Cours de Probabilité et Statistique élémentaires à l'usage des étudiants en biologie.

Hélène Guérin
IRMAR, Université Rennes 1
Campus de Beaulieu, 35042 Rennes Cedex FRANCE.
helene.guerin@univ-rennes1.fr

Septembre 2008

Table des matières

Préambule

Le but de ce cours est de donner les outils nécessaires à la modélisation stochastique. Qu'est-ce que la modélisation stochastique ? Tout d'abord un modèle est une représentation simplifiée de la réalité. Il peut être exprimé sous forme d'équations, de schémas, ou être un ajustement de données. Un modèle a en général pour but de mieux comprendre notre environnement, voir de prédire l'avenir. Il a aussi pour but de communiquer, notamment entre scientifiques. Il existe deux principales formes de modèles (en sachant qu'on peut les combiner) : les modèles déterministes et les modèles stochastiques. Dans le modèle déterministe, partant d'un état initial on a qu'un seul état final, il n'y a pas d'incertain. Dans un modèle stochastique on fait intervenir une part d'aléa. Dans la réalité il y a toujours de l'aléa, on ne peut pas tout contrôler (panne de réveil, retard du bus, rencontrer un ami...) et heureusement car sinon la vie serait triste ! Du coup dans la réalité partant d'un état initial il y a plusieurs états finaux possibles. On parle de stochasticité pour tout ce qui échappe au déterministe. Il y a différentes sources de stochasticité :
- des erreurs de mesures,
- des phénomènes inconnus non modélisés (état initial non exact, fluctuation de l'environnement, variation individuelle comme par exemple génétique)
- de l'aléa intrinsèque (systèmes chaotiques).

Comment identifier la stochasticité ? On peut faire un relevé de données et chercher à partir de celui-ci une loi de probabilité qui régit au mieux le phénomène.
Avant toute modélisation stochastique, on a donc besoin d'avoir un minimum de connaissance en probabilité afin d'ajuster au mieux notre modèle à la réalité.

Première partie

Introduction aux Probabilités

Chapitre 1

Qu'est-ce qu'une probabilité ?

1 Expérience aléatoire

1.1 Ensemble fondamental d'une expérience aléatoire

Une expérience aléatoire est une expérience dont on ne peut pas prédire le résultat. L'**ensemble fondamental d'une expérience aléatoire** est l'ensemble de tous les résultats possibles de l'expérience. Cet ensemble est en général noté Ω. Il peut être fini, dénombrable, ou infini non dénombrable. Un point $\omega \in \Omega$ est une réalisation de l'expérience.

Exemple 1.1 1. On lance un dé, on s'intéresse au chiffre obtenu. L'ensemble fondamentale est alors fini, plus précisément

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

2. On jette une pièce autant de fois que nécessaire pour obtenir une fois "face". L'ensemble fondamentale est alors infini dénombrable. Si on note F pour "face" et P pour "pile", on a

$$\Omega = \{F, PF, PPF, PPPF, \ldots, PPP \cdots PPF, \ldots\}.$$

3. On s'intéresse à la durée de vie d'une bactérie. L'ensemble fondamental est alors infini, non dénombrable : $\Omega = [0, +\infty[$.

Remarque 1.2 Une expérience doit être décrite de façon précise afin son ensemble fondamentale soit défini de façon claire. Lorsqu'on lance un dé il peut se passer beaucoup de choses, il peut attérir sur une face, sur la tranche, dans un trou...

1.2 Évènement d'une expérience aléatoire

Définition 1.3 Un **événement d'une expérience aléatoire** est une combinaison de résultats possibles.

Exemple 1.4 On lance deux dés et on regarde les chiffres obtenus. L'ensemble fondamental est

$$\Omega = \{(1, 1), (1, 2), (1, 3) \ldots, (2, 1), (2, 2), \ldots, (6, 5), (6, 6)\}.$$

On peut considérer les événements suivants :
- la somme des points est 6 : $A = \{(1,5),(5,1),(2,4),(4,2),(3,3)\}$,
- la somme des points est un multiple de 3 :
$B = \{(1,2),(2,1),(1,5),(5,1),(2,4),(4,2),(3,3),(3,6),(6,3),(4,5),(5,4),(6,6)\}$.

Toute expérience aléatoire comprend un **événement certain** et un **événement impossible**. L'événement impossible, noté $\emptyset$, est le sous-ensemble qui ne contient pas d'éléments. L'événement certain est le sous-ensemble qui contient tous les éléments, c'est l'ensemble fondamental Ω lui-même.

Exemple 1.5 On lance un dé.
Si A_1 est l'événement "le nombre de points est supérieur à 7", alors $A_1 = \emptyset$.
Si A_2 est l'événement "le nombre de points est inférieur à 7", alors $A_2 = \{1,2,3,4,5,6\} = \Omega$.

On appelle **événement simple** un événement qui ne contient qu'un seul élément.

1.3 Opérations sur les événements

Soit A et B deux évènements.
- **Négation** : $\overline{A}$ est l'évènement contraire de A, ou complémentaire de A par rapport à Ω : $\overline{A} = \Omega \backslash A$. $\overline{A}$ est l'événement qui se réalise lorsque A ne se réalise pas.
- **Intersection** : L'événement $C = A \cap B$ combiné à partir des événements A et B est réalisé lorsque A et B sont réalisés simultanément.
- **Union** : L'événement $D = A \cup B$ combiné à partir des événements A et B est réalisé lorsque l'un des événements A ou B est réalisé.

Deux événements A et B sont dits incompatibles ou disjoints si $A \cap B = \emptyset$.
La relation "l'événement A implique l'événement B" signifie que si A se réalise alors B se réalise aussi. On a $A \subset B$.

Exemple 1.6 On jette deux dés. On considère les évènements suivants : $A =$"la somme des points vaut 6", $B =$"la somme des points est un multiple de 3", $C =$"avoir au moins un 6". On a alors

$$\begin{aligned} A &= \{(1,5),(5,1),(2,4),(4,2),(3,3)\} \\ B &= \{(1,2),(2,1),(1,5),(5,1),(2,4),(4,2),(3,3),(3,6),(6,3),(4,5),(5,4),(6,6)\} \\ C &= \{(1,6),(6,1),(2,6),(6,2),(3,6),(6,3),(4,6),(6,4),(5,6),(6,5),(6,6)\} \end{aligned}$$

On a $A \subset B$, $A \cap C = \emptyset$ et $\overline{C} =$"n'avoir aucun 6".

<u>Extension des définitions</u> :
On considère une suite infinie d'évènements $A_1, A_2, \ldots$.
- L'union $\cup_{i=1}^{\infty} A_i$ est l'événement qui regroupe tous les points des événements $A_1, A_2, \ldots$: on a "$a \in \cup_{i=1}^{\infty} A_i$" ssi "il existe un entier $i \geq 1$ tel que $a \in A_i$".
- L'intersection $\cap_{i=1}^{\infty} A_i$ est l'événement composé des points communs à tous les événements $A_1, A_2, \ldots$: on a "$a \in \cap_{i=1}^{\infty} A_i$" ssi "pour tous les entiers $i \geq 1$ on a $a \in A_i$".

2 Notion de mesure de probabilité

Considérons une expérience dont l'ensemble fondamental est Ω.

Définition 2.1 Une **probabilité** (ou **mesure de probabilité**) $\mathbb{P}$ sur Ω est une application sur l'ensemble des événements telle que

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

(ii) pour tout événement A, $0 \leq \mathbb{P}(A) \leq 1$,

(iii) pour toute suite $A_1, A_2, \ldots$ d'événements disjoints (i.e., pour tout $i \neq j$, $A_i \cap A_j = \emptyset$), on a

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

Une probabilité est une mesure dans le sens où plus un événement est grand plus sa probabilité est importante.

Propriétés 2.2 On a alors

1. $\mathbb{P}(\emptyset) = 0$.
2. Si $A_1, A_2, \ldots, A_n$ est une suite finie d'événements disjoints, on a

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

3. Pour tout événement A, $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
4. Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$ et $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$.

Preuve.

1. On pose $A_1 = \Omega$ et $A_i = \emptyset$ pour $i \geq 2$, alors d'après 2. $P(\Omega) = P(\Omega) + \sum_{i=2}^{\infty} \mathbb{P}(\emptyset)$ et donc $\mathbb{P}(\emptyset) = 0$.
2. On pose $A_i = \emptyset$ pour $i \geq n+1$, alors d'après 2.

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{i=n+1}^{\infty} \mathbb{P}(\emptyset) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

   On conclut en remarquant que $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$.
3. Les événements $A_1 = A$ et $A_2 = \overline{A}$ sont disjoints et $A_1 \cup A_2 = \Omega$, donc $\mathbb{P}(\Omega) = 1 = \mathbb{P}(A) + \mathbb{P}(\overline{A})$.
4. Si $A \subset B$, les événements $A_1 = A$ et $A_2 = B \backslash A$ sont disjoints et $A_1 \cup A_2 = B$, alors $\mathbb{P}(B) = \mathbb{P}(B \backslash A) + \mathbb{P}(A)$.

$\triangle$

Proposition 2.3 Soit A et B deux événements quelconques, alors

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

De plus si $(A_n)_{n \geq 0}$ est une suite d'événements quelconque, on a

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

Preuve. Il suffit d'écrire $A \cup B$ comme l'union disjointe $(A\backslash(A \cap B)) \cup (B\backslash(A \cap B)) \cup (A \cap B)$ pour la première relation et de généraliser cela pour la seconde en posant $B_1 = A_1$, $B_2 = A_2\backslash(A_1 \cap A_2)$, ..., $B_i = A_i\backslash(A_1 \cap \ldots \cap A_i)$. Les événements B_i sont disjoints avec $B_i \subset A_i$ et $\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i$. △

Cas particulier d'un ensemble fondamental discret
Considérons une expérience aléatoire dont l'ensemble fondamental Ω est fini ou dénombrable : $\Omega = \{\omega_1, \ldots, \omega_n, \ldots\}$ où les points ω_i sont distincts.
Une probabilité P sur cet espace est entièrement déterminée par les probabilités $\mathbb{P}(\{\omega_i\})$, pour $i \geq 1$. En effet, si A est un événement on a

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

On a forcément que la somme totale des probabilités est égale à 1 :

$$\mathbb{P}(\Omega) = \sum_{i=1}^{+\infty} \mathbb{P}(\{\omega_i\}) = 1$$

Exemple important : l'équiprobabilité
Si $\Omega = \{\omega_1, \ldots, \omega_n\}$ est un espace fini où les points ω_i sont distincts, on définit la **probabilité uniforme** comme la probabilité qui associe à chaque singleton $\{\omega_i\}$ la même valeur, c'est à dire $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\{\omega_j\})$ pour tout i, j. Comme $\mathbb{P}(\Omega) = 1 = \sum_{i=1}^{n} \mathbb{P}(\{\omega_i\})$, on a forcément $\mathbb{P}(\{\omega_i\}) = \frac{1}{n} = \frac{1}{card(\Omega)}$. On dit alors qu'on est dans une situation d'**équiprobabilité**.

Exemple 2.4 On jette un dé : $\Omega = \{1, 2, \ldots, 6\}$. On définit la probabilité de tomber sur une face du dé. On suppose que le dé est bien équilibré (non truqué), c'est à dire que chaque face à la même chance d'apparaître. On est en situation d'équiprobabilité et donc $\mathbb{P}(\{i\}) = 1/6$ pour $i \in \{1, 2, \ldots, 6\}$.
Soit A l'événement "obtenir un chiffre pair", on a alors

$$\mathbb{P}(A) = \mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/2.$$

Soit B l'événement "obtenir au moins deux". On remarque que $\overline{B}$ ="tomber sur 1", d'où

$$\mathbb{P}(B) = 1 - \mathbb{P}(\overline{B}) = 1 - 1/6 = 5/6$$

3 Probabilités conditionnelles

En 2005, en France, le nombre de porteurs du virus de l'hépatite C (VHC_+) est d'environ 300 000 personnes [1]. Un médecin avant de recevoir un patient inconnu en consultation peut donc estimer que ce patient a une probabilité de 0.5% d'être VHC_+ en supposant que sa clientèle ressemble globalement à l'ensemble de la population française (qui comporte environ 60 millions de personnes). Si en consultant le dossier du patient, il s'aperçoit que son patient a 10 ans, alors la probabilité que ce patient soit VHC_+ chute pour devenir inférieur à 10^{-4} (vu le mode de transmission du virus). Par contre si en consultant le dossier, il s'aperçoit que le patient est toxicomane par voie intraveineuse depuis 10 ans, les données épidémiologiques

1. Voir travaux de *C. Meffre and al.* de l'Institut de veille sanitaire présentés en 2005.

montrent que la probabilité que ce patient soit VHC_+ est alors supérieure à 50%. Si dans le dossier, il est seulement indiqué que le patient est asthmatique, alors le médecin n'aura aucune information supplémentaire vu qu'il n'y a aucun lien entre l'asthme et le fait d'être porteur de VHC. Le médecin estimera donc que son patient a toujours une probabilité de 0.5% d'être VHC_+. Le fait d'avoir l'asthme influe pas sur la probabilité d'être VHC_+, on dit que ces événements sont indépendants.
On voit bien dans cet exemple qu'une information complémentaire peut influer ou non la probabilité d'être VHC_+. La notion de probabilité conditionnelle permet de prendre en compte des information complémentaires dans le calcul des probabilités.

Dans l'exemple suivant, on quantifie exactement de le rôle d'une information complémentaire.

Exemple 3.1 On jette deux dés distincts. On a $\Omega = \{(1,1),(1,2),\ldots,(5,6),(6,6)\}$. On a donc 36 événements élémentaires qui ont chacun la même probabilité d'apparaître, soit 1/36.
On veut connaitre la probabilité de l'événement A : "la somme des dés vaut 8". On remarque que $A = \{(2,6),(6,2),(3,5),(5,3),(4,4)\}$, d'où $\mathbb{P}(A) = 5/36$.
Maintenant on cherche toujours la même probabilité, mais on sait déjà que le premier dé donne un 3. On note B l'événement : "le premier dé vaut 3".
Que vaut alors la probabilité que la somme vaut 8 sachant que le premier dé vaut 3 ? La probabilité recherchée est appelée probabilité conditionnelle de A sachant B et est notée $P_B(A)$.
Sachant que le premier dé vaut 3, on ne regarde plus la même expérience, on n'est plus sur le même espace car on ne considère plus que le deuxième dé (c'est comme-ci on ne lancait plus qu'un dé). Il y a alors 6 résultats dans cette expérience : $\{1,2,3,4,5,6\}$ et chacune a la même probabilité d'apparaître, soit 1/6. Par conséquent la probabilité que la somme des deux dés soit égale à 8 sachant que le premier dé vaut 3 est $P_B(A) = 1/6$ (il n'y a qu'une possibilité, le deuxième dé doit valoir 5).
On remarque que l'événement $A \cap B$ est l'événement "le premier dé donne 3 et la somme des chiffres vaut 8 lorsqu'on lance deux dés", on a alors $A \cap B = \{(3,5)\}$ dans l'ensemble $\Omega = \{(1,1),(1,2),\ldots,(5,6),(6,6)\}$ de cardinal 36.
D'autre part, on a $B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$.
D'où $\mathbb{P}(A \cap B) = 1/36$ et $\mathbb{P}(B) = 6/36 = 1/6$. On remarque que

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

Définition 3.2 Soient A et B deux événements avec $\mathbb{P}(B) \neq 0$. La **probabilité conditionnelle** $P_B(A)$, noté aussi $\mathbb{P}(A|B)$, de A sachant B est définie par :

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

3.1 Formule de Bayes

On considère deux événements A et B quelconques.
On peut A écrire comme l'union de deux événements disjoints : $A = (A \cap B) \cup (A \cap \overline{B})$, d'où

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \\ &= \mathbb{P}_B(A)\mathbb{P}(B) + \mathbb{P}_{\overline{B}}(A)\mathbb{P}(\overline{B}) \qquad (1.1)\end{aligned}$$

L'équation (1.1) s'appelle **formule des probabilités totales**.

Exemple 3.3 Une compagnie d'assurance répartit ses clients en deux classes : ceux qui sont enclins aux risques cardio-vasculaires (ie., à haut risque) et ceux qui ne le sont pas (i.e., à risque modéré). Elle constate que ceux à haut risque ont une probabilité de 0.4 d'avoir un accident dans l'année, tandis que ceux à risque modéré ont une probabilité de 0.2. On suppose que 30% de la population est à haut risque. Quelle est alors la probabilité qu'un nouvel assuré ait un accident cario-vasculaire durant l'année qui suit le signature de son contrat ?
On note A l'événement "l'assuré a un accident pendant l'année" et B l'événement "l'individu est à haut risque". Alors d'après la formule des probabilités totales (1.1), on a

$$\mathbb{P}(A) = 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26.$$

La formule de Bayes permet de calculer les probabilités a postériori d'un événement en fonction des probabilités a priori de cet événement, i.e. connaitre $\mathbb{P}_A(B)$ quand on connait $\mathbb{P}_B(A)$ et $\mathbb{P}_{\overline{B}}(A)$.

Proposition 3.4 Soient A et B deux événements. La **formule de Bayes** s'écrit :

$$\mathbb{P}_A(B) = \frac{\mathbb{P}_B(A)\mathbb{P}(B)}{\mathbb{P}_B(A)\mathbb{P}(B) + \mathbb{P}_{\overline{B}}(A)\mathbb{P}(\overline{B})}$$

Preuve. Il suffit d'écrire $\mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}_B(A)\mathbb{P}(B)}{\mathbb{P}(A)}$ et d'utiliser la formule des probabilités totales. △

Attention ! $\mathbb{P}_{\overline{B}}(A) \neq 1 - \mathbb{P}_B(A)$. Par contre, on a bien $\mathbb{P}_B(\overline{A}) = 1 - \mathbb{P}_B(A)$. La fonction $A \to \mathbb{P}_B(A)$ est une probabilité.

Exemple 3.5 On reprend l'exemple 3.3 sur la compagnie d'assurance. On veut connaitre la probabilité qu'un assuré à bas risque ait un accident cadio-vasculaire, i.e. la probabilité $\mathbb{P}_A(C)$, où C est l'événement "l'assuré n'est pas enclin aux accidents". On remarque que $C = \overline{B}$. D'où

$$\mathbb{P}_A(C) = \frac{\mathbb{P}_C(A)\mathbb{P}(C)}{\mathbb{P}(A)} = \frac{0.2 \times 0.7}{0.26} = 0.54.$$

4 Indépendance

Dans le langage courant, on dit de deux événements qui ne sont pas liés entre eux qu'ils sont indépendants. Par conséquent, on a tendance à dire que si A et B sont indépendants, la connaissance de B ne donne aucune information utile pour la connaissance de l'événement A et donc de manière naturelle $\mathbb{P}_B(A)$ doit être égale à $\mathbb{P}(A)$. On va donner à partir de cette constatation la définition mathématique de l'indépendance :

Définition 4.1 Deux événements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Attention ! Les notions d'événements indépendants et d'événements disjoints n'ont aucun rapport. Par exemple, si on jette deux dés et on considère les événements $A =$"le premier dé vaut 3" et $B =$"le deuxième dé vaut 1". Les événements sont indépendants car on jette les dés de façon indépendante, mais les événements ne sont pas disjoints car l'événement $A \cap B = \{$le premier dé vaut 3 et le deuxième vaut 1$\}$ est possible.

Proposition 4.2 Si A et B sont deux événements indépendants, alors A et $\overline{B}$ le sont aussi.

Preuve. On écrit A comme l'union de deux événements disjoints : $A = (A \cap B) \cup (A \cap \overline{B})$. D'où $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A \cap \overline{B})$. Ce qui implique que $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(\overline{B})$ et donc A et $\overline{B}$ sont indépendants. △

4.1 Indépendance entre plusieurs événements

Définition 4.3 On considère n événements $A_1, \dots, A_n$. Les événements sont **mutuellements indépendants** si pour tout $k \leq n$, pour tout $i_1, \dots, i_k \in \{1, \dots, n\}$ distincts on a $\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$.

Remarque 4.4 Si des événements sont mutuellements indépendants, alors ils sont indépendants deux à deux, mais la réciproque est fausse.
Par exemple, on jette deux dés, l'un rouge l'autre bleu et on considère les évènements suivants :

$$\begin{aligned} A &= \{\text{le dé rouge donne un chiffre impair}\} \\ B &= \{\text{le dé bleu donne un chiffre impair}\} \\ C &= \{\text{la somme des deux dés est un chiffre impair}\}. \end{aligned}$$

A, B et C sont deux à deux indépendants mais pas mutuellement indépendants.

5 Petits rappels d'analyse combinatoire

Dans le cas fini, lorsque les événements élémentaires sont équiprobables (par exemple, lorsque l'on lance des dés non truqués), on peut calculer la probabilité d'un événement A par :

$$\mathbb{P}(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

L'analyse combinatoire est l'étude des différents manières de "ranger" des objets.

5.1 Permutations

On appelle **permutation** un rangement ordonné de n objets. Le nombre de permutations est alors égal à $n!$.
Par exemple, si on considère trois objets a, b, c, il y a $3! = 6$ façons de les ordonner :

$$abc\,, acb\,, bac\,, bca\,, cab\,, cba.$$

5.2 Permutations avec répétition

Le nombre de permutations possible lorsque certains objets sont identiques est plus faible que lorsque tous les objets sont distincts. Si on considère n objets comprenant respectivement $n_1, n_2, \dots, n_k$ termes identiques, le nombre de permutations est égal à $\dfrac{n!}{n_1! n_2! \dots n_k!}$.

Si on veut ranger 3 boules vertes et 2 boules rouges, on a en tout 5 éléments. Le nombre de classements est alors $\dfrac{5!}{3!2!} = 10$.

5.3 Arrangements

On veut ranger k éléments parmi n en tenant compte de l'ordre. On n'autorise pas les répétitions d'objet (on ne peut pas utiliser plusieurs fois le même élément). Le nombre de possibilités est alors :

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

Par exemple si on veut savoir combien de nombres à 3 chiffres peuvent être formés avec l'ensemble $1, 3, 5, 7, 9$. On trouve dans ce cas $A_5^3 = 60$.

5.4 Combinaisons

On veut ranger k éléments parmi n mais cette fois-ci l'ordre ne nous intéresse pas. On n'autorise toujours pas les répétitions (expérience sans remise). Le nombre de possibilités est alors :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Ces coefficients satisfont une relation bien utile (triangle de Pascal) :

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

La formule du binôme de Newton est donnée par

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Chapitre 2

Variables Aléatoires

Après avoir réalisé une expérience, on ne s'intéresse bien souvent à une certaine fonction du résultat et non au résultat en lui-même. Lorsqu'on regarde une portion d'ADN, au lieu de vouloir connaitre tout la suite de nucléotides, on peut vouloir juste connaitre le nombre d'apparition d'un “”mot“”. Ces grandeurs (ou fonctions) auxquelles on s'intéresse sont en fait des fonctions réelles définies sur l'ensemble fondamental et sont appelées **variables aléatoires**.

On considère un ensemble Ω muni d'une probabilité $\mathbb{P}$.

Définition 0.1 Une **variable aléatoire** X est une fonction de l'ensemble fondamental Ω à valeurs dans $\mathbf{R}$, $X : \Omega \to \mathbf{R}$.
Lorsque la variable X ne prend que des valeurs discrètes, on parle de **variable aléatoire discrète**.
Un **vecteur aléatoire** $X : \Omega \to \mathbf{R}^d$ est une fonction $X = (X_1, \ldots, X_d)$ à valeurs dans $\mathbf{R}^d$ telle que les coordonnées X_i soient des variables aléatoires.

Pour tout intervalle $[a, b] \subset \mathbf{R}$, l'ensemble $\{X \in [a, b]\} = \{\omega \in \Omega : X(\omega) \in [a, b]\}$ est un événement.

Exemple 0.2

1. On jette deux dés distincts et on s'intéresse à la somme des points. On note X cette variable aléatoire, elle est définie par

$$\begin{array}{rccl} X : & \Omega & \to \mathbf{R} & \text{avec } \Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\} \\ & (\omega_1, \omega_2) & \mapsto \omega_1 + \omega_2. & \end{array}$$

   L'ensemble des valeurs possibles de X est $\{2, 3, \ldots, 12\}$.

2. On lance toujours deux dés, mais cette fois on s'intéresse au plus grand chiffre Y obtenu. On a alors

$$\begin{array}{rccl} Y : & \Omega & \to \mathbf{R} & \text{avec } \Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\} \\ & (\omega_1, \omega_2) & \mapsto \max(\omega_1, \omega_2). & \end{array}$$

   La variable Y est à valeurs dans $\{1, 2, \ldots, 6\}$.

