

L2 Informatique HLMA303

STATISTIQUE DESCRIPTIVE ET PROBABILITÉS

Ce cours est une introduction aux méthodes statistiques pour l'étude des **variables quantitatives** (que l'on peut mesurer, compter) qui sont très souvent les variables d'intérêt dans les applications associées à de nombreux domaines. L'économie, la médecine, l'industrie, les sciences sociales (psychologie etc) et les sciences du vivant en sont des exemples. Les méthodes statistiques y sont utilisées pour décrire, analyser et interpréter des phénomènes, faire des sondages, tester des hypothèses avec contrôle du risque de se tromper, simuler des données...

Ces méthodes sont complétées par l'utilisation d'outils de probabilité. Par exemple, un phénomène étudié (la météo...), peut être modélisé par un modèle probabiliste. Cependant, à l'échelle de l'application, l'outil informatique est indispensable! En effet, face au volume croissant des données récoltées à analyser, toutes ces méthodes sont en pratique développées sur des logiciels ou implémentées de manière spécifique par des ingénieurs.

Le plan du cours est le suivant :

1. **Statistique Descriptive** : ensemble de méthodes qui fournit un résumé des données étudiées (présentation synthétique, résumés numériques et graphiques).
2. **Dénombrement et Lois de Probabilités** : A l'issue de l'étude de statistique descriptive, on peut aller plus loin pour estimer certaines caractéristiques de la population avec un certain degré de confiance, valider certaines hypothèses sur la population. Pour cela, on a besoin d'outils de probabilités.
3. **Statistique Inférentielle** : elle permet d' "extrapoler" certaines caractéristiques, conclusions sur la population entière (en général inobservable dans son ensemble) à partir d'un échantillon observé. On introduira la notion d'estimateurs ponctuels et d'estimation par intervalle de confiance d'un paramètre. Cette partie s'appuie sur le calcul des probabilités. Chaque partie débute par une liste de mots-clés qu'il faut savoir définir.

1 Statistique Descriptive Univariée et Généralités

Mots-clés :

- Population, échantillon, individus ou unité statistique, variable ou caractère,
- Différents types de variables : quantitatives (continues ou discrètes), qualitatives nominales ou ordinales.
- Données brutes et données regroupées en classe.
- Distribution statistique Observée (DO) des effectifs, des fréquences, des fréquences cumulées croissantes et décroissantes.
- Histogramme des fréquences, polygone des fréquences, courbe des fréquences cumulées croissantes et décroissantes.
- Paramètres de position : moyenne, mode, médiane, quartiles, déciles, quantiles.
- Paramètres de dispersion : variance et écart-type, étendue, écart interquartile, coefficient de variation.
- Boîte à moustaches ou boxplot.

1.1 Vocabulaire de la statistique descriptive

Définition 1 (Population et unité statistique) En statistique, la **population** désigne un ensemble d'unités statistiques. Les **unités statistiques** ou **individus** sont les entités qui représentent des personnes, des animaux, des végétaux ou des objets.

Dans la pratique, il est souvent impossible d'étudier la population entière, on a alors recours à un sous-ensemble de la population que l'on peut observer, appelé **échantillon**.

Remarque 1 *L'opération servant à choisir cette fraction de la population est l'opération d'échantillonnage.*

Remarque 2 *La population est qualifiée de population-parent.*

Remarque 3 *L'inférence statistique fait l'hypothèse que les informations (variance, moyenne...) tirées de l'étude de l'échantillon sont représentatives de celles de la population d'étude toute entière. Un échantillon représentatif est dit non biaisé. Pour cela, il faut bien choisir son échantillon ("ne pas prendre que des personnes particulières"). La précaution principale à prendre est de ne pas être sélectif et de procéder à un échantillonnage aléatoire simple. C'est à dire que chaque individu de la population entière doit avoir la même probabilité d'être sélectionné lors du tirage.*

Définition 2 (Variable ou Caractère) Chacun des individus qui constituent la population étudiée est décrit par une ou plusieurs **variables** ou **caractères**. Une variable (ou caractère) peut être **quantitative** ou **qualitative**.

Définition 3 (Variable Quantitative/Qualitative) Lorsque les valeurs d'une variable sont numériques (que l'on peut mesurer, compter) et ordonnées, on dit que la variable est **quantitative**. Si les valeurs d'une variable appartiennent à un intervalle de \mathbb{R} ou à \mathbb{R} tout entier, on dit que la variable est **quantitative continue**. Si la variable quantitative provient d'un comptage (nombre de ...), on dit qu'elle est **quantitative discrète**.

Lorsque les valeurs d'une variable ne correspondent ni à une mesure, ni à un comptage, on dit que la variable est **qualitative**. Si ces valeurs peuvent être ordonnées, on dit que la variable est **qualitative ordinale** sinon on parle de variable **qualitative nominale**.

Exemple 1 Le tableau suivant contient 7 variables observées sur un échantillon de 100 écrevisses de Californie dite écrevisse signal d'une rivière : la taille, le nombre de jours d'observations, le nombre de taches, la taille de la pince, le poids, le sexe, la couleur des taches.

	Taille en mm	Nb jours observation	Nb taches	Taille pince en mm	Poids en gr	Femelle M/F	Couleur des taches Bleu/Orangé/Foncé
1	120	284	1	27	100	M	Bleu
2	113	282	2	33	135	M	Bleu
3	128	279	1	28	115	F	Orangé
4	108	282	1	23	125	F	Bleu
5	136	286	4	25	93	M	Orangé
6	138	244	4	33	178	M	Orangé
7	132	245	2	23	140	M	Foncé
8	120	289	3	25	125	M	Foncé
9	143	299	3	30	136	F	Bleu
10	140	351	2	27	120	M	Bleu
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
95	119	274	3	33	120	M	Bleu
96	107	280	4	36	117	F	Orangé
97	119	273	1	24	108	F	Orangé
98	133	279	5	37	140	F	Orangé
99	155	287	2	33	143	M	Bleu
100	126	273	2	22	150	M	Orangé

Le tableau donne les valeurs de 7 variables pour 100 individus :

- le sexe ou la couleur des taches de l'abdomen sont des caractères qualitatifs,
- le poids, la taille et la taille des pinces sont des variables quantitatives continues.
- le nombre de jours d'observations, le nombre de taches sont des variables quantitatives discrètes.

Lorsque les données sont présentées comme dans l'exemple 1 où la ligne i fournit les valeurs observées de chacune des variables pour l'individu i , on dit qu'il s'agit des **données brutes**.

Définition 4 (Modalités d'une variable) Les modalités d'une variable sont les diverses caractéristiques possibles de celle ci telles qu'elles soient :

- incompatibles : un individu n'appartient pas à deux modalités,
- exhaustives : toutes les situations sont prévues,
- sans ambigüité : ne pas faire des erreurs de classement.

Exemple 2 Les modalités pour deux exemples de variable :

- La variable "sexe" a deux modalités : masculin et féminin.
- La variable "résultat lors d'un lancé de dé" a six modalités : 1, 2, 3, 4, 5, 6.

1.2 Distribution Statistique Observée (DO) des effectifs ou des fréquences

On considère la taille des écrevisses. La lecture des données brutes directement dans le tableau est peu informative.

Exemple 3 Considérons la taille en mm des 100 écrevisses :

120 113 128 108 136 138 132 120 143 140 144 141 110 114 115 92 115 144 119 105 115 137 122 131 103 146 114 125 114 122
 93 130 119 113 134 107 134 122 129 110 111 87 143 155 110 122 145 115 108 102 143 146 124 124 145 106 75 107 124 122 101
 128 104 97 137 103 142 130 156 133 120 91 127 153 121 120 99 149 129 139 114 138 138 131 125 128 134 114 85 135 87 125
 105 120 119 107 119 133 155 126

Une façon de traiter ce problème est de présenter les données brutes sous la forme d'un tableau synthétique appelé **Distribution Observée** de la variable taille. On regroupe les données par classes ou intervalles. On parle alors de **données groupées en classes**.

Définition 5 (Distribution Observée des effectifs et des fréquences) La donnée du nombre d'individus ou **effectif** n_k pour chaque classe k s'appelle la **distribution observée (DO) des effectifs**. Elle est donnée sous forme d'un tableau où l'on juxtapose la colonne des classes avec celles des effectifs. De même, la **distribution observée (DO) des fréquences** donne pour chaque classe k , la proportion d'individus ou **fréquence** $f_k = n_k/n$ de la classe k .

