamental Ω est une fonction qui à tout événement E associe ses "chances" de se réaliser $P(E)$, elle doit vérifier les trois axiomes suivants :

- **Axiome 1**

$P(E)$ est un réel compris entre 0 et 1.

- **Axiome 2**

Pour l'événement certain, $P(\Omega) = 1$

Pour l'événement impossible, $P(\emptyset) = 0$.

- **Axiome 3 : Axiome d'additivité**

Si E_1 et E_2 sont deux événements incompatibles ($E_1 \cap E_2 = \emptyset$) alors

$$P(E_1 \cup E_2) = P(E_1) + P(E_2).$$

Plus généralement si (E_i) est une suite d'évènements deux à deux incompatibles alors

$$P\left(\bigcup E_i\right) = \sum P(E_i).$$

Propriétés

Si E_1 et E_2 sont deux évènements quelconques :

- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$,
- $P(\bar{E}) = 1 - P(E)$.

2.3.3 Équiprobabilité

On dit qu'on est dans le cas d'équiprobabilité quand l'espace fondamental Ω est fini et que tous les évènements élémentaires ont la même probabilité. Dans ce cas il est très simple de calculer la probabilité de n'importe quel évènement en faisant le ratio des cas favorables sur tous les cas possibles. Soit

$$P(E) = \frac{\text{Card}(E)}{\text{Card}(\Omega)}.$$

Trouver la probabilité d'un évènement donné est donc équivalent à trouver son nombre d'éléments, ce qui justifie les résultats de dénombrement précédents.

L'exemple suivant va permettre de comprendre pourquoi il n'est pas utile de dénombrer dans le cas d'un tirage sans ordre avec remise.

Exemple : On lance deux fois un dé équilibré. Quelle est la probabilité d'obtenir une paire ?

Soit E l'évènement "obtenir une paire", il s'agit donc de tirer **avec remise** deux boules dans une urne de 6 boules. L'ordre n'a pas d'importance, on peut donc décider de travailler avec ou sans. Les deux tableaux ci-dessous contiennent tous les résultats possibles, avec en gras les résultats favorables.

avec ordre

dé	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$P(E) = \frac{6}{36} = \frac{1}{6}$$

sans ordre

dé	1	2	3	4	5	6
1	{1,1}	{1,2}	{1,3}	{1,4}	{1,5}	{1,6}
2		{2,2}	{2,3}	{2,4}	{2,5}	{2,6}
3			{3,3}	{3,4}	{3,5}	{3,6}
4				{4,4}	{4,5}	{4,6}
5					{5,5}	{5,6}
6						{6,6}

$$P(E) = \frac{6}{21} = \frac{2}{7}$$

On obtient deux résultats différents, il y a donc une erreur quelque part. Le problème vient du cas sans ordre : tous les évènements élémentaires n'ont pas la même probabilité. Par exemple il y a deux façons d'obtenir $\{1, 2\}$: $(1, 2)$ ou $(2, 1)$; tandis qu'il n'y a qu'une seule façon d'obtenir $\{1, 1\}$: $(1, 1)$!

Ainsi s'il y a bien dans ce cas 6 tirages favorables et 21 tirages possibles, l'absence d'équiprobabilité nous interdit d'utiliser la formule $P(E) = \frac{\text{Card}(E)}{\text{Card}(\Omega)}$, dénombrer ne nous est donc d'aucune utilité.

Il ne faut donc jamais se placer dans le cas d'un tirage sans ordre avec remise.

2.3.4 Probabilités conditionnelles

Définitions

On appelle probabilité conditionnelle de E_1 sachant E_2 la probabilité que l'évènement E_1 se réalise sachant que l'évènement E_2 s'est réalisé. Cela s'écrit $P(E_1|E_2)$ et on a la formule :

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

On en déduit la formule suivante, souvent utile pour calculer la probabilité d'une intersection :

$$P(E_1 \cap E_2) = P(E_1|E_2) * P(E_2) = P(E_2|E_1) * P(E_1)$$

Exemple :

Quelle est la probabilité d'obtenir un nombre inférieur 4 au cours du jet unique d'un dé

1. sans autre information
2. en sachant que le résultat obtenu est impair

On dit que deux évènements E_1 et E_2 sont indépendants si $P(E_1 \cap E_2) = P(E_1) * P(E_2)$. En utilisant la formule précédente, on voit qu'alors $P(E_1|E_2) = P(E_1)$ et que $P(E_2|E_1) = P(E_2)$, c'est-à-dire que la réalisation d'un des deux évènements n'a aucune incidence sur les chances de réalisation du second.

Théorème de Bayes

Soient E_1, E_2, \dots, E_n des évènements deux à deux incompatibles dont l'union est l'ensemble fondamental Ω . Alors si A est un évènement quelconque on a :

$$P(A) = \sum_{k=1}^n P(A|E_k)P(E_k)$$

Exemple :

Une enquête établie dans la population des automobilistes français a révélé que 60% des personnes interrogées ne connaissent qu'imparfaitement le code de la route (événement I) et que parmi celles-ci, 25% ont déjà provoqué un accident. 30% des personnes interrogées connaissent le code, mais ne l'appliquent pas, notamment sur les limitations de vitesse (événement V). Parmi celles-ci, 40% ont déjà causé un accident. Enfin, le reste des personnes interrogées connaît et respecte le code (événement C), mais parmi elles, 10% ont eu un accident. Un automobiliste cause un accident (événement A). Quelle est la probabilité qu'il fasse partie de la première catégorie ?

Chapitre 3

Variables aléatoires et applications

3.1 Variables aléatoires

3.1.1 Définition

On appelle variable aléatoire toute expérience aléatoire dont le résultat est **numérique**, c'est à dire un nombre réel. Les variables aléatoires sont dites **discrètes** si elles peuvent prendre un nombre fini ou dénombrable de valeurs, **continues** dans le cas contraire. Par exemple :

1. Le résultat du lancé d'un dé est une variable aléatoire discrète car elle ne peut prendre qu'un nombre fini de valeurs.
2. Le nombre de recherches google sur une journée est une variable aléatoire discrète. Son ensemble de valeurs possibles \mathbb{N} est infini mais dénombrable.
3. La durée de vie en heures d'une ampoule est une variable aléatoire continue. Son ensemble de valeurs possibles \mathbb{R}_+ est infini non-dénombrable.

On dit que deux variables aléatoires X et Y sont **indépendantes** si tout évènement associé à la variable aléatoire X (par exemple $(X = 3)$ ou $(X > 5)$) est indépendant de tout évènement associé à la variable aléatoire Y . Concrètement, cela signifie que les deux variables aléatoires n'ont aucun lien l'une avec l'autre. Par exemple :

1. Soit X la variable aléatoire mesurant le nombre de personnes dans la file d'attente d'un bureau de poste et soit Y la variable aléatoire mesurant le nombre de guichets ouverts dans ce même bureau. Ces deux variables sont liées, elles ne sont pas indépendantes.
2. Soit X la variable aléatoire mesurant le nombre de personnes dans la file d'attente d'un bureau de poste et soit Y la variable aléatoire mesurant la somme d'un lancé de deux dés. Assez clairement, ces deux variables n'ont aucun lien, elles sont indépendantes.
3. L'exemple suivant est essentiel et on l'utilisera énormément quand on fera des statistiques. On répète exactement la même expérience aléatoire n fois, X_1 est la variable aléatoire correspondant au résultat de la première expérience, X_2 est la variable aléatoire correspondant au résultat de la deuxième expérience etc... Alors ces variables sont deux à deux indépendantes, et elles ont de plus la même loi de probabilité (cf sous-section suivante).

3.1.2 Loi de probabilités

Cas discret

Soit X une variable aléatoire discrète soit $(x_i)_{i \in I}$ la liste de toutes ses valeurs possibles. On appelle **loi de probabilités** de X , ou **distribution de probabilités** de X l'ensemble des probabilités $p_i = P(X = x_i)$ pour tout $i \in I$.

Exemple :

Dans une boîte, il y a 4 boules rouges et 5 boules blanches. On tire des boules jusqu'à obtenir une boule blanche et soit X le nombre de boules rouges tirées. Déterminez la loi de X dans le cas où le tirage est sans remise, puis dans le cas où le tirage est avec remise.