3. On observe deux bactéries et on s'intéresse à la durée de vie T de la bactérie qui disparaîtra la première. L'ensemble fondamental est $\Omega = [0, +\infty[\times[0, +\infty[$. La variable T s'écrit alors

$$\begin{array}{rccl} T: & \Omega & \longrightarrow & \mathbf{R} \\ & (\omega_1, \omega_2) & \mapsto & \inf\{\omega_1, \omega_2\}. \end{array}$$

Seul le dernier exemple n'est pas une variable discrète.

1 Loi de probabilité, Fonction de répartition

La loi de probabilité d'une variable aléatoire permet de connaitre les chances d'apparition des différentes valeurs de cette variable.
On se place sur l'espace de probabilité $(\Omega, \mathbb{P})$.

Définition 1.1 Soit X une variable aléatoire. La **loi de probabilité** de X est définie par la fonction F_X, appelée **fonction de répartition** de la variable X, définie par

$$\begin{array}{rccl} F_X: & \mathbf{R} & \rightarrow & [0,1] \\ & x & \mapsto & \mathbb{P}(X \leq x). \end{array}$$

On dit que deux variables aléatoires X et Y ont la **même loi** si elles ont la même fonction de répartition $F_X = F_Y$.

Remarque 1.2 Soit I un intervalle de $\mathbf{R}$. L'événement $\{X \leq x\}$ représente l'ensemble des valeurs $\omega \in \Omega$ telles que $X(\omega)$ soit inférieur à x, i.e.$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$.

La loi de X est en générale notée $\mathcal{L}(X)$ ou $\mathcal{L}oi(X)$.

Remarque 1.3 On a $\mathbb{P}(X \in \mathbf{R}) = 1$, car $\mathbb{P}(X \in \mathbf{R}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \mathbf{R}\}) = \mathbb{P}(\Omega) = 1$.

Propriétés 1.4 La fonction de répartition est une fonction croissante à valeur dans $[0, 1]$ telle que $\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to+\infty} F_X(x) = 1$, mais elle n'est pas forcément continue.

Remarque 1.5 Soit $a \leq b$, on a $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a)$.

1.1 Loi d'une variable discrète

La fonction de répartition d'une variable discrète est constante par morceaux. Si X est une variable discrète à valeurs dans $\{x_1, \ldots, x_n\}$ avec $x_1 < \ldots < x_n$ alors pour $x \in \mathbf{R}$

$$F_X(x) = \sum_{i=1}^{k} \mathbb{P}(X = x_i) \quad \text{avec } k \text{ tel que } x_k \leq x < x_{k+1}.$$

De même, si X prend une infinité de valeurs $\{x_1, \ldots, x_n, \ldots\}$ avec $x_1 < \ldots < x_n \ldots$, on a pour $x \in \mathbf{R}$

$$F_X(x) = \sum_{i=1}^{k} \mathbb{P}(X = x_i) \quad \text{avec } k \text{ tel que } x_k \leq x < x_{k+1}.$$

Les sauts de la fonction de répartition F_X ont lieu en les points x_i et la hauteur du saut au point x_i est égale à $\mathbb{P}(X = x_i)$. Il suffit donc de calculer la fonction de répartition aux points x_i.

Proposition 1.6 Si X est à valeurs discrètes dans $\{x_1, \ldots, x_n\}$ (ou $\{x_1, \ldots, x_n, \ldots\}$), la loi de X est entièrement caractérisée par $\{\mathbb{P}(X = x_i) : i \geq 1\}$. (En effet, voir p. 12)

On remarque que

1. pour tout $i \geq 1$, $\mathbb{P}(X = x_i) \in [0, 1]$,
2. $\sum_{i \geq 1} \mathbb{P}(X = x_i) = 1$. (En effet, $1 = \mathbb{P}(X \in \mathbf{R}) = \sum_{i \geq 1} \mathbb{P}(X = x_i)$.)

Exemple 1.7 Reprennons les deux premiers exemples précédents, qui décrivent effectivement des variables discrètes. Pour trouver la loi de ces variables on utilise la proposition 1.6.

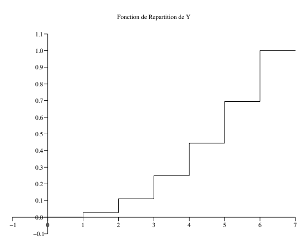
1. X est à valeurs dans $\{2, 3, \ldots, 12\}$, donc $\mathbb{P}(X = k) = 0$ pour $k \notin \{2, 3, \ldots, 12\}$.

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

2. Y est à valeurs dans $\{1, 2, \ldots, 6\}$, donc $\mathbb{P}(Y = k) = 0$ pour $k \notin \{1, 2, \ldots, 6\}$.

k	1	2	3	4	5	6
$\mathbb{P}(Y = k)$	1/36	3/36	5/36	7/36	9/36	11/36
$F_Y(k)$	1/36	4/36	9/36	16/36	25/36	1

La fonction de répartition de Y est



1.2 Lois discrètes usuelles

Loi de Bernoulli, $\mathcal{B}(p)$, avec $p \in]0, 1[$.

Une variable aléatoire X de Bernoulli est une variable qui ne prend que deux valeurs : l'échec (au quel on associe la valeur 0) et le succès (auquel on associe la valeur 1) d'une expérience. Cette expérience est appelée **épreuve de Bernoulli**. Par exemple, on souhaite savoir si une cellule est atteinte par un virus. On associe la valeur 1 si elle est atteinte (succès) et la valeur 0 si elle est saine (échec).
La loi est donnée par : $P(X = 1) = p \quad P(X = 0) = 1 - p$.

Loi Binomiale, $\mathcal{B}(n, p)$, avec $p \in]0, 1[$, $n \in \mathbb{N}$, $n \geq 1$.

Exemple 1.8 On étudie les mésanges bleues parmi la population de mésange en Ille et Vilaine. La mésange n'est pas un oiseau migrateur et vit en groupes formés de plusieurs espèces de mésanges. On poste alors des observateurs à différents endroits dans le département. On note p la proportion de mésanges bleues parmi la population de mésange. Quelle est la probabilité que le nombre de mésanges bleues sur 1000 mésanges étudiées soit égal à k ?

La loi Binomiale est utilisée pour modéliser un “sondage avec remise“. C'est la loi du nombre de succès lorsqu'on renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p. On note X le nombre de succès obtenus à l'issue des n épreuves. Sa loi s'appelle loi Binomiale de paramètres n et p. On a

$$P(X=k)=\begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \text{ avec } k \in \{0,1,..,n\}.$$

On peut écrire le nombre de succès X à l'aide des résultats de chaque épreuve de Bernoulli. On note X_i le résultat de la $i^{\text{ème}}$ expérience :

$$X_i = \begin{cases} 1 \text{ si la } i^{\text{ème}} \text{ expérience est réussie} \\ 0 \text{ si la } i^{\text{ème}} \text{ expérience est un échec} \end{cases}$$

On a alors $X = X_1 + \ldots + X_n$. Toute variable de Bernoulli peut s'écrire de cette façon.

Loi Hypergéométrique $H(N,m,n)$, avec $N \geq 1$, $(m,n) \in \{1,..,N\}^2$.

Exemple 1.9 Un lac contient N poissons. On en pêche m qu'on marque et qu'on remet à l'eau. On pêche à nouveau n poissons. Quelle est la probabilité de pêcher k poissons marqués ?

La loi hypergéométrique est utilisée pour modéliser un ”sondage sans remise“. C'est le cas de pratiquement tous les sondages (notamment lorsqu'on veut étudier la conformité d'un lot de médicaments, étudier le nombre de cellules atteintes par un virus, etc...).

La loi est donnée par : $P(X=k) = \dfrac{\begin{pmatrix} m \\ k \end{pmatrix}\begin{pmatrix} N-m \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$ si $k \in \{0,..,\min(m,n)\}$.

Remarque 1.10 Reprenons, l'exemple des poissons. Lorsque la taille N de la population de poissons dans le lac est très grande, le fait d'enlever 1 poisson ne changera quasiment pas la proportion $p = m/N$ de poissons marqués. De même, si on en enlève un nombre fini n. Par conséquent, chaque poisson pêché a à peu près la même probabilité p d'être marqué. On peut donc approcher la loi Hypergéométrique par la loi binomiale $\mathcal{B}(n,p)$ où p est la proportion de ”poissons marqués“.

De manière générale, lorsque la taille de la population est grande, on utilise souvent la binomiale même pour un ”sondage sans remise”.

Loi Géométrique, $\mathcal{G}(p)$, avec $p \in]0,1[$.

La loi géométrique est la loi du premier succès.

Exemple 1.11 On lance une pièce truquée jusqu'à ce qu'on obtienne une fois "Pile". On note p la probabilité de tomber sur “Pile”. On veut connaitre la probabilité d'avoir “Pile” au premier lancer, au deuxième, ..., au $k^{\text{ième}}$ lancer, ... On note X le nombre de lancers nécessaires pour avoir un succès.

La loi est donnée par : $P(X=k) = p(1-p)^{k-1}$ avec $k \in \mathbb{N}$, $k \geq 1$.
Une formule utile quand on veut faire des calculs avec la loi géométrique :

$$1 + x^2 + \ldots + x^n = \frac{1-x^{n+1}}{1-x}.$$

On peut calculer explicitement la fonction de répartition : $F(i) = 1 - (1-p)^i$ avec $i \geq 1$.

Loi de Poisson, $\mathcal{P}(\lambda)$, avec $\lambda > 0$ un réel.
La loi de Poisson est utilisée pour modéliser le comptage d'événements rares, c'est à dire des événements ayant une faible probabilité de réalisation : maladies rares, accidents mortels rares, le titrage d'une solution virale, mutations ou recombinaisons dans une séquence génétique, pannes, radioactivité...

Exemple 1.12 Une plaque d'Agar est une boîte de Pétri stérile qui contient un milieu de culture utilisée pour la culture des micro-organismes. Elles sont traitées avec un antibiotique. Les bactéries qui sont réparties sur les plaques ne peuvent se multiplier sauf les rares résistantes aux antibiotiques. Elles forment des colonies. Le décompte des colonies suit la loi de Poisson.

La loi est donnée par $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ avec $k \in \mathbb{N}$.
Une formule utile :

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

1.3 Loi d'une variable aléatoire à densité (ou continues)

Considérons la durée de vie d'une bactérie. On conçoit facilement que la probabilité que cette durée de vie vaille exactement une certaine valeur est nulle. Par exemple, il est quasiment impossible qu'une bactérie vive exactement 1 an 0 mois, 0 heure, 0 minute, ... La fonction de répartition d'une telle variable est par conséquent continue. On peut par contre s'intéresser à la probabilité que la bactérie vive moins d'un an.
On ne verra dans ce cours que des variables qui sont soit discrètes soit continues même s'il existe des variables plus complexes.

Définition 1.13 Une variable aléatoire X est **à densité**, ou **continue**, s'il existe une fonction f définie sur $\mathbf{R}$ telle que la fonction de répartition de X s'écrit

$$\forall x \in \mathbf{R} \quad F_X(x) = \int_{-\infty}^{x} f(t)dt$$

où f est une fonction intégrable sur $\mathbf{R}$ satisfaisant les conditions suivantes :

1. $f(t) \geq 0$ pour tout $t \in \mathbf{R}$,
2. $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = 1$.

Une fonction qui vérifie les conditions 1. et 2. est appelée **densité de probabilité**.

Propriétés 1.14 Soit X une variable aléatoire à densité.
Alors pour tout $x \in \mathbf{R}$, $\mathbb{P}(X = x) = 0$.

Preuve. Soit $x \in \mathbf{R}$. On considère l'intervalle réduit à un point $I = \{x\}$. On a

$$\mathbb{P}(X = x) = \mathbb{P}(X \in I) = \int_x^x f(t)dt = 0.$$

$\triangle$

Remarque 1.15

1. La probabilité $\mathbb{P}(X \in [a, b]) = \int_a^b f(t)dt$ correspond à l'aire de la surface comprise entre la courbe de f et l'axe des abcisses sur l'intervalle $[a, b]$.
2. La fonction de répartition d'une variable à densité est continue.

Proposition 1.16 Soit X une variable aléatoire de fonction de répartition F_X. Si F_X est continue sur $\mathbf{R}$ et dérivable sur $\mathbf{R}$ (sauf peut-être en un nombre fini de points), alors X est une variable à densité f donnée par $f(x) = F'_X(x)$.

1.4 Lois à densité usuelles

Loi Uniforme, $\mathcal{U}([a, b])$, avec $a, b \in \mathbb{R}$, $a < b$.
Cette loi est l'analogue continu de l'équiprobabilité dans le cas discret. Elle permet de modéliser le tirage d'un nombre aléatoire dans l'intervalle $[a, b]$.

Exemple 1.17 Lors d'une étude du comportement animal, on a relaché des oiseaux dont l'orientation a été rendue très difficile. On s'attend alors à ce que les oiseaux choisissent au hasard leur direction. On peut modéliser la direction prise par un oiseau de la façon suivante. On considère X l'angle entre le nord et la direction prise par l'oiseau (selon le sens des aiguilles d'une montre). La variable X suit une loi uniforme entre 0 et 360 degrés.

Densité :

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

Fonction de répartition :

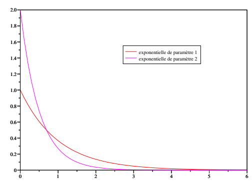
$$\begin{aligned} F(x) &= 0 \qquad \text{si } x < a \\ &= \frac{x-a}{b-a} \quad \text{si } x \in [a, b] \\ &= 1 \qquad \text{si } x > b \end{aligned}$$

Loi Exponentielle, $\mathcal{E}(\lambda)$, avec $\lambda > 0$ un réel.
Cette densité de probabilité permet en général de modéliser des durées de vie d'êtres non soumis au vieillissement (par exemple, la durée de vie d'une bactérie) ou des temps d'attente (par exemple, le temps d'attente entre deux signaux synaptiques).

Exemple 1.18 Dans une substance radioactive, la désintegration des noyaux se fait de façon spontanée. Le nombre de désintegration sur un intervalle de temps fixé suit une loi de Poisson. Par contre le temps d'attente entre deux désintégrations est modélisé par une loi exponentielle.

Densité :

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} \quad \text{si } x \geq 0 \\ &= 0 \quad \text{si } x < 0 \end{aligned}$$



Densité de lois exponentielles

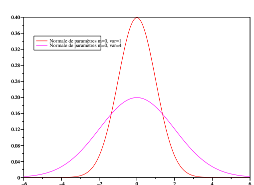
Fonction de répartition :

$$\begin{aligned} F(x) &= 0 \quad \text{si } x < 0 \\ &= 1 - e^{-\lambda x} \quad \text{si } x \geq 0 \end{aligned}$$

La loi exponentielle est la seule loi continue qui vérifie la propriété d'absence de mémoire :
Si $X \sim \mathcal{E}(\lambda)$, alors pour tout $s, t > 0$ $\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$.
La loi exponentielle est celle de la mortalité des êtres qui ne seraient pas soumis au veillissement : à chaque moment ils ont la même probabilité de mourrir dans l'unité de temps qu'il leur reste, quelque soit leur âge.

Loi Normale (ou loi Gaussienne), $\mathcal{N}\left(m, \sigma^2\right)$, avec $m \in \mathbb{R}$, $\sigma > 0$ réel.
La loi Normale est une loi centrale dans la théorie des probabilités. Elle est notamment très utilisée en statistique. Une grandeur influencée par un grand nombre de paramètres indépendants est souvent modélisée par une loi normale (par exemple, les erreurs de mesures lors d'une expérience).
Densité :

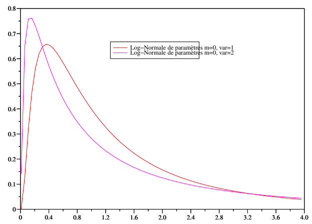
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-m)^2}{2\sigma^2}} \quad \text{avec } x \in \mathbb{R}.$$



Densité de lois normales

Loi Log-Normale, $LN\left(m, \sigma^2\right)$, avec $m \in \mathbb{R}$, $\sigma > 0$ réel.
Une variable X suit une loi Log-Normale $LN\left(m, \sigma^2\right)$ si $\ln(X)$ suit une loi normale $\mathcal{N}\left(m, \sigma^2\right)$.
Cette loi permet de modéliser par exemple le temps de survie des bactéries en présence de désinfectant, le dosage de certains médicaments ...
Densité :

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\dfrac{(\ln(x)-m)^2}{2\sigma^2}} \quad \text{avec } x \in> 0.$$



Densité d'une loi Log-Normale

Loi de Weibull, $W(a, b)$, avec $a > 0$, $b > 0$ réels.
La loi de Weibull est utilisée en démographie pour modéliser le veillissement et en épidémiologie pour modéliser la distribution de probabilité de la durée d'incubation d'une maladie infectieuse.
Densité :

$$f(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^{a}\right) \text{ avec } x \in> 0.$$

Fonction de répartition :

$$F(x) = 1 - \exp\left(-\left(\frac{x}{b}\right)^{a}\right) \text{ avec } x \in> 0.$$

On constate que pour $a = 1$, on retrouve la loi exponentielle. La loi exponentielle est celle de la mortalité des être vivants qui ne seraient pas soumis au veillissement : à chaque moment ils ont la même probabilité de mourrir dans l'unité de temps qu'il leur reste, quelque soit leur âge. Plus a est grand plus le veillissement se fait pesant (la mortalité augmente avec l'âge). Le cas $a < 1$ correspond à un monde dans lequel plus on veillirait, moins forte serait la probabilité de mourir dans l'unité de temps qui vient.

On va maintenant définir les paramètres de position et de dispersion d'une loi.

2 Espérance et variance d'une variable aléatoire

2.1 L'espérance

L'idée intuitive de l'espérance puise son origine dans les jeux de hasard. Considérons le jeu suivant : on lance un dé plusieurs fois de suite. Supposons que pour une mise de 1 euro, on gagne 1 euro si le résultat obtenu est pair, 2 euros si le résultat est 1 ou 3, et on perd 3 euros si le résultat est 5. Est-il intéressant de jouer à ce jeu ? Quel peut-être le gain moyen ? Soit X la variable aléatoire correspondant au nombre d'euros gagnés ou perdus. La loi de X est

k	-3	1	2
$\mathbb{P}(X = k)$	$1/6$	$1/2$	$1/3$

L'espérance de gain, noté $\mathbb{E}[X]$, est alors $\mathbb{E}[X] = -3 \cdot 1/6 + 1 \cdot 1/2 + 2 \cdot 1/3 = 2/3$. Le joueur gagne donc en moyenne $2/3$ d'euros pour une mise de 1 euro...

Définition 2.1 L'**espérance** d'une variable aléatoire X est notée $\mathbb{E}[X]$. Elle représente la valeur moyenne prise par la variable X.

1. Si X est une variable discrète à valeurs dans $\mathbb{D} = \{x_1, \ldots, x_n\}$, son espérance est

$$\mathbb{E}[X] = x_1\mathbb{P}(X = x_1) + \ldots + x_n\mathbb{P}(X = x_n) = \sum_{i=1}^{n} x_i\mathbb{P}(X = x_i).$$

2. Si X est une variable discrète à valeurs dans l'ensemble infini $\mathbb{D} = \{x_i : i \geq 1\}$, lorsque la somme est bien définie, son espérance est

$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i\mathbb{P}(X = x_i).$$

3. Si X est une variable à densité f, lorsque l'intégrale est bien définie, son espérance est

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

Lorsqu'une variable X vérifie $\mathbb{E}[X] = 0$, on dit que la variable est **centrée**.

Propriétés 2.2
1. L'espérance est linéaire : soient a et $b \in \mathbf{R}$, deux variables aléatoires X et Y d'espérance finie alors

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

2. Si $X \geq 0$, alors $\mathbb{E}[X] \geq 0$.
3. Si $X \leq Y$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Exemple 2.3 Supposons que la durée de vie T d'une bactérie est modélisée par la loi exponentielle de densité $f(t) = \lambda \exp(-\lambda t)$ pour $t \geq 0$ pour une certaine valeur de λ. Alors sa durée de vie moyenne est $E(T) = 1/\lambda$.

2.2 Espérance d'une fonction

On considère une variable aléatoire X et une fonction h. On aimerait connaitre l'espérance de la nouvelle variable $Y = h(X)$.

Proposition 2.4 Soit X une variable aléatoire, h une fonction définie sur $\mathbf{R}$. Alors :

1. si X est une variable discrète à valeurs dans $\mathbb{D} = \{x_1, \ldots, x_n\}$, on a

$$E\left[h(X)\right] = h(x_1)\mathbb{P}(X = x_1) + \ldots + h(x_n)\mathbb{P}(X = x_n) = \sum_{i=1}^{n} h(x_i)\mathbb{P}(X = x_i).$$

2. si X est une variable discrète à valeurs dans l'ensemble infini $\mathbb{D} = \{x_i : i \geq 1\}$, lorsque la somme est bien définie, on a

$$E\left[h(X)\right] = \sum_{i=1}^{+\infty} h(x_i)\mathbb{P}(X = x_i).$$

3. si X est une variable à densité f, lorsque l'intégrale est bien définie, on a

$$E\left[h(X)\right] = \int_{\mathbf{R}} h(x)f(x)dx.$$

Exercice 2.5 Soit X une v.a. de densité f définie par $f(x) = 0$ si $x < 0$ et $f(x) = xe^{-x^2/2}$ sinon.

A-1 Vérifier que f est bien une densité de probabilité.

A-2 Chercher la loi et l'espérance de $Y = X^2$.

On s'intéresse maintenant aux précipitations à Brest. On suppose que la durée (en heure) T entre deux précipitations vérifie la relation $T = \lambda Y$, où λ est une constante. En consultant les relevés météo, on se rend compte qu'il s'écoule au moins 24h consécutives sans pluie avec une probabilité 0.09.

B-1 Déterminer λ.

B-2 Combien de temps s'écoule-t-il en moyenne entre deux averses ?

B-3 Il est midi. La pluie vient de s'arrêter. Quelle est la probabilite qu'il pleuve avant 14h ?

Corrigé :

A-1 évident.

A-2 on calcule la fonction de répartition de Y : pour $x > 0$, comme X est positive on a $F_Y(x) = \mathbb{P}(X^2 \leq x) = \mathbb{P}(X \leq \sqrt{x}) = 1 - e^{-x/2}$ et pour $x \leq 0$, comme Y est positive $F_Y(x) = 0$. Donc Y suit la loi exponentielle de paramètre 1/2.

B-1 $\mathbb{P}(\lambda Y > 24) = 0.09 \Rightarrow e^{-24/2\lambda} = 0.09$ d'où la valeur de λ.

B-2 $E[T] = \lambda E[Y] = 2\lambda$.

B-3 $\mathbb{P}(T \leq 2) = 1 - e^{-1/\lambda}$.

⋆

2.3 Variance, Écart type

On a vu que l'espérance correspondait à la valeur moyenne d'une variable aléatoire. L'écart type représente l'écart moyen (la distance moyenne) entre la variable et sa moyenne. Elle mesure la dispersion d'une variable, plus l'écart-type est grand plus la variable prend des valeurs qui peuvent être éloignées les unes des autres, plus l'écart-type est petit plus la variable prend des valeurs proches de sa moyenne.

Définition 2.6 La **variance** d'une variable aléatoire X, notée $Var(X)$, est définie par

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

L'**écart type** est la racine carrée de la variance :

$$\sigma(X) = \sqrt{Var(X)}.$$

Lorsqu'une variable X vérifie $Var(X) = 1$, on dit que la variable est **réduite**.

Remarque 2.7 La variance s'écrit aussi $Var(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.

Preuve. Il suffit de développer le carré et d'utiliser la linéarité de l'espérance. △

Propriétés 2.8
1. $Var(X) = 0$ ssi X est constante.
2. Soient a et $b \in \mathbf{R}$, alors $Var(aX + b) = a^2 Var(X)$.

Preuve. Comme $E[aX+b] = aE[X]+b$, on a par définition $Var(aX+b) = E[(aX-aE[X])^2] = a^2Var(X)$. △

Exemple 2.9 Supposons que la durée de vie T d'une bactérie est modélisée par la loi exponentielle de densité $f(t) = \lambda \exp(-\lambda t)$ pour $t \geq 0$ pour une certaine valeur de λ. La variance de la durée de vie de la bactérie étudiée est $Var(T) = 1/\lambda^2$.

3 Une fonction remarquable : la transformée de Laplace

Cette fonction est surtout utile en évolution des populations. En effet, si on considère une population dont chaque génération évolue de façon indépendante de la précédente mais selon la même loi, pour connaître s'il y a un risque d'extinction de la population, on utilise la transformée de Laplace.

Définition 3.1 Soit X une variable aléatoire à valeurs dans $\mathbf{R}$. Pour $t \in \mathbf{R}$, sa **transformée de Laplace** est définie par

$$L_X(t) = \mathbb{E}[e^{tX}]$$

pour $t \in \mathbf{R}$.