Par convention, on choisit d'exclure la borne droite de l'intervalle définissant chaque classe.

Exemple 4 La plus petite écrevisse mesure 75 mm et la plus grande 156 mm. Ainsi, on choisit de regrouper les observations en 6 classes.

Taille en mm	Effectifs n_k	Fréquence $f_k = n_k/n$
[75,90[4	0.04
[90,105[12	0.12
[105,120[30	0.30
[120,135[30	0.30
[135,150[20	0.20
[150,165[4	0.04
Somme	100	1

La somme des fréquences est égale à 1.

1.3 Histogramme

Définition 6 L'**histogramme** est un ensemble de rectangles, chacun correspondant à une classe de la distribution observée des fréquences :

- La base du rectangle est placée sur l'axe des abscisses et est délimitée par les bornes de classe,
- l'aire du rectangle est égale à la fréquence f_k de la classe : La hauteur du k^e rectangle vaudra donc $h_k = f_k/l_k$ où l_k est la longueur de l'intervalle correspondant à la k^e classe.

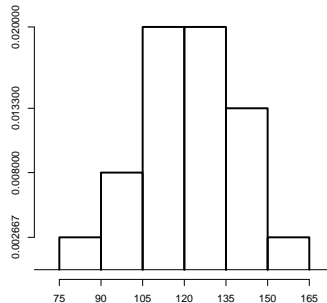
La hauteur h_k elle est appelée la **densité de fréquence** .

L'histogramme est une représentation graphique de la Distribution Observée des fréquences. Par construction, l'aire totale de l'histogramme vaut 1.

Taille en mm	l_k	Fréq. f_k	Hauteur du rectangle h_k
[75,90[15	0.04	0.002667
[90,105[15	0.12	0.008
[105,120[15	0.30	0.02
[120,135[15	0.30	0.02
[135,150[15	0.20	0.0133
[150,165[15	0.04	0.002667

Le premier rectangle a une base entre 75 et 90 et une surface égale à $f_1 = 0.04$. Ainsi, $h_1 \times l_1 = 0.04$. Et donc

$$h_1 = \frac{0.04}{15} = 0.002667$$



Si on avait choisi un autre découpage en classes, par exemple en regroupant les deux dernières classes, le résultat aurait été différent :

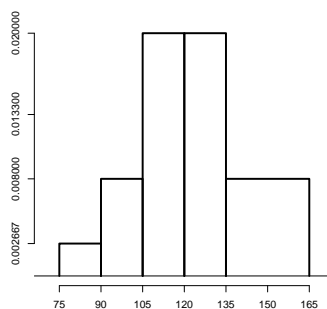
Taille en <i>mm</i>	l_k	Fréq. f_k	Hauteur du rectangle h_k
[75,90[15	0.04	0.002667
[90,105[15	0.12	0.008
[105,120[15	0.30	0.02
[120,135[15	0.30	0.02
[135,165[30	0.24	0.008

La construction de l'histogramme est identique au cas précédent pour les 4 premiers rectangles. Ensuite, on recalcule la hauteur du dernier rectangle

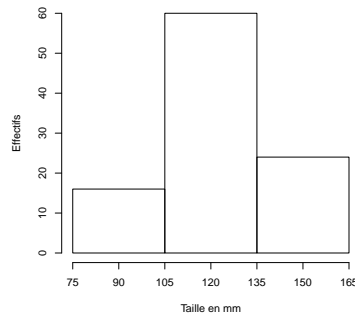
$$h_5 \times l_5 = 0.24$$

. Et donc

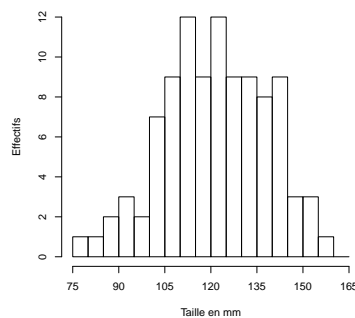
$$h_5 = \frac{0.24}{30} = 0.008$$



Si on avait choisi un nombre de classe plus faible, par exemple les trois suivantes : [75,105], [105,135], [135,165], le résultat serait le suivant :



Cependant, si on avait augmenté le nombre de classes, le résultat serait encore différent :



Remarque 4 Lorsque toutes les classes sont de même longueur (tous les l_k sont égaux), on peut prendre des rectangles de hauteur égale à l'effectif n_k ou à la fréquence f_k mais alors l'aire de l'histogramme n'est plus égale à 1.

L'aspect d'un histogramme dépend fortement du nombre de classes choisies comme on peut le voir avec les deux histogrammes ci-dessus (à gauche avec $K = 3$ classes et à droite avec $K = 17$ classes). Un trop faible nombre de classes conduit à une représentation trop lissée, alors qu'un trop grand nombre de classes donne un histogramme très dentelé avec possiblement des "trous" (classes vides).

Il existe des règles qui donnent une indication pour le choix du nombre de classes K en fonction de la taille n de l'échantillon observé. La règle de Sturges fournit une méthode pour choisir le nombre de classes, valide lorsque la distribution observée est symétrique :

$$K = 1 + \frac{10}{3} \ln(n)$$

1.4 Distribution observée des fréquences cumulées

Définition 7 La distribution observée des **fréquences cumulées croissantes** donne la proportion d'observations inférieures à la borne supérieure de chaque classe. Si n_k est l'effectif de la k^e classe, la fréquence cumulée croissante F_k de la classe k est égale à :

$$F_k = \frac{n_1 + \dots + n_k}{n} = f_1 + \dots + f_k.$$

Taille en mm	Fréq. f_k	Fréq. cumulée F_k
[75,90[0.04	0.04
[90,105[0.12	0.16
[105,120[0.3	0.46
[120,135[0.3	0.76
[135,150[0.2	0.96
[150,165[0.04	1

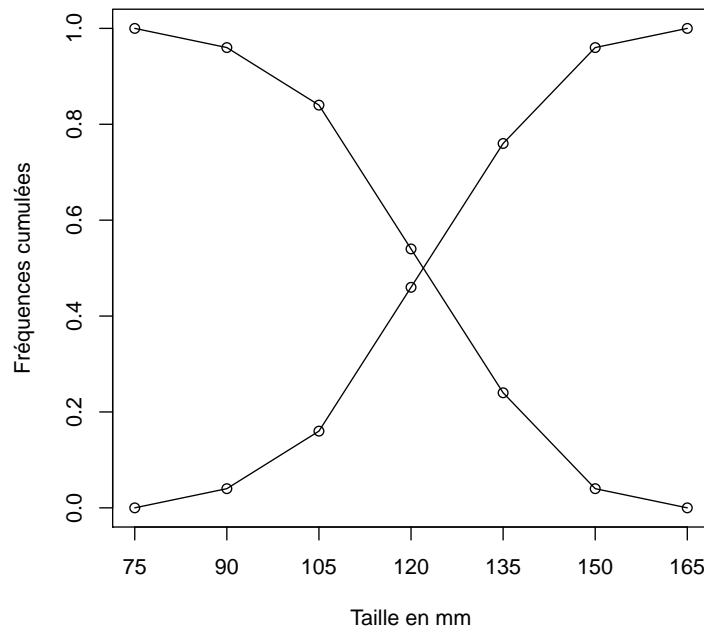
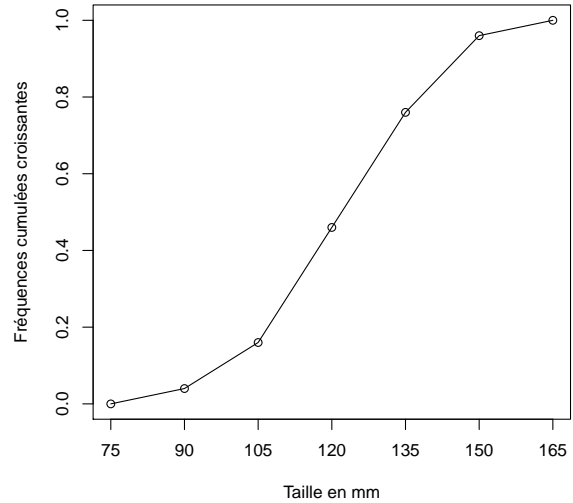
Ainsi, on peut lire que 46% des écrevisses observées ont une taille inférieure à 120 mm et 96% une taille inférieure à 150 mm .