Réponse :

Si le tirage est sans remise, la variable X peut donc prendre les valeurs 0,1,2,3 et 4. L'évènement $X = 0$ est donc "la première boule est blanche", ainsi :

$$P(X = 0) = \frac{5}{9}$$

$$P(X = 1) = \frac{4}{9} \frac{5}{8} \text{ etc...}$$

Ainsi on trouve :

x_i	0	1	2	3	4
p_i	$\frac{5}{9}$	$\frac{5}{18}$	$\frac{5}{42}$	$\frac{5}{126}$	$\frac{1}{126}$

Si le tirage est avec remise, X peut alors prendre n'importe quelle valeur entière. On a alors, pour tout $n \in \mathbb{N}$, $P(X = n) = \left(\frac{4}{9}\right)^n \frac{5}{9}$

Cas continu

Pour des variables aléatoires continues, cela n'a pas de sens de chercher la probabilité d'obtenir une valeur exacte (sauf dans le cas des lois dites "exotiques" que nous n'aborderons pas dans ce cours). On cherche plutôt la probabilité d'obtenir un résultat compris entre deux valeurs. Par exemple, la probabilité que la durée de vie d'une ampoule soit exactement de 120 heures, 36 minutes, 12 secondes, 45 centièmes etc... n'est pas ce qui nous importe (cette probabilité est de toute façon nulle), par contre on aimerait savoir quelle est la probabilité qu'elle soit par exemple comprise entre 120 et 121 heure.

Connaître la loi d'une variable aléatoire continue X , c'est savoir calculer la probabilité que X appartienne à I , notée $P(X \in I)$, pour tout intervalle I . En général c'est une question très difficile mais que l'on sait traiter pour les variables aléatoires à **densité**.

Soit X une variable aléatoire continue et soit f une fonction de \mathbb{R} dans \mathbb{R} . On dit que f est la **densité de probabilité** de X si :

1. f est positive : pour tout $x \in \mathbb{R}$, $f(x) \geq 0$.
2. Pour tout intervalle I de \mathbb{R} , $P(X \in I) = \int_{x \in I} f(x).dx$.

Le deuxième point entraîne donc que $\int_{x \in \mathbb{R}} f(x).dx = 1$.

Exercice :

Soit X une variable aléatoire continue de densité f avec $f(x) = 0$ si $x \leq 0$, $f(x) = ce^{-2x}$ si $x > 0$, où c est une constante.

- Déterminez c .
- Déterminez la probabilité que X soit comprise entre 2 et 10.

3.1.3 Fonction de répartition

Soit X une variable aléatoire, discrète ou continue. On appelle **fonction de répartition** de X la fonction F définie de \mathbb{R} dans \mathbb{R} par

$$F(x) = P(X \leq x).$$

La fonction de répartition est donc une fonction croissante vérifiant de plus :

1. Pour tout $a \in \mathbb{R}$, $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$.
2. Pour tout intervalle $]a, b]$, $P(X \in]a, b]) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$.

Le calcul explicite de la fonction de répartition dépend de la nature de la variable aléatoire (discrète ou continue).

• **Cas discret**

Soit X une variable aléatoire discrète et soit $(x_i)_{i \in I}$ la liste de ses valeurs possibles. Pour tout $x \in \mathbb{R}$ on a

$$F(x) = \sum_{x_i \leq x} P(X = x_i).$$

• **Cas continu**

Soit X une variable aléatoire continue et soit f sa densité. Pour tout $x \in \mathbb{R}$ on a

$$F(x) = \int_{-\infty}^x f(t)dt.$$

Cette égalité a deux conséquences immédiates :

1. Pour tout intervalle $]a, b]$, $P(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t)dt$.
2. Pour tout $x \in \mathbb{R}$, $F'(x) = f(x)$.

Exemple :

Reprenons l'exemple de l'urne à 9 boules dans le cas d'une tirage sans remises. Déterminez la fonction de répartition de X .

Réponse :

On a : si $x < 0$ alors $F(x) = 0$,

si $0 \leq x < 1$ alors $F(x) = 5/9$,

si $1 \leq x < 2$ alors $F(x) = F(0) + F(1) = 5/9 + 5/18 = 5/6$,

si $2 \leq x < 3$ alors $F(x) = F(0) + F(1) + F(2) = 5/9 + 5/18 + 5/42 = 20/21$,

si $3 \leq x < 4$ alors $F(x) = 125/126$,

si $4 \leq x$ alors $F(x) = 1$.

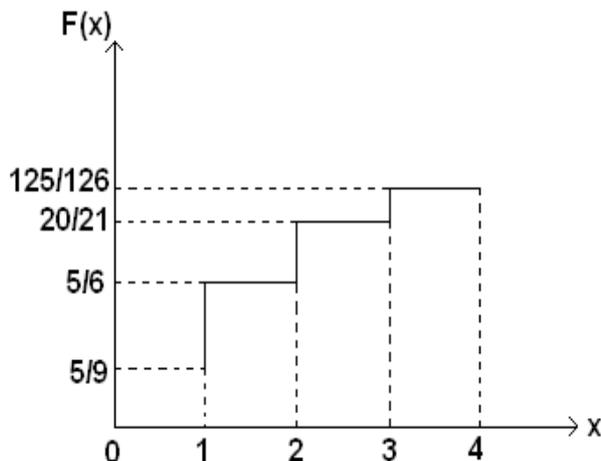


FIGURE 3.1 – Fonction de répartition

3.1.4 Espérance

Définition

Soit X une variable aléatoire, son espérance, notée $E(X)$, est sa valeur moyenne. Le calcul de l'espérance dépend du type de la variable :

- **Cas discret**

Soit X une variable aléatoire discrète et soit $(x_i)_{i \in I}$ la liste de ses valeurs possibles. Alors

$$E(X) = \sum_{i \in I} P(X = x_i)x_i.$$

- **Cas continu**

Soit X une variable aléatoire continue et soit f sa densité. Alors

$$E(X) = \int_{-\infty}^{+\infty} f(x)x dx.$$

Exemple :

Un joueur lance un dé. Il gagne 20 euros s'il tire un 2, 40 euros s'il tire un 4 et perd 30 euros s'il tire un 6, le coup étant nul pour tous les autres chiffres tirés. Calculez l'espérance de ses gains (le gain moyen).

Réponse :

Soit X la variable aléatoire de ses gains. X peut prendre les valeurs -30 , 0 , 20 et 40 avec les probabilités $1/6$, $3/6$, $1/6$ et $1/6$. Donc $E(X) = -30 * 1/6 + 0 * 3/6 + 20 * 1/6 + 40 * 1/6 = 5$ euros.

Exercice :

Soit X une variable aléatoire continue de densité f avec $f(x) = 0$ si $x \leq 0$, $f(x) = 2e^{-2x}$ si $x > 0$. Déterminez $E(X)$.

Réponse :

$$E(X) = \int_{-\infty}^{+\infty} f(x)xdx = \int_0^{+\infty} 2xe^{-2x}dx = 2 \int_0^{+\infty} xe^{-2x}dx.$$

En intégrant par partie on trouve $E(X) = 2 \left(\left[\frac{xe^{-2x}}{-2} \right]_0^{+\infty} - \int_0^{+\infty} \frac{e^{-2x}}{-2} dx \right)$.

Ainsi on obtient $E(X) = 2 \left(\left[\frac{xe^{-2x}}{-2} \right]_0^{+\infty} - \left[\frac{e^{-2x}}{4} \right]_0^{+\infty} \right) = \frac{1}{2}$.

Propriétés

L'espérance vérifie un certain nombre de propriétés qui pour la plupart sont intuitives si l'on garde à l'esprit qu'un calcul d'espérance est un calcul de moyenne.

1. Soit X une variable aléatoire et soit c une constante, alors

$$E(cX) = cE(X).$$

2. Soit X une variable aléatoire et soit b une constante, alors

$$E(X + b) = E(X) + b.$$

3. Plus généralement si X une variable aléatoire et si b et c sont deux constantes, alors

$$E(cX + b) = cE(X) + b.$$

4. Si X et Y sont deux variables aléatoires alors

$$E(X + Y) = E(X) + E(Y).$$

5. Si X et Y sont deux variables aléatoires **indépendantes** alors

$$E(XY) = E(X)E(Y).$$

3.1.5 Variance et écart-type

L'espérance est rarement suffisant pour approcher raisonnablement une variable aléatoire. Par exemple si X est une v.a. prenant les valeurs -1 ou 1 avec une probabilité $1/2$, et si Y est une v.a. prenant les valeurs -10 et 10 avec une probabilité $1/2$, alors ces deux variables ont la même espérance : 0 . Par contre, si on décide d'approcher ces variables par leur espérance, on ne commettra pas la même erreur : on aura une erreur de 1 pour X et de 10 pour Y . Pour différencier ces deux cas, on a besoin d'outils mesurant la **dispersion**, c'est-à-dire l'écart entre la variable et sa moyenne. De nombreux outils jouent ce rôle (étendue, distance inter-quartile,...), le plus précis d'entre eux étant la variance.