Proposition 3.2 La transformée de Laplace L_X d'une variable X vérifie les propriétés suivantes

1. L_X est à valeurs dans $[0, +\infty]$, $L_X(0) = 1$,
2. L_X caractérise la loi de X (i.e., si deux variables ont la même transformée de Laplace, alors elles ont la même loi),
3. si $\mathbb{E}[|X|^k] < \infty$ alors L_X est dérivable k fois en 0 et la dérivée $k^textime$ en zéro vaut $L_X^{(k)}(0) = \mathbb{E}[X^k]$.

Exercice 3.3 Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ de loi de Poisson $\mathcal{P}(\lambda)$:

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \text{ avec } k \in \mathbb{N}$$

1. Calculer la fonction génératrice de X.
2. Donner l'expression de la transformée de Laplace de X.
3. En déduire les moments de X.

4 Récapitulatif des lois usuelles

4.1 Lois de probabilité discrètes

Loi de Bernoulli, $\mathcal{B}(p)$, avec $p \in]0,1[$:
$P(X=1) = p \quad P(X=0) = 1-p$.

Loi Binomiale, $\mathcal{B}(n,p)$, avec $p \in]0,1[$, $n \in \mathbb{N}^*$:
$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ avec $k \in \{0,1,..,n\}$.

Loi Géométrique, $\mathcal{G}(p)$, avec $p \in]0,1[$:
$P(X=k) = p(1-p)^{k-1}$ avec $k \in \mathbb{N}^*$.

Loi de Poisson, $\mathcal{P}(\lambda)$, avec $\lambda > 0$:
$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ avec $k \in \mathbb{N}$.

Loi	**Notation**	**Espérance**	**Variance**	**Transformée de Laplace**
Bernoulli	$\mathcal{B}(p)$	p	$p(1-p)$	$1-p+pe^t$
Binomiale	$\mathcal{B}(n,p)$	np	$np(1-p)$	$(1-p+pe^t)^n$
Géométrique	$\mathcal{G}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\dfrac{pe^t}{1-(1-p)e^t}$
Poisson	$\mathcal{P}(\lambda)$	λ	λ	$e^{\lambda(e^t-1)}$

4.2 Lois de probabilité à densité

Loi	**Notation**	**Densité**	**Espérance**	**Variance**	**Transformée de Laplace**
Uniforme	$\mathcal{U}([a,b])$	$\dfrac{1}{b-a}$ pour $x \in [a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Exponentielle	$\mathcal{E}(\lambda)$, $\lambda > 0$	$\lambda e^{-\lambda x}$ pour $x \geq 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	
Normale	$\mathcal{N}(m,\sigma^2)$, $\sigma^2 > 0$	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-m)^2}{2\sigma^2}}$	m	σ^2	
Cauchy	$\mathcal{C}$	$\dfrac{1}{\pi}\dfrac{1}{1+x^2}$	pas d'espérance	pas de variance	

Chapitre 3

Loi de probabilités usuelles

1 Lois Discrètes

Loi de Bernoulli, $\mathcal{B}(p)$, avec $p \in]0, 1[$.

Une variable aléatoire X qui ne prend que deux valeurs suit la loi de Bernoulli : l'échec (au quel on associe la valeur 0) et le succès (auquel on associe la valeur 1). Par exemple, on souhaite savoir si une cellule est atteinte par un virus. On associe la valeur 1 si elle est atteinte (succès) et la valeur 0 si elle est saine (échec). On a alors :
Loi : $P(X = 1) = p \quad P(X = 0) = 1 - p$.
Espérance : p
Variance : $p(1 - p)$.
Cette expérience est appelée **épreuve de Bernoulli**.

Loi Hypergéométrique

Exemple 1.1 Un lac contient N poissons. On en pêche m qu'on marque et qu'on remet à l'eau. On pêche à nouveau n poissons. Quelle est la probabilité d'avoir pêcher k poissons marqués ?

La loi hypergéométrique, notée $H(N, m, n)$, avec $N \geq 1$, $(m, n) \in \{1, .., N\}^2$, est utilisée pour modéliser un "sondage avec remise". cette loi est utilisée par exemple lorsqu'on comptabilise certains poissons dans un lac, ou lorsque qu'on comptabilise à la jumelle les mésanges bleues parmi la population de mésange en Ille et Vilaine (la mésange n'étant pas un oiseau migrateur).

Loi : $P(\{k\}) = \dfrac{\begin{pmatrix} m \\ k \end{pmatrix}\begin{pmatrix} N-m \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$ si $k \in \{0, .., \min(m, n)\}$.
Espérance : np avec $p = \frac{m}{N}$
Variance : $\dfrac{N-n}{N-1} np(1-p)$ avec $p = \frac{m}{N}$.

Loi Binomiale,

Exemple 1.2 Afin de vérifier la conformité d'un lot de comprimés, on prélève un échantillon de taille n avec remise. On note p la probabilité d'avoir un comprimé non conforme. Quelle est la probabilité que le nombre de comprimés non conforme soit égal à k.

La loi Binomiale est notée $\mathcal{B}(n,p)$, avec $p \in]0,1[$, $n \in \mathbb{N}$, $n \geq 1$.

On renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p. La loi du nombre X de succès obtenus à l'issue des n épreuves s'appelle loi Binomiale de paramètres n et p. On a
Loi : $P(X=k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$ avec $k \in \{0,1,..,n\}$.
Espérance : np
Variance : $np(1-p)$.

<u>Loi Géométrique</u>, $\mathcal{G}(p)$, avec $p \in]0,1[$.

Un exemple classique où apparait cette loi est la recherche du premier succès : on lance une pièce truquée jusqu'à ce qu'on obtienne une fois "Pile" (voir l'exemple 0.3-**??** du Chapitre 2.).
Loi : $P(\{k\}) = p(1-p)^{k-1}$ avec $k \in \mathbb{N}$, $k \geq 1$.
Espérance : $\dfrac{1}{p}$
Variance : $\dfrac{1-p}{p^2}$.
Fonction de répartition : $F(i) = 1-(1-p)^i$ avec $i \geq 1$.
Cette loi vérifie une propriété d'absence de mémoire : Soit X une variable aléatoire de loi géométrique $\mathcal{G}(p)$, alors pour tout $i,j \in \mathbb{N}^*$ $\mathbb{P}(X > i+j | X > j) = \mathbb{P}(X > i)$.

<u>Loi de Poisson</u>, $\mathcal{P}(\lambda)$, avec $\lambda \in \mathbb{R}_+^*$.
Cette loi décrit en général des événements rares. Elle a d'innombrables applications en biologie (par exemple pour le titrage d'une solution virale) et épidémiologie (pour modéliser la survenue aléatoire de morts dans de petits territoires au cours de périodes courtes).

Loi : $P(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}$ avec $k \in \mathbb{N}$.
Espérance : λ
Variance : λ.
<u>Loi Multinomiale</u> $\mathcal{M}_d(n,p)$, avec $n \in \mathbb{N}^*$, $n \geq 1$, $p \in]0,1[^d$ tel que $p_1 + p_2 + .. + p_d = 1$.
On considère une expérience qui peut mener à d événements élémentaires de probabilité respectives $p_1, \ldots, p_d$. On répète l'expérience n fois et on compte le nombre de fois où chaque événement élémentaire est apparu. Par exemple, dans une population, les génotypes AA, Aa et aa sont respectivement en proportion $p_1 = 5/10$, $p_2 = 4/10$ et $p_3 = 1/10$ *(ici* $d = 3$*)*. On peut alors se poser la question de savoir quelle est la probabilité d'avoir 5 AA, 4 Aa et un aa dans un échantillon de 10 individus *(*$n = 10$*)*.
Loi : $P(\{k_1, k_2, ..., k_d\}) = \frac{n!}{k_1!k_2!...k_d!} p_1^{k_1} p_2^{k_2} ... p_d^{k_d}$ avec $\sum_{i=i}^d k_i = n$.
Espérance : $(np1, \cdots, np_d)$.

<u>Loi Binomiale Négative</u>, avec $n \in \mathbb{N}^0$, $n \geq 1$, $p \in]0,1[$.

Loi : $P(\{k\}) = \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} p^n (1-p)^{k-n}$ si $k \geq n$.
Espérance : $\dfrac{n}{p}$
Variance : $n\dfrac{1-p}{p^2}$.

2 Lois sur $\mathbb{R}$ à densité de probabilité

<u>Loi Uniforme</u>, $\mathcal{U}([a,b])$, avec $a, b \in \mathbb{R}$, $a < b$.
Densité :

$$f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x) \text{ avec } x \in \mathbb{R}.$$

Espérance : $\dfrac{a+b}{2}$
Variance : $\dfrac{(b-a)^2}{12}$.
Fonction de répartition :

$$\begin{aligned} F(t) &= 0 && \text{si } t < a \\ &= \frac{x-a}{b-a} && \text{si } t \in [a,b] \\ &= 1 && \text{si } t > b \end{aligned}$$

<u>Loi Exponentielle</u>, $\mathcal{E}(\lambda)$, avec $\lambda \in \mathbb{R}_+^*$.
Densité :

$$f(x) = \lambda e^{-\lambda x} \mathbb{I}_{[0,+\infty[}(x).$$

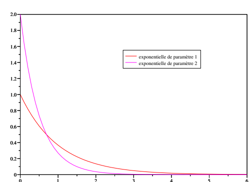


Figure 1. Densité de lois exponentielles

Espérance : $\dfrac{1}{\lambda}$
Variance : $\dfrac{1}{\lambda^2}$.
Fonction de répartition :

$$\begin{aligned} F(t) &= 0 && \text{si } t < 0 \\ &= 1 - e^{-\lambda t} && \text{si } t \geq 0 \end{aligned}$$

La loi géométrique est la seule loi continue qui vérifie la propriété d'absence de mémoire :
Si $X \sim \mathcal{E}(\lambda)$, alors pour tout $s, t > 0$ $\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$.
Cette loi modélise en général des durées de vie ou des temps d'attente.

<u>Loi Normale (ou loi Gaussienne)</u>, $\mathcal{N}(m, \sigma^2)$, avec $m \in \mathbb{R}$, $\sigma \in \mathbb{R}_+^*$.
Densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \text{ avec } x \in \mathbb{R}.$$

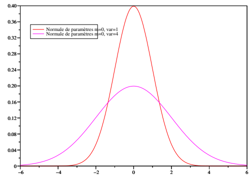


Figure 2. Densité de lois normales

Espérance : m
Variance : σ^2.
La loi Normale est une loi centrale dans la théorie des probabilités. Elle est notamment très utilisée en statistique.

Loi de Cauchy
Densité :

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2} \text{ avec } x \in \mathbb{R}.$$

Son espérance et sa variance ne sont pas définies (intégrales divergentes sur $\mathbb{R}$).

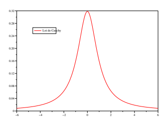


Figure 3. Densité de la loi de Cauchy

2.1 Quelques autres lois sur $\mathbb{R}$ à densité

Loi Gamma, $G(a, \lambda)$, avec $a > 0$ et $\lambda > 0$.
Densité :

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \text{ avec } x \in]0, +\infty[$$

Espérance : $\frac{a}{\lambda}$
Variance : $\frac{a}{\lambda^2}$.
On remarque que la loi exponentielle $\mathcal{E}(\lambda)$ est aussi la loi Gamma $G(1, \lambda)$.

Loi Beta, $\mathcal{B}(a, b)$, avec $a > 0$ et $b > 0$.
Densité :

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \text{ avec } x \in]0, 1[$$

Espérance : $\frac{a}{a+b}$
Variance : $\frac{ab}{(a+b)^2(a+b+1)}$.

à finir
Loi Lognormale, $LN\left(m, \sigma^2\right)$, avec $m \in \mathbf{R}$ et $\sigma > 0$.
Une variable X à valeurs positives est distribuée selon la loi Lognormale si $Y = \ln(X)$ suit une loi normale.

Loi de Weibull, $\mathcal{W}(a, b)$, La loi de Weibull est utiliséée en démographie pour modéliser le veillissement et en épidémiologie pour modéliser la durée d'incubation d'une maladie infectieuse.

Loi de Gumbel,

Loi du Chi-deux, $\chi^2(n)$,

Chapitre 4

Lois jointes, indépendance

On va la plupart du temps se limiter à l'étude de couple de variables aléatoires, on peut bien sûr étendre les notions introduites à des n-uplets de variables.
On considère deux variables aléatoires X et Y. On aimerait connaitre s'il y a influence d'une variable sur l'autre et la quantifier.

Exemple 0.1 On peut se poser la question de l'influence des catastrophes météorologiques (tempêtes, ouragans, tsunamis, ...) sur les cours de la bourse. La variable X modélisera alors les catastrophes météorologiques et la variable Y le cours de la bourse.

1 Loi jointe

Définition 1.1 Soient X et Y deux variables aléatoires. La **loi jointe** de (X,Y) est définie par sa **fonction de répartition** $F_{(X,Y)}$:

$$\begin{array}{rcl} F_{(X,Y)} : \mathbf{R} \times \mathbf{R} & \to & [0,1] \\ (u,v) & \to & \mathbb{P}(X \leq u, Y \leq v) \end{array}$$

Les **lois marginales** du couple (X,Y) sont la loi $\mathcal{L}(X)$ de X et la loi $\mathcal{L}(Y)$ de Y.

Remarque 1.2 À partir de la loi du couple, on retrouve facilement les lois marginales. En effet, soit $u \in \mathbf{R}$, $F_X(u) = \mathbb{P}(X \leq u) = \mathbb{P}(X \leq u, Y \in \mathbf{R}) = \lim_{v \to +\infty} F_{(X,Y)}(u,v)$.
Par contre, des lois marginales on ne peut pas déduire la loi du couple, car elles ne rendent pas compte des liens qui existent entre les variables.

1.1 Cas des variables discrètes

Propriétés 1.3 Soient X et Y deux variables discrètes, X à valeurs dans $\mathbb{D}_X$ et Y à valeurs dans $\mathbb{D}_Y$. La loi du couple (X,Y) est définie par l'ensemble des probabilités :

$$\mathbb{P}(X = x, Y = y) \quad \text{avec } x \in \mathbb{D}_X \text{ et } y \in \mathbb{D}_Y.$$

Remarque 1.4 On obtient les lois marginales de façon simple. Soit $x \in \mathbb{D}_X$, on a

$$\begin{array}{rcl} \mathbb{P}(X = x) & = & \mathbb{P}(X = x \text{ et } Y \in \mathbb{D}_Y) \\ & = & \sum_{y \in \mathbb{D}_Y} \mathbb{P}(X = x, Y = y) \end{array}$$

De même, pour $y \in \mathbb{D}_Y$, on a $\mathbb{P}(Y=y) = \sum_{x \in \mathbb{D}_X} \mathbb{P}(X=x, Y=y)$

Dans le cas où les variables sont discrètes et prennent un petit nombre de valeurs, on écrit en général la loi du couple sous la forme d'un tableau :

$Y \backslash X$	$\ldots$	Somme des colonnes
$\vdots$	$\mathbb{P}(X=x, Y=y)$	$\mathbb{P}(Y=y)$
Somme des lignes	$\mathbb{P}(X=x)$	

Exemple 1.5 1. On lance une pièce truquée 3 fois. La probabilité de tomber sur "Pile" est $2/3$. Soit X le nombre de "Face" obtenu dans les deux premiers jets et Y le nombre de "Face" obtenu dans les deux derniers jets. La loi de (X, Y) est donnée par

$y \backslash x$	0	1	2	$\mathbb{P}(Y=y)$
0	$\left(\frac{2}{3}\right)^3 = \frac{8}{27}$	$\frac{1}{3}\left(\frac{2}{3}\right)^2 = \frac{4}{27}$	0	$4/9$
1	$\frac{1}{3}\left(\frac{2}{3}\right)^2 = \frac{4}{27}$	$\frac{1}{3}\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{6}{27}$	$\left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{2}{27}$	$4/9$
2	0	$\left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{2}{27}$	$\left(\frac{1}{3}\right)^3 = \frac{1}{27}$	$1/9$
$\mathbb{P}(X=x)$	$4/9$	$4/9$	$1/9$	

2. L'université de Rennes 1 veut évaluer l'effet de l'offre MIPE sur le campus et voir quel système d'exploitation est apprécié des étudiants. Les proportions collectées sont résumées dans un tableau :

Système d'exploitation / Filière	Windows	Mac OS	Linux
Biologie	0.07	0.05	0.02
Droit/Économie	0.08	0.02	0
Informatique	0.25	0.13	0.09
Mathématiques	0.21	0.04	0.04

On déduit de ce tableau les proportions d'élèves qui ont profité de l'offre MIPE en fonction des filières, ainsi que la répartition des systèmes d'exploitation sur le campus.

Exercice 1.6 On effectue une suite infinie de lancers indépendants d'un dé équilibré. On note les lancers à partir de 1. On définit les deux variables aléatoires :

X est égale au numéro du lancer qui donne le premier 6,
Y est égale au nombre de 5 obtenus avant le premier 6.

Déterminer la loi du couple (X, Y).

Corrigé : Le couple est à valeurs dans $\mathbb{N}^* \times \mathbb{N}$ avec $Y < X$. Par conséquent si $k \geq n$, $\mathbb{P}(X=n, Y=k) = 0$ et si $k < n$,

$$\begin{aligned} \mathbb{P}(X=n, Y=k) &= \mathbb{P}(\text{“k fois 5 et pas de 6 sur les } n-1 \text{ premiers lancers et un 6 au } n^{\text{ème}} \text{ lancer.”}) \\ &= \binom{n-1}{k} \left(\frac{1}{6}\right)^k \left(\frac{4}{6}\right)^{n-1-k} \frac{1}{6}. \end{aligned}$$

⋆

1.2 Cas des variables à densité

Définition 1.7 La loi du couple de v.a. (X, Y) est dite à densité s'il existe une fonction $f_{(X,Y)}$ de deux variables telle que le fonction de répartition du couple vérifie pour tout $(u, v) \in \mathbf{R}^2$

$$\mathbb{P}(X \leq u, Y \leq v) = \int\int_{]-\infty,u]\times]-\infty,v]} f_{(X,Y)}(x, y)dxdy$$

satisfaisant les conditions suivantes :

1. $f_{(X,Y)}(x, y) \geq 0$ pour tout $(x, y) \in \mathbf{R}^2$,
2. $\displaystyle\int\int_{\mathbf{R}^2} f_{(X,Y)}(x, y)dxdy = 1$.

On peut facilement retrouver la densité à partir de la fonction de répartition. En effet, si le couple (X, Y) a pour densité $f_{(X,Y)}$, sa fonction de répartition s'écrit :

$$F_{(X,Y)}(u, v) = \int_{-\infty}^{v}\int_{-\infty}^{u} f_{(X,Y)}(x, y)dxdy.$$

En dérivant une fois par rapport à chacune des variables la fonction de répartition, on obtient

$$f_{(X,Y)}(u, v) = \frac{\partial^2}{\partial u \partial v} F_{(X,Y)}(u, v).$$

Remarque 1.8 De même que pour les variables discrètes, on peut retrouver facilement les lois marginales. Soit $u \in \mathbf{R}$, on a

$$\begin{aligned} F_X(u) &= \mathbb{P}(X \leq u) \\ &= \mathbb{P}(X \in]-\infty, u] \text{ et } Y \in \mathbf{R}) = \mathbb{P}((X, Y) \in]-\infty, u] \times \mathbf{R}) \\ &= \int_{-\infty}^{u} \left(\int_{\mathbf{R}} f_{(X,Y)}(x, y)dy \right) dx. \end{aligned}$$

La variable X est une variable continue de densité $f_X(x) = \displaystyle\int_{\mathbf{R}} f_{(X,Y)}(x, y)dy$.

De même, la variable Y est une variable continue de densité $f_Y(y) = \displaystyle\int_{\mathbf{R}} f_{(X,Y)}(x, y)dx$.

Exemple 1.9 1. Considérons le couple (X, Y) de densité $f_{(X,Y)}(x, y) = 3/8(x^2 + xy/2)\mathbb{I}_{[0,1]\times[0,2]}(x, y)$.

Cette fonction est bien une densité de probabilité. On en déduit la densité de X :

$$\begin{aligned} f_x(x) &= \int_{\mathbf{R}} f_{(X,Y)}(x, y)dy = \int_0^2 \frac{3}{8}(x^2 + \frac{xy}{2})dy \\ &= \frac{3}{4}x(x + 2) \quad \text{si } x \in [0, 1] \end{aligned}$$

On peut calculer par exemple $\mathbb{P}(X > Y)$:

$$\begin{aligned} \mathbb{P}(X > Y) &= \int\int_{x<y} f_{(X,Y)}(x, y)dxdy = \int_0^1 \left(\int_0^x \frac{3}{8}(x^2 + \frac{xy}{2})dy \right) dx \\ &= \frac{3}{8}\int_0^1 2x^3 dx = \frac{1}{4}. \end{aligned}$$

2. Considérons le couple (X, Y) de densité

$$f_{(X,Y)}(x, y) = c(y^2 - x^2)e^{-y} \quad si - y < x < y,\ y > 0$$

Cette fonction est bien une densité de probabilité lorsque $c = 1/8$. Les densités des lois marginales sont :

$$\begin{aligned} f_X(x) &= \frac{1}{4}(|x| + 1)e^{|x|} \quad \text{pour } x \in \mathbf{R} \\ \text{et } f_Y(y) &= \frac{1}{6}y^3e^{-y} \quad \text{pour } y > 0 \end{aligned}$$

2 Covariance

La covariance permet d'estimer la dépendance entre deux variables aléatoires.

Définition 2.1 La **covariance** de deux v.a. X et Y est

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

L'espérance $E(XY)$ est calculée à partir de la loi jointe de (X, Y) :

1. dans le cas discret, lorsque la somme a un sens,

$$E(XY) = \sum_{x \in \mathbb{D}_X, y \in \mathbb{D}_Y} xy\mathbb{P}(X = x, Y = y)$$

2. dans le cas continu, lorsque l'intégrale a un sens,

$$E(XY) = \int\int_{\mathbf{R}^2} xyf_{(X,Y)}(x, y)dxdy.$$

Remarque 2.2 Soient X et Y deux v.a. Alors

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

Preuve. En développant le carré, on obtient le résultat :

$$\begin{aligned} Var(X + Y) &= E[(X + Y - E[X + Y])^2] = E[(X - E[X] + Y - E[Y])^2] \\ &= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E[(X - E[X])(Y - E[Y])] \\ &= Var(X) + Var(Y) + 2E[XY] - 2E[X]E[Y] \\ &= Var(X) + Var(Y) + 2Cov(X, Y). \end{aligned}$$

△

Exemple 2.3 Une étude médicale sur l'effet du tabac est menée dans un hopital. Les 2278 patients sont divisés en deux groupes : ceux atteints d'un cancer pulmonaire ($X = 1$) et les autres ($X = 0$). Les membres de chaque groupe sont ensuite répartis selon le nombre Y de paquets de cigarettes fumés par jour.

Cancer pulmonaire	Nombre de paquets de cigarettes					Total
	0	1	2	3	4	
0	1247	492	319	58	9	2125
1	66	50	28	6	3	153
Total	1313	542	347	64	12	2278

On souhaite étudier l'association entre cancer pulmonaire et la consommation de cigarette en calculant la covariance.
La proportion de personnes atteintes d'un cancer pulmonaire est 6.72%, le nombre moyen de paquets de cigarettes consommés est 0.65, on obtient

$$\begin{aligned} Cov(X,Y) &= \frac{1\times 1\times 50 + 1\times 2\times 28 + 1\times 3\times 6 + 1\times 4\times 3}{2278} - 0.0672\times 0.65 \\ &= 0.02 \end{aligned}$$

La covariance est positive, le résultat indique qu'il y a un lien positif entre la déclaration du cancer et la consommation de cigarettes.

3 Variables aléatoires indépendantes

Définition 3.1 Deux variables aléatoires X et Y sont indépendantes si pour tout intervalle A et B de $\mathbf{R}$ on a

$$\mathbb{P}(X\in A, Y\in B) = \mathbb{P}(X\in A)\mathbb{P}(Y\in B).$$

Proposition 3.2 (Admis)
Deux v.a. X et Y sont indépendantes

ssi la fonction de répartition du couple vérifie pour tout $(x,y)\in\mathbf{R}^2$

$$F_{(X,Y)}(x,y) = F_X(x)F_Y(y),$$

ssi dans le cas discret pour tout $(x,y)\in\mathbf{R}^2$,

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

ssi dans le cas continu, la densité du couple vérifie pour tout $(x,y)\in\mathbf{R}^2$

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y).$$

Proposition 3.3 (Hors Programme)
Deux v.a. X et Y sont indépendantes

ssi pour toutes fonctions h,g positives,

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)],$$

ssi la transformée de Laplace du couple vérifie pour tout (u,v),

$$L_{(X,Y)}(u,v) = L_X(u)L_Y(v) \quad \text{où } L_{(X,Y)}(u,v) = E[e^{uX+vY}].$$

Remarque 3.4 Si X et Y sont indépendantes, alors $Cov(X,Y)=0$. D'où $Var(X+Y) = Var(X)+Var(Y)$.
Attention La réciproque est fausse !