La **courbe des fréquences cumulées croissantes** relie successivement par des segments de droite les points d'abscisse égale à la borne supérieure de la classe k et d'ordonnée égale à la fréquence cumulée croissante F_k associée :

abs.	75	90	105	120	135	150	165
ord.	0	0.04	0.16	0.46	0.76	0.96	1

La courbe démarre du point (75, 0) : aucune écrevisse n'a été observée ne mesurant moins de 75 mm) et elle se termine au point (165, 1) : toutes les écrevisses observées ont une taille inférieure à 165 mm .

De même, on peut définir les fréquences cumulées décroissantes $F'_k = 1 - F_k$ et la courbe des fréquences cumulées décroissantes. Les deux courbes de fréquences cumulées croissantes et décroissantes peuvent être représentées sur le même graphique.



1.5 Paramètres de position ou de tendance

On va maintenant essayer de synthétiser l'information contenue dans les données au moyen d'un petit nombre de caractéristiques ou résumés numériques. On présente les caractéristiques de **tendance centrale** (ou de position ou de localisation) qui vont nous renseigner sur une valeur moyenne ou centrale des données.

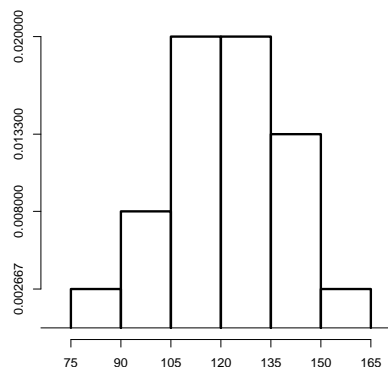
• Le mode :

Définition 8 Le **mode** désigne la valeur la plus représentée d'une variable quantitative.

- Lorsque les données sont regroupées en classes d'amplitudes égales, la **classe modale** désigne celle qui a le plus grand effectif.
- Si les classes sont d'amplitudes différentes, la classe modale est celle qui a la plus forte densité de fréquence f_k/l_k c'est-à-dire celle qui correspond au rectangle le plus haut de l'histogramme. Le **mode** est alors le **centre de la classe modale**.

Le mode n'est pas unique : une distribution observée peut être unimodale ou bimodale (deux modes) ou multimodale (plusieurs modes).

Cas des données continues : le mode est le centre de la classe où la hauteur du rectangle de l'histogramme est maximale (classe modale). Dans l'exemple de la taille des écrevisses, on a obtenu l'histogramme ci-contre. Il y a donc deux classes modales : $[105;120[$ et $[120;135[$ et donc deux modes 112.5 et 127.5 : le mode n'est pas unique et il dépend du choix du regroupement en classes.



Cas des données discrètes : Considérons le nombre de taches sur la face dorsale de l'abdomen : c'est une variable quantitative discrète. On calcule la distribution observée des effectifs.

Exemple 5

Nombre de tâches	1	2	3	4	5	6	7	10
Effectifs n_k	24	28	19	10	10	5	2	2

La valeur que l'on observe le plus souvent est le 2, avec un effectif $n_2 = 28$, donc le mode de cette variable est égal à 2.

• La moyenne

Définition 9 (Moyenne) Si l'on dispose des données brutes x_1, \dots, x_n , la **moyenne observée sur l'échantillon**, notée m_e se calcule de la façon suivante :

$$m_e = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

Exemple 6 Pour la taille des écrevisses, la moyenne peut se calculer directement avec les données brutes :

$$m_e = \frac{1}{100} (120 + 113 + \dots + 155 + 126) = \frac{12228}{100} = 122.28 \text{ mm.}$$

La taille moyenne des écrevisses est d'environ 122.3 mm.

Si l'on ne dispose que des données regroupées en K classes, on peut calculer une valeur approchée de la **moyenne** m_e :

Classe	Centre de la classe c_k	Effectifs
$[a_1, a_2[$	c_1	n_1
$[a_2, a_3[$	c_2	n_2
\vdots	\vdots	\vdots
$[a_K, a_{K+1}[$	c_K	n_K

Avec m_k le centre de la classe $[a_k, a_{k+1}[$, le calcul approché de la moyenne avec les données regroupées en classes vaut :

$$m_e \simeq \frac{1}{n} \sum_{k=1}^K n_k c_k = \frac{1}{n} (n_1 c_1 + \dots + n_K c_K)$$

Exemple 7 La taille moyenne des écrevisses calculée sur la distribution des effectifs regroupés en classes nous donne :

$$\begin{aligned} m_e &\simeq \frac{1}{100} (4 \times 82.5 + 12 \times 97.5 + 30 \times 112.5 + 30 \times 127.5 + 20 \times 142.5 + 4 \times 157.5) \\ &= \frac{12180}{100} \simeq 121.8 \text{ mm} \end{aligned}$$

On a donc une taille moyenne de 121.8 mm (au lieu de 122.3 mm calculée sur les données brutes).

Le calcul sur données regroupées en classe est moins précis (perte d'information) que le calcul sur les données brutes : tout se passe comme si les observations dans une classe étaient toutes concentrées au milieu de la classe.

Cas d'une variable discrète : La **moyenne** se calcule de la façon suivante :

$$m_e = \frac{1}{n} \sum_{i=1}^K n_i x_i = \frac{1}{n} (n_1 x_1 + \dots + n_K x_K)$$

Exemple 8 Le nombre moyen de taches observées sur la partie dorsale de l'abdomen est calculée :

Nb taches	1	2	3	4	5	6	7	10
Effectifs	24	28	19	10	10	5	2	2

Le nombre moyen de taches est égal à :

$$m_e = \frac{1}{100} (24 \times 1 + 28 \times 2 + 19 \times 3 + \dots + 2 \times 10) = \frac{291}{100} = 2.91$$

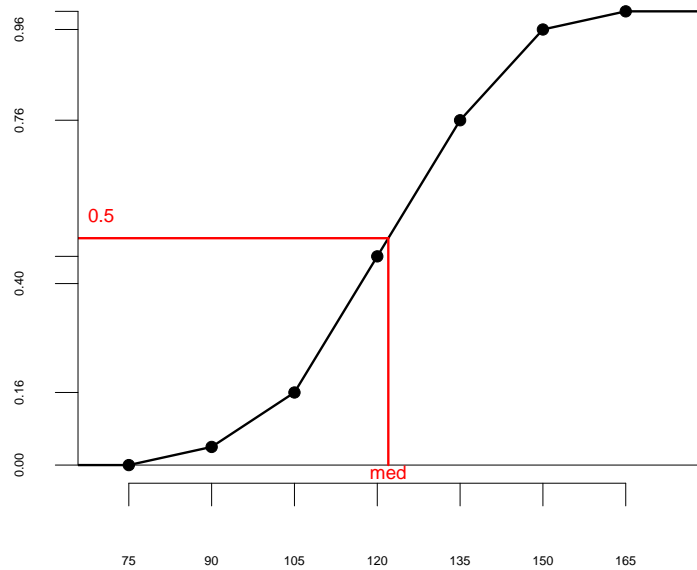
La variable nombre de taches étant discrète, on arrondira à 3 la valeur du nombre moyen de taches observées.

Ici la médiane du Nombre de tâches est égal à 2.

2^e Méthode de calcul approché lorsque les données ont été regroupées en classes

La **médiane** peut être approchée par l'**abscisse** *med* du point de la courbe de fréquences cumulées croissantes dont l'ordonnée vaut 0.5.