Définitions

La **variance** est la "moyenne des écarts à la moyenne au carré", c'est à dire pour une v.a. X :

$$Var(X) = E[(X - E(X))^2].$$

Remarquons qu'avec cette définition, la variance de X est toujours positive, et qu'elle est nulle si et seulement la probabilité que X soit différente de son espérance est nulle. On comprend bien que la moyenne des écarts donne une indication sur la dispersion de la v.a. Par contre l'intérêt d'étudier les carrés des écarts au lieu d'étudier directement les écarts n'est pas évident et mérite discussion :

1. Prendre les carrés permet tout d'abord de ne travailler qu'avec des nombres **positifs** et donc d'avoir une moyenne d'écart non nulle. En effet si on ne regarde que les écarts à la moyenne, et bien il y a "autant" d'écarts positifs que de négatifs, et on trouverait ainsi une moyenne d'écarts nulle quelque soit la variable.
2. Pour éviter l'écueil du point précédant, on pourrait faire autrement que prendre les carrés, par exemple en calculant la moyenne de la valeur absolue des écarts (c'est d'ailleurs un autre outil mesurant la dispersion : l'**écart moyen**). Néanmoins on se priverait d'une propriété importante du carré : le carré d'un nombre a plus petit que 1 est "beaucoup" plus petit que a , le carré d'un nombre b plus grand que 1 est "beaucoup" plus grand que b . Dans le calcul de la variance, on "gonfle" les gros "écarts" tandis qu'on "diminue" les petits. Ainsi la variance indique mieux que l'écart moyen les chances d'avoir un gros écart (voir l'inégalité de Tchebichev).
3. On pourrait amplifier cette information en prenant au lieu des carrés des puissances de 4, de 6 etc... Les petits écarts seraient ainsi encore plus petits, les gros encore plus gros. Cependant travailler avec les carrés permet de conserver une structure mathématique très utile, celle d'espace de Hilbert, ce qui n'est pas le cas des autres puissances. En travaillant avec les carrés, on a ainsi une foule d'outils mathématiques très puissants et utiles (le produit scalaire et ses conséquences). Ceci dépasse largement le cadre de ce cours et nous ne pousserons pas plus loin la discussion.

Enfin l'**écart-type** $\sigma(X)$ d'une v.a. X est la racine carré de la variance de X (pour ceux qui pousseront l'étude des probabilités/statistiques plus loin, il est bon de remarquer alors que $\sigma(X)$ est en fait la norme L_2 de $X - E(X)$).

Le calcul de l'espérance dépend du type de la variable :

- **Cas discret**

Soit X une variable aléatoire discrète et soit $(x_i)_{i \in I}$ la liste de ses valeurs possibles. Alors

$$V(X) = \sum_{i \in I} P(X = x_i)(x_i - E(X))^2.$$

- **Cas continu**

Soit X une variable aléatoire continue et soit f sa densité. Alors

$$E(X) = \int_{-\infty}^{+\infty} f(x)(x - E(X))^2 dx.$$

Propriétés

La variance vérifie un certain nombre de propriétés qui pour la plupart sont intuitives si l'on garde à l'esprit qu'un calcul de variance est un calcul de moyenne d'écarts au carré. Les autres se déduisent des propriétés de l'espérance.

1. Soit X une variable aléatoire et soit c une constante, alors

$$V(cX) = c^2 V(X).$$

2. Soit X une variable aléatoire et soit b une constante, alors

$$V(X + b) = V(X).$$

3. Plus généralement si X une variable aléatoire et si b et c sont deux constantes, alors

$$V(cX + b) = c^2V(X).$$

4. La formule suivante est extrêmement utile pour calculer simplement une variance, elle indique que la variance est aussi “la moyenne des carrés moins le carré de la moyenne”.
Si X est une variable aléatoire alors

$$V(X) = E(X^2) - E(X)^2.$$

5. Si X et Y sont deux variables indépendantes alors

$$V(X + Y) = V(X) + V(Y).$$

6. Enfin la variance permet d'estimer les probabilités d'écarts à la moyenne. Ce résultat très important que l'on donne sans démonstration porte le nom d'inégalité de Tchebichev.
Soit X une v.a. et soit $\varepsilon > 0$, alors

$$P(|X - E(X)| > \varepsilon) \leq \frac{\sigma(X)^2}{\varepsilon^2}.$$

Exemple :

1. Évaluer l'espérance de la somme des points au cours de jets d'une paire de dés.
2. Évaluer la variance de la somme des points au cours de jets d'une paire de dés.
3. Évaluer l'écart-type de la somme des points au cours de jets d'une paire de dés.

Réponse :

Soient X et Y les variables aléatoires correspondants au résultat de chacun des dés. Ces deux variables sont ainsi indépendantes et la somme des dés est donc la v.a. $X + Y$.

1. On a donc : $E(X) = E(Y) = 1 * (1/6) + 2 * (1/6) + \dots + 6 * (1/6) = 7/2$.
On en déduit $E(X + Y) = E(X) + E(Y) = 7$.
2. $E(X^2) = E(Y^2) = 1^2 * (1/6) + 2^2 * (1/6) + \dots + 6^2 * (1/6) = 91/6$.
Donc $Var(X) = Var(Y) = 91/6 - (7/2)^2 = 35/12$.
Par indépendance de X et Y on a donc $Var(X+Y) = Var(X)+Var(Y) = 35/6$.
3. $\sigma_{X+Y} = \sqrt{Var(X + Y)} = \sqrt{35/6}$.

Variables aléatoires centrées réduites

On dit qu'une variable aléatoire X est **centrée** si $E(X) = 0$ (sa moyenne est nulle), **réduite** si $V(X) = 1$. Ce n'est pas le cas en général mais on peut toujours se ramener à ce cas de figure.

Soit X une v.a. de moyenne μ et d'écart-type σ , alors la v.a. Y définie par

$$Y = \frac{X - \mu}{\sigma} \text{ est centrée réduite.}$$

La démonstration de ce résultat est laissée en exercice, il s'agit d'une application directe des propriétés de l'espérance et de la variance.

Quand on travaille avec une variable centrée réduite, les calculs sont beaucoup plus simple que dans le cas général. Il est ainsi souvent plus facile quand on est face à une v.a X d'introduire la v.a. centrée réduite Y comme ci-dessus, de travailler sur cette nouvelle variable, puis de d'utiliser les informations obtenues sur Y pour en déduire des informations sur X que de travailler directement sur X .

3.2 Lois de probabilités classiques

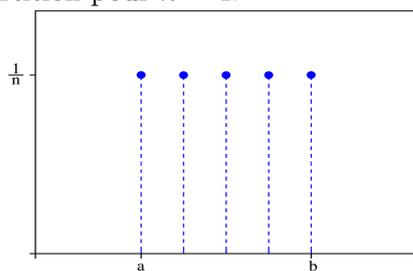
3.2.1 Variables aléatoires discrètes

Soit X une variable aléatoire pouvant prendre un nombre n **fini** de valeurs distinctes : x_1, x_2, \dots, x_n . On dit que X suit la loi uniforme si elle prend chaque valeur avec la même probabilité :

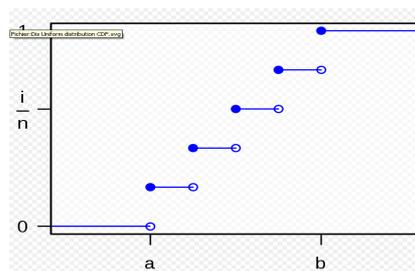
$$P(X = x_i) = \frac{1}{n}, \forall i \in \mathbb{N}, 1 \leq i \leq n.$$

Dans cas, $E(X) = \frac{x_1 + x_2 + \dots + x_n}{n}$.