Définition 3.5 Les variables aléatoires $X_1,\dots,X_n$ sont indépendantes si pour tout intervalles $A_1,\dots,A_n$ de $\mathbf{R}$ on a

$$\mathbb{P}(X_1\in A_1,\dots,X_n\in A_n) = \prod_{i=1}^n \mathbb{P}(X_i\in A_i).$$

Une suite de variables $(X_n)_{n\geq}$ indépendantes est une suite telle que pour toute sous partie finie $I\subset\mathbb{N}$, les variables $(X_i)_{i\in I}$ sont indépendantes.

Remarque 3.6 Si les v.a. $X_1, \ldots, X_n$ sont indépendantes, alors elles sont indépendantes deux à deux.
Attention La réciproque est fausse ! Par exemple, soient X et Y deux variables indépendantes de même loi : $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. On considère $Z = XY$. Les variables sont deux à deux indépendantes, mais pas mutuellement indépendantes.

Proposition 3.7 (Admis)
Soit $X_1, \ldots, X_n$ des variables indépendantes, alors pour tout $p \in \{1, \ldots, n\}$, toute fonction de $X_1, \ldots, X_p$ est indépendante de toute fonction de $X_{p+1}, \ldots, X_n$.

Exemple 3.8 (Important)
Une loi binomiale $\mathcal{B}(n, p)$ de paramètres n et p correspond au modèle suivant :
On renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p. On compte le nombre de succès obtenus à l'issue des n épreuves et on appelle S la variable aléatoire correspondant à ce nombre de succès. Donc si on note X_i le résultat du i$^{\text{ème}}$ lancer, les variables $X_1, \ldots, X_n$ sont des variables de Bernoulli de paramètre p, indépendantes. On a $S = X_1 + \ldots + X_n$.
On en déduit facilement que si $S_1 \sim \mathcal{B}(n_1, p)$ et $S_2 \sim \mathcal{B}(n_2, p)$ indépendantes, alors $S_1 + S_2 \sim \mathcal{B}(n_1 + n_2, p)$.

Exemple 3.9 On considère deux serveurs informatiques. On suppose que les durées de vie T_1 et T_2 de chaque serveur informatique suivent des lois exponentielles de paramètre respectif α_1 et α_2. On souhaite connaitre la loi du premier instant de panne $T = \min(T_1, T_2)$, i.e. le premier instant où l'un des deux serveurs cesse de fonctionner. En calculant la fonction de répartition de T, on montre que T suit la loi exponentielle de paramètre $\alpha_1 + \alpha_2$.

Exercice 3.10 Soient X et Y deux variables indépendantes, de loi de Poisson de paramètre respectif λ et μ. Montrer que $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

Corrigé : On peut le montrer facilement en calculant la transformée de Laplace ou la fonction génératrice de $X + Y$. ⋆

Exercice 3.11 Deux personnes se donnent rendez-vous. L'heure d'arrivée de chacune de ces deux personnes sur les lieux est une variable uniforme entre midi et une heure. Ces deux variables sont indépendantes. Quelle est la probabilité qu'ils arrivent au même instant ? Quelle est la probabilité que le premier arrivé doive attendre plus de 10 minutes ? Si les deux personnes se donnent un rendez-vous plus précis, à midi exactement par exemple. La loi uniforme est-elle adaptée au problème ? Quelle autre loi peut-on utiliser ?

Corrigé : Notons X, Y les heures d'arrivées des personnes, elle ont pour loi à densité $f(x) = \mathbb{I}_{[12,13]}(x)$. Par indépendance, la densité du couple est alors $f_{X,Y}(x, y) = \mathbb{I}_{[12,13]}(x)\mathbb{I}_{[12,13]}(y)$.
La probabilité qu'ils arrivent au même instant est nulle car le couple est à densité. Par ailleurs $\mathbb{P}(\max(X, Y) > \min(X, Y) + 1/6) = 25/36$. si elles se donnent rendez-vous à midi, il faut une loi qui charge plus du coté de midi. ⋆

4 Loi conditionnelle pour des variables discrètes

On se place sur $(\Omega, \mathcal{F}, P)$ un espace probabilisé.

Définition 4.1 On considère deux variables aléatoires X et Y discrétes à valeurs respectivement dans $\mathbb{D}_X$ et $\mathbb{D}_Y$. La loi de X sachant que $Y = y$ pour $y \in \mathbb{D}_Y$ est donnée par

$$\mathbb{P}_{Y=y}(X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \quad \text{pour } x \in \mathbb{D}_X.$$

Exercice 4.2

1. Soient X et Y deux variables indépendantes, de loi de Poisson de paramètre respectif λ et μ. Déterminer la loi conditionnelle de X sachant $\{X + Y = n\}$.
2. Si $X_1, ..., X_r$ sont indépendantes de lois de Poisson de paramètres respectifs $\lambda_1, ..., \lambda_r$, quelle est la loi conditionnelle de $(X_1, ..., X_r)$ sachant $\{X_1 + ... + X_r = n\}$?

Corrigé :

1. D'après l'exercice 3.10, on sait déjà que $X+Y \sim \mathcal{P}[\lambda+\mu)$. Soit $k \in \mathbb{N}$, si $k > n$ $\mathbb{P}_{X+Y=n}(X = k) = 0$ et si $k \leq n$

$$\begin{aligned} \mathbb{P}_{X+Y=n}(X = k) &= \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y = n)} \overset{\text{Indép.}}{=} \frac{\mathbb{P}(X = k)\mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)} \\ &= \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu}\right)^k \left(\frac{\mu}{\lambda + \mu}\right)^{n-k}. \end{aligned}$$

   On retrouve la loi $\mathcal{B}(n, \frac{\lambda}{\lambda+\mu})$.

2. On obtient une loi multinomiale $\mathcal{M}(n, p)$ avec $p = (\frac{\lambda_1}{\lambda_1+\cdots+\lambda_r}, \ldots, \frac{\lambda_r}{\lambda_1+\cdots+\lambda_r})$.

⋆

Chapitre 5

Théorèmes limites

On va définir ici la convergence de suite de variables aléatoires. On va voir qu'on peut définir plusieurs notions de convergence, mais on n'abordera que deux notions de convergence : la convergence en loi et la convergence en probabilité, alors qu'el en existe d'autre. On donnera ensuite les deux principaux théorèmes de convergence en probabilité.

1 Quelques notions sur les variables gaussiennes

Définition 1.1 Une variable aléatoire **gaussienne** (ou **normale**) X d'espérance m et de variance σ^2 a pour densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad \text{avec } x \in \mathbf{R}.$$

On dit que X suit la loi $\mathcal{N}(m, \sigma^2)$.

Proposition 1.2

1. La transformée de Laplace d'une variable aléatoire X de loi $\mathcal{N}(m, \sigma^2)$ est :
$$L_X(t) = E[e^{tX}] = e^{mt} e^{\frac{\sigma^2 t^2}{2}} \quad \text{pour } t \in \mathbf{R}.$$
2. Si $a, b \in R$, X une variable gaussienne $\mathcal{N}(m, \sigma^2)$, alors $Y = aX + b$ est une variable gaussienne $\mathcal{N}(am + b, a^2\sigma^2)$.
3. Soit $X_1, \ldots, X_n$ des variables gaussiennes réelles indépendantes d'espérance m_i et de variance σ_i^2 respectives. Soient $a_1, \ldots, a_n$ des réels, alors $Y = a_1X_1 + \ldots + a_nX_n$ est une variable gaussienne $\mathcal{N}(\sum_{i=1}^{n} a_i m_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$.

Preuve.

1. Calculons la transformée de Laplace d'une variable gaussienne :
$$\begin{aligned} L_X(t) &= E[e^{tX}] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} e^{tx} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m-\sigma^2 t)^2 + m^2 - (m+\sigma^2 t)^2}{2\sigma^2}} dx \\ &= e^{\frac{\sigma^4 t^2 + 2\sigma^2 mt}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m-\sigma^2 t)^2}{2\sigma^2}} dx = e^{mt} e^{\frac{\sigma^2 t^2}{2}}. \end{aligned}$$
On reconnaît la densité de $\mathcal{N}(m+\sigma^2 t, \sigma^2)$ dans l'intégrale, on a donc $\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{(x-m-\sigma^2 t)^2}{2\sigma^2}} dx = 1$.

2. On calcule la transformée de Laplace de Y :

$$\begin{aligned} L_Y(t) &= E[e^{atX+bt}] = e^{bt}E[e^{(at)X}] = e^{bt}e^{mat}e^{\sigma^2 a^2 t^2/2} \\ &= e^{(am+b)t}e^{(a\sigma)^2 t^2/2} \end{aligned}$$

   On reconnaît la transformée de Laplace de la loi gaussienne $\mathcal{N}(am+b, a^2\sigma^2)$, et comme la transformée de Laplace caractérise la loi, on a le résultat.

3. On utilise encore la transformée de Laplace. Comme les variables X_i sont indépendantes, la transformée de Laplace de la somme est le produit des transformées de Laplace :

$$g_Y(t) = E[e^{ta_1X_1+\ldots+ta_nX_n}] = \prod_{i=1}^{n} E[e^{ta_iX_i}] = \prod_{i=1}^{n} e^{\sum_{i=1}^{n} a_i m_i t} e^{\sum_{i=1}^{n} a_i^2\sigma_i^2 t^2/2}$$

   La variable Y suit par conséquent la loi gaussienne $\mathcal{N}(\sum_{i=1}^{n} a_i m_i, \sum_{i=1}^{n} a_i^2\sigma_i^2)$.

$\triangle$

Remarque 1.3 1. La loi gaussienne est bien connue. N'importe quel ordinateur donne les valeurs de la fonction de répartition.

2. Soit X une variable de loi $\mathcal{N}(0,1)$, alors les valeurs de X sont "concentrées" entre -4 et 4. En effet, la probabilité $\mathbb{P}(|X| > 4) = 5.10^{-5}$ est très faible.

3. Du fait de la symétrie de la loi Normale $\mathcal{N}(0,1)$ par rapport à 0, on a

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \geq -x) = 1 - \mathbb{P}(X \leq x), \\ \text{et donc } \mathbb{P}(|X| \leq x) &= 2\mathbb{P}(X \leq x) - 1 = 2F_X(x) - 1. \end{aligned}$$

On donne maintenant la table de la loi gaussienne qui permet de connaitre la valeur de la fonction de répartition F d'une loi $\mathcal{N}(0,1)$ en fonction de la valeur de x.

Table de la loi gaussienne

La table ci-dessous comporte les valeurs de la fonction de répartition de la loi normale, à savoir les valeurs de :

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9986	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Utilisation

On lit les décimales dans les lignes, et les centièmes en colonnes. Par exemple, la valeur de F(1.65) se trouve à l'intersection de la ligne 1.6 et de la colonne 0.05 - on trouve F(1.65)=0.9505, à 10^{-4} près. Pour les valeurs négatives de x, on utilise la relation F(-x)=1-F(x).

2 Inégalités importantes

L'inégalité de Markov
Soit X une variable positive et un réel $a > 0$, alors

$$\mathbb{P}(X > a) \leq \frac{E[X]}{a}.$$

Preuve. La variable X étant positive, on a

$$E[X] = E[X\mathbb{I}_{X>a}] + E[X\mathbb{I}_{X\leq a}] \geq E[X\mathbb{I}_{X>a}] \geq a\mathbb{P}(X > a).$$

△

L'inégalité de Bienaymé-Tchebichev
Soit X une variable dans L^2 et un réel $a > 0$, alors

$$P\left(|X - E[X]| > a\right) \leq \frac{Var(X)}{a^2}.$$

Preuve. On applique l'inégalité de Markov à la variable $|X - E[X]|^2$ et on obtient le résultat car $\mathbb{P}(|X - E[X]| > a) = \mathbb{P}(|X - E[X]|^2 > a^2)$. △

Exercice 2.1 Estimer en utilisant l'inégalité de Bienaymé-Tchebichev le nombre de lancers nécessaires d'une pièce bien équilibrée pour que la moyenne du nombre de piles obtenu soit dans l'intervalle $[1/2 - 0.01, 1/2 + 0.01]$ avec une probabilité supérieure ou égale à 0.96 ?

3 Différentes notions de convergence

On considère une suite de variables $(X_n)_{n\geq 1}$ et la variable X.

• Convergence en probabilité
La suite $(X_n)_{n\geq 1}$ converge en probabilité vers X ssi

$$\forall \varepsilon > 0 \quad \lim_{n\to+\infty} P\left(|X_n - X| > \varepsilon\right) = 0.$$

On note alors $X_n \underset{n\to+\infty}{\overset{proba}{\longrightarrow}} X$.
La convergence en probabilité signifie que la probabilité que X_n soit loin de X (à distance plus grande que ε pour $\varepsilon > 0$) converge vers 0 quand n tend vers $+\infty$.

Propriétés 3.1 Soient $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ deux suites de v.a. qui convergent respectivement vers X et Y en probabilité. Alors

1. si f est continue sur $\mathbf{R}$ alors $(f(X_n))_{n\geq 1}$ converge en probabilité vers $f(X)$,
2. $(X_nY_n)_{n\geq 1}$ converge en probabilité vers XY,
3. si $(a_n)_{n\geq 1}$ est une suite de réels convergeant vers a, alors $(X_n + a_nY_n)_{n\geq 1}$ converge en probabilité vers $X + aY$.

Preuve. A faire.... △

• Convergence en loi

La suite $(X_n)_{n\geq 1}$ converge en loi vers X ssi

$$\forall I \text{ intervalle} \quad \mathbb{P}(X_n \in I) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(X \in I).$$

On note alors $X_n \underset{n\to+\infty}{\overset{loi}{\longrightarrow}} X$.

La convergence en loi signifie que la loi de X_n converge vers la loi de X quand n tend vers $+\infty$. Par conséquent l'ensemble des valeurs de X_n converge vers l'ensemble des valeurs de X.

Proposition 3.2 $X_n \underset{n\to+\infty}{\overset{loi}{\longrightarrow}} X$
ssi il y a convergence des fonctions de répartition : $F_{X_n}(x) \underset{n\to+\infty}{\longrightarrow} F_X(x) \quad \forall x$ point de continuité de F_X
ssi dans le cas discret, $\forall x \in \mathbb{D}_X$, $\mathbb{P}(X_n = x) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(X = x)$.

La proposition suivante est *hors programme* pour les étudiants préparant le capes.

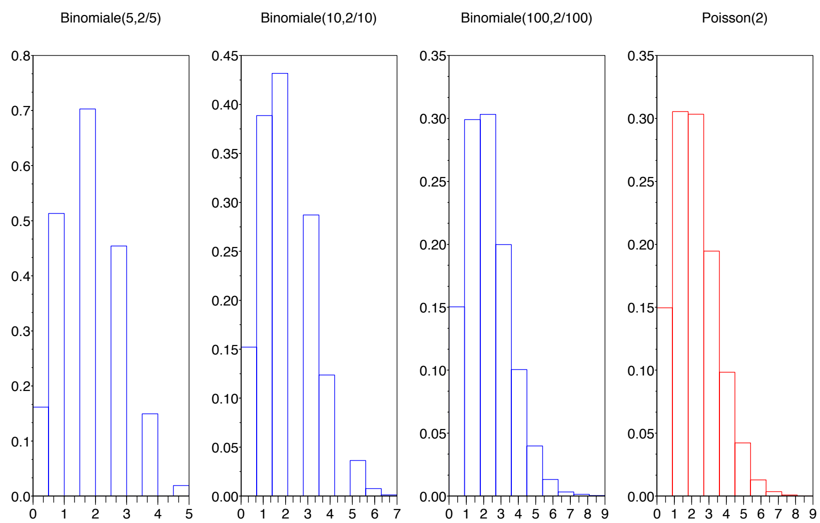
Proposition 3.3 $X_n \underset{n\to+\infty}{\overset{loi}{\longrightarrow}} X$
ssi il y a convergence des transformées de Laplace : $L_{X_n}(t) \underset{n\to+\infty}{\longrightarrow} L_X(t)$ pour tout t
ssi $\forall h$ continue bornée $E[h(X_n)] \underset{n\to+\infty}{\longrightarrow} E[h(X)]$.

Exemple 3.4

1. Si X_m suit une loi hypergéométrique $H(N, m, n)$ avec $p_m = \frac{m}{N} \underset{m\to+\infty}{\longrightarrow} p \in]0,1[$, alors la loi de X_m converge vers la loi Binomiale de paramètre (n, p).

2. Si X_n suit une loi Binomiale $\mathcal{B}(n, p_n)$ avec $np_n \underset{n\to+\infty}{\longrightarrow} \lambda > 0$, alors la loi de X_n converge vers la loi de Poisson $\mathcal{P}(\lambda)$.

   On peut illustrer la convergence de la loi Binomiale vers la loi de Poisson, en regardant l'évolution de l'histogramme d'une Binomiale $\mathcal{B}(n, 2/n)$ par exemple :



Preuve.

1. On rencontre la loi $H(N, m, n)$ lorsque l'on fait un sondage sans remise. N représente la taille de la population, m est le nombre d'individu qui ont le caractère étudiée dans la population totale et n est la taille de la sous-population sur laquelle on fait le sondage. Soit X_m le nombre d'individu qui ont le caractère étudié parmi les m individus sondés, X_m suit la loi $H(N, m, n)$. La quantité $p_m = \frac{m}{N}$ représente la proportion d'individus ayant le caractère étudié dans la population totale.
   Comme $p_m \underset{m\to+\infty}{\longrightarrow} p$, on a $N \underset{m\to+\infty}{\sim} \frac{m}{p}$.
   Par ailleurs, X_m est à valeurs dans $\{0, 1, \ldots, \min(m, n)\}$. Soit $k \leq n$ et $m \geq n$, on a
   $$\mathbb{P}(X_m = k) \quad = \quad \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \binom{n}{k} \frac{m!}{(m-k)!} \frac{(N-m)!}{(N-m-n+k)!} \frac{(N-n)!}{N!}$$
   On a $\frac{m!}{(m-k)!} \underset{m\to+\infty}{\sim} m^k$, $\frac{(N-n)!}{N!} \underset{m\to+\infty}{\sim} N^{-n}$ et comme $N-m \underset{m\to+\infty}{\longrightarrow} +\infty$ on a aussi $\frac{(N-m)!}{(N-m-n+k)!} \underset{m\to+\infty}{\sim}$ $(N-m)^{n-k}$. Par conséquent, $\forall k \in \{0, \ldots, n\}$
   $$\mathbb{P}(X_m = k) \underset{m\to+\infty}{\longrightarrow} \binom{n}{k} p^k (1-p)^{n-k}.$$

2. Soit $K \geq 0$ fixé, pour n assez grand ($n \geq k$) on a

$$\mathbb{P}(X_n = k) = \frac{1}{k!}\frac{n!}{(n-k)!}p_n^k(1-p_n)^{n-k} \underset{m\to+\infty}{\sim} \frac{1}{k!}(np_n)^k e^{-(n-k)p_n}.$$

D'où le résultat.

△

Propriétés 3.5
1. Soit $(X_n)_{n\geq 1}$ une suite qui converge en loi vers X. Si f est continue sur $\mathbf{R}$ alors $(f(X_n))_{n\geq 1}$ converge en loi vers $f(X)$.
2. Soit f une fonction continue sur $\mathbf{R}^2$.
Si $(X_n, Y_n) \xrightarrow[n\to+\infty]{loi} (X, Y)$, alors $f(X_n, Y_n) \xrightarrow[n\to+\infty]{loi} f(X, Y)$.

Attention, si on a juste f continue, $X_n \xrightarrow[n\to+\infty]{loi} X$ et $Y_n \xrightarrow[n\to+\infty]{loi} Y$, on ne peut rien conclure sur $f(X_n, Y_n)$. Notamment on ne peut rien dire sur la convergence de $(X_nY_n)_{n\geq 1}$ ou de $(X_n + Y_n)_{n\geq 1}$.

Preuve. Il suffit de montrer le premier résultat.... à faire... △

• Lien entre les deux notions de convergence

"convergence en proba" ⇒ "convergence en loi"

Attention la réciproque est fausse ! Cependant, on a le résultat suivant :

Remarque 3.6 Si $X_n \xrightarrow[n\to+\infty]{loi} a$ et a est constante, alors $X_n \xrightarrow[n\to+\infty]{proba} a$

Preuve. A faire.... △

Proposition 3.7 (admis)
Soit $X_n \xrightarrow[n\to+\infty]{loi} X$ et $Y_n \xrightarrow[n\to+\infty]{proba} Y$, alors $(X_n, Y_n) \xrightarrow[n\to+\infty]{loi} (X, Y)$.

3.1 La loi faible des grands nombres

On s'intéresse à la variable aléatoire $\overline{X}_n = \frac{X_1+X_2+...+X_n}{n}$ appelée **moyenne empirique** et on souhaite connaître son comportement quand n grandit. On a le théorème suivant :

Théorème 3.8 Soit X une v.a. dans L^1.
Soient $(X_n)_{n\geq 0}$ une suite de variables indépendantes et de même loi que X. Alors

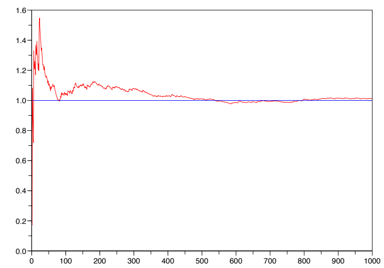
$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \underset{n\to+\infty}{\longrightarrow} E[X] \quad \text{en proba.}$$

Preuve. Preuve dans le cas où $X \in L^2$. On suppose $E[X] = 0$ (sinon il suffit de travailler avec la suite $(X_n - E[X])_{n\geq 0}$). Soit $\varepsilon > 0$. On a

$$P\left(\frac{|X_1 + \ldots + X_n|}{n} \geq \varepsilon\right) \overset{\text{Bienaymé -Tchebichev}}{\leq} \frac{Var(X_1 + \ldots + X_n)}{n^2\varepsilon^2} \overset{\text{Indépendance}}{\leq} \frac{Var(X)}{n\varepsilon^2}.$$

D'où le résultat. △

Exemple 3.9 On peut retrouver le résultat par simulation. Prenons l'exemple de la loi exponentielle $\mathcal{E}(1)$. On simule des variables $X_1, \ldots, X_n$ suivant la loi $\mathcal{E}(1)$ et on trace l'évolution de la moyenne empirique $\overline{X}_n = \frac{X_1+X_2+\ldots+X_n}{n}$ en fonction de n. On observe bien la convergence vers 1.



3.2 Le théorème limite centrale

On se demande maintenant à quelle vitesse la moyenne empirique converge vers l'espérance.

Théorème 3.10 Soit X une v.a. dans L^2. On pose $m = E[X]$ et $\sigma^2 = Var(X)$. Soient X_1, X_2, ..., X_n des variables indépendantes et de même loi que X. Alors

$$\frac{\sqrt{n}}{\sigma}\left(\frac{X_1+X_2+...+X_n}{n} - m\right) \underset{n\to+\infty}{\longrightarrow} Z \quad \text{en loi, avec } \mathcal{L}(Z) = \mathcal{N}(0,1).$$

Donc $\forall a, b \in \mathbf{R}$, $a < b$,

$$\mathbb{P}(a \leq \frac{\sqrt{n}}{\sigma}\left(\frac{X_1+X_2+...+X_n}{n} - m\right) \leq b) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(a \leq Z \leq b) = \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$$

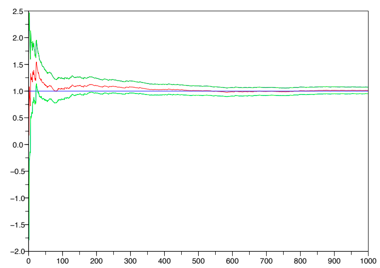
La vitesse de convergence de la moyenne empirique est donc assez lente car elle est en $\frac{\sqrt{n}}{\sigma}$ (bien sûr, plus σ est petit plus la vitesse est rapide).