Courbe des fréquences cumulées de la taille de écrevisses



En pratique, on repère la première classe $[a_k, a_{k+1}[$ qui a une fréquence cumulée F_k croissante supérieure ou égale à 0.5. La médiane vaut alors :

	Taille en <i>mm</i>	Fréq. cum. F_k
$[a_1, a_2[$	$[75, 90[$	$0.04 = F_1$
$[a_2, a_3[$	$[90, 105[$	$0.16 = F_2$
$[a_3, a_4[$	$[105, 120[$	$0.46 = F_3$
$[a_4, a_5[$	$[120, 135[$	$0.76 = F_4$
$[a_5, a_6[$	$[135, 150[$	$0.96 = F_5$
$[a_6, a_7[$	$[150, 165[$	$1 = F_6$

$$med = a_4 + (0.5 - F_3) \times \frac{a_5 - a_4}{F_4 - F_3}$$

Exemple 12 La première classe avec une fréquence cumulée croissante supérieure à 0.5 est $[120, 135[$. Donc la médiane *med* appartient à cette classe et les 3 points de coordonnées $A(120, 0.46)$, $B(135, 0.76)$ et $M(med, 0.5)$ sont **alignés**. Cela se traduit par l'égalité des coefficients directeurs des droites (AB) et (AM) par exemple :

$$\frac{y_B - y_A}{x_B - x_A} = \frac{y_M - y_A}{x_M - x_A}$$

En remplaçant par les valeurs, on obtient :

$$\frac{0.76 - 0.46}{135 - 120} = \frac{0.5 - 0.46}{med - 120} \iff med = 120 + (0.5 - 0.46) \times \frac{135 - 120}{0.76 - 0.46} = 122 \text{ mm.}$$

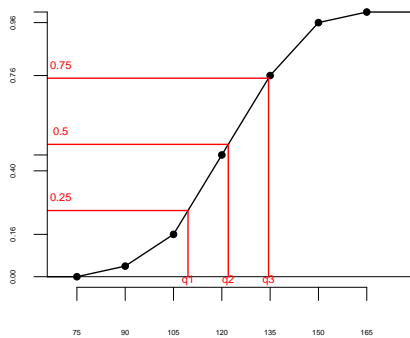
• Les quartiles

Définition 11 (Quartiles) Le **premier quartile** est la plus petite valeur Q_1 de la série de données telle que 25% au moins des autres valeurs lui sont inférieures ou égales.

Pour une variable continue, regroupée en classes, on peut l'approcher par l'abscisse du point dont l'ordonnée vaut 0.25 sur la courbe des fréquences cumulées croissantes.

Le **troisième quartile** est la plus petite valeur Q_3 de la série de données telle que 75% au moins des autres valeurs lui sont inférieures ou égales.

On peut l'approcher par l'abscisse du point dont l'ordonnée vaut 0.75 sur la courbe des fréquences cumulées croissantes.



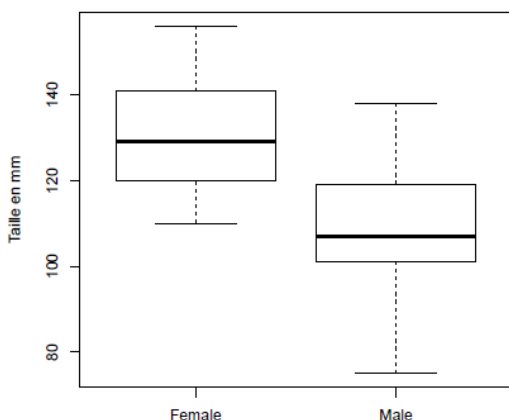
Les **quartiles** Q_1 et Q_3 peuvent être approchés par les abscisses des points de la courbe des fréquences cumulées croissantes dont l'ordonnée est respectivement égale à 0.25 et 0.75. Ainsi, on repère que Q_1 est dans la classe $[105; 120[$ et que Q_3 est dans $[120; 135[$ et on calcule :

$$Q_1 = 105 + (0.25 - 0.16) \times \frac{120 - 105}{0.46 - 0.16} = 109.5 \text{ mm.}$$

$$Q_3 = 120 + (0.75 - 0.46) \times \frac{135 - 120}{0.76 - 0.46} = 134.5 \text{ mm.}$$

Boîtes à moustaches ("boxplot") Les trois **quartiles** d'une variable statistique servent à mesurer la symétrie de la DO du caractère étudié mais aussi la concentration.

On représente les informations données par les quartiles à l'aide de la boîte à moustaches en y ajoutant les valeurs du minimum et du maximum :



Ce type de graphique donne une idée de la distribution des observations et permet de juger de la symétrie et de la concentration des données. Il est notamment très utile lorsqu'on souhaite **comparer plusieurs distributions observées** : on peut dans ce cas représenter les distributions par des boîtes à moustaches en parallèles et repérer facilement les similitudes et/ou les différences de comportement. On peut comparer la DO de la taille des écrevisses mâles et femelles. Il apparaît clairement que les mâles sont plus petits que les femelles.

- Les **quantiles**

Définition 12 (Quantiles) Soit $0 < \alpha < 1$. On appelle **quantile d'ordre α** la plus petite valeur Q_α de la série de données telle qu'au moins $\alpha\%$ des autres valeurs lui sont inférieures ou égales. On peut approcher le quantile Q_α par l'abscisse du point d'ordonnée α sur la courbe des fréquences cumulées croissantes.

Les quartiles sont des quantiles particuliers. Les déciles correspondent aux quantiles d'ordre 0.1, 0.2, ..., les centiles aux quantiles d'ordre 0.01, 0.02, etc.

1.6 Paramètres de dispersion

On a vu comment définir des mesures de **tendance** qui indiquent la position des données. On va maintenant présenter des mesures de **dispersion** qui vont nous indiquer comment se répartissent les données autour des paramètres de position. Ces mesures de dispersion sont **positives**, et seront petites ou grandes suivant que les données sont très concentrées ou au contraire très étalées.

- **L'étendue**

Définition 13 (Étendue) L'**étendue** est l'indicateur le plus simple de la dispersion des données. On observe x_1, \dots, x_n que l'on range en ordre croissant : $x_{(1)}, \dots, x_{(n)}$.

L'étendue est l'écart entre la valeur maximum et la valeur minimum observées :

$$x_{(n)} - x_{(1)} = x_{max} - x_{min}.$$

Exemple 13 La plus petite écrevisse mesure 75 mm et la plus grande 156 mm. L'étendue vaut

$$156 - 75 = 81 \text{ mm}.$$

- **L'écart interquartile :**

Définition 14 (Ecart Interquartile) Une façon de regarder si les données sont très étalées est de regarder la différence entre le premier Q_1 et le troisième Q_3 quartile. C'est ce qu'on appelle l'**écart interquartile** $Q_3 - Q_1$.

Exemple 14 Pour la variable taille, on avait calculé :

Le premier quartile vaut :

$$Q_1 = 105 + (0.25 - 0.16) \times \frac{120 - 105}{0.46 - 0.16} = 109.5 \text{ mm}$$

Le troisième quartile se calcule :

$$Q_3 = 120 + (0.75 - 0.46) \times \frac{153 - 120}{0.76 - 0.46} = 134.5 \text{ mm}$$

L'écart interquartile vaut donc

$$Q_3 - Q_1 = 134.5 - 109.5 = 25 \text{ mm}$$

Cela s'interprète de la façon suivante : la moitié des écrevisses de l'échantillon observé a une taille comprise entre $Q_1 = 109.5$ et $Q_3 = 134.5$ mm : l'écart interquartile est de 25 mm, il donne une mesure de la dispersion des 50% des observations centrales autour de la médiane. L'intervalle interquartile n'est pas forcément symétrique autour de la médiane.

• **La variance et l'écart-type :**

Définition 15 (Variance) La **variance**, notée σ_e^2 , calculée sur les données brutes de l'échantillon observé, est donnée de la façon suivante :

$$\sigma_e^2 = \frac{1}{n} ((x_1 - m_e)^2 + \dots + (x_n - m_e)^2) = \frac{1}{n} \sum_{i=1}^n (x_i - m_e)^2$$

où m_e est la moyenne observée sur l'échantillon.

Exemple 15 On calcule la variance de la variable taille. La moyenne observée est $m_e = 122.28$ mm. Avec la formule ci-dessus, la variance vaut alors :

$$\begin{aligned} \sigma_e^2 &= \frac{1}{100} ((120 - 122.28)^2 + (113 - 122.28)^2 + \dots + (126 - 122.28)^2) \\ &= \frac{28784}{100} = 287.84 \text{ mm}^2 \end{aligned}$$

Définition 16 (Ecart-type) L'**écart-type** est la racine carrée de la variance :

$$\sigma_e = \sqrt{\sigma_e^2}$$

Dans l'exemple, on a $\sigma_e = \sqrt{287.84} = 16.97$ mm. L'écart-type s'exprime avec la même unité de mesure que les observations : pour cette raison, il est plus facile à interpréter que la variance.