On trouvera ci-dessous une représentation de la loi de probabilité et de la fonction de répartition pour $n = 4$.



loi de probabilité



Fonction de répartition

Loi de Bernoulli

Soit X une variable aléatoire ne pouvant prendre que deux valeurs, que l'on note arbitrairement 0 et 1 (ou bien *rate* et *russie*, ou encore *pile* et *face*,...). Alors on a

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

On dit que X suit la loi de Bernoulli de paramètre p , notée $\mathcal{B}(p)$.

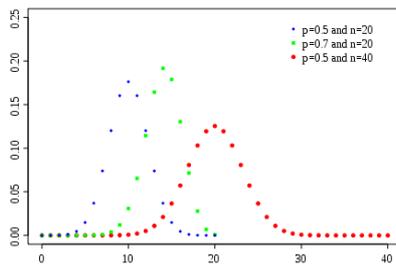
Loi Binomiale

Soit une épreuve aléatoire n'ayant que deux issues : réussie avec la probabilité p , ratée avec la probabilité $q = 1 - p$. On reproduit à l'identique n fois cette expérience et on compte dans la variable X le nombre de fois où elle a réussi (les valeurs possibles de X sont donc $0, 1, 2, \dots, n$). Alors X suit la loi Binomiale de paramètre n et p , notée $\mathcal{B}(n, p)$.

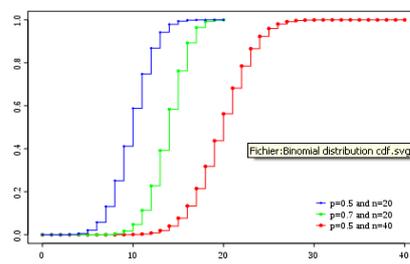
$$P(X = i) = C_n^i p^i q^{n-i}, \forall i \in \mathbb{N}, 0 \leq i \leq n.$$

Dans ce cas, $E(X) = np$.

Une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n, p)$ est donc la somme de n variables deux à deux indépendantes suivant toutes la loi $\mathcal{B}(p)$.



loi de probabilité



Fonction de répartition

Exemple :

Une chaîne de fabrication produit 40 000 fours dont 80% sont en bon état de marche, les 20% restant étant défectueux.

Le service de contrôle-qualité qui ne connaît pas ces chiffres, prélève un échantillon aléatoire de 10 fours pour estimer la qualité de l'ensemble de la fabrication. Quelle est la probabilité que dans cet échantillon on obtienne 5 fours défectueux ? En moyenne, combien de fours sont défectueux parmi 10 fours ?

Réponse :

Soit X la variable aléatoire comptant les fours défectueux parmi les 10 de l'échantillon. Chaque four ayant la même probabilité 20% d'être défectueux, X suit donc la loi $\mathcal{B}(10, 0.2)$. Ainsi $P(X = 5) = C_{10}^5 (0.2)^5 (0.8)^5 = 0.026 \approx 3\%$. Enfin, la moyenne de fours défectueux sur 10 fours est donc $E(X) = 10 \times 0.2 = 2$.

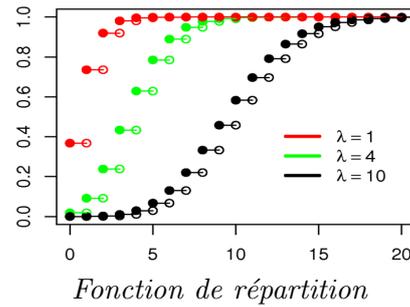
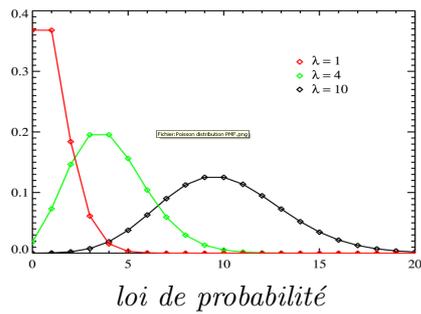
Loi de Poisson

Soit λ un nombre réel strictement positif. Une variable aléatoire X à valeur dans \mathbb{N} (c'est à dire pouvant prendre les valeurs $0, 1, 2, \dots$) suit la loi de Poisson de paramètre λ , notée $\mathcal{P}(\lambda)$ si :

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

On a alors $E(X) = \lambda$.

On utilise très souvent la loi de Poisson pour modéliser un nombre d'événements dans un laps de temps donné (nombre de requêtes sur un serveur informatique durant une heure, nombre d'appels à un S.A.V sur une journée etc...). Elle entretient de plus un lien important avec la loi exponentielle.



3.2.2 Variables aléatoires continues

Loi uniforme

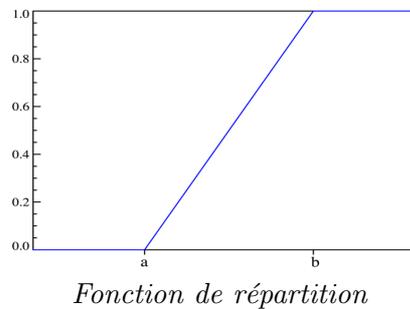
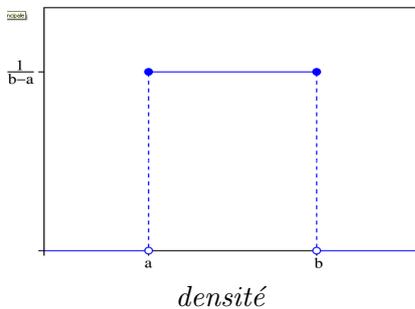
On dit que la v.a X sur la loi uniforme sur l'intervalle $[a, b]$, ou qu'elle est uniformément distribuée sur $[a, b]$, si elle prend ses valeurs dans l'intervalle $[a, b]$ avec "les mêmes probabilités". Dans ce cas, sa densité f vérifie :

- $f(x) = 0$ si $x < a$,
- $f(x) = \frac{1}{b-a}$ si $a \leq x \leq b$,
- $f(x) = 0$ si $x > b$.

Sa fonction de répartition F vérifie alors :

- $F(x) = 0$ si $x < a$,
- $F(x) = \frac{x-a}{b-a}$ si $a \leq x \leq b$,
- $F(x) = 1$ si $x > b$.

On a alors $E(X) = \frac{a+b}{2}$.



Loi exponentielle

Soit λ un nombre réel strictement positif. On dit que la variable aléatoire X suit la loi exponentielle de paramètre λ , notée $\mathcal{E}(\lambda)$, si sa densité f vérifie :

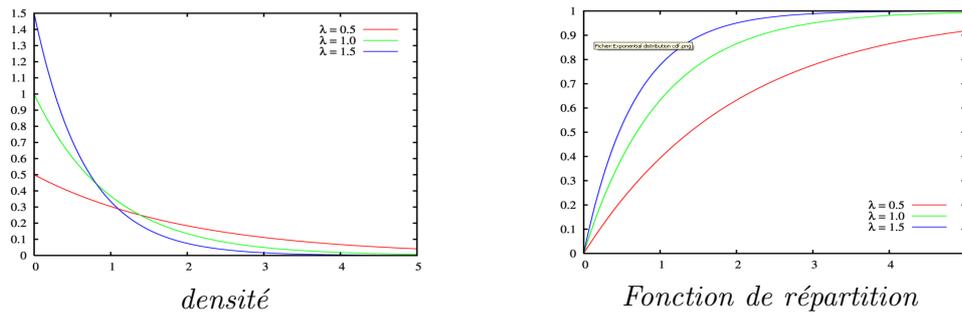
- $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$ ($\lambda > 0$),
- 0 sinon.

Sa fonction de répartition F est alors :

1. $F(x) = 1 - e^{-\lambda x}$ si $x \geq 0$,
2. $F(x) = 0$ sinon.

On a alors $E(X) = \frac{1}{\lambda}$.

On utilise souvent la loi exponentielle pour mesurer l'intervalle de temps entre deux évènements comptés par une loi de Poisson (temps entre deux requêtes informatiques, entre deux arrivées de clients à un guichet de poste,...).



Loi normale

Ces lois jouent un rôle essentiel car elles nous permettront (voir chapitre suivant) d'approcher toutes autres lois.

Loi normale centrée réduite $N(0, 1)$

La variable aléatoire X suit la loi $N(0, 1)$ si sa densité f vérifie :

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

Dans ce cas $E(X) = 0$ et $V(X) = 1$ ce qui justifie l'appellation "centrée réduite".