Preuve. A faire.... △

Exemple 3.11 Reprenons l'exemple de la loi exponentielle $\mathcal{E}(1)$. On simule des variables $X_1, \ldots, X_n$ suivant la loi $\mathcal{E}(1)$. En utilisant la table de la loi gaussienne p.45, on remarque que $\mathbb{P}(|Z| \leq 1.96) = 0.95$ où $\mathcal{L}(Z) = \mathcal{N}(0,1)$. Comme $m = 1$ et $\sigma = 1$, d'après le théorème central limite on a donc

$$P\left(\sqrt{n}\left|\overline{X}_n - 1\right| \leq 1.96\right) \underset{n\to+\infty}{\longrightarrow} 0.95.$$

D'où en réécrivant l'événement, on a $P\left(1 \in \left[\overline{X}_n - \frac{1.96}{\sqrt{n}}, \overline{X}_n + \frac{1.96}{\sqrt{n}}\right]\right) \underset{n\to+\infty}{\longrightarrow} 0.95$. Ceci signifie que pour n assez grand, on a environ 95% de chance que la valeur 1 soit dans l'intervalle $\left[\overline{X}_n - \frac{1.96}{\sqrt{n}}, \overline{X}_n + \frac{1.96}{\sqrt{n}}\right]$.
On trace en rouge l'évolution de la moyenne empirique $\overline{X}_n$ et en vert les bornes de l'intervalle $\left[\overline{X}_n - \frac{1.96}{\sqrt{n}}, \overline{X}_n + \frac{1.96}{\sqrt{n}}1.96\right]$ en fonction de n :



Exemple 3.12 1. Si X_i sont des v.a. indépendantes de loi $\mathcal{B}(p)$, alors $S_n = \sum_{i=1}^n X_i$ suit une loi $\mathcal{B}(n,p)$. Par conséquent, pour n grand on approche

$$\frac{S_n - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0,1)$$

2. Si X_i sont des v.a. indépendantes de loi $\mathcal{P}(\lambda)$, alors $S_n = \sum_{i=1}^n X_i$ suit une loi $\mathcal{P}(n\lambda)$. Par conséquent, pour n grand on approche

$$\frac{S_n - n\lambda}{\sqrt{n\lambda}} \sim \mathcal{N}(0,1)$$

Règles d'usage
1. On approche la loi $\mathcal{P}(\lambda)$ par $\mathcal{N}(\lambda,\lambda)$ dès que $\lambda \geq 10$.
2. On approche la loi $\mathcal{B}(n,p)$ par $\mathcal{N}(np, np(1-p))$ dès que $np \geq 10$ et $n(1-p) \geq 10$.
3. Si $np < 10$ et $p \leq 0.1$, on approche la loi $\mathcal{B}(n,p)$ par $\mathcal{P}(np)$.

Exercice 3.13 Un restaurateur peut servir 75 repas, uniquement sur réservation. La pratique montre que 20% des clients ayant réservés ne viennent pas.
Combien le restaurateur doit-il accepter de réservations indépendantes pour avoir une probabilité supérieure ou égale à 0.9 de pouvoir servir tous les clients qui se présenteront ?

Exercice 3.14 Une compagnie d'assurance se propose d'assurer 100000 clients contre le vol. Les sommes en euros (la plupart du temps nulles) $X_1, \ldots, X_{100000}$ qu'aura à rembourser chaque année la compagnie aux clients sont des v.a. indépendantes d'espérance 75 et d'écart type 750.
Quelle somme cette compagnie d'assurance doit-elle faire payer à chaque client par an pour que ses frais évalués à $1,5$ millions d'euros soient couverts avec une probabilité supérieure ou égale à 0.999 ?

Exercice 3.15 Un restaurateur peut servir 75 repas, uniquement sur réservation. La pratique montre que 20% des clients ayant réservés ne viennent pas.
Combien le restaurateur doit-il accepter de réservations indépendantes pour avoir une probabilité supérieure ou égale à 0.9 de pouvoir servir tous les clients qui se présenteront ?

Exercice 3.16 Une compagnie d'assurance se propose d'assurer 100000 clients contre le vol. Les sommes en euros (la plupart du temps nulles) $X_1, \ldots, X_{100000}$ qu'aura à rembourser chaque année la compagnie aux clients sont des v.a. indépendantes d'espérance 75 et d'écart type 750.

Quelle somme cette compagnie d'assurance doit-elle faire payer à chaque client par an pour que ses frais évalués à $1,5$ millions d'euros soient couverts avec une probabilité supérieure ou égale à 0.999 ?
(On utilisera sans justification l'approximation par la loi Normale.)

Deuxième partie

Quelques notions de statistique descriptive

Chapitre 6

Vocabulaire

La statistique descriptive a pour but d'étudier une **population** à travers des **observations** (aussi appelées **données**). Cette description se fait à travers leur présentation (la plus synthétique possible), leur représentation graphique et le calcul de résumés numériques. Il n'est pas besoin pour cette étude de modèles probabilistes.
On parle de **recensement** lorsque l'on fait une étude exhaustive d'une **population**. Lorsqu'on n'étudie qu'une sous-population, on parle alors d'**échantillon**.
Les éléments de ses populations sont appelés **individus** ou **unités statistiques**. Cette terminologie, que la statistique a hérité de son premier champ d'action, la démographie, s'applique aussi bien à des ensembles d'objets concrets ou abstrait : personnel d'un établissement, clientèle d'un magasin, production d'un atelier, ensemble des accidents intervenus au cours d'une période.

Il convient de définir avec précision les ensembles que l'on étudie et notamment leurs fontières. Voici une version allégée de la définition d'une agglomération au sens de l'INSEE :

Exemple 0.1 Définition d'une agglomération selon l'INSEE
Les limites entre territoire urbain et rural sont redéfinies à l'occasion de chaque recensement. Leur tracé fait intervenir la notion d'agglomération de population définie comme un ensemble d'habitations. Dans cet ensemble, qui doit abriter au moins 2000 habitants, aucune habitation ne doit être séparée de la plus proche de plus de 200 mètres. Les frontières de ces zones coincident dans tous les cas avec les limites communales. En revanche, les limites des autres circonscriptions administratives (canton, arrondissements, départements) ne sont pas prises en compte lors de leur délimitation. Une même unité urbaine peut s'étendre sur deux départements. Si l'agglomération de population s'étend sur plusieurs communes, l'ensemble de ces communes forment une agglomération urbain. Si l'agglomération s'étend sur une seule commune, celle-ci est une ville isolée. Toutes ces communes sont considérées comme urbaines. Les autres communes sont classées communes rurales.

A chaque individu de la population sont associés des **caractères**, appelés aussi **variables**. Ainsi le personnel d'une entreprise peut être décrit selon divers caratères : âge, sexe, qualification, ancienneté dans l'entreprise, commune de résidence...De même un lot de pièce mécanique peut être décrit suivant le poids, le diamètre, la matière...

1 Modalités

Chacun des caractères étudiés peut présenter deux ou plusieurs **modalités**. Les modalités sont les différentes situations où les individus peuvent se trouver à l'égard du caractères considéré. Les modalités d'un même caractères doivent être incompatibles et exhaustives : chaque individu de la population présente une et une seule des modalités du caractère étudié. Un salarié peut être de sexe masculin ou féminin, une personne peut-être mariée, célibataire, en union libre, veuf ou divorcée.

2 Les caractères qualitatifs

Un caractère est dit **qualitatif** si ses modalités ne sont pas mesurables : la couleur des cheveux, le sexe, la présence de maladie. Dans la pratique on associe des valeurs à chaque classe d'individus (0 : cheveux noirs, 1 : cheveux bruns...), mais il faut cependant rester vigilant car faire des opérations (par exple, la moyenne) sur ce type de donnée n'a aucun sens.

Exemple 2.1 Nomenclature des professions et des catégories socioprofessionnelles :

1. Niveau agrégé (8 postes)

Code	Libellés
1	Agriculteurs exploitants
2	Artisans, commerçants, chefs d'entreprise
3	Cadres et professions intellectuelles supérieures
4	Professions intermédiaires
5	Employés
6	Ouvriers
7	Retraités
8	Autres sans activité professionnelle

2. Niveau de publication courante (24 postes)...extrait

Code	Libellés
10	Agriculteurs exploitants
21	Artisans
22	Commerçants et assimilés
23	Chefs d'entreprise de 10 salariés ou plus
31	Professions libérales
32	Cadres de la fonction publique, professions intellectuelles et artistiques
36	Cadres d'entreprise
41	Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés
46	Professions intermédiaires administratives et commerciales des entreprises
47	Techniciens
48	Contremaîtres, agent de maîtrise
51	Employés de la fonction publique

Un caractère qualitatif est dit **ordinal** si ses modalités peuvent être rangées dans un ordre naturel, il sera **nominal** sinon.

Exemple 2.2 La commune de résidence est un caractère nominal, tandis que le grade militaire est un caractère ordinal.

3 Caractères quantitatifs

Par opposition à la notion de qualitatif, un caractère est dit **quantitatif** s'il est mesurable, si chacune de ces modalités correspondent à un nombre. On parle alors de **variable statistique** : les modalités d'un caractère quantitatif sont les différentes valeurs possibles de la variable statistique.

On en distingue de deux types : les discrètes et les continues.
Une **variable statistique discrète** ne peut prendre que des valeurs isolées d'un intervalle réel : le nombre d'enfant dans un couple, l'âge d'une personne en années révolues, le nombre de sinistres dûs au feu ...
Une **variable statistique continue** est quand a elle suceptible de prendre toutes les valeurs appartenant à un intervalle réel : la taille, les poids des nouveau nés, la durée de fonctionnement d'un moteur...

La distinction entre variables discrètes et variables continues est claire sur le plan théorique, mais sur le plan pratique même si une grandeur est par nature continue elle est mesurée par un instrument qui ne fournit que des valeurs discrètes (une balance ne fournira pas le poids exact d'un nouveau né mais son poids arrondi au gramme près). On appellera donc variable statistique continue une variable qui peut prendre un très grand nombre de valeurs susceptibles d'être très proches les unes des autres.

L'objet de la statistique descriptive est l'étude des caractères, qu'ils soient qualitatifs ou quantitatifs. L'analyse univariée consiste à étudier une seule variable, l'analyse multivariée cherche à analyser les liaisons entre plusieurs variables (par exple, la taille des agglomérations et les secteurs d'activités).

Chapitre 7

Distribution statistique à un caractère

Le but de cette partie est de rendre lisible un jeu de données. Si ce jeu est important, il est évident que présenter les données brutes est illisible. On va introduire ici les outils élémentaires, adaptés à la variable observée, permettant de présenter une série de données de façon synthétique et d'en résumer les principales caractéristiques.

1 L'analyse des caractères quantitatifs

L'analyse des caractères quantitatifs cherche à mettre en évidence les caractéristiques d'une distribution. Pour analyser ces caractéristiques, on a recours à des **statistiques**. Formellement on observe une population de taille N sur laquelle on observe une variable X (par exemple, le poids). On note X_j la valeur de la variable pour l'individu j. Une statistique n'est alors rien d'autre qu'une fonction dépendant des X_j, i.e. :

$$\begin{array}{rcl} \mathbb{R}^N & \to & \mathbb{R}^P \text{avec } P \leq N \\ (X_1, \ldots, X_N) & \to & T(X_1, \ldots, X_N) \end{array}$$

Au vu de la définition il existe une infinité de statistiques, mais il est préférable de la choisir de façon la plus optimale possible.
Le statisticien Yule (1945) avait défini des propriétés souhaitables : une *bonne* statistique doit :

1. *être définie de manière objective* : Les règles opératoires de détermination de la caractéristique doivent donc être suffisamment précises pour que deux personnes différentes parviennent au même résultat.
2. *dépendre de toutes les observations* : si on fait varier une observation la statistique doit rendre compte de cette variation. La question se pose toutefois de savoir si l'on doit tenir compte d'une observation très différente des autres ou si l'on doit l'écarter, considérant que sa valeur exceptionnelle est due à un facteur quelconque étranger au reste de la série comme une erreur d'observation par exemple.
3. *avoir une signification concrète* : la caractéristique ne doit pas avoir un caractère mathématique trop abstrait pour être sujet à une interprétation simple et immédiate. ceci est d'autant plus indispensable si les résultats sont présentés à des personnes non familiarisées avec la technique statistique.

4. *être simple à calculer* : il arrive que l'on préfère des caractéristiques moins efficace que d'autres, mais d'un calcul plus rapide ou plus commode. Cependant à notre époque, vu la performance de calcul des ordinateurs, cette préoccupation est minime.
5. *se prêter facilement au calcul algébrique* : cette condition est très importante. Si, par exemple, on dispose de la distribution des salaires dans plusieurs établissements composant une même entreprise, la tendance centrale des salaires de l'entreprise doit pouvoir être exprimée simplement en fonction de celle de chacun des établissements. De ce point de vue, la moyenne arithmétique possède des propriétés plus simples que la médiane. En effet, il sera plus facile de trouver le salaire moyen de l'entreprise en connaissant les effectifs et les salaires moyens de chacun des établissements que de trouver le salaire médian dans cette entreprise.
6. *être peu sensible aux fluctuations d'échantillonnage* : cette condition est essentielle lorsqu'il s'agit d'informations collectées par sondage sur un échantillon de la population étudiée. Ce que l'on recherche alors, à travers les résultats de cette observation partielle, ce sont les caractéristiques de l'ensemble.

En réalité très peu de statistiques satisfont toutes les conditions. Cela vient en parti du fait que certaines d'entre elles sont incompatibles (par exple 2. et 6.). Il est cependant utile de connaitre leur limite afin de se prémunir d'emblée contre des interprétations hasardeuses.

2 Les données

2.1 Données observées

On parle de données observées lorsqu'on étudie les données réelles de chaque individu.
Dans le cas de variables discrètes, il est fréquent que le nombre de modalités différentes que prend la variable X soit faible, disons p. Il est alors possible de présenter les donnèes sous la forme du Tableau 7.1 où x_i est la $i^{\text{ème}}$ modalité de la variable X et N_i le nombre d'individus ayant la modalités x_i de X. Il vient immédiatement que dans ce cas, la moyenne se transforme en une *moyenne pondérée* s'écrivant sous la forme :

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{p} N_i x_i \quad \text{avec} \sum_{i=1}^{p} N_i = N.$$

Modalités	Effectifs
x_1	N_1
x_2	N_2
$\vdots$	$\vdots$
x_i	N_i
$\vdots$	$\vdots$
x_{p-1}	N_{p-1}
x_p	N_p

Table 7.1 – Cas de données observées.

Attention *Dans ce cours X_j représente la valeur du $j^{\text{ème}}$ individu pour le caractère étudié X et x_i est l'une modalités (i.e. valeur possible) du caractère X. De manière général, le $i^{\text{ème}}$ individu n'a aucun raison d'avoir la modalité x_i.*

Exemple 2.1 Age des élèves préparant le Capes à Rennes 1 en 2007/2008 :

Age	Effectifs
20	2
21	12
22	20
23	11
24	7
25	1
26	4
27	3
29	2
31	1
39	1
49	1

2.2 Données groupées

Par opposition aux données observées, on introduit la notion de données groupées (ou classées). On ne connait pas pour chaque individu i la valeur mais juste un intervalle dans lequel elle se trouve. C'est en général le cas lorsque la variable étudiée est continue.
Soit K le nombre d'intervalles qui découpe la plage de variation de la variable X et $\{e_{k-1}, e_k\}$ les bornes de l'intervalle k, les données se présentent sous la forme du Tableau 7.2

Classes	Effectifs	Centre de classes : c_k
e_0 à e_1	N_1	$\frac{e_0+e_1}{2}$
e_1 à e_2	N_2	$\frac{e_1+e_2}{2}$
$\vdots$	$\vdots$	$\vdots$
e_{i-1} à e_i	N_i	$\frac{e_{i-1}+e_i}{2}$
$\vdots$	$\vdots$	$\vdots$
e_{K-2} à e_{K-1}	N_{p-1}	$\frac{e_{K-2}+e_{K-1}}{2}$
e_{K-1} à e_K	N_p	$\frac{e_{K-1}+e_K}{2}$

Table 7.2 – Cas de données groupées.

Exemple 2.2 Taille des éléves préparant le Capes à Rennes 1 en 2007/2008 :

Taille	Effectifs	Centre de classes : c_k
1m50à 1m55	2	152.5
1m55 à 1m60	5	157.5
1m60 à 1m65	11	162.5
1m65 à 1m70	16	167.5
1m70 à 1m75	9	172.5
1m75 à 1m80	5	177.5
1m80 à 1m85	13	182.5
1m85 à 1m90	3	187.5
1m90 à 1m95	1	192.5

3 Les caractéristiques à tendance centrale

Les caractéristiques à tendance centrale visent à résumer les N valeurs X_i d'un échantillon en une seule quantité. La plus connue est la **moyenne** (MEAN), on ne regardera ici que la moyenne arithmétique.

3.1 La moyenne arithmétique

Définition 3.1 On appelle moyenne arithmétique (ou moyenne empirique), notée $\overline{X}$ de l'échantillon $\{X_i\}_{i=1,\ldots,N}$ la quantité suivante :

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Cette moyenne représente la valeur qui est la plus proche de l'ensemble des observations $\{X_i\}_{i=1,\ldots,N}$, au sens de la distance euclidienne.
Dans le cas de données groupées, la *vraie* moyenne $\overline{X}$ est impossible à connaitre et on l'approche par la quantité

$$\overline{X} \simeq \frac{1}{N}\sum_{i=1}^{K} N_i c_i \quad \text{avec } \sum_{i=1}^{K} N_i = N$$

où N_i est le nombre d'individu dans la classe i, c'est à dire dans l'intervalle $[e_{i-1}, e_i[$ et c_i est le centre de l'intervalle $[e_{i-1}, e_i[$.
Cette approximation est d'autant meilleure que le nombre de classes K est grand.

Remarque 3.2 La moyenne arithmétique a la qualité d'être facile à calculer et simple d'interprétation. En revanche, elle est extrêment sensible aux valeurs extrêmes. Un des indicateurs concurrents de la moyenne le plus utilisé est la **médiane**. L'une des approches classiques pour la définir de manière analytique est via la fonction de répartition.

3.2 La fonction de répartition et les quantiles

La fonction de répartition

Définition 3.3 On appelle **fonction de répartition** de l'échantillon (*la distribution*) $(X_i)_{i\in\{1,\ldots,N\}}$ la fonction F telle que

$$F : \begin{cases} \mathbf{R} & \to \quad [0,1] \\ x & \to \quad F(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{X_i \le x} \end{cases}$$

La quantité $F(x)$ correspond à la proportion de la population pour laquelle la variable est inférieure ou égale à x.

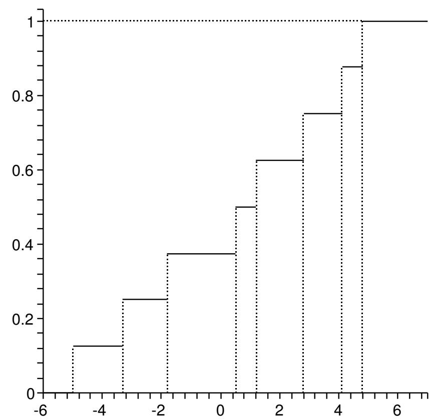


Figure 7.1 – Fonction de répartition de l'échantillon $\{-5, 1.2, -3.3, 4.8, 4.1, -1.8, 0.5, 2.8\}$

Propriétés 3.4

1. La fonction de répartition est une fonction croissante en escalier. Dans le cas où les X_i sont distincts, la hauteur des sauts est $\frac{1}{N}$.
2. Pour $x < \min(X_i)$, $F(x) = 0$ *(pour x assez petit)*
3. Pour $x \geq \max(X_i)$, $F(x) = 1$ *(pour x assez grand)*

Les quantiles

Définition 3.5 On appelle **fractile** ou bien **quantile** d'ordre α une solution X_α de l'équation : $F(x) = \alpha$.

Les quantiles les plus utilisés sont sont :

1. **la médiane** correspondant à l'équation $F(med) = \frac{1}{2}$
2. **les quartiles** correspondant aux équations $F(Q_k) = \frac{k}{4}$ pour $k \in \{1, 2, 3\}$
3. **les déciles** correspondant aux équations $F(d_k) = \frac{k}{10}$ pour $k \in \{1, \ldots, 9\}$
4. **les centiles** correspondant aux équations $F(c_k) = \frac{k}{100}$ pour $k \in \{1, \ldots, 99\}$

La médiane correspond à la valeur de la variable qui partage la population en deux sous populations d'effectifs égaux. Cette quantité est par conséquent moins sensible que la moyenne aux valeurs extrêmes. *(Il y a autant d'individus dont la valeur de la modalité est supérieure à* med *que d'individus dont la valeur de la modalité est inférieure à* med.*)*

Remarque 3.6 La valeur de la médiane dépend de la parité de N. On ordonne les valeurs $X_1, \ldots, X_N$ de la façon suivante : $X_{(1)} \leq \cdots \leq X_{(N)}$ (on a $X_{(1)} = \min(X_i)$ et $X_{(N)} = \sup(X_i)$).

- si N est impair : $N = 2p+1$ la médiane correspondra à la $p+1^{\text{ème}}$ plus grande valeur de la série notée $X_{(p+1)}$
- si N est pair : $N = 2p$ la médiane sera définie comme $med = \frac{X_{(p)}+X_{(p+1)}}{2}$.

Cas d'une distribution observée

Dans le cas de distribution observée, lorsque les différentes modalités prises par le caractère sont faibles, disons p, on regroupe les données où x_i représente la $i^{\text{ème}}$ modalité du caractère X et N_i le nombre d'individus ayant la modalités x_i de X.
On définit la **fréquence** de la $i^{\text{ième}}$ modalité par :

$$f_i = \frac{N_i}{N}.$$

La fonction de répartition s'écrit alors

$$F(x) = \frac{1}{N}\sum_{i=1}^{p} N_i \mathbb{I}_{x_i \leq x} = \sum_{i=1}^{p} f_i \mathbb{I}_{x_i \leq x}$$

En ordonnant les modalités, i.e. $x_1 < x_2 < \cdots < x_p$, on écrit ces données sous forme de tableau :

Modalitès	Effectifs	Fréquence	Fréquence cumulée
x_1	N_1	f_1	$F_1 = f_1$
x_2	N_2	f_2	$F_2 = f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	N_i	f_i	$F_i = f_1 + f_2 \cdots + f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{p-1}	N_{p-1}	f_{p-1}	$F_1 = f_1 + f_2 \cdots + f_{p-1}$
x_p	N_p	f_1	$F_p = f_1 + f_2 \cdots + f_p = 1$

Table 7.3 – Cas d'une distribution observée.

S'il existe i tel que $F_i = 1/2$ alors la médiane est $\frac{x_i + x_{i+1}}{2}$.

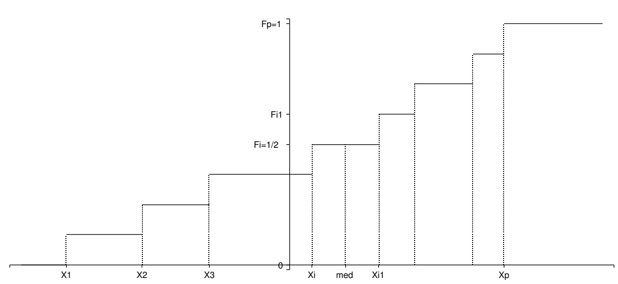


Figure 7.2 – 1er exemple de médiane dans le cas d'une distribution observée.

S'il existe i tel que $F_{i-1} < 1/2 < F_i$ alors la médiane est x_i.

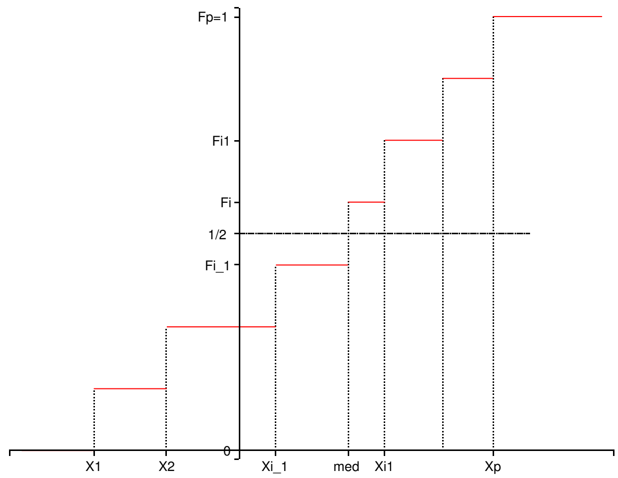


Figure 7.3 – 2ème exemple de médiane dans le cas d'une distribution observée.

Cas d'une distribution groupée

On procède de la même façon pour les distributions groupées.