2ème formule pour calculer la variance :

$$\sigma_e^2 = \underbrace{\frac{1}{n} \sum_{i=1}^n x_i^2}_{\text{moyenne des carrés}} - \underbrace{\left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2}_{\text{carré de la moyenne}}$$

Exemple 16 moyenne $m_e = 122.28$, somme des observations au carré $\sum_{i=1}^{100} x_i^2 = 1524024$

$$\sigma_e^2 = \frac{1}{100} \times 1524024 - (122.28)^2 = 287.84$$

Données continues regroupées en classes Comme pour la moyenne, quand les données sont regroupées en classes, on peut utiliser une formule approchée pour le calcul de la variance basée sur les centres des classes et les effectifs.

$$\sigma_e^2 = \frac{1}{n} \sum_{k=1}^K n_k (c_k - m_e)^2$$

avec c_k centre de la classe k et m_e moyenne approchée à partir des données groupées en classes.

Taille en mm	Centre c_k	Effectifs n_k
[75,90[82.5	4
[90,105[97.5	12
[105,120[112.5	30
[120,135[127.5	30
[135,150[142.5	20
[150,165[157.5	4

Exemple 17 On reprend le calcul de la variance de la variable taille. On a vu que la moyenne calculée sur les données regroupées en classes vaut 121.8 mm. Il reste à calculer :

$$\frac{1}{100} \sum_{k=1}^6 n_k (c_k - m_e)^2 = \frac{1}{100} (4 \times (82.5 - 121.8)^2 + 12 \times (97.5 - 121.8)^2 + \dots + 4 \times (157.5 - 121.8)^2) = 305.01$$

Et donc

$$\sigma_e^2 = 305.01 \text{ mm}^2$$

on en déduit la valeur de l'écart-type $\sigma_e = \sqrt{305.01} = 17.47 \text{ mm}$.

Lorsque cela est possible, il est préférable d'utiliser les données brutes pour faire les calculs des paramètres de position et de dispersion.

Cas des données discrètes : si la variable est discrète, prenant K valeurs distinctes, la formule de la variance devient :

x_k	x_1	x_2	x_3	\dots	\dots	x_K
n_k	n_1	n_2	n_3	\dots	\dots	n_K

$$\sigma_e^2 = \frac{1}{n} \sum_{k=1}^K n_k (x_k - m_e)^2 = \frac{1}{n} \sum_{k=1}^K n_k x_k^2 - (m_e)^2$$

Exemple 18 On a déjà vu que le nombre moyen de tâches est $m_e = 2.91$. Utilisons la 2ème formule pour calculer la variance :

Nb tâches	1	2	3	4	5	6	7	10
Effectifs	24	28	19	10	10	5	2	2

$$\frac{1}{100} (24 \times 1^2 + 28 \times 2^2 + 19 \times 3^2 + 10 \times 4^2 + 10 \times 5^2 + 5 \times 6^2 + 2 \times 7^2 + 2 \times 10^2) = 11.95.$$

La variance vaut donc :

$$\sigma_e^2 = 11.95 - (2.91)^2 = 3.48$$

et l'écart-type $\sigma_e = \sqrt{\sigma_e^2} = 1.87$

• **Coefficient de variation** :

Définition 17 Le **coefficient de variation** est défini comme le rapport entre l'écart-type observé σ_e et la moyenne observée m_e :

$$C_v = \frac{\sigma_e}{m_e}$$

Ce nombre est sans unité. Il est utilisé pour comparer deux séries de données d'unités différentes ou d'ordre de grandeur différent.

Exemple 19 Si l'on veut comparer la dispersion de la taille des écrevisses signal de Californie et celle des écrevisses autochtones à pattes blanches des rivières de l'Aveyron, on peut évaluer leur coefficient de variation :

	Écrevisse signal	Écrevisse autochtone
m_e en mm	122.28	77.2
σ_e en mm	16.97	10.8
C_v	0.139	0.140

L'écart-type de la taille de l'écrevisse signal (16.97 mm) est une fois et demi plus grand que celui de l'écrevisse autochtone (10.8 mm). Mais, les coefficients de variations sont presque égaux. Donc, relativement aux valeurs des moyennes observées, on ne peut pas dire que la dispersion est plus importante chez l'écrevisse signal. En fait, les dispersions sont équivalentes.

1.7 Les outils descriptifs pour les variables non continues

1.7.1 "Distribution Observée" des effectifs et des fréquences

- **Le cas des variables quantitatives discrètes :**

Dans le cas de variables quantitative discrètes, les données ne sont pas regroupées en classes mais selon les modalités de la variable.

Exemple 20 Une variable quantitative discrète : la variable nombre de taches.

Nb taches	Effectifs n_k	Fréquence $f_k = n_k/n$
1	24	0.24
2	28	0.28
3	19	0.19
4	10	0.10
5	10	0.10
6	5	0.05
7	2	0.02
10	2	0.02
Somme	100	1

- **Le cas des variables qualitatives ordinales :**

Exemple 21 Une variable qualitative ordinale : Département de naissance d'enfants hospitalisés d'un hôpital parisien durant l'année 1998.

Département de naissance	Effectifs n_k	Fréquence $f_k = n_k/n$
05	5	0.02
06	18	0.08
75	80	0.34
77	64	0.27
91	70	0.29
Somme	237	1

- **Le cas des variables qualitatives nominales :**

Exemple 22 Une variable qualitative nominale : la variable sexe.

Sexe	Effectifs n_k	Fréquence $f_k = n_k/n$
M	40	0.24
F	60	0.28
Somme	100	1

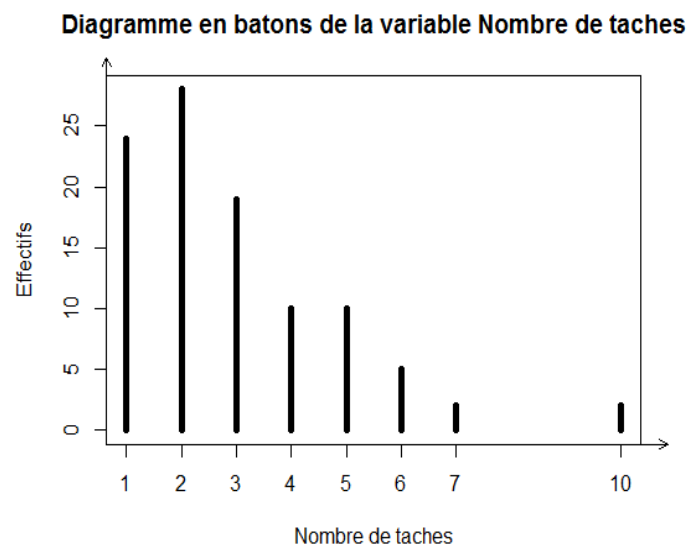
1.7.2 Représentations graphiques des distributions observées de variables non continues

Dans le cas de variables aléatoires continues, nous avons vu que la distribution observée de celle-ci pouvait être représentée graphiquement par un histogramme. Cependant le choix de la représentation graphique dépend de la nature de la variable étudiée. Il convient de choisir adéquatement le mode de représentation graphique d'une variable adapté à son type (quantitative continue, quantitative discrète, qualitative ordinale, qualitative nominale). En effet, le type d'une variable se traduit par des caractéristiques particulières d'un graphique donné. Par exemple, mettre une flèche au bout d'un axe des abscisses indique un ordre sur les modalités correspondantes. Ainsi, par convention on indiquera le sens de l'axe des modalités pour toutes les variables sauf pour les variables qualitatives nominales.

- **Le cas des variables quantitatives discrètes :**

Dans le cas de variables quantitatives discrètes, il est plus adapté d'utiliser un diagramme en bâtons.