Son graphe est une courbe en cloche nommée **Gaussienne**, symétrique par rapport la droite $x = 0$ et dont les points d'inflexion sont $x = -1$ et $x = 1$.

Sa fonction de répartition, notée $\Pi(x)$ ou $\Phi(x)$, ne s'exprime malheureusement pas simplement. On a toutefois des tables numériques très précises qui nous permettent d'approcher ses valeurs.

Notons qu'on a $\Phi(-x) = 1 - \Phi(x)$ (pour s'en convaincre il suffit de regarder l'aire sous la Gaussienne, et on peut le prouver facilement à l'aide un changement de variable dans l'intégrale).

Loi normale $N(\mu, \sigma)$

On dit que la variable aléatoire X suit la loi normale $N(\mu, \sigma)$ si elle s'écrit sous la forme $X = \sigma X_0 + \mu$ où X_0 est une variable aléatoire normale centrée réduite .

On en déduit directement à l'aide des propriétés de l'espérance et de la variance que dans ce cas,

$$E(X) = \mu \text{ et } V(X) = \sigma^2.$$

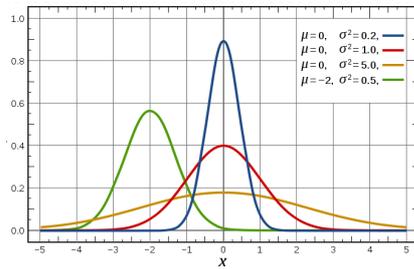
On peut facilement exprimer sa fonction de répartition F à l'aide de Φ , la fonction de répartition de X_0 :

$$F(x) = P(X \leq x) = P(\sigma X_0 + \mu \leq x) = P(X_0 \leq \frac{x - \mu}{\sigma}) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

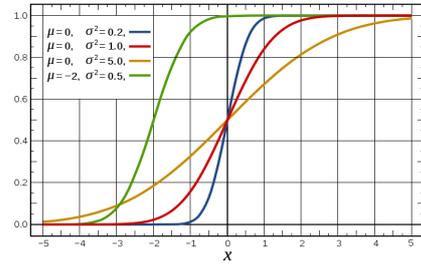
Enfin en dérivant cette dernière expression, on prouve que la densité f de X vérifie :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2}.$$

Son graphe est donc une Gaussienne symétrique par rapport la droite $x = \mu$ et dont les points d'inflexion sont $x = \mu - \sigma$ et $x = \mu + \sigma$.



densité



Fonction de répartition

Chapitre 4

Applications des probabilités aux statistiques

4.1 Échantillon et série statistique

Soit X une v.a d'espérance μ et de variance σ^2 à priori inconnu. Le but des statistiques est de retrouver des informations sur X à partir d'informations parcellaires : **l'échantillon**.

4.1.1 Définition

Soit $n \in \mathbb{N}^*$ et soient X_1, X_2, \dots , et X_n n variables aléatoires. On appelle échantillon de taille n toute suite (x_1, x_2, \dots, x_n) de réalisations des n variables aléatoires. La façon d'obtenir cet échantillon se nomme **l'échantillonnage**.

Exemple :

1. Reprenons l'exemple des fours pris précédemment. Une chaîne de fabrication produit 40 000 fours dont 20% sont défectueux, on note D un four défectueux, M un four qui fonctionne. Le service de contrôle-qualité qui ne connaît pas ces chiffres, teste aléatoirement 10 fours pour estimer la qualité de l'ensemble de la fabrication. Il obtient l'échantillon $(D, M, M, D, M, M, M, D, M, M)$.
2. Un autre exemple pourrait être de vouloir connaître la taille de 12 000 étudiants adultes. En mesurant seulement 100 étudiants, on obtient un échantillon de taille 100.

4.1.2 Échantillonnage

Soit une urne contenant des boules numérotées, on souhaite obtenir des informations sur sa contenance par échantillonnage. Pour cela, on va piocher n boules dans l'urne, X_1 sera la variable aléatoire contenant le résultat du premier tirage, X_2 celle contenant le résultat du deuxième tirage etc...

On imposera ici à l'échantillon d'être aléatoire, c'est-à-dire qu'il n'y aura pas plus de raisons de tirer une boule qu'une autre (ce n'est pas le cas des sondages par exemple, où on impose des conditions de revenus, socio-culturelle etc...). Il y a ensuite deux façons de procéder : soit on remet la boule dans l'urne entre chaque tirage, on

parle alors d'**échantillon non-exhaustif**, soit on ne la remet pas, on parle alors d'**échantillon exhaustif**.

Intuitivement, l'échantillonnage exhaustif semble meilleur car il apporte plus d'information. Ainsi si l'urne contient 100 boules, un échantillon exhaustif de taille 100 correspondra en fait au contenu total de l'urne (chaque boule est tirée une fois), tandis qu'avec un non-exhaustif, certaines boules auront été tirées plusieurs fois et d'autre aucune. Ce raisonnement n'est cependant pas pertinent car on fait des statistiques justement quand les données sont trop grandes pour être énumérées, l'échantillon est nécessairement "petit" comparé à tout les cas possibles.

L'échantillonnage non-exhaustif a lui un intérêt majeur : puisque chaque tirage est absolument identique au précédent, toutes les variables aléatoire X_i **sont indépendantes et de même loi**. L'échantillon est donc en fait une liste de n réalisations de la même variable aléatoire X , ici le tirage d'une boule, ce qui nous permettra d'utiliser de puissants théorèmes probabilistes.

Dans la pratique, avec une taille d'échantillon négligeable par rapport à la masse de données, les résultats des deux approches sont très similaires (si par exemple notre urne contient 4 milliards de boules, en enlever une ne change quasiment pas les probabilités).

Dans la suite, nous ne travaillerons plus qu'avec des échantillons aléatoires non-exhaustifs.

4.1.3 Espérance et variance empirique

Soit X une v.a d'espérance μ et de variance σ^2 , soient X_1, X_2, \dots, X_n n variables aléatoires indépendantes de même loi que X , et soit (x_1, x_2, \dots, x_n) l'échantillon de taille n associé.

On appelle **moyenne empirique** de l'échantillon $\tilde{\mu}$ la quantité :

$$\tilde{\mu} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

On appelle **variance empirique** de l'échantillon $\tilde{\sigma}^2$ la quantité :

$$\tilde{\sigma}^2 = \frac{(x_1 - \tilde{\mu})^2 + (x_2 - \tilde{\mu})^2 + \dots + (x_n - \tilde{\mu})^2}{n}.$$

En développant le numérateur, on trouve la formule ci-dessous qui, si elle est moins porteuse de sens, est beaucoup plus pratique pour les calculs. Il s'agit de "la moyenne des carrés moins le carré de la moyenne" :

$$\tilde{\sigma}^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \tilde{\mu}^2.$$

Exemple :

Soit une urne contenant 10 boules numérotées de 1 à 10 et soit X la variable aléatoire contenant le résultat d'un tirage. On considère un échantillon aléatoire non-exhaustif de taille 5 : (1, 4, 6, 9, 5). La moyenne empirique de l'échantillon est :

$$\tilde{\mu} = \frac{1 + 4 + 6 + 9 + 5}{5} = 5.$$

Sa variance empirique est

$$\tilde{\sigma}^2 = \frac{5^2 + 4^2 + 6^2 + 9^2 + 5^2}{5} - 5^2 = \frac{34}{5} = 6,8.$$

Si on calcule à présent l'espérance μ et la variance σ^2 de X on trouve :

$$\mu = 1 \times \frac{1}{10} + 2 \times \frac{1}{10} + \dots + 10 \times \frac{1}{10} = 5,5,$$

$$\sigma^2 = 1^2 \times \frac{1}{10} + 2^2 \times \frac{1}{10} + \dots + 10^2 \times \frac{1}{10} - 5,5^2 = 8,25.$$

On remarque ainsi que la moyenne empirique est proche de μ , tandis que la variance empirique est proche de σ^2 . Ce n'est (presque) pas du au hasard comme nous le verrons dans la section suivante.

4.2 Théorèmes de convergence et applications

4.2.1 La loi forte des grand nombres

Théorème 1 Soit (X_n) une suite de variables aléatoires indépendantes de même loi d'espérance μ . Alors $\frac{X_1 + X_2 + \dots + X_n}{n}$ converge **presque-sûrement** vers μ .