Classes	Effectifs	Centre de classes : c_k	Fréquence	Fréquence cumulée
e_0 à e_1	N_1	$\frac{e_0+e_1}{2}$	f_1	$F_1 = f_1$
e_1 à e_2	N_2	$\frac{e_1+e_2}{2}$	f_2	$F_2 = f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_{i-1} à e_i	N_i	$\frac{e_{i-1}+e_i}{2}$	$f_i = \frac{N_i}{N}$	$F_i = f_1 + f_2 \cdots + f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_{K-2} à e_{K-1}	N_{p-1}	$\frac{e_{K-2}+e_{K-1}}{2}$	f_{p-1}	$F_1 = f_1 + f_2 \cdots + f_{p-1}$
e_{K-1} à e_K	N_p	$\frac{e_{K-1}+e_K}{2}$	f_p	$F_p = f_1 + f_2 \cdots + f_p = 1$

Table 7.4 – Cas d'une distribution groupée.

On approche par interpolation linéaire la fonction de répartition :

$$F(x) \simeq F_{k-1} + f_k \frac{x - e_{k-1}}{e_k - e_{k-1}} \quad \text{pour } x \in [e_{k-1}, e_k[.$$

On définit la médiane de la façon suivante : on cherche l'indice k_0 tel que $F_{k_0-1} < 1/2$ et $F_{k_0} > 1/2$. Par interpolation linéaire on obtient :

$$med = e_{k_0-1} + \frac{0.5 - F_{k_0-1}}{F_{k_0} - F_{k_0-1}}(e_{k_0} - e_{k_0-1}).$$

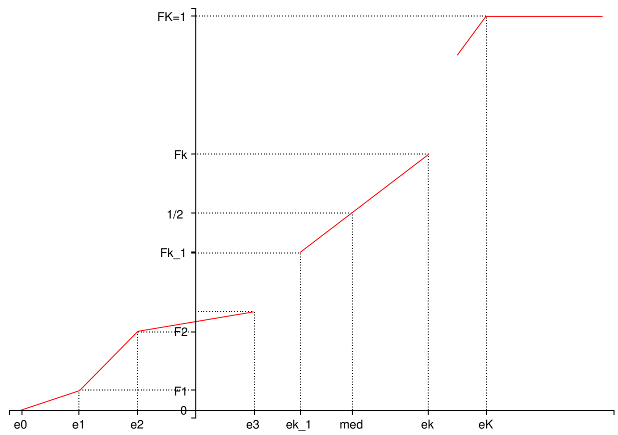


Figure 7.4 – Fonction de répartition dans le cas d'une distribution groupée.

3.3 Le mode

Le mode est défini comme étant la modalité du caractère la plus souvent prise dans la population. On s'aperçoit immédiatement des limites d'un tel indicateur :

1. il n'a de sens que dans le cas d'un faible nombre de modalité,
2. il peut exister plusieurs mode.

4 Les caractéristiques de dispersions

Les caractéristiques centrales visent à résumer l'information contenue dans l'échantillon $(X_i)_{i\in\{1,\ldots,N\}}$ par une seule valeur. Ce résumé ne permet pas cependant de comparer deux échantillons différents.
Par exemple, $\{499, 500, 501\}$ et $\{0{,}500{,}1000\}$ ont la même moyenne, la même médiane, pourtant ils sont sensiblement différents. Dans la première série les chiffres sont proches les uns des autres, alors que dans la seconde c'est le contraire.
On perçoit intuitivement que la notion de dispersion est liée à la notion d'écart à une tendance centrale. La première série de chiffres de notre exemple est moins dispersée car ses valeurs fluctuent moins autour de la moyenne. Il est naturel d'introduire les statistiques de dispersions comme des fonctions des écarts à la moyenne.
La statistique la plus simple pourrait être la moyenne des écarts à la moyenne, soit

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X}).$$

Seulement cette expression est nulle, ce qui apporte peut d'information sur la dispersion !!
L'idée est de rendre positif les termes des écarts à la moyenne et on considérera plutôt des statistiques du genre :

$$\frac{1}{N}\sum_{i=1}^{N}\phi(X_i - \overline{X}) \quad \text{avec } \phi \text{ une fonction à valeurs positives.}$$

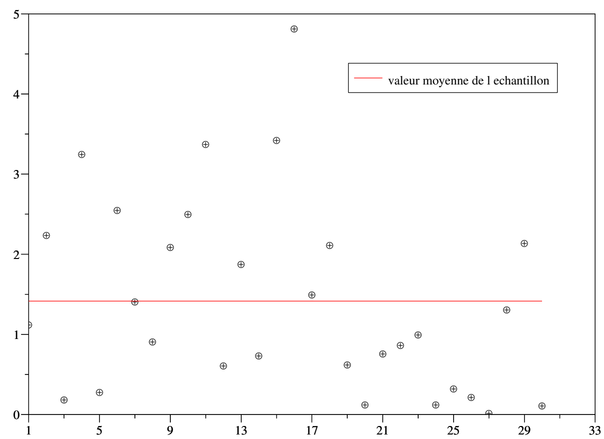


FIGURE 7.5 – Position des points d'un échantillon par rapport à sa moyenne

4.1 La variance et ses dérivés

Définition 4.1 La **variance** des données $(X_i)_{i\in\{1,\dots,N\}}$ est la quantité :

$$V(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2.$$

Propriétés 4.2
1. $V(X) \geq 0$,
2. $V(X) = 0$ ssi les données sont constantes : $\forall i : X_i = \text{constante} = \overline{X}$,
3. pour tout réel a, $V(aX) = a^2V(X)$,
4. $V(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2 = \frac{1}{N}\sum_{i=1}^{N}X_i^2 - \overline{X}^2$.

Les dérivés de la variance

Définition 4.3 L'**écart-type** est défini par

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2}.$$

L'écart-type et la variance sont deux statistiques permettant de mesurer la dispersion d'une distribution de données. Quand on dispose de deux distributions ces indicateurs permettent également de comparer les dispersions à condition que les statistiques de tendance centrale soient sensiblement les mêmes, sinon elles sont à utiliser avec précautions.
Par exemple, les séries suivantes $\{9, 10, 11\}$ et $\{9999, 10000, 10001\}$ ont la même variance et le même écart-type. Mais intuitivement on aura tendance à dire que la seconde est moins dispersée que la première.
Il faut donc comparer la dispersion absolue, mesurée par la variance et l'écart-type, à la localisation (la grandeur) des objets considérés. On introduit la dispersion relative mesurée

par le **coefficient de variation** noté $CV(X)$ et défini par :

$$CV(X) = \frac{\sigma(X)}{\overline{X}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2}}{\frac{1}{N}\sum_{i=1}^{N} X_i}$$

Le coefficient de variation est une grandeur sans unité.
On adapte facilement ces formules au cas des distributions observées et et à celui des données groupées. On obtient :

- Données observées $V(X) = \frac{1}{N}\sum_{i=1}^{p} N_i(x_i - \overline{X})^2$
- Données groupées $V(X) = \frac{1}{N}\sum_{k=1}^{K} N_k(c_k - \overline{X_K})^2$ avec $\overline{X_K} = \frac{1}{N}\sum_{k=1}^{K} N_k c_k \simeq \overline{X}$.

La variance et ses dérivés sont des statistiques peu sensibles aux fluctuations d'échantillonnage et/ou aux valeurs extrêmes.

Exemple 4.4 Le tableau suivant rend compte de la variation des indicateurs calculés sur la population des communes fançaises.

Statistiques	Avec Paris (a)	Sans Paris (b)	$\frac{a}{b}$
Variance	194 136 445	70 766 343	2.47
Ecart-type	13 933	8 412	1.65
Coefficient de variation	870.81	545.56	1.56

Une solution pour avoir des indicateurs de dispersions encore moins sensibles aux variations d'échantillonnage est d'utiliser des fonctions ϕ moins sensibles ou bien de construire des indicateurs à partir des quantiles.

4.2 L'écart absolu moyen, l'intervalle interquartile, l'étendue

L'**écart absolu moyen** est défini par

$$EAM = \frac{1}{N}\sum_{i=1}^{N}|X_i - \overline{X}|.$$

Il est peu utilisé en général car il se prète assez mal au calcul algébrique.
On regarde plutôt l'**écart interquartile** (**Interquartile range**) :

$$IQ = Q_3 - Q_1$$

Il s'agit de l'étendue qui sépare les premier et troisème quartiles.
Dans la population des communes françaises, il vaut 752. Ce qui veut dire que 50% des communes française ont une différence de population inférieure à 752 habitants.
L'**étendue** est la plus grande valeur moins la plus petite valeur d'une distribution :

$$\max_{i\in\{1,\ldots,N\}}(X_i) - \min_{i\in\{1,\ldots,N\}}(X_i)$$

Cette statistique est évidemment très sensible aux extrêmes.

5 Les représentations graphiques

5.1 Le box-plot

Il est possible de résumer une partie des statistiques précédentes sous forme d'un seul graphique : le Box Plot ou la boite à moustaches ou encore la boite à pattes.
Le box-plot est une représentation adaptée aux variables continues.

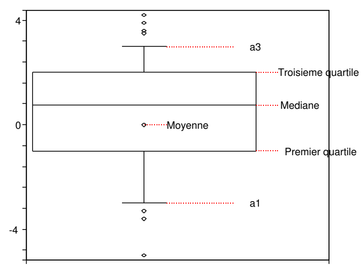


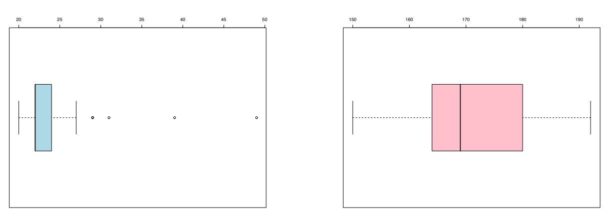
Figure 4. Exemple de boite à moustache

Les moustaches sont définies par

- a_1 est la plus petite valeur supérieure à $Q_1 - 1.5 \times IQ$
- a_3 est la plus grande valeur inférieure à $Q_3 + 1.5 \times IQ$,

où Q_1 est le premier quartile, Q_3 le troisième quartile et $IQ = Q_3 - Q_1$ l'intervalle interquartile.
Les valeurs en dehors de ses bornes sont des valeurs "extrêmes" qui sont représentée par des points.

Exemple 5.1 Étude des étudiants préparant le capes à Rennes 1 en 2007/2008 :

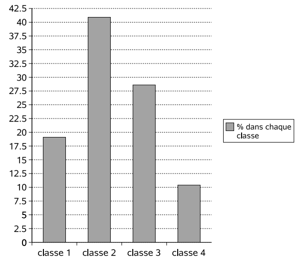


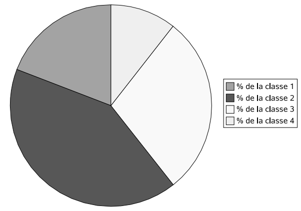
Age Taille

En général, lorsque la moyenne est supérieure à la médiane c'est le signe que la distribution est étalée vers la droite (et inversement).

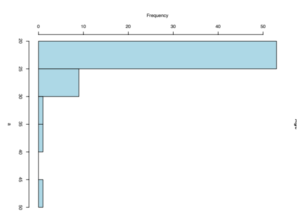
5.2 L'histogramme et le camembert

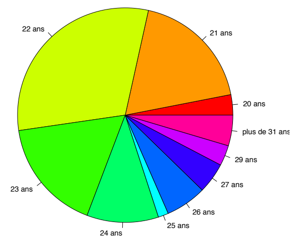
Pour les variables discrètes, on utilisera plutôt l'histogramme (ou diagramme en tuyaux d'orgues) ou le camembert.
Dans le cas discret :
Pour le diagramme en tuyaux d'orgues, la hauteur des barres est proportionnelle à la fréquence de la modalité.
Dans le camembert, c'est la surface du secteur alloué à la modalité qui est proportionnelle à la fréquence.





Exemple 5.2 Age des étudiants préparant le Capes à Rennes 1 en 2007/2008 :





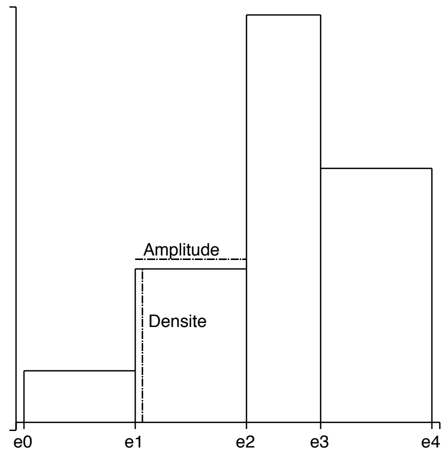
<u>Dans le cas continu</u> :
L'adaptation du diagramme en tuyaux d'orgues à une variable continue se fait dans le cadre du regroupement de données par classe. On définit l'amplitude a_i de la $i^{\text{ième}}$ par :

$$a_i = e_i - e_{i-1}.$$

On représente alors la densité défini comme la fréquence divisée par l'amplitude :

Classes	Effectifs	Centre de classes : c_k	Fréquence	Amplitude	Densité
e_0 à e_1	N_1	$\frac{e_0+e_1}{2}$	f_1	$a_1 = e_1 - e_0$	$d_1 = \frac{f_1}{a_1}$
e_1 à e_2	N_2	$\frac{e_1+e_2}{2}$	f_2	$a_2 = e_2 - e_1$	$d_2 = \frac{f_2}{a_2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_{i-1} à e_i	N_i	$\frac{e_{i-1}+e_i}{2}$	$f_i = \frac{N_i}{N}$	$a_i = e_i - e_{i-1}$	$d_i = \frac{f_i}{a_i}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_{K-2} à e_{K-1}	N_{p-1}	$\frac{e_{K-2}+e_{K-1}}{2}$	f_{p-1}	$a_{p-1} = e_{p-1} - e_{p-2}$	$d_{p-1} = \frac{f_{p-1}}{a_{p-1}}$
e_{K-1} à e_K	N_p	$\frac{e_{K-1}+e_K}{2}$	f_p	$a_p = e_p - e_{p-1}$	$d_p = \frac{f_p}{a_p}$

Table 7.5 – Construction d'un histogramme pour une distribution groupée.


Amplitude
Densite
e0
e1
e2
e3
e4

Chapitre 8

Distribution statistique à deux caractères

Lorsque l'on étudie deux caractères simultanément, notre but est d'évaluer le lien entre les caractères, leur dépendance. On peut étudier par exemple, l'age et la taille, le chiffre d'affaire et la production, le niveau d'étude et le salaire, le cours de la bourse et la météo. Les deux caractères ne sont pas forcément de même nature (qualitatif, quantitatif). Selon la nature des caractères on n'utilisera pas les mêmes techniques pour quantifier la dépendance.

1 Présentation dans le cas discret : Les tableaux statistiques

Considérons une population de N individus décrits simultanéments suivant deux caractères X et Y.

En regroupant les individus selon les valeurs des modalités des caractères, on peut résumer les données sous la forme d'un tableau statistique à double entrée (en lignes : les modalités de X, et en colonne : les modalités de Y).
Désignons par $x_1, \ldots, x_p$ les p modalités du caractères X,
et par $y_1, \ldots, y_q$ les q modalités du caractères Y.
Soit N_{ij} le nombre d'individus de la population présentant à la fois la modalité x_i du caractère X et la modalité y_j du caractère Y.
Du fait que les modalités sont incompatibles et exhaustives, la somme des N_{ij} est égale à la taille de la population

$$\sum_{i=1}^{p} \sum_{j=1}^{q} N_{ij} = N$$

On note de la façon suivante les effectifs marginaux :

- effectifs de la population ayant la modalité x_i du caractère X :

$$N_{i\bullet} = \sum_{j=1}^{q} N_{ij}$$

- effectifs de la population ayant la modalité x_j du caractère Y :

$$N_{\bullet j} = \sum_{i=1}^{p} N_{ij}$$

Modalités du caractère X	Modalités du caractère Y						Distribution marginale de X
	y_1	y_3	$\dots$	y_j	$\dots$	y_q	
x_1	N_{11}	N_{12}	$\dots$	N_{1j}	$\dots$	N_{1q}	$N_{1\bullet}$
x_2	N_{21}	N_{22}	$\dots$	N_{2j}	$\dots$	N_{2q}	$N_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	N_{i1}	N_{i2}	$\dots$	N_{ij}	$\dots$	N_{iq}	$N_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_p	N_{p1}	N_{p2}	$\dots$	N_{pj}	$\dots$	N_{pq}	$N_{p\bullet}$
Distribution de Y	$N_{\bullet 1}$	$N_{\bullet 2}$	$\dots$	$N_{\bullet j}$	$\dots$	$N_{\bullet q}$	$N_{\bullet\bullet} = N$

Table 8.1 – Tableau statistique d'une étude simultanée de deux caractères.

On appelle fréquence du couple de modalités x_i et x_j (ou encore fréquence totale) la proportion d'individus qui présentent simultanément les modalités X_i et Y_j :

$$f_{ij} = \frac{N_{ij}}{N}$$

1.1 Distributions marginales

Les effectifs $N_{i\bullet}$ définissent ce qu'on appelle la distribution marginale selon le caractère X seul. La **fréquence marginale** de la modalité x_i est

$$f_{i\bullet} = \frac{N_{i\bullet}}{N}.$$

De même pour la caractère Y, on définit les fréquences par

$$f_{\bullet j} = \frac{N_{\bullet j}}{N}.$$

1.2 Distributions conditionnelles

Considérons les $N_{\bullet j}$ individus qui présentent la modalité y_j de Y. La j$^{\text{ème}}$ colonne du tableau 8.1 décrit cette sous population suivant le caractère X : N_{ij} individus sur $N_{\bullet j}$ présentent la modalité x_i. La fréquence conditionnelle de la modalité x_i liée à y_j (ou sachant y_j) est

$$f_i^j = \frac{N_{ij}}{N_{\bullet j}}$$

(*lire "f" "i" sachant "j"*).
De même, on définit la distribution conditionnelle sachant x_i, les fréquences sont :

$$f_j^i = \frac{N_{ij}}{N_{i\bullet}}.$$

Remarque 1.1 $f_{ij} = \frac{N_{ij}}{N} = f_{i\bullet} f_j^i = f_{\bullet j} f_i^j$.

2 Indépendance et liaison fonctionnelle dans le cas discret

2.1 Indépendance

On dit que le caractère X est indépendant du caractère Y si les distributions conditionnelles $(X|y_j)$ sont identiques entres elles : f_i^j ne dépend pas de j. Dans ce cas les distributions conditionnelles $(X|y_j)$ sont identiques à la distribution de X : $f_i^j = f_{i\bullet}$ pour tout j.
Par conséquent les colonnes du tableau 8.1 sont proportionelles entre elles.
L'indépendance de X par raport à Y implique l'indépendance de Y par rapport à X. Par conséquent les lignes du tableau 8.1 sont elles aussi proportionnelles.
Lorsque les caractères X et Y sont indépendants, le fait de savoir qu'un individu possède la modalité x_i du caractère X ne fournit aucune information sur la valeur du caractère Y.

Exemple 2.1 Exemple de caractères indépendants

Modalités du caractère X	Modalités du caractère Y			
	y_1	y_3	y_3	y_4
x_1	3	5	2	4
x_2	6	10	4	8
x_3	12	20	8	16

2.2 Liaison fonctionnelle

On dit que le caractère Y est lié fonctionnellement au caractère X si à chaque modalité X_i de X correspond une seule modalité possible de Y (dans chaque ligne un seul terme est non nul). *Il existe une fonction φ telle que* $Y = \varphi(X)$. Par conséquent connaitre la valeur du caractère X d'un individu permet de connaitre la valeur du caractère Y.
Inversement le caractère X est lié fonctionnellement au caractère Y si à chaque modalité Y_j de Y correspond une seule modalité possible de X (dans chaque colonne un seul terme est non nul).
Il n'y a pas équivalence entre les deux notions.

Exemple 2.2 Exemple où X est lié fonctionnellement à Y, mais Y n'est pas lié fonctionnellement à X.

Modalités du caractère X	Modalités du caractère Y				
	y_1	y_3	y_3	y_4	y_5
x_1	4	0	7	0	0
x_2	0	6	0	0	2
x_3	0	0	0	9	0

2.3 Cas général

L'indépendance et la liaison fonctionnelle sont deux cas extrêmes que l'on rencontre rarement à l'état pur dans la pratique. Notre but va être de mesurer l'intensité de la liaison entre deux caractères X et Y.

2.4 Les indices de liaison : Le Chi-deux et ses dérivés

Cette partie ne concerne pas les élèves préparant le capes.
Le Chi-deux permet de comparer le tableau des effectifs relevés à ce qu'il aurait du être si les caractères avaient été indépendants.

Définition 2.3 Le Chi-deux est défini par

$$\chi^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(N_{ij} - \frac{N_{i\bullet} N_{\bullet j}}{N})^2}{\frac{N_{i\bullet} N_{\bullet j}}{N}} = N \left[\sum_{i=1}^{p} \sum_{j=1}^{q} \frac{N_{ij}^2}{N_{i\bullet} N_{\bullet j}} - 1 \right]$$

Propriétés 2.4
1. Les caractères X et Y sont indépendants si et seulement si $\chi^2 = 0$.
2. $\chi^2 \geq 0$. Il est d'autant plus grand que la liaison entre X et Y est forte.

Le problème du Chi-deux est qu'il dépend de la taille de la population N et des nombres de modalités p et q. Que signifie grand dans ce cas ? Pour palier à ce défaut on introduit de nouvelles quantités plus robustes.

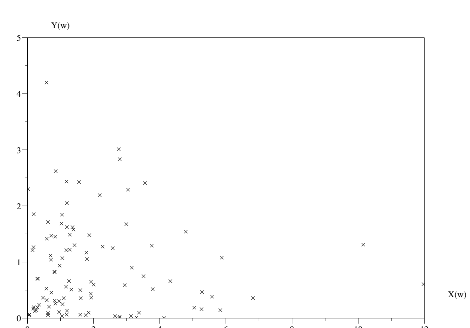
Autres indicateurs liés au Chi-deux :
1. Phi-deux : $\Phi^2 = \frac{\chi^2}{N}$ (dépend encore de p et q !)
2. Le coefficient de Tschuprow : $T = \sqrt{\frac{\Phi^2}{\sqrt{(p-1)(q-1)}}}$. On a $0 \leq T \leq 1$.
3. Le coefficient de Cramer : $C = \sqrt{\frac{\Phi^2}{d-1}}$, avec $d = \min(p, q)$. On a $0 \leq T \leq C \leq 1$.

Toutes les notions que l'on a introduites jusqu'à présent sont bien définies quelque soit la nature des caractères. On va maintenant voir des notions spécifiques selon la nature des caractères.

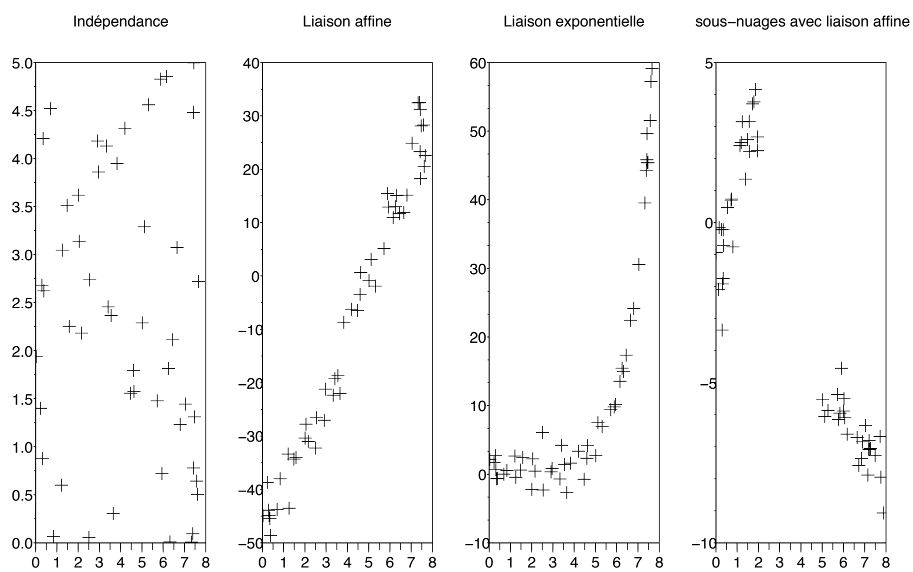
3 Description de distribution à deux caractères quantitatifs

3.1 Représentation graphique : le nuage de points

On considère une population de N individus décrits simultanéments suivant deux caractères X et Y. On trace sur un graphique l'ensemble des points de coordonnées (X_k, Y_k) correspondant à chacun des individus, $k \in \{1, \ldots, N\}$: (X_k, Y_k) est la valeur du $k^{\text{ième}}$ individu concernant les caractère X et Y.