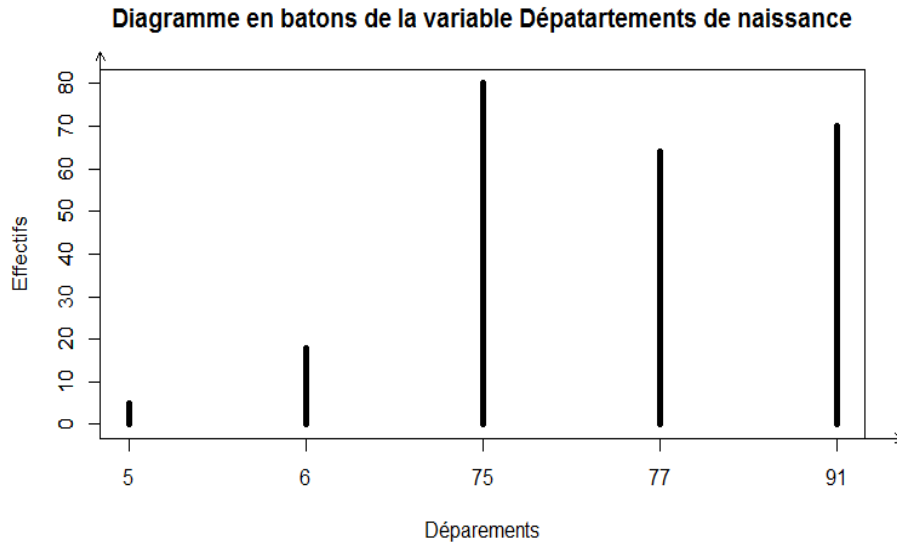
Exemple 23 La variable nombre de taches :



- **Le cas des variables qualitatives ordinales :**

Dans le cas d'une variable aléatoire qualitative ordinale, le diagramme en bâtons reste une représentation graphique adaptée.

Exemple 24 La variable département de naissance :



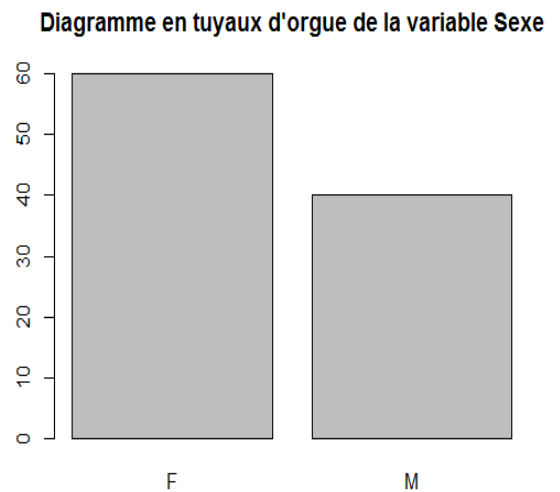
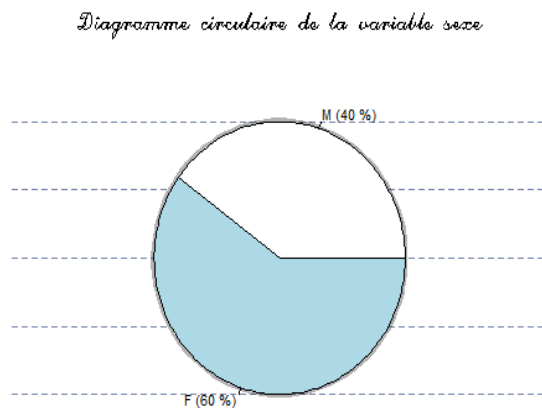
• **Le cas des variables qualitatives nominales** : Dans le cas de variables aléatoires qualitatives nominales, on utilisera un diagramme circulaire ou a tuyaux d'orgues. Pour ce dernier, les tuyaux sont représentés par hauteur décroissante.

Pour une variable qualitative nominale à K modalités la procédure du construction du diagramme en tuyaux d'orgue est la suivante :

- On trace K segments verticaux côte à côte par modalité en abscisse.
- Les hauteurs de chaque bande k est choisie proportionnelle au $k - ime$ effectif n_k .

Concernant la construction du diagramme circulaire, ce dernier est un disque divisé en K secteurs pour les K modalités. Chaque modalité k est représentée par un secteur dont l'aire est proportionnelle à l'effectif n_k ou la fréquence f_k .

Exemple 25 La variable sexe :



Remarque 5 *Pour une modalité fixée, si la fréquence est de 10% (c'est à dire de 0.1), alors le secteur du disque associé sera formé par un angle de 10% de 360 degrés que forme le disque complet.*

Remarque 6 *Pour les diagrammes à tuyaux d'orgue, on peut prendre des segments de 1 cm pour chacune des bandes. Ainsi, lorsque les hauteurs de bande sont égales aux fréquences, l'aire totale des bandes du diagramme sera de 1 (100%).*

2 Probabilités élémentaires et Généralités

Introduction

La théorie des Probabilités est la branche des mathématiques qui étudie les phénomènes aléatoires, c'est-à-dire soumis au hasard. Notons que le mot *aléatoire* vient du latin *alea*, le dé, la chance, et que *hasard* vient de l'arabe *al-zahr*, le dé.

Issue de l'étude des jeux de hasard, qui lui ont fourni ses notions de base (probabilité d'un événement, espérance de gain), la théorie des Probabilités est aujourd'hui une branche des mathématiques au même titre que la Géométrie, l'Algèbre et l'Analyse. Elle entretient naturellement des liens étroits avec les applications : en Sciences physiques, en Biologie (où le hasard est omniprésent, comme source de la diversité biologique des individus), dans les Sciences économiques et sociales et en technologie.

Mots-clés :

- Expérience aléatoire, ensemble fondamental ou univers, événement aléatoire, variable aléatoire.
- Tribu, espace probabilisable, espace probablisé ou espace de probabilité, espace probabilisable discret, probabilité.
- Événements indépendants deux à deux, événements indépendants dans leur ensemble, variables aléatoires indépendantes.
- Probabilité conditionnelle, système complet d'événements.

2.1 Expérience aléatoire

Pour parler du hasard en termes mathématiques, il est nécessaire d'isoler une situation d'intérêt (exemple : lancer de dé, distribution de cartes...). A une telle situation isolée, que l'on peut répéter à l'envi, on donne le nom d'expérience aléatoire.

Définition 18 (Expérience aléatoire) Une *expérience aléatoire* est une expérience renouvelable, au moins en théorie, dont il est impossible de connaître le résultat avant de l'avoir observé et qui, renouvelée dans des conditions identiques, ne donne pas à chaque essai le même résultat.

Voici quelques exemples :

Exemple 26 (Expériences aléatoires) – lancer de deux dés ;

- distribuer les cartes au tarot, c'est-à-dire répartir le paquet de 78 cartes entre les 3, 4 ou 5 joueurs ;
- observer la formation d'un caractère génétique d'un individu à partir des caractères correspondants de ses parents ;
- observer la désintégration d'un noyau atomique radioactif ;
- attendre le tram à la station Universités, à partir de 18h.

Attention ! Quand nous parlons de conditions identiques, nous parlons du point de vue de l'observateur. Le hasard apparaît ainsi souvent comme un nom que nous donnons à notre ignorance de certaines conditions de l'expérience. Il y a donc notamment du hasard lorsque le nombre de facteurs est si grand qu'il est impossible de tous les contrôler. Mais on parle également de hasard lorsque toute imprécision sur les conditions initiales se trouve tellement amplifiée durant l'expérience qu'il devient impossible de faire une prévision correcte : par exemple, les lois de la mécanique des solides devraient nous permettre d'anticiper le résultat d'un lancer de

dé mais les incertitudes sur les conditions initiales (manière de lancer) nous interdisent toute prévision.

Exemple 27 Voici une liste de phénomènes aléatoires qui font l'objet de recherches scientifiques ou technologiques :

- l'apparition d'objets défectueux dans une chaîne de fabrication (fiabilité) ;
- les effets d'un engrais sur la croissance des céréales (agronomie) ;
- les appels reçus à un standard téléphonique (files d'attente) ;
- le mouvement d'une particule dans un liquide, observé par Robert Brown en 1827, et appelé depuis mouvement brownien (mathématiques financières) ;
- les incertitudes sur la trajectoire d'un véhicule spatial (contrôle stochastique).