Dire que $\frac{X_1 + X_2 + \dots + X_n}{n}$ converge presque-sûrement vers μ , c'est dire "qu'à l'infini il n'y a plus de hasard" : quand n vers l'infini, $\frac{X_1 + X_2 + \dots + X_n}{n}$ peut prendre la valeur μ avec une probabilité 1, et toute autre valeur avec la probabilité 0.

Application à l'échantillonnage : approche par valeur moyenne

Soit X une v.a d'espérance μ et de variance σ^2 toutes deux inconnues, et soit (x_1, x_2, \dots, x_n) un échantillon aléatoire non-exhaustif de taille n .

Alors la moyenne empirique $\tilde{\mu} = \frac{x_1 + x_2 + \dots + x_n}{n}$ est une réalisation de la variable aléatoire $\frac{X_1 + X_2 + \dots + X_n}{n}$ où X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes de même loi que X . D'après la loi forte des grands nombres, $\tilde{\mu}$ tend vers μ quand n tend vers l'infini. Ainsi quand la taille de l'échantillon est "très grand", la moyenne empirique est "très proche" de la moyenne théorique, on peut donc se servir de la moyenne empirique pour approcher la moyenne. Un raisonnement identique sur la variance permet d'énoncer le résultat suivant :

Quand la taille n de l'échantillon est grand, $\tilde{\mu}$ est une valeur approchée de μ et $\tilde{\sigma}^2$ est une valeur approchée de σ^2 .

En choisissant les "bonnes" variables aléatoires, on peut ainsi approcher la probabilité de n'importe quel évènement.

Exemple :

Soit X la variable aléatoire mesurant le temps d'attente avant d'être mis en relation avec un service après vente. On souhaite évaluer le temps d'attente moyen ainsi

que la probabilité d'attendre cinq minutes ou plus. Pour cela on mesure aléatoirement 20 temps d'attente, on a donc un échantillon aléatoire non-exhaustif (x_1, \dots, x_{20}) de taille 20 de la durée d'attente X :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	6	2	18	1	8	9	2	14	13	6	0	15	4	14	2	5	5	28	5	0

Si on calcule la moyenne empirique de l'échantillon, on aura d'après la loi forte des grands nombres une valeur approchée de la moyenne de X , c'est-à-dire une approximation du temps moyen d'attente. Ainsi,

$$E(X) \simeq \frac{6 + 2 + \dots + 5 + 0}{20} = 7,85.$$

Pour répondre à la deuxième question, on introduit une nouvelle variable aléatoire Y qui vaut 1 si le temps d'attente est de 5 minutes ou plus, 0 sinon. Y suit donc une loi de Bernoulli $B(p)$ où $p = P(X \geq 5)$ est la probabilité d'attendre 5 minutes ou plus, son espérance vaut donc p .

L'échantillon précédent nous permet d'obtenir un échantillon de même taille pour Y :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	6	2	18	1	8	9	2	14	13	6	0	15	4	14	2	5	5	28	5	0
y_i	1	0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0

On en déduit d'après la loi forte des grands nombres une approximation de la probabilité p d'attendre 5 minutes ou plus :

$$p \simeq \frac{1 + 0 + \dots + 1 + 0}{20} = \frac{16}{20} = 0,8.$$

4.2.2 Le théorème central limite

L'approche précédente souffre d'un grave défaut : rien ne permet de savoir à partir de quelle taille d'échantillon on a une bonne approximation, ni quelle erreur d'approximation est commise. Le théorème suivant permet de résoudre en grande partie ce problème.

Théorème 2 Soit (X_n) une suite de variables aléatoires indépendantes de même loi d'espérance μ et de variance σ^2 . Alors $Z = \frac{(X_1 - \mu) + (X_2 - \mu) + \dots + (X_n - \mu)}{\sigma\sqrt{n}}$ converge en loi vers la loi normale centrée réduite $N(0, 1)$.

Dire que Z converge en loi vers $N(0, 1)$, c'est dire que la loi de probabilité de Z devient de plus en plus proche de $N(0, 1)$ quand n tend vers l'infini. On pourra ainsi faire des calculs de probabilité sur Z en remplaçant sa loi par $N(0, 1)$. Dans la pratique, on considère que quand on a une taille d'échantillon $n \geq 30$, l'erreur commise en remplaçant la loi de Z par $N(0, 1)$ est négligeable.

Application à l'échantillonnage : approche par intervalle de confiance

Soit X une v.a d'espérance μ inconnue et de variance σ^2 connue ou non, et soit (x_1, x_2, \dots, x_n) un échantillon aléatoire non-exhaustif de taille n correspondant des variables aléatoires indépendantes X_1, X_2, \dots, X_n de même loi que X . On va chercher **des intervalle de confiance** pour μ , c'est-à-dire des intervalles ayant un certain pourcentage de chance de contenir la valeur de μ .

On introduit deux variables aléatoires Y et Z de la façon suivante :

$$Y = \frac{X_1 + X_2 + \dots + X_n}{n},$$

$$Z = \frac{(X_1 - \mu) + (X_2 - \mu) + \dots + (X_n - \mu)}{\sigma\sqrt{n}}.$$

La moyenne empirique $\tilde{\mu}$ est donc une réalisation de Y et Z suit la loi $N(0, 1)$ d'après le T.C.L. De plus ces deux nouvelles variables vérifient la relation

$$\frac{\sigma Z}{\sqrt{n}} = Y - \mu.$$

Soit α un petit nombre positif (par exemple 0,01), on peut lire dans la table de la loi normale le nombre u_α (avec $\alpha = 0,01$, on a $u_\alpha = 2,5758$) tel que :

$$P(|Z| > u_\alpha) = \alpha.$$

On en déduit donc que $P\left(\left|\frac{\sigma Z}{\sqrt{n}}\right| > \frac{\sigma}{\sqrt{n}}u_\alpha\right) = \alpha$. Puisque $\frac{\sigma Z}{\sqrt{n}} = Y - \mu$ et que $\tilde{\mu}$ est une réalisation de Y , on finalement la relation :

$$P\left(|\tilde{\mu} - \mu| > \frac{\sigma}{\sqrt{n}}u_\alpha\right) = \alpha.$$

Ceci se réécrit :

$$P\left(|\tilde{\mu} - \mu| \leq \frac{\sigma}{\sqrt{n}}u_\alpha\right) = 1 - \alpha.$$

Ainsi le segment

$$I_\alpha = \left[\tilde{\mu} - \frac{\sigma}{\sqrt{n}}u_\alpha, \tilde{\mu} + \frac{\sigma}{\sqrt{n}}u_\alpha\right]$$

contient μ avec une probabilité de $1 - \alpha$. En reprenant $\alpha = 0,01$, μ a donc 99% de chance d'être dans le segment

$$I_{0,01} = \left[\tilde{\mu} - \frac{2,5758\sigma}{\sqrt{n}}, \tilde{\mu} + \frac{2,5758\sigma}{\sqrt{n}}\right].$$

$I_{0,01}$ est un intervalle de confiance pour μ à 99%. Reste à remplacer σ par sa valeur (ou par $\tilde{\sigma}$ si on ne la connaît pas) pour déterminer exactement l'intervalle de confiance. La

différence majeur avec l'approche par valeur moyenne est que l'on maîtrise à présent l'erreur commise en approchant μ par $\tilde{\mu}$: on fait une erreur d'approximation $\leq \frac{\sigma}{\sqrt{n}}u_\alpha$ avec une probabilité de $1 - \alpha$.

Exemple :

En reprenant l'exemple précédent, donner un intervalle de confiance à 99% du temps moyen d'attente ainsi que de la probabilité d'attendre plus de 5 minutes. Même question avec un intervalle de confiance à 95%.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	6	2	18	1	8	9	2	14	13	6	0	15	4	14	2	5	5	28	5	0
y_i	1	0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0

Temps d'attente moyen.