Lorsque deux ou plusieurs points sont confondus, soit on trace un point dont la taille est proportionnelle à sa fréquence, soit on indique à coté de chaque point sa fréquence d'apparition.
L'ensemble de ces points donne en général une assez bonne idée de la variation conjointe des deux variables et est appelé **nuage** ou **diagramme de dispersion** (scatter-plot).
On souhaite à partir de ce nuage trouver un lein entre les variables X et Y. Plus le nuage est dispersé, plus on peut penser que les variables sont indépendantes. Voici quelques exemples de nuages



Remarque 3.1 Le choix des échelles est délicat. Dans le cas de deux variables homogènes (représentant la même grandeur exprimée dans la même unité), on choisira la même échelle sur les deux axes. Dans le cas de deux variables hétérogènes, soit on représentera la variables centrées réduites $(\frac{X-\overline{X}}{\sigma(X)}, \frac{Y-\overline{Y}}{\sigma(Y)})$, soit on choisira des échelles adéquates.

3.2 Les caractéristiques des distributions marginales

Les caractéristiques marginales de X :

$$\begin{aligned}
\overline{X} &= \frac{1}{N}\sum_{k=1}^{N} X_k \overset{\text{cas discret}}{=} \sum_{i=1}^{p} f_{i\bullet} x_i, \\
V(X) &= \frac{1}{N}\sum_{k=1}^{N} (X_k-\overline{X})^2 \overset{\text{cas discret}}{=} \sum_{i=1}^{p} f_{i\bullet} (x_i-\overline{X})^2.
\end{aligned}$$

Écart-type : $\sigma(X) = \sqrt{Var(X)}$.

On peut de même calculer les caractéristiques marginales de Y.

3.3 La covariance et le coefficient de corrélation linéaire

L'objectif est de définir un indice rendant compte numériquement de la manière dont les deux variables considérées varient simultanément. C'est ce qu'on appelera la corrélation linéaire.

Définition 3.2 La covariance est définie par

$$\begin{aligned} Cov(X,Y) = c_{X,Y} &= \frac{1}{N}\sum_{k=1}^{N}(X_k - \overline{X})(Y_k - \overline{Y}) \\ &= \frac{1}{N}\sum_{k=1}^{N} X_k Y_k - \overline{X}\,\overline{Y} \\ &\overset{\text{cas discret}}{=} \sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij} x_i y_j - \overline{X}\,\overline{Y} \end{aligned}$$

Propriétés 3.3
1. la covariance est symétrique : $Cov(X,Y) = Cov(Y,X)$,
2. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$,
3. $Cov(X,Y)^2 \leq Var(X)Var(Y)$.

Le problème avec la covariance c'est qu'elle dépend des unités de mesure. On définit par conséquent la corrélation en normalisant la covariance.

Définition 3.4 Le coefficient de corrélation linéaire est défini par

$$Corr(X,Y) = r_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}.$$

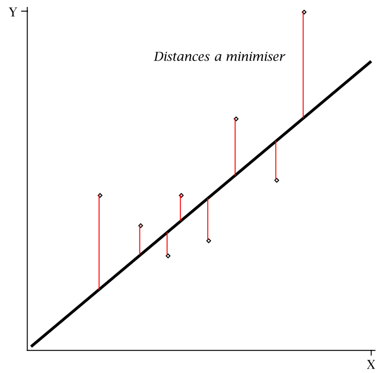
Propriétés 3.5
1. $Corr(X,Y) = Cov(\frac{X-\overline{X}}{\sigma_X}, \frac{Y-\overline{Y}}{\sigma_Y})$,
2. la corrélation linéaire est symétrique : $r_{X,Y} = r_{Y,X}$,
3. $r_{X,Y} \in [-1,1]$
4. $|r_{X,Y}| = 1$ si et seulement si il existe une liaison linéaire entre X et Y ($\exists a,b,c \in \mathbf{R}$ tels que $aX + bY + c = 0$),
5. Si X et Y sont indépendantes alors $r_{X,Y} = 0$.

3.4 Regression linéaire

Quand deux variables sont très corrélées ($|r_{X,Y}|$ proche de 1), il est naturel de penser que l'une des variables influe sur l'autre (par exemple X est cause de Y) et donc on souhaite trouver la fonction de X approchant "le mieux possible" Y. La régression linéaire consiste à chercher des fonctions affines (du type $aX + b$).
Connaissant les valeurs de X, on souhaite trouver la droite $y = ax+b$ qui approche le mieux le nuage de points, c'est à dire trouver les coefficients a et b qui minimisent la distance entre le nuage de points et la droite d'équation $y = ax+b$ au sens de la distance euclidienne, c'est à dire qui minimisent

$$S(a,b) = \sum_{k=1}^{N}\left[Y_k - (aX_k + b)\right]^2$$

Un dessin est parfois plus clair : les traits rouges sont les distances à minimiser

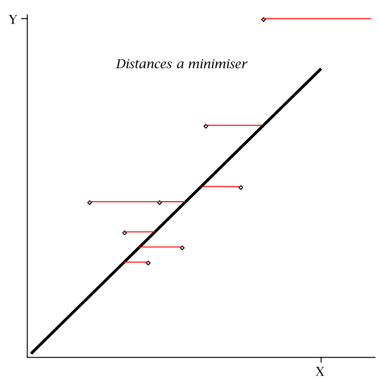


Soit en dérivant par rapport à chacune des variables, soit en séparant les variables, on obtient

$$\hat{a} = \frac{c_{X,Y}}{Var(X)} \quad \text{et} \quad \hat{b} = \overline{Y} - \hat{a}\overline{X}. \tag{8.1}$$

Propriétés 3.6
1. La droite d'équation $y = \hat{a}x + \hat{b}$ est appelée **droite de regression de Y sur X**. Elle passe par le barycentre du nuage de points ω_k de coordonnées $(\overline{X}, \overline{Y})$.
2. Les valeurs $\hat{Y}_k = \hat{a}X_k + \hat{b}$ sont appelées les **valeurs ajustées**. Elles ont la même moyenne $\overline{Y}$ que Y.
3. Les valeurs $\hat{E}_k = Y_k - \hat{Y}_k$ sont appelées les **résidus**. Ils sont de moyenne nulle et de variance $\frac{1}{N}S(\hat{a}, \hat{b})$.
4. La variable causale $X = (X_1, \ldots, X_N)$ et la variable résiduelle $\hat{E} = (\hat{E}_1, \ldots, \hat{E}_N)$ sont non corrélées : $Corr(X, \hat{E}) = 0$.

Attention Si $y = \hat{a}x + \hat{b}$ est la droite de regression de Y sur X, la droite de regression de X sur Y n'est pas $x = \frac{1}{\hat{a}}(y - \hat{b})$, ce qui confirmé par les formules (8.1). Ceci s'explique par le fait qu'on doive minimiser les distances suivantes :



Remarque 3.7
1. On peut remarquer qu'il existe une liaison entre X et Y mais qui ne soit pas affine (par exemple : exponentielle, logarithmique...). Pour cela, il suffit de faire un changement d'échelle (par exemple : logarithmique, exponentielle) afinc de pouvoir exprimer Y en fonction de $\exp(X)$ ou $\ln(X)$... etc...
2. Dans le cas, où le nuage de points n'est pas "connexe" et se décompose en plusieurs sous nuages, on étudie chaque sous nuage de façon indépendante.

4 Distribution avec un caractère quantitatif et un caractère qualitatif

Cette section est hors programme pour les étudiants préparant le capes.

4.1 Représentation graphique

Supposons que X soit une variable qualitative à p modalités et Y soit une variable quantitative de moyenne $\overline{Y}$ et de variance $Var(Y)$.
On peut faire des histogrammes parallèles ou des boites à moustaches parallèles. Chaque histogramme ou boite à moustaches est la représentatation graphique des lois conditionnelles $(Y|X = x_i)$, $i \in \{1, \ldots, p\}$.

Exemple 4.1 Une étude a été menée pour évaluer l'influence de la vitamine C sur la croissance des dents de 10 cobayes selon la quantité (trois doses ont été administrée : 0.5, 1 et 2 mg) et selon le mode de délivrance (jus d'orange ou ascide ascorbique)

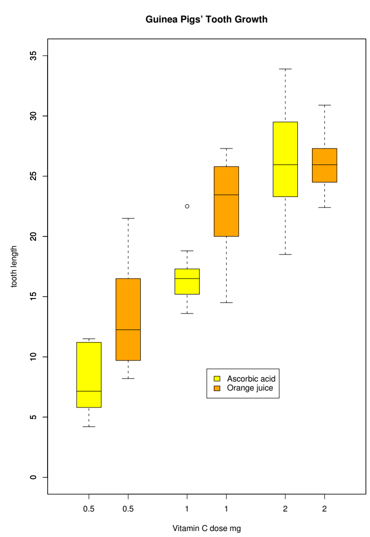


Figure 8.1 – Exemples de boites à moustaches parallèles.

4.2 Le rapport de corrélation

Caractéristiques de la variable conditionnelle $Y|X = x_i$:
On fait la moyenne des individus qui ont la modalité x_i, ce qui revient à faire la moyenne de la $j^{\text{ème}}$ colonne du tableau 8.1.
Moyenne :

$$\overline{Y}_i = \frac{1}{N_{i\bullet}} \sum_{k:X_k=x_i} Y_k \overset{\text{cas discret}}{=} \sum_{j=1}^{p} f_j^i y_j,$$

Variance :

$$V_i(Y) = \frac{1}{N_{i\bullet}} \sum_{k:X_k=x_i} (Y_k - \overline{Y}_i)^2 \overset{\text{cas discret}}{=} \sum_{j=1}^{p} f_j^i (y_j - \overline{Y}_i)^2.$$

Remarque 4.2

$$\begin{aligned}
\overline{Y} &= \sum_{i=1}^{p} f_{i\bullet} \overline{Y_i}, \\
V(Y) &= \underbrace{\sum_{i=1}^{p} f_{i\bullet} V_i(Y)}_{\sigma_R^2\text{: la Variance résiduelle}} + \underbrace{\sum_{i=1}^{p} f_{i\bullet} (\overline{Y}_i - \overline{Y})^2}_{\sigma_E^2\text{: la variance expliquée par la partition, i.e. par } X}
\end{aligned}$$

On note parfois **Variance** $=$ **Variance Intra** $+$ **Variance Inter**.

La variance est égale à la moyenne des variances plus la variance des moyennes.
La quantité $\sum_{i=1}^{p} f_{i\bullet} V_i(Y)$ est la moyenne des dispersions à l'intérieur des strates de la population définies par la population X. Il s'agit de population "intra-strates" d'où le terme de **variance intra**.
La quantité $\sum_{i=1}^{p} f_{i\bullet} (\overline{Y}_i - \overline{Y})^2$ est la dispersion des moyennes entre les strates. Il s'agit de dispersion inter-strates, d'où le terme de **variance inter**.

Définition 4.3 Le rapport de corrélation est un indice de liaison entre les deux variables X et Y :

$$s_{Y|X} = \sqrt{\frac{\sigma_E^2}{V(Y)}} = \sqrt{\frac{\text{Variance Inter}}{\text{Variance}}}.$$

Proposition 4.4

1. Le rapport de corrélation n'est pas symétrique (X et Y ne sont pas de même nature !)
2. $0 \leq s_{Y|X} \leq 1$,
3. Les variables sont indépendantes si et seulement si $s_{Y|X} = 0$.
4. La relation est parfaite entre X et Y si et seulement si $s_{Y|X} = 0$.

Chapitre 9

Conclusion

IL faut être vigilant et attentif quand travaille sur des données statistiques. Il faut être le plus exhaustif possible afin de rendre compte au mieux de la réalite. En ce que concerne les graphes et les diagrammes, l'échelle choisie est primordiale car elle peut influencer la lecture des données.
Voivi un exemple où chacun donne son interprétation. Suite à une conférence de presse de la FNAIM donnée le 5 janvier 2006, les journaux nationaux publient leur interprétation des données.

LEFIGARO·fr |Entreprises

| Entreprises |

Les prix de l'immobilier grimperont en 2006

LOGEMENT Le marché a du mal à se calmer. En 2006, les hausses seront de l'ordre de 7 à 10% sur l'ensemble de la France. Il y aura de très fortes disparités selon les villes et les arrondissements.

Odile Coupé

[06 janvier 2006]

LeMonde.fr

Immobilier : après quatre années de hausse, les prix se stabilisent
LEMONDE.FR | 05.01.06 | 19h04 " Mis à jour le 05.01.06 | 19h17

Libération.fr

Evénement

Immobilier
Propriétaire, un rêve bientôt accessible ?
Après des années de folie spéculative, le marché des biens immobiliers semble à la veille de revenir à la raison. Enquête.

Par Tonino SERAFINI
vendredi 06 janvier 2006

Le dernier mot revient au Canard Enchaîné (désolée pour la qualité) :

Le Canard enchaîné

Journal satirique paraissant le mercredi

Après son voyage à Santiago

Ségolène lance un nouveau courant :

PIERRE QUI ROULE

FAUT-IL acheter ? S'endetter pour trente ans ? Attendre que ça redescende ? Craindre que ça ne continue à monter ? Que fait la pierre ? La pierre est imprévisible. La pierre n'en fait qu'à sa tête. La preuve : s'il faut en croire « Le Figaro » (6/1), le journal des petites annonces immobilières, *« le marché a du mal à se calmer »*, et cette année ça va continuer à grimper vigoureusement de 7 % à 10 %.

> Les prix de l'immobilier grimperont en 2006

Mais « Le Monde » (7/1) l'entend autrement : certes, la pierre va encore grimpoter, mais plutôt mollo, *« du 6 % ou 7 % »*.

> Immobilier : les prix continueront à croître en 2006, mais moins fort

Quant à « Libé » (6/1), il a vu tout autre chose dans sa boule de cristal, et l'explique à la une : *« Après sept ans de hausse ininterrompue, le marché semble entrer dans une phase "charnière". A Paris et dans les métropoles régionales, les vendeurs révisent leurs prétentions à la baisse. »*

> **Immobilier Chute de pierre ?**

Cela nous rappelle que, comme disait l'autre, il est difficile de faire des prévisions, surtout concernant l'avenir.

M 00708 - 4446 - F: 1,20 €

Troisième partie

Introduction à la statistique

Chapitre 10

L'estimation

Maximum de vraisemblance à aborder !

Exemple initial Une entreprise reçoit un lot important de pièces fabriquées en série. L'entreprise n'accepte la livraison que si la proportion de pièces défectueuses est inférieure à 5%. Il est impossible d'inspecter chaque pièce, on extrait alors du lot un échantillon de 200 pièces. Sur ces 200 pièces, on dénombre 15 défectueuses. L'entreprise doit-elle accepter ce lot ?
Ce cours a pour but de répondre à cette question. Au vu du nombre de pièce défectueuses dans l'échantillon, on a va essayer d'estimer au mieux la proportion de pièces défectueuses de tout le lot ($\Rightarrow$ estimation, intervalle de confiance). Il faudra ensuite répondre à la question ($\Rightarrow$ problème de test).

1 Notion d'échantillon

Définition 1.1 Un **n-échantillon** (ou échantillon de taille n) est un n-uplet de v.a. $(X_1, ..., X_n)$ où les X_i sont indépendantes et identiquement distribuées (i.i.d.).
Un **échantillon** (ou échantillon de taille infini) est une suite de v.a. $(X_n)_{n\geq 0}$ où les X_i sont indépendantes et identiquement distribuées (i.i.d.).

Dans la suite, par abus de langage on ne parlera que d'échantillon même lorsqu'il s'agit de n-échantillon.

Concrètement
Les valeurs X_i sont le résultat d'observations ou d'un sondage. Les valeurs sont donc connues. Dans l'exemple intitial, X_i vaut par exemple 1 si la $i^{\text{ème}}$ pièce inspectée est défectueuse et 0 sinon.

Exemple 1.2

1. Sondage avec remise : On souhaite évaluer la proportion de mésanges bleues parmi la population de mésanges en Ille et Vilaine. Pour cela, on poste des observateurs à différents endroits dans le département (la mésange n'étant pas un oiseau migrateur). Chaque observateur compte le nombre de mésanges bleues ainsi que le nombre total de mésanges obervées à la jumelle. *On souhaite connaitre la proportion p de mésanges bleues.*
On associe à la variable X_i la valeur 1 si la $i^{\text{ème}}$ mésange observée est bleue et 0 si la mésange n'est pas bleue. On limite notre étude à n oiseaux (par exemple $n = 1000$

mésanges). On suppose que les X_i sont indépendants (en effet, les mésanges vivent en groupes formés de plusieurs espèces de mésanges), et de même loi de Bernoulli $\mathcal{B}(p)$.
Par conséquent *le nombre de mésanges bleues* $S_n = \sum_{i=1}^n X_i$ *sur* n *mésanges observées suit la loi Binomiale* $\mathcal{B}(n,p)$.

Chercher la mésange bleue !

2. Sondage sans remise : On reprend l'exemple initial. Une entreprise reçoit un lot important de pièces fabriquées en série. On connait le nombre total de pièces, il vaut N. L'entreprise souhaite connaître le nombre n_1 de pièces défectueuses dans ce lot.

   On extrait du lot un échantillon de n pièces choisies au hasard. On associe à la variable X_i la valeur 1 si la $i^{\text{ème}}$ pièce est défectueuses, et la valeur 0 sinon. *Le nombre de pièces défectueuses* $S_n = \sum_{i=1}^n X_i$ *parmi les* n *pièces étudiées suit la loi Hypergéométrique* $H(N, n_1, n)$:

$$P(S_n = k) = \frac{\begin{pmatrix} n_1 \\ k \end{pmatrix}\begin{pmatrix} N - n_1 \\ n - k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \quad \text{pour } k \in \{0, .., \min(n_1, n)\}.$$

Remarque 1.3 On sait que lorsque la taille de la population initiale N est très grande, on peut approcher la loi Hypergéométrique $H(N, n_1, n)$ par la loi Binomiale $\mathcal{B}(n, \theta)$, où θ est la proportion de pièces défectueuses dans tout le lot ($\theta = \frac{n_1}{N}$).
La plupart des études réalisées sur une population divisée en deux classes (il n'y a que deux réponses possible pour chaque individu) sont des sondages sans remise, on utilise pourtant la loi binomiale en se basant cette remarque.

2 Estimation Ponctuelle

On considère un échantillon $(X_i)_{1 \leq i \leq n}$ de même loi P_θ qui dépend d'un paramètre inconnu θ. Les X_i sont connus, le but est d'estimer la valeur de θ à partir des X_i.

On notera $E_\theta[Y]$ l'espérance d'une variable Y lorsque les $(X_i)_{1\leq i\leq n}$ suivent la loi P_θ. On verra quand on abordera les tests pourquoi il est important de spécifier la loi sous laquelle on travaille.

Exemple 2.1 On reprend les exemples 1.2

1. le paramètre inconnu est $\theta = p$.
2. le paramètre inconnu est $\theta = (n_1, n_2)$.

Définition 2.2 Soit $\theta \in \mathbb{R}$ inconnu. Un **estimateur** $\hat{\theta}_n$ de θ est une fonction de l'échantillon, i.e. une variable aléatoire de la forme $\hat{\theta}_n = f(X_1, X_2, ..., X_n)$.

ATTENTION : Un estimateur de θ ne doit pas dépendre de θ ! !

Evidement, il faut choisir la fonction f de façon à "bien" estimer θ.

Exemple 2.3 On reprend l'exemple des pièces défectueuses. On note θ la proportion de pièces défectueuses.
Les variables $\hat{\theta} = 10$, $\hat{\theta} = X_1$, $\hat{\theta} = \sqrt{X_1 X_2}$ sont des estimateurs de θ. Cependant, un estimateur naturel de θ est la fréquence de *succès* :

$$\hat{\theta} = \frac{X_1 + ... + X_n}{n} = \frac{\sharp\text{pièces défectueuses}}{\sharp\text{pièces étudiées}}.$$

L'entreprise ne considère qu'un échantillon de 200 pièces, par conséquent une réalisation de $\hat{\theta}$ est $\frac{15}{200} = 0.075$.

Définition 2.4 Un estimateur $\hat{\theta}_n$ de θ est dit **sans biais** si $E_\theta(\hat{\theta}_n) = \theta$.
Sinon il est dit **biaisé**, et on définit **le biais** par $b(\hat{\theta}_n, \theta) = E_\theta(\hat{\theta}_n) - \theta$.

La valeur moyenne d'un estimateur sans biais est l'inconnue θ, c'est à dire qu'il prend des valeurs autour de θ, ce qui a priori est une bonne qualité.

Définition 2.5 L'estimateur $\hat{\theta}_n$ est un **estimateur convergent** de θ si $\hat{\theta}_n \underset{n\to+\infty}{\overset{proba}{\longrightarrow}} \theta$.

Plus l'échantillon est grand, plus l'estimateur convergent est proche de l'inconnue θ.

Exemple 2.6 On reprend l'exemple initial, $\overline{X}_n$ est un estimateur sans biais de $p = E_p(X_1)$ et gràce à la loi des grands nombres, il est convergent.

2.1 Comparaison d'estimateurs

Définition 2.7 On appelle **risque quadratique** de l'estimateur $\hat{\theta}_n$ par rapport à θ :

$$R(\hat{\theta}_n, \theta) = E_\theta\left[(\hat{\theta}_n - \theta)^2\right]$$

Plus le risque est faible, plus l'estimateur est proche de l'inconnue θ en moyenne. On souhaite donc avoir un risque le plus faible possible.

Définition 2.8 Si $R(\hat{\theta}_n, \theta) \leq R(\tilde{\theta}_n, \theta)$ $\forall\theta$, on dit que l'estimateur $\hat{\theta}_n$ est meilleur que $\tilde{\theta}_n$.

Attention : on ne peut pas toujours comparer deux estimateurs !

Remarque 2.9 En développant le carré dans l'expression du risque, on a

$$R(\hat{\theta}_n, \theta) = Var_\theta(\hat{\theta}_n) + b^2(\hat{\theta}_n, \theta)$$

Par conséquent si l'estimateur est sans biais, le risque vaut $R(\hat{\theta}_n, \theta) = var_\theta(\hat{\theta}_n)$.

3 Estimateurs des moments

3.1 Moyenne empirique

On considère un échantillon $(X_1, ..., X_n)$ de même loi P_θ dont l'espérance $\theta = E_\theta(X)$ est inconnue.
La moyenne empirique est définie par : $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
La moyenne empirique est un estimateur sans biais et convergent de l'espérance $E_\theta(X)$ (grâce à la loi des grands nombres). Son risque est $R(\overline{X}_n, \theta) = \frac{Var_\theta(X_1)}{n}$.

Remarque 3.1
1. Pour un échantillon de loi de Bernoulli $\mathcal{B}(\theta)$, alors $S_n = \sum_{i=1}^n X_i$ suit la loi Binomiale $\mathcal{B}(n, \theta)$.
2. Pour un échantillon de loi normale $\mathcal{N}(\theta, \sigma^2)$, alors $\overline{X}_n$ suit une loi normale $\mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$.

3.2 Principe général

On considère un échantillon $(X_1, ..., X_n)$ de même loi P_θ. On veut estimer θ.
- Si θ s'écrit $\theta = E_\theta[f(X_1)]$, avec f continue. Alors $T = \frac{1}{n}\sum_{i=1}^n f(X_i)$ est un estimateur sans biais et convergent.
- Si θ s'écrit $\theta = g(E_\theta[X_1])$, avec g continue. Alors $R = g(\overline{X}_n)$ est un estimateur convergent, mais en général biaisé.

3.3 Exemples : variance et covariance empiriques

Comme $\sigma(X)^2 = E(X^2) - E(X)^2$, il est naturel de considérer l'estimateur

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

C'est une estimateur congervent, mais biaisé : $E[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2$.
Du coup, $S^2 = \frac{n}{n-1}\hat{\sigma}^2$ est un estimateur convergent et sans biais de la variance.

De même pour la covariance, si on considère un échantillon $(X_i, Y_i)_{i \le n}$, on utilise l'estimateur $\hat{C}_n = \frac{1}{n}\sum_{i=1}^n X_i Y_i - \overline{X}_n \overline{Y}_n$ de la covariance $Cov(X, Y) = E(XY) - E(X)E(Y)$.

Exercice 3.2 Le but de cet exercice est de montrer que la moyenne empirique n'est pas forcément le meilleur estimateur de l'espérance.
Soit $(X_1, ..X_n)$ un échantillon de la loi uniforme sur $[0, \theta]$, $\theta > 0$. On veut estimer θ.

a) Donner un estimateur de θ par la méthode des moments. Calculer son biais et son risque.

b) On considère $M = \max(X_1, ..X_n)$. Trouver la loi de M. Calculer son espérance et sa variance.

c) Déduire de ce qui précède que $\tilde{\theta}_n = \frac{n+1}{n} M$ est un estimateur sans biais de θ meilleur que la moyenne empirique.