Pour étudier un phénomène dépendant du hasard, le praticien essaiera d'isoler une expérience aléatoire et il en construira un modèle probabiliste lui permettant de faire certaines prévisions ; disons pour être concret qu'il pourra calculer la probabilité de certains résultats ou de certains événements. La première étape de la modélisation mathématique d'une expérience aléatoire consiste à spécifier **l'ensemble des résultats possibles** (ou réalisations) de cette expérience, que l'on note Ω .

Définition 19 (Ensemble fondamental) *L'ensemble de tous les résultats d'une expérience aléatoire est appelé **ensemble fondamental** ou **univers**, noté Ω . Les éléments de Ω , c'est-à-dire les résultats possibles, sont eux notés ω .*

Exemple 28 1) lancer de deux dés à 6 faces : les résultats possibles sont les couples d'entiers entre 1 et 6 donc $\Omega = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\} = \{1, 2, \dots, 6\}^2$.

2) désintégration d'un noyau radioactif : si on s'intéresse à la durée de vie du noyau, on pose $\Omega =]0, +\infty[$, l'ensemble des réels strictement positifs.

Attention ! Une même expérience peut être décrite de plusieurs façons différentes : si deux joueurs lancent deux dés chacun à leur tour, on peut poser $\Omega = \{1, 2, \dots, 6\}^4$ ou $\Omega = \{2, 3, \dots, 12\}^2$. L'objectif de l'étude doit entrer en ligne de compte pour déterminer Ω .

2.2 Événements et variables aléatoires

2.2.1 Événement aléatoire

Un **événement aléatoire** est un événement lié à une expérience aléatoire : il peut se réaliser ou non, et sa réalisation, ou sa non-réalisation, dépend exclusivement du résultat ω de cette expérience. Mathématiquement, un événement aléatoire se représente par la donnée de l'ensemble des éléments ω de Ω pour lesquels cet événement est réalisé.

Exemple 29 (Lancer de deux dés à 6 faces) L'événement "le résultat est inférieur ou égal à 4" est représenté par le sous-ensemble

$$\{(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1)\}$$

de $\Omega = \{1, 2, \dots, 6\}^2$.

2.2.2 Représentation ensembliste

On a l'habitude, en Probabilités, d'identifier un événement et le sous-ensemble de Ω qui le représente. Cette identification joue un rôle fondamental : les notions et opérations logiques que l'on définit sur les événements correspondent à celles définies en théorie des ensembles. Cette correspondance est résumée dans le tableau suivant, où A et B sont deux événements.

Terminologie probabiliste	Terminologie ensembliste
résultat possible	ω , élément de Ω
événement	A , sous-ensemble de Ω
A est réalisé	$\omega \in A$
A implique B	$A \subset B$
A ou B	$A \cup B$
A et B	$A \cap B$
contraire de A	A^c , complémentaire de A
événement impossible	\emptyset
événement certain	Ω
A et B sont incompatibles	$A \cap B = \emptyset$

Attention ! On ne considère pas en général tout sous-ensemble de Ω comme un événement. Par exemple, si on lance deux dés indiscernables, le résultat $\omega = (1, 2)$ n'est pas distinguable de $\omega = (2, 1)$ donc le sous-ensemble $\{(1, 2)\}$ n'est pas un événement.

2.2.3 Propriétés

2.2.3.a Réunion et intersection d'événements

On note $A \cup B$ l'événement réalisé si et seulement si l'un au moins des événements A et B est réalisé. L'événement $A \cup B$ pourrait se lire A ou B si le ou en français ne signifiait pas tantôt "soit l'un, soit l'autre", tantôt "au moins l'un des deux". On préférera donc dire "A union B".

On note $A \cap B$ l'événement réalisé si et seulement si les deux événements A et B sont réalisés. On le note parfois également $\{A; B\}$.

L'événement $A \cap B$ se lit A et B ou bien "A inter B".

Voici quelques propriétés :

$$\begin{aligned}
 A \cup A^c &= \Omega, A \cap A^c = \emptyset \\
 A \cup \Omega &= \Omega, A \cap \Omega = A \\
 A \cup \emptyset &= A, A \cap \emptyset = \emptyset \\
 A \cup A &= A, A \cap A = A \\
 (A \cap B)^c &= A^c \cup B^c, (A \cup B)^c = A^c \cap B^c
 \end{aligned}$$

Les opérations \cap et \cup sont **commutatives** et **associatives** :

$$\begin{aligned}
 A \cup B &= B \cup A, A \cap B = B \cap A \\
 (A \cup B) \cup C &= A \cup (B \cup C), (A \cap B) \cap C = A \cap (B \cap C)
 \end{aligned}$$

Les opérations \cap et \cup sont **distributives** l'une vis à vis de l'autre :

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

Ces propriétés se généralisent à une famille quelconque d'événements.

2.2.3.b Implication

On écrit $A \subseteq B$ si la réalisation de A implique celle de B, et on dit que A implique B ou que A est contenu dans B.

Si $A \subseteq B$ et $B \subseteq A$ alors $A = B$.

Les propositions $A \subseteq B^c$ ou $B \subseteq A^c$ équivalent à $A \cap B = \emptyset$: on dit dans ce cas que A et B sont **incompatibles** ou **disjoints**.

2.2.3.c Différence et différence symétrique

$B \cap A^c$ est un événement réalisé si B est réalisé sans que A le soit. En particulier, si $A \subseteq B$, il est commode d'écrire

$$B \cap A^c = B - A$$

la différence de B et de A.

La différence symétrique de deux événements A et B est l'événement réalisé si l'un des deux seulement est réalisé.

$$A \Delta B = (A \cap B^c) \cup (A^c \cap B) = (A \cup B) - (A \cap B).$$

Cela correspond au "ou exclusif" en français. Si $A \subseteq B$, on a évidemment $A \Delta B = B - A$.

2.2.4 Variable aléatoire

Quand on étudie un phénomène aléatoire, on est amené à étudier des grandeurs numériques (ou vectorielles) liées à celui-ci. En termes vagues, une variable aléatoire est un nombre (variable) lié à une expérience aléatoire, dont la valeur dépend exclusivement du résultat $\{\omega\}$ de cette expérience. Mathématiquement, il s'agit donc tout simplement d'une fonction sur l'ensemble Ω . Si cette fonction est à valeurs dans \mathbb{R} , on parle de **variable aléatoire réelle**. Nous donnerons une définition plus formelle par la suite.

Exemple 30 1) lancer de deux dés à 6 faces : notons X la somme obtenue sur les deux dés, qui n'est autre que la fonction

$$\begin{aligned} X : \quad \Omega &\mapsto \{2, \dots, 12\} \\ \omega = (\omega_1, \omega_2) &\mapsto \omega_1 + \omega_2 \end{aligned}$$

2) Si A est un événement aléatoire, la variable aléatoire indicatrice de A, notée $\mathbb{1}_A$, vaut 1 si A est réalisé, 0 sinon :

$$\begin{aligned} \mathbb{1}_A : \quad \Omega &\mapsto \{0, 1\} \\ \omega &\mapsto \mathbb{1}_{\{\omega \in A\}} \end{aligned}$$

2.3 Tribus et probabilités

Passons maintenant à la théorie mathématique proprement dite en commençant par introduire les notions de tribu et d'espace probabilisable, avant la définition des probabilités.

Définition 20 (Tribu) Une famille \mathcal{A} de sous-ensembles d'un ensemble Ω est une **tribu** ou **σ -algèbre** sur Ω si elle satisfait aux trois axiomes suivants :

- (i) $\Omega \in \mathcal{A}$.
- (ii) Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.
- (iii) Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Exemple 31 1) Si A est un sous-ensemble de Ω , la famille $\{A, A^c, \Omega, \emptyset\}$ est une tribu sur Ω dite **tribu engendrée** par l'événement A .

2) La tribu $\{\Omega, \emptyset\}$ est appelée **tribu triviale** sur Ω .

Définition 21 (Espace probabilisable) On appelle **espace probabilisable** un couple (Ω, \mathcal{A}) où Ω est un ensemble et \mathcal{A} une tribu sur l'ensemble Ω . Quand un espace probabilisable est fixé, on dit que \mathcal{A} est **la tribu des événements**.

La modélisation d'un phénomène aléatoire et d'une famille d'événements commence par le choix d'un espace probabilisable qui rend compte de l'ensemble des réalisations envisagées et des événements qui peuvent être sujets de l'étude. Nous pouvons maintenant définir la notion de probabilité.