Donner un intervalle de confiance du temps d'attente moyen, c'est donc donner un intervalle de confiance de la moyenne μ de la variable aléatoire X . Ils sont de la forme (on remplace σ qui est inconnu par $\tilde{\sigma}$) :

$$\left[\tilde{\mu} - \frac{\tilde{\sigma}}{\sqrt{n}} u_\alpha, \tilde{\mu} + \frac{\tilde{\sigma}}{\sqrt{n}} u_\alpha \right].$$

Ici $n = 20$ et il faut donc commencer par calculer $\tilde{\mu}$ et $\tilde{\sigma}$:

$$\tilde{\mu} = \frac{6 + 2 + \dots + 5 + 0}{20} = 7,85.$$

$$\tilde{\sigma} = \sqrt{\frac{(6 - 7,85)^2 + (2 - 7,85)^2 + \dots + (5 - 7,85)^2 + (0 - 7,85)^2}{20}} = 7,01.$$

1. Pour un intervalle de confiance à 99%, on a $\alpha = 1\% = 0,01$ et donc $u_\alpha = 2,5758$. On trouve donc un intervalle de confiance :

$$\left[7,85 - \frac{7,01}{\sqrt{20}} 2,5758; 7,85 + \frac{7,01}{\sqrt{20}} 2,5758 \right] = [3,81; 11,89].$$

Il y a donc 99% de chance que le temps d'attente moyen soit dans l'intervalle $[3,81; 11,89]$.

2. Pour un intervalle de confiance à 95%, on a $\alpha = 5\% = 0,05$ et donc $u_\alpha = 1,96$. On trouve donc un intervalle de confiance :

$$\left[7,85 - \frac{7,01}{\sqrt{20}} 1,96; 7,85 + \frac{7,01}{\sqrt{20}} 1,96 \right] = [4,78; 10,92].$$

Il y a donc 95% de chance que le temps d'attente moyen soit dans l'intervalle $[4,78; 10,92]$.

Probabilité d'attendre plus de cinq minutes.

Donner un intervalle de confiance de la probabilité d'attendre plus de 5 minutes, c'est donc donner un intervalle de confiance de la moyenne μ de la variable aléatoire Y . Ils sont de la forme (on remplace σ qui est inconnu par $\tilde{\sigma}$) :

$$\left[\tilde{\mu} - \frac{\tilde{\sigma}}{\sqrt{n}} u_\alpha, \tilde{\mu} + \frac{\tilde{\sigma}}{\sqrt{n}} u_\alpha \right].$$

Ici $n = 20$ et il faut donc commencer par calculer $\tilde{\mu}$ et $\tilde{\sigma}$:

$$\frac{1 + 0 + \dots + 1 + 0}{20} = \frac{16}{20} = 0,8.$$

$$\tilde{\sigma} = \sqrt{\frac{(1 - 0,8)^2 + (0 - 0,8)^2 + \dots + (1 - 0,8)^2 + (0 - 0,8)^2}{20}} = 0,4.$$

On trouve alors :

1. Pour un intervalle de confiance à 99%, on a $\alpha = 1\% = 0,01$ et donc $u_\alpha = 2,5758$.
On trouve donc un intervalle de confiance :

$$\left[0,8 - \frac{0,4}{\sqrt{20}} 2,5758; 0,8 + \frac{0,4}{\sqrt{20}} 2,5758 \right] = [0,57; 1,03].$$

Il y a donc 99% de chance que la probabilité d'attendre plus de 5 minutes soit dans l'intervalle $[0,57; 1,03]$.

2. Pour un intervalle de confiance à 95%, on a $\alpha = 5\% = 0,05$ et donc $u_\alpha = 1,96$.
On trouve donc un intervalle de confiance :

$$\left[0,8 - \frac{0,4}{\sqrt{20}} 1,96; 0,8 + \frac{0,4}{\sqrt{20}} 1,96 \right] = [0,62; 0,98].$$

Il y a donc 95% de chance que la probabilité d'attendre plus de 5 minutes soit dans l'intervalle $[0,62; 0,98]$.

4.3 Tests d'hypothèses

4.3.1 Décisions statistiques

Dans la pratique, nous sommes très souvent amenés à prendre des décisions concernant des variables aléatoires sur la base d'informations recueillies sur des échantillons. De telles décisions sont appelées **décisions statistiques**.

Exemple

- Décider, sur la base de données d'échantillonnage, qu'un sérum est efficace pour combattre une maladie donnée.
- Décider qu'une pièce de monnaie est truquée.
- Décider qu'une méthode pédagogique est meilleure qu'une autre.

4.3.2 Hypothèses statistiques

Pour prendre des décisions, des hypothèses sur les variables aléatoires concernées doivent être faites. Ces hypothèses sont dites **hypothèses statistiques**, elles peuvent être vraies ou non et sont en général des énoncés relatifs aux distributions de ces variables.

Dans de nombreux cas, l'hypothèse statistique n'est formulée que dans le but d'être annulée ou rejetée. Ce genre d'hypothèse est désignée sous le nom d'**hypothèse nulle** et est notée H_0 . L'hypothèse complémentaire de l'hypothèse nulle H_0 est appelée **hypothèse alternative** et est notée H_1 .

Exemple

- Si nous souhaitons décider qu’une pièce de monnaie est truquée, nous formulons l’hypothèse nulle H_0 : “elle est équilibrée”, c’est-à-dire que $p = 0.5$ où p est la probabilité de faire face. H_1 est donc $p \neq 0,5$.
- Si nous voulons montrer qu’un sérum est efficace, nous formulons l’hypothèse nulle $H_0 : m_s \geq m$, où m_s est le temps de guérison moyen avec le sérum, m le temps moyen de guérison sans le sérum. H_1 est donc $m_s < m$.

4.3.3 Tests d’hypothèses**Définition**

Si, sur la base de la supposition qu’une hypothèse donnée est vraie, nous trouvons que les résultats observés sur un échantillon aléatoire diffèrent de manière notable de ceux que l’on pourrait espérer théoriquement, on est amené à rejeter l’hypothèse proposée. Par exemple si au cours de 20 lancers d’une pièce on obtient 16 fois “face”, on aura tendance à rejeter l’hypothèse “la pièce est équilibrée”, il est pourtant possible qu’elle le soit.

On appelle **test d’hypothèse de H_0 relativement à H_1** toute expérience qui, s’il elle est réussie, conduit à accepter l’hypothèse H_0 , si elle est ratée à rejeter H_0 , c’est-à-dire accepter H_1 .

Erreurs de première et de deuxième espèce

Lors d’un test d’hypothèse de H_0 relativement à H_1 , on dit que l’on commet une **erreur de première espèce** si l’on rejette à tort H_0 ; on dit que l’on commet une **erreur de deuxième espèce** si l’on conserve à tort H_0 . Il y a eu dans les deux cas une décision erronée, c’est-à-dire que le test nous a induit en erreur, mais l’interprétation de ces deux types d’erreurs est très différente.

Si par exemple on essaie d’évaluer l’efficacité du système judiciaire de la façon suivante : l’hypothèse H_0 est “le prévenu est coupable”, l’hypothèse H_1 est alors “le prévenu est non-coupable”. On prend pour test de H_0 relativement à H_1 “le prévenu a été condamné” (s’il n’a pas été condamné alors il est innocent, s’il a été condamné il est coupable).

Dans ce cadre, une erreur de première espèce serait de rejeter à tort H_0 : cela correspond à la libération d’un coupable; une erreur de seconde espèce serait de conserver à tort H_0 : cela correspond à condamner un innocent. On devine intuitivement qu’il est donc difficile de concilier les deux types de risque : on peut imposer à un test de limiter les risques d’un des types d’erreurs, mais pas des deux.

On cherchera à construire des test limitant le risque de première espèce, c’est-à-dire limitant le risque de rejeter à tort H_0 (pour limiter les risques de seconde espèce il suffit d’inverser les deux hypothèses). Cela nous permettra d’écarter H_0 avec une faible chance de se tromper lorsque le test échoue.

Niveau de signification

On appelle **niveau de signification** du test d’hypothèse de H_0 relativement à H_1 , la probabilité maximale de faire une erreur de première espèce. En pratique, un

niveau de signification de 0.05 ou 0.01 est chose courante, bien que l'on utilise aussi d'autres valeurs.

4.3.4 Réalisation de tests à l'aide du T.C.L

Nous allons présenter ici deux types de tests concernant des hypothèses sur la moyenne d'une variable aléatoire. Pour utiliser ces tests pour évaluer une probabilité, il faudra procéder comme dans le cas des intervalles de confiance : introduire une nouvelle variable aléatoire dont l'espérance est la probabilité recherchée.