Exemple 3.3 Retour à l'exemple initial :
Sur un échantillon de 200 pièces, l'entreprise dénombre 15 défectueuses. Elle va estimer la proportion de pièces défectueuses dans le lot en utilisant les données de son échantillon. On note X_i la variable aléatoire suivante : $X_i = 1$ si la i^{me} pièce est défectueuse et $X_i = 0$ sinon. Les pièces étant choisies au hasard, les variables X_i sont indépendantes et de même loi. Leur loi commune est la loi de Bernoulli $\mathcal{B}(\theta)$, où θ est la proportion de pièce défectueuses dans le lot entier. θ est l'inconnue.
On estime θ par $\frac{\sum_{i=1}^{200} X_i}{200}$. Dans le cas présent, une réalisation de $\hat{\theta}$ est $\frac{15}{200} = 0.075$.
La question est de savoir si cette estimation est bonne, si la vraie valeur a de grande chance d'être proche de 0.075. Pour répondre à cette question, on va introduire la notion d'intervalle de confiance.

4 Estimation par intervalle de confiance

On considère un échantillon $(X_1, ..., X_n)$ le loi $\mathbb{P}_\theta$. Le paramètre θ est inconnu.
On sait estimer l'inconnu θ de façon ponctuelle, mais il serait intéressant de connaître la *distance* entre θ et son estimateur $\hat{\theta}$ afin de savoir si l'approximation est bonne ou pas. Pour cela, on va chercher un intervalle tel que θ soit dans cet intervalle avec un certain niveau de confiance (de préférence assez fort).

4.1 Définitions

Définition 4.1 Un **intervalle de confiance** I au niveau $1-\alpha$ est un intervalle aléatoire qui dépend de l'échantillon $X_1, ..., X_n$, mais pas de θ, tel que

$$P_\theta(\theta \in I) = 1 - \alpha \quad \forall \theta.$$

Quelque soit la valeur de l'inconnu θ, on sait qu'il se trouve dans l'intervalle de confiance I avec probabilité $1-\alpha$. Le nombre α représente le taux d'erreur maximal que l'on accepte de prendre (de l'ordre en général de 1% ou 5%). L'intervalle I est connu car il ne dépend que de l'échantillon.

Comment construire des intervalles de confiance ?
Prendre un estimateur $\hat{\theta}$ raisonnable de θ. Lorsque l'on peut calculer cette probabilité, il *suffit* de trouver des réels a, b indépendants de θ tels que $P(\hat{\theta} \in [\theta - a, \theta + b]) = 1 - \alpha$. Alors $I = \left[\hat{\theta} - b, \hat{\theta} + a\right]$ est un intervalle de confiance pour θ au niveau $1-\alpha$.
Le problème est qu'il n'est pas simple en général de connaître la loi de $\hat{\theta}$ et encore moins de trouver les valeurs a et b.
L'exemple suivant utilise fortement les propriétés des lois gaussiennes, il ne se généralise malheureusement pas aux autres lois.

Exemple 4.2 On considère un échantillon $X_1, ..., X_n$ de loi $\mathcal{N}(\theta, 1)$. D'après les propriétés de la loi gaussienne, $\overline{X}_n$ suit la loi $\mathcal{N}(\theta, \frac{1}{n})$. Par conséquent, $\sqrt{n}(\overline{X}_n - \theta)$ suit la loi $\mathcal{N}(0, 1)$, donc pour $t > 0$ on a

$$P_\theta(\sqrt{n}|\overline{X}_n - \theta| \leq t) = P(|Z| \leq t) \quad \text{où } \mathcal{L}(Z) = \mathcal{N}(0, 1).$$

Pour α donné, il suffit alors de trouver t_α tel que $P(|Z| \leq t) = 1 - \alpha$ pour conclure que $I = [\overline{X}_n - \frac{t_\alpha}{\sqrt{n}}, \overline{X}_n + \frac{t_\alpha}{\sqrt{n}}]$ est un intervalle de confiance de niveau $1 - \alpha$.
Par exemple, en regardant la table gaussienne p.45, pour $\alpha = 1\%$, $t_\alpha = 2.57$ et pour $\alpha = 5\%$, $t_\alpha = 1.96$.
Posons $\alpha = 5\%$, alors $I_1 = [\overline{X}_n - 1.96/n, \overline{X}_n + 1.96/n]$ est un intervalle de confiance au niveau de confiance 95%. On se rend compte que $I_2 = [\overline{X}_n - 1.645/n, +\infty[$ est aussi un intervalle de confiance au niveau de confiance 95%. Lequel choisir ?

Propriétés 4.3 Comment comparer deux intervalles de confiance ? De manière naturelle, le meilleur est celui de longueur la plus petite.

On va maintenant introduire la notion d'intervalle de confiance asymptotique et on verra dans les exemples qu'il est plus simple de trouver des intervalles asymptotiques.

Définition 4.4 Un **intervalle de confiance asymptotique** au niveau α est un intervalle I_n qui ne dépend que de l'échantillon et tel que $\lim_{n \to +\infty} P_\theta(\theta \in I_n) = 1 - \alpha \ \forall \theta$.

Par abus de langage, on parle souvent d'intervalle de confiance même lorsqu'ils sont asymptotiques.

4.2 Exemple : estimation par intervalle de confiance de la moyenne

On considère un échantillon $(X_1, ..., X_n)$ d'espérance θ et de variance σ^2. On souhaite estimer θ.

Estimation de θ à σ^2 connu.

La moyenne empirique $\overline{X}_n$ est un estimateur sans biais et convergent de θ. De plus, d'après le théorème de la limite centrale, on a

$$\forall t > 0 \quad P\left(\frac{\sqrt{n}}{\sigma}|\overline{X}_n - \theta| \leq t\right) \underset{n \to +\infty}{\longrightarrow} P\left(|Z| \leq t\right) \quad \text{où } \mathcal{L}(Z) = \mathcal{N}(0, 1).$$

En utilisant la table gaussienne p.45, on choisit t_α tel que $P\left(|Z| \leq t_\alpha\right) = \alpha$.
Par conséquent

$$I_n = \left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} t_\alpha, \overline{X}_n + \frac{\sigma}{\sqrt{n}} t_\alpha\right]$$

est un intervalle de confiance asymptotique de l'espérance θ au niveau $1 - \alpha$, quand la variance est connue.

Estimation de θ avec σ^2 inconnu.

• Si σ s'écrit comme une fonction de θ : $\sigma = f(\theta)$ avec f continue, on peut estimer σ par $\hat{\sigma}_n = f(\overline{X}_n)$.
D'après la loi faible des grands nombres, on a $\hat{\sigma}_n \xrightarrow[n\to+\infty]{\text{Proba}} \sigma$.
• De manière générale, on peut toujours estimer la variance σ^2 par la **variance empirique** :

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

D'après la loi faible des grands nombres, on a $\hat{\sigma}_n^2 \xrightarrow[n\to+\infty]{\text{Proba}} \sigma^2$.
Comme

$$\begin{array}{rcl} \hat{\sigma}_n & \xrightarrow{proba} & \sigma \quad \text{et} \\ \frac{\sqrt{n}}{\sigma}(\overline{X}_n - m) & \xrightarrow{loi} & Z \text{ avec } Z \sim \mathcal{N}(0,1) \end{array}$$

d'après la proposition 3.7 du chapitre 5 on a

$$\frac{\sqrt{n}}{\hat{\sigma}_n}\left(\overline{X}_n - m\right) \xrightarrow{loi} Z \text{ avec } Z \sim \mathcal{N}(0,1).$$

C'est une version améliorée du théorème central limite.
On procède alors de la même façon que précédemment : on choisit t_α tel que $P\left(|Z| \leq t_\alpha\right) = \alpha$ et alors l'intervalle

$$I_n = \left[\overline{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}}t_\alpha, \overline{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}}t_\alpha\right]$$

est un intervalle de confiance asymptotique de l'espérance θ au niveau $1-\alpha$ quand la variance est inconnue.

Remarque 4.5 Dans certains cas, la variance s'exprime en fonction de l'espérance, on n'est pas alors obligé d'utiliser la variance empirique comme estimateur de la variance.

Exemple 4.6 Soit $(X_1, ..., X_n)$ est un échantillon de loi $\mathcal{E}(\lambda)$, avec λ inconnu. On remarque que $\theta = 1/\lambda$ et $\sigma = 1/\lambda$.
En utilisant le théorème de la limite centrale, pour n assez grand, on a $P\left(|\overline{X}_n - \frac{1}{\lambda}| \leq \frac{1.96}{\sqrt{n}}\frac{1}{\lambda}\right) \simeq P\left(|Z| \leq 1.96\right) = 0.95$ avec $\mathcal{L}(Z) = \mathcal{N}(0,1)$, d'où $I_n = \left[\frac{\overline{X}_n}{1+\frac{1.96}{\sqrt{n}}}, \frac{\overline{X}_n}{1-\frac{1.96}{\sqrt{n}}}\right]$ est un intervalle de confiance asymptotique au niveau 95%.

Exemple 4.7 On reprend l'exemple initial. On voudrait avoir un intervalle de confiance à 95% de la proportion de pièces défectueuses.
Les variables X_i de l'échantillon suivent la loi de Bernoulli $\mathcal{B}(\theta)$, où l'inconnue θ est la proportion de pièces défectueuses dans le lot entier.
On a $E[X_i] = \theta$ et $Var(X_i) = \theta(1-\theta)$. On estime alors l'espérance par $\overline{X}_n = \frac{1}{200}\sum_{i=1}^{200} X_i$ (dont la réalisation ici est 0.075) et la variance par $\hat{\sigma}_n^2 = \frac{1}{200}\sum_{i=1}^{200} X_i^2 - (\frac{1}{200}\sum_{i=1}^{200} X_i)^2$. Comme les X_i sont des variables de Bernoulli, on a $\hat{\sigma}_n^2 = \overline{X}_n(1-\overline{X}_n)$ (dont la réalisation est 0.069).

On a donc le résultat suivant, la proportion de pièces défectueuses du lot est avec probabilité 0.95 dans l'intervalle

$$I_n = \left[\overline{X}_n - t_\alpha \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + t_\alpha \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$$

de réalisation $\left[0.075 - \sqrt{\frac{0.069}{200}}1.96, 0.075 + \sqrt{\frac{0.069}{200}}1.96\right] = [0.038, 0.111]$. L'entreprise accepte le lot si la proportion de pièces défectueuses est inférieure à 5%. Que faire du lot ? On va maintenant tester si la proportion est en effet inférieure à 5% et donc qui va nous permettre de répondre à la question.

Exercice 4.8 Lors d'une enquête d'opinion, on a dénombré 81 personnes satisfaites d'un produit sur 1681 interrogées.
En admettant que les personnes de l'échantillon ont été prises au hasard dans une grande population, donner l'intervalle de confiance pour de la proportion p de personnes satisfaites dans la population totale, avec une probabilité de confiance de 0.98.

Chapitre 11

Les tests

On va baser ce chapitre le même exemple : une entreprise reçoit un lot important de pièces fabriquées en série. L'entreprise n'accepte la livraison que si la proportion de pièces défectueuses est inférieure à 5%. Il est impossible d'inspecter chaque pièce, on extrait alors du lot un échantillon de 200 pièces. Sur ces 200 pièces, on dénombre 15 défectueuses. L'entreprise doit-elle accepter ce lot ?

1 Généralités

On considère un échantillon $X_1, X_2, \ldots$ dont la loi P_θ dépend du paramètre inconnu θ.

Définition 1.1 Faire un **test** de l'**hypothèse nulle** H_0 contre l'**hypothèse alternative** H_1 au niveau de risque α (de l'ordre de 5%), c'est fixer une **zone de rejet** $D = D(X_1, ..., X_n)$ ne dépendant que de l'échantillon, telle que

$$P_{H_0}(D) = \sup_{\theta \in H_0} P_\theta(D) = \alpha.$$

Alors la conclusion du test est la suivante :

si D est réalisé, on rejette l'hypothèse H_0
si D n'est pas réalisé, on accepte l'hypothèse H_0.

Remarque 1.2 Il y a deux erreurs possibles :

1. **Erreur de $1^{\text{ère}}$ espèce** : Rejetter à tord l'hypothèse H_0.
   La probabilité de cet événement est $P_{H_0}(D)$, c'est à dire α.
2. **Erreur de 2^{nde} espèce** : Accepter H_0 alors que l'hypothèse est fausse.
   La probabilité de cet événement est $\beta = P_{H_1}(\overline{D}) = \sup_{\theta \in H_1} P_\theta(D)$.

La première erreur α étant en générale fixée, une bonne zone de rejet minimise la valeur de β.

Remarque 1.3 Plus on diminue le niveau de risque α, plus on diminue la probabilité de se tromper, i.e. de rejeter H_0 alors que H_0 est vrai, plus on diminue la taille de la zone de rejet.

Comment trouver une zone de rejet ?
On va dans la suite essayer de répondre à la question.
On limite ce chapitre au cas où le paramètre inconnu θ est la moyenne $\theta = E[X_1]$ et où $\sigma^2 = Var(X_1)$ existe.

2 Tests avec une hypothèse simple

<u>Idée</u> :
On considère un échantillon $(X_1, ..., X_n)$ dont la loi commune P_θ dépend d'un paramètre inconnu θ.
On estime un paramètre inconnu θ à l'aide des observations $(X_1, ..., X_n)$. Notre intuition ou un autre échantillon, nous laisse à penser que la valeur réelle de θ est une certaine valeur connue θ_0. La question qui se pose alors est : est-ce que cette valeur théorique θ_0 est proche de la réalité ? Pour y répondre, on va tester à l'aide de l'échantillon $(X_1, ..., X_n)$ si l'hypothèse $H_0 : \theta = \theta_0$" est réaliste. Ce test permet de comparer la valeur liée à l'expérience à une certaine valeur désirée θ_0.

Exemple 2.1 On veut connaitre la probabilité de naissance d'un garçon. Notre intuition nous dit que cette proportion est 1/2.
On fait un sondage, sur 429440 naissances on dénombre 221023 garçons. On se demande si la proportion de garçons est compatible avec l'hypothèse d'équiprobabilité de naissance des garçons et des filles.

Définition 2.2 Faire un test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau de risque α (de l'ordre de 5%), c'est fixer une zone de rejet $D = D(X_1, ..., X_n)$ ne dépendant que de l'échantillon, telle que $P_{H_0}(D) = P_{\theta_0}(D) = \alpha$.
Alors la conclusion du test est la suivante :

si D est réalisé, on rejette l'hypothèse $H_0 : \theta = \theta_0$
si D n'est pas réalisé, on accepte l'hypothèse $H_0 : \theta = \theta_0$.

Exemple 2.3 *Test par intervalle de confiance*
Supposons que I est un intervalle de confiance pour θ au niveau de confiance $1 - \alpha$. On a alors, $\forall \theta \; P_\theta(\theta \in I) = 1 - \alpha$. Posons $D = \{\theta_0 \notin I\}$.
Alors D est un test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau de risque α.

Exemple 2.4 Reprennons l'exemple 2.1 sur les naissances.
On pose $n = 429440$. On note X_i le résultat de la i$^{\text{ème}}$ naissance : $X_i = 1$ si c'est un garçon et $X_i = 0$ sinon. On estime la proportion de naissance de garçon p par la proportion de l'échantillon, soit par $\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i$.
Il est évident qu'on va rejetter l'hypothèse d'équiprobabilité si la proportion estimée est trop petite ou trop grande, donc on va chercher un zone de rejet de la forme $D = \{\hat{p} \notin [a, b]\}$. Pour ce faire, on va construire un intervalle de confiance de niveau $1 - \alpha$. En estimant la variance de l'échantillon par $\hat{p}(1 - \hat{p})$, et en utilisant le théorème central limite, l'intervalle

$$I = [\hat{p} - t_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + t_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

est un intervalle de confiance de p au niveau $1 - \alpha$ si t_α vérifie $P(|Z| \leq t_\alpha) = 1 - \alpha$ avec $Z \sim \mathcal{N}(0, 1)$.

On pose

$$D = \{1/2 \notin I\} = \{1/2 \notin [\hat{p} - t_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + t_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]\}.$$

On fixe $\alpha = 5\%$, d'où $t_\alpha = 1.96$. Au vu des résultats du sondage, on a $\hat{p} = 0.515$, l'intervalle est donc $I = [0.512, 0.517]$. Comme $1/2 \notin I$, D est réalisé. On est dans la zone de rejet. Donc on rejette l'hypothèse d'équiprobabilité des naissances.

3 Tests entre hypothèses composites

Reprennons l'exemple initiale de l'entreprise qui reçoit un lot de pièces mécaniques et qui ne l'accepte que si la proportion de pièces défectueuses est inférieure à 5%. Si on note θ la proportion de pièces défectueuses dans tout le lot, on souhaite donc tester $H_0 : \theta \leq 5\%$ contre $H_1 : \theta > 5\%$.

Définition 3.1 Faire un test de $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$ au niveau de risque α (de l'ordre de 5%), c'est fixer une zone de rejet $D = D(X_1, ..., X_n)$ ne dépendant que de l'échantillon, telle que $P_{H_0}(D) = \sup_{\theta \leq \theta_0} P_\theta(D) = \alpha$.
Alors la conclusion du test est la suivante :

si D est réalisé, on rejette l'hypothèse $H_0 : \theta \leq \theta_0$
si D n'est pas réalisé, on accepte l'hypothèse $H_0 : \theta > \theta_0$.

Cette définition s'étend de façon évidente aux cas où
- $H_0 : \theta \geq \theta_0 \quad \Rightarrow \quad \alpha = \sup_{\theta \geq \theta_0} P_\theta(D)$
- $H_0 : \theta \in [\theta_1, \theta_2] \quad \Rightarrow \quad \alpha = \sup_{\theta \in [\theta_1, \theta_2]} P_\theta(D)$...

Remarque 3.2 Il n'est en général pas simple de trouver une zone de rejet qui vérifie $P_{H_0}(D) = \alpha$. Dans ce cas, on cherche une zone de rejet qui vérifie $P_{H_0}(D) \leq \alpha$, par conséquent qui a une erreur de première espèce inférieure.

Exemple 3.3 Reprennons l'exemple initiale de l'entreprise. On veut tester $H_0 : \theta \leq 5\%$ contre $H_1 : \theta > 5\%$. Fixons le niveau de risque à $\alpha = 1\%$.
Il est naturel de rejetter l'hypothèse H_0 si la proportion de pièces défectueuses de l'échantillon est trop importante. On va donc chercher une zone de rejet de la forme $D = \{\overline{X}_n \geq d\}$, où $\overline{X}_n$ est le proportion de pièces défectueuses de l'échantillon.
En utilisant le théorème central limite, on a

$$P_\theta\left(\sqrt{\frac{n}{\theta(1-\theta)}}\left(\overline{X}_n - \theta\right) \leq t\right) \underset{n\to+\infty}{\longrightarrow} P(Z \leq t) \text{ avec } Z \sim \mathcal{N}(0,1).$$

On cherche d tel que $P_{H_0}(D) = \sup_{\theta \leq 0.05} P_\theta(D) = 0.01$. En fait on va trouver d tel que

$\sup_{\theta \leq 0.05} P_\theta(D) \leq 0.01$. En effet,

$$\begin{aligned}
\sup_{\theta \leq 0.05} P_\theta(D) &= \sup_{\theta \leq 0.05} P_\theta\Big(\sqrt{\frac{n}{\theta(1-\theta)}}\left(\overline{X}_n - \theta\right) \geq \sqrt{\frac{n}{\theta(1-\theta)}}\,(d-\theta)\Big) \\
&\simeq \sup_{\theta \leq 0.05} P\Big(Z \geq \sqrt{\frac{n}{\theta(1-\theta)}}\,(d-\theta)\Big) \\
&\leq \sup_{\theta \leq 0.05} P\Big(Z \geq \sqrt{\frac{n}{\theta(1-\theta)}}\,(d-0.05)\Big) \\
&\leq P\Big(Z \geq \sqrt{\frac{n}{0.05 \times 0.95}}\,(d-0.05)\Big) \\
&\quad \text{du fait de la décroissance de la fonction } \theta \to \frac{1}{\theta(1-\theta)} \text{ sur l'intervalle } [0, 0.05]
\end{aligned}$$

On choisit alors d tel que $P(Z \geq \sqrt{\frac{n}{0.05\times 0.95}}\,(d-0.05)) = 0.01$, d'où en utilisant la table gaussienne p.45

$$\sqrt{\frac{n}{0.05 \times 0.95}}\,(d-0.05) = 2.33 \Rightarrow d = 0.05 + 2.33 \times \sqrt{\frac{0.05 \times 0.95}{n}}$$

Comme $n = 200$, on obtient $d = 0.086$ et donc $D = \{\overline{X}_n > 0.086\}$. Comme la réalisation de $\overline{X}_n$ est 0.075, on n'est pas dans la zone de rejet. L'entreprise accepte le lot.

4 Égalité des moyennes pour des échantillons non indépendants

On considère n individus sur lesquels on réalise deux séries de mesures concernant les variables X et Y : par exemple X est la quantité observée avant traitement et Y est la même quantité après traitement.
On se demande si le traitement a eu un effet sur la moyenne. On veut par conséquent effectuer le test suivant : $H_0 : E[X] = E[Y]$ contre $H_1 : E[X] \neq E[Y]$.
Les variables X et Y étant mesurées sur les mêmes individus, elles sont dépendants. On se ramène au cadre d'un échantillon de variables indépendantes et de même loi en considérant la variable Z définie par

$$Z_1 = Y_1 - X_1, \ldots, Z_n = Y_n - X_n$$

On effectue alors le test : $H_0 : E[Z] = 0$ contre $H_1 : E[Z] \neq 0$ en utilisant les notions introduites dans la Section 1.

Remarque 4.1 Si on veut tester $H_0 : E[X] \leq E[Y]$ contre $H_1 : E[X] > E[Y]$, alors on effectue le test $H_0 : E[Z] \leq 0$ contre $H_1 : E[Z] > 0$ en utilisant les notions introduites dans la Section 2.

Exemple 4.2 Dans un échantillon de 100 sujets extrait d'une population, on mesure le rythme cardiaque, noté X avant administration d'un médicament et Y après. On obtient les résultats suivants :

$$\sum_{i=1}^{100} x_i = 7983 \qquad \sum_{i=1}^{100} y_i = 8210$$
$$\sum_{i=1}^{100} x_i^2 = 667600 \qquad \sum_{i=1}^{100} y_i^2 = 731160 \qquad \sum_{i=1}^{100} x_i y_i = 689900$$

Peut-on considérer au niveau de risque de test 0.05 que le médicament a une action sur le rythme cardiaque ?

5 Comparaison de 2 moyennes pour des échantillons indépendants

On veut comparer deux séries statistiques observées sur deux populations distinctes, ou bien aux résultats obtenus dans deux expériences réalisées indépendamment sur une même population.
On a donc deux échantillons indépendants :

$X_1, \dots, X_{n_1}$ des variables indépendantes et de même loi, on pose $E[X_1] = m_1, Var(X_1) = \sigma_1^2$
$Y_1, \dots, Y_{n_2}$ des variables indépendantes et de même loi, on pose $E[Y_2] = m_2, Var(Y_2) = \sigma_2^2$

On veut tester $H_0 : m_1 = m_2$ contre $H_1 : m_1 \neq m_2$ au niveau de risque α. D'après le théorème centrale limite, on a

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \underset{n \to +\infty}{\longrightarrow} Z \text{ avec } Z \sim \mathcal{N}(0,1)$$

On choisit t_α tel que $P(|Z| > t_\alpha) = \alpha$, et on a donc

$$P_{H_0}\left(\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2} - (m_1 - m_2)|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > t_\alpha\right) = P_{H_0}\left(\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > t_\alpha\right) \simeq \alpha$$

Si σ_1^2 et σ_2^2 sont connus, on choisit comme zone de rejet

$$D = \left\{\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > t_\alpha\right\}$$

Si σ_1^2 et σ_2^2 sont inconnus, on les estime par les variances empiriques $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$ et on choisit la zone de rejet

$$D = \left\{\frac{|\overline{X}_{n_1} - \overline{Y}_{n_2}|}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}} > t_\alpha\right\}$$

Exemple 5.1 On se demande si la sensibilité aux intoxications professionnelles est identique pour les populations urbaine et rurale. Les observations sur deux échantillons ont montré :
123 personnes sensibles sur 276 en milieu urbain
145 personnes sensibles sur 295 en milieu rural
La différence est-elle significative au niveau $\alpha = 0.05$? $\alpha = 0.1$? Conclusion.

Bibliographie

[1] Yadolah DODGE, *Premiers pas en statistique*, Ed. Springer.

[2] Sheldon M. ROSS, *Initiation aux probabilités*, Ed. Presses Polytech. et Univ. Romandes.

[3] Philippe TASSI, *Méthodes statistiques*, Ed. Economica.