Soit X une v.a d'espérance μ inconnue et de variance σ^2 connue ou non, et soit (x_1, x_2, \dots, x_n) un échantillon aléatoire non-exhaustif de taille n correspondant des variables aléatoires indépendantes X_1, X_2, \dots, X_n de même loi que X .

On réintroduit les mêmes variables Y et Z . On rappelle que la variable aléatoire Z suit la loi normale $N(0, 1)$, que ces deux variables vérifient la relation $\frac{\sigma Z}{\sqrt{n}} = Y - \mu$, et que puisque $\tilde{\mu}$ est une réalisation de Y , $\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - \mu)$ est une réalisation de Z .

Test de $H_0 : \mu = m$ relativement à $H_1 : \mu \neq m$ avec un niveau de signification α

La variable Z suit la loi $N(0, 1)$, on lit donc dans la table de la loi normale le nombre u_α tel que $P(|Z| > u_\alpha) = \alpha$. Ainsi si $\mu = m$, $|\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - m)| > u_\alpha$ ne se réalise qu'avec une probabilité α . On a donc le test suivant :

Si $|\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - m)| \leq u_\alpha$ on conserve H_0 , sinon on rejette H_0 .

Ce test a bien un niveau de signification α , et σ est inconnu, on le remplace par $\tilde{\sigma}$. On parle dans ce cas de **test bilatéral**.

Test de $H_0 : \mu \leq m$ relativement à $H_1 : \mu > m$ avec un niveau de signification α

La variable Z suit la loi $N(0, 1)$, on lit donc dans la deuxième table de la loi normale le nombre U_α tel que $P(Z \leq U_\alpha) = 1 - \alpha$. Ainsi si $\mu \leq m$,

$$\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - \mu) \geq \frac{\sqrt{n}}{\sigma}(\tilde{\mu} - m).$$

Donc si $\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - m)$ est strictement supérieur à U_α , c'est aussi le cas de $\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - \mu)$.

Or $\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - m)$ n'est strictement supérieur à U_α qu'avec une probabilité α . On a donc le test suivant :

Si $\frac{\sqrt{n}}{\sigma}(\tilde{\mu} - m) \leq U_\alpha$ on conserve H_0 , sinon on rejette H_0 .

Ce test a bien un niveau de signification α , et σ est inconnu, on le remplace par $\tilde{\sigma}$. On parle dans ce cas de **test unilatéral**.

4.4 Statistique bidimensionnelle

On s'intéresse désormais non plus à obtenir des informations sur une variable aléatoire X , mais sur le lien potentiel entre deux variables aléatoires X et Y . Par exemple y-a-t'il un lien entre taille et poids dans une population donnée? Entre niveau de richesse et longévité etc... Pour répondre à cela, on relève simultanément un échantillon (x_1, x_2, \dots, x_n) et (y_1, y_2, \dots, y_n) des deux variables aléatoires, et on va construire un test statistique permettant d'y répondre.

Afin de ne pas confondre les moyennes et les écarts-types empiriques des deux échantillons, on notera μ_X et σ_X ceux de l'échantillon (x_1, x_2, \dots, x_n) , μ_Y et σ_Y ceux de l'échantillon (y_1, y_2, \dots, y_n) . On abandonne de plus la notation $\tilde{\cdot}$ car on ne travaille plus que sur les échantillons, on n'a donc plus besoin de distinguer les valeurs empiriques des théoriques.

4.4.1 coefficient de corrélation

On appelle **covariance empirique** des échantillons la quantité σ_{XY} définie par la formule :

$$\sigma_{XY} = \frac{\sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)}{n}.$$

En développant, on arrive à une nouvelle formule la encore beaucoup plus pratique pour les calculs. Il s'agit de "la moyenne des produits moins le produits des moyennes" :

$$\sigma_{XY} = \frac{\sum_{i=1}^n x_i y_i}{n} - \mu_X \mu_Y.$$

On appelle **coefficient de corrélation empirique** des échantillons la quantité ρ_{XY} définie par la formule :

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

On termine avec un théorème qui nous permettra de construire un test statistique : **l'inégalité de Cauchy-Schwarz**.

Théorème 3 Inégalité de Cauchy-Schwarz

Pour toutes suites finies de nombres réels (x_1, x_2, \dots, x_n) et (y_1, y_2, \dots, y_n) , on a :

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\sum_{i=1}^n x_i^2\right) \left(\sum_{i=1}^n y_i^2\right)$$

De plus cette inégalité est une égalité si et seulement si les deux suites sont colinéaires, c'est à dire si et seulement si il existe $a \in \mathbb{R}$ tel que $y_i = ax_i$ pour toutes les valeurs de i .

Démonstration

Considérons deux cas :

- $x_i = 0, \forall i \in \mathbb{R}$ l'inégalité devient $0 \leq 0$ et est donc vérifiée.
- Les x_i ne sont pas tous nuls. Soit $\lambda \in \mathbb{R}$, posons :

$$f(\lambda) = \sum_{i=1}^n (\lambda x_i + y_i)^2$$

f est une fonction réelle d'une variable réelle définie sur \mathbb{R} et à valeurs dans \mathbb{R}^+ , développons :

$$f(\lambda) = \sum_{i=1}^n (\lambda^2 x_i^2 + 2\lambda x_i y_i + y_i^2) = \left(\sum_{i=1}^n x_i^2 \lambda^2 \right) + 2 \left(\sum_{i=1}^n x_i y_i \right) \lambda + \sum_{i=1}^n y_i^2$$

f est donc un polynôme du second degré défini sur \mathbb{R} et de signe ≥ 0 , son discriminant est donc ≤ 0 soit :

$$\Delta' = \left(\sum_{i=1}^n x_i y_i \right)^2 - \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right) \text{ et } \Delta' \leq 0 \text{ d'où}$$

$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right)$$

- Si de plus $\Delta' = 0$ alors $\exists a$, un réel tel que $f(a) = 0$ soit $\sum_{i=1}^n (ax_i + y_i)^2 = 0$, on obtient alors $ax_i + y_i = 0, \forall i = 1, \dots, n$. (x_1, x_2, \dots, x_n) et (y_1, y_2, \dots, y_n) sont donc colinéaires. \square

4.4.2 Test de linéarité et droite d'ajustement

Grâce à tous les outils introduits précédemment, nous sommes en mesure d'établir un test de l'hypothèse "X et Y sont liées linéairement" ce qui est donc équivalent à "Y=aX+b".

Dire que les deux variables aléatoires sont liées linéairement, c'est dire que la représentation graphique de Y en fonction de X est une droite. Or nous avons à notre disposition un échantillon double de X et Y, c'est à dire des points de la représentation graphique. On en déduit donc un test assez simple :

- Si les points ne sont pas alignés, il ne peuvent pas par définition se trouver sur une même droite. On peut donc rejeter l'hypothèse "X et Y sont liées linéairement".
- Si les points sont alignés, il est possible que la courbe dont ils sont extraits soit une droite.

Il reste à trouver un outil performant pour tester l'alignement de nos points : le coefficient de corrélation. En effet, en appliquant l'inégalité de Cauchy-Schwarz aux deux suites $(x_1 - \mu_X, x_2 - \mu_X, \dots, x_n - \mu_X)$ et $(y_1 - \mu_Y, y_2 - \mu_Y, \dots, y_n - \mu_Y)$ on obtient les résultats suivants :

- $-1 \leq \rho_{XY} \leq 1$,
- $|\rho_{XY}| = 1$ si et seulement si tous les points (x_i, y_i) sont parfaitement alignés.

On a donc un test naturel assez simple : si le coefficient de corrélation est proche de 1 en valeur absolue on pourra espérer une relation linéaire entre X et Y ; on pourra rejeter cette hypothèse dans le cas contraire.

Lorsque l'on soupçonne une relation linéaire entre X et Y , on cherche naturellement à savoir laquelle. On cherche donc la droite $Y = aX + b$ "la plus proche possible" de notre nuage de points par la méthode des moindres carrés, c'est à dire celle qui minimise la quantité suivante :

$$D(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

Théorème 4 *Il existe une et une seule droite qui minimise l'expression $D(a, b)$, cette droite d'équation $y = ax + b$ passe par le point moyen (\bar{X}, \bar{Y}) et a pour pente*

$$a = \frac{\sigma_{XY}}{\sigma_X^2} = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$$

Il s'agit de la droite de régression de Y par rapport à X .