Définition 22 (Probabilité) Soit (Ω, \mathcal{A}) un espace probabilisable. Une **probabilité** \mathbb{P} sur cet espace est une application de \mathcal{A} dans \mathbb{R}^+ satisfaisant :

- (i) $\mathbb{P}(\Omega) = 1$.
- (ii) **Propriété de σ -additivité** : pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments disjoints deux à deux de \mathcal{A} ,

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé **espace probabilisé** ou **espace de probabilité**.

Propriété 2 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On peut montrer les propriétés suivantes :

- (i) On a : $\mathbb{P}(\emptyset) = 0$.
- (ii) Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments disjoints deux à deux de \mathcal{A} ,

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

- (iii) Si $A, B \in \mathcal{A}$ et si $A \subseteq B$, on a

$$\mathbb{P}(A) \leq \mathbb{P}(B) \quad \text{et} \quad \mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A).$$

- (iv) Pour tout $A \in \mathcal{A}$, on a : $\mathbb{P}(A) \in [0, 1]$.
- (v) Pour tout $A \in \mathcal{A}$, on a : $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- (vi) Pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , c'est-à-dire tels que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$, on a :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

(vii) Pour toute suite décroissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} , c'est-à-dire tels que $A_{n+1} \subset A_n$ pour tout $n \in \mathbb{N}$, on a :

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Propriété 3 (Formule de Poincaré) Si $n \geq 2$, pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'éléments de \mathcal{A} , on a :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n).$$

Démonstration. On démontre cette propriété par récurrence.

– Montrons d'abord qu'elle est vraie à l'ordre 2. On a $\Omega = \Omega \cap \Omega = (A \cup A^c) \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \cup (A^c \cap B^c)$. Ces quatre événements étant incompatibles deux à deux, on a, d'après le point (ii) de la Proposition 1 : $\mathbb{P}(\Omega) = 1 = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A^c \cap B^c)$. Or, $\mathbb{P}(A^c \cap B^c) = \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B)$, ce qui nous donne finalement :

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

De plus, $A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) \Rightarrow \mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.

De même, $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$ et finalement :

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

– Supposons qu'elle est vraie à l'ordre n . Soit une suite $(A_i)_{1 \leq i \leq n+1}$ d'éléments de \mathcal{A} . On a : $\bigcup_{i=1}^{n+1} A_i = (\bigcup_{i=1}^n A_i) \cup A_{n+1}$. Ces deux termes étant des éléments de \mathcal{A} , on peut leur appliquer la formule de Poincaré à l'ordre 2 :

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left[\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1}\right] \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right). \end{aligned}$$

En appliquant la formule de Poincaré à l'ordre n aux premier et dernier termes, puis en regroupant convenablement tous les termes, on obtient la formule à l'ordre $n + 1$.

2.4 Espaces probabilisés discrets

2.4.1 Définition

Un ensemble **dénombrable** est, par définition, un ensemble qui peut être mis en bijection avec une partie de \mathbb{N} : un ensemble fini est donc dénombrable.

Définition 23 (Espace probabilisable discret) On appelle **espace probabilisable discret** un espace probabilisable (Ω, \mathcal{A}) où Ω est dénombrable et où $\mathcal{A} = \mathcal{P}(\Omega)$, l'ensemble des parties (ou sous-ensembles) de Ω .

Une probabilité \mathbb{P} sur un espace probabilisable discret $(\Omega, \mathcal{P}(\Omega))$ est définie dès que l'on connaît les probabilités des réalisations élémentaires $\{\omega\}$. En effet, puisque ces réalisations sont incompatibles deux à deux, on a :

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

pour tout sous-ensemble A de Ω .

2.4.2 Probabilité uniforme sur un ensemble fini

La probabilité uniforme sur un ensemble fini Ω est celle qui attribue la même valeur à la probabilité de chaque réalisation élémentaire $\{\omega\}$:

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|},$$

où $|\Omega|$ désigne le cardinal (nombre d'éléments) de Ω . Par conséquent, on a :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

pour tout événement $A \subset \Omega$.

Généralement, quand on parle de choix "au hasard" sur un ensemble fini, on sous-entend que ce choix est fait au moyen de la probabilité uniforme, c'est-à-dire en donnant à chaque élément de l'ensemble les mêmes chances d'être choisi. On s'aperçoit également qu'il est impossible de définir une probabilité uniforme sur un ensemble non fini.

Exemple 32 (Modélisation) On lance n fois un dé à 6 faces équilibré et on cherche la probabilité d'obtenir k fois ($k \leq n$) le nombre 6. Une réalisation est une succession de n entiers compris entre 1 et 6 donc $\Omega = \{1, \dots, 6\}^n$ et $|\Omega| = 6^n$. L'événement "obtenir k fois le nombre 6" est représenté par le sous-ensemble

$$A_k = \{(x_1, \dots, x_n) | x_i = 6 \text{ pour } k \text{ indices exactement}\}.$$

Pour obtenir une réalisation appartenant à A_k , il nous faut d'abord choisir les k indices des x_i valant 6, puis affecter aux x_i restants un entier entre 1 et 5. On a donc $|A_k| = C_n^k 5^{n-k}$ et

$$\mathbb{P}(A_k) = C_n^k \frac{5^{n-k}}{6^n},$$

où $C_n^k = \frac{n!}{k!(n-k)!}$ est le nombre de combinaisons de k objets parmi n .

Si $n = 3$, on a $A_0 = \frac{5^3}{6^3} = \frac{125}{216}$, $A_1 = 3 \times \frac{5^2}{6^3} = \frac{75}{216}$, $A_2 = 3 \times \frac{5}{6^3} = \frac{15}{216}$ et $A_3 = \frac{1}{216}$. On vérifie que

$$\mathbb{P}(\Omega) = \mathbb{P}(A_0 \cup A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_0) + \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) = 1.$$

2.5 Indépendance d'événements et de variables aléatoires

2.5.1 Événements indépendants

Définition 24 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

(i) Deux événements $A \in \mathcal{A}$ et $B \in \mathcal{A}$ sont **indépendants** si :

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

(ii) Soit $(A_i)_{i \in I}$ une famille quelconque d'événements. Ces événements sont dits **indépendants** (dans leur ensemble) si, pour toute partie non vide J de I , on a :

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Remarque 7 1) L'indépendance des événements A_1, A_2, \dots, A_n représente $2^n - n - 1$ conditions (nombre de parties de $I = \{1, \dots, n\}$ de cardinal ≥ 2).

2) Si n événements sont indépendants dans leur ensemble, ils sont indépendants deux à deux. Mais la réciproque n'est pas nécessairement vraie.

3) On peut avoir :

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

sans que les événements soient indépendants deux à deux.

Exemple 33 1) On considère l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, où $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ est un ensemble à quatre éléments muni de la tribu de ses parties et où \mathbb{P} est la probabilité uniforme sur Ω . On considère les événements

$$A = \{\omega_1, \omega_2\}, \quad B = \{\omega_1, \omega_3\}, \quad C = \{\omega_1, \omega_4\}.$$

On a :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$$

et

$$\mathbb{P}(A) \mathbb{P}(B) = \mathbb{P}(A) \mathbb{P}(C) = \mathbb{P}(B) \mathbb{P}(C) = \frac{1}{4}.$$

Les événements A , B et C sont donc indépendants deux à deux, mais

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \text{ et } \mathbb{P}(A) \mathbb{P}(B) \mathbb{P}(C) = \frac{1}{8}$$

donc les événements A , B et C ne sont pas indépendants dans leur ensemble.

2) On considère l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, où $\Omega = \Omega_1 \times \Omega_2$, avec $\Omega_1 = \Omega_2 = \{1, 2, \dots, 6\}$. L'ensemble Ω est muni de la tribu de ses parties et \mathbb{P} est la probabilité uniforme sur Ω . On considère les trois événements :

$$A = \Omega_1 \times \{1, 2, 5\}, \quad B = \Omega_1 \times \{4, 5, 6\}, \quad C = \{(i, j) \in \Omega \mid i + j = 9\}.$$

En fait, on a $C = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$ donc $\mathbb{P}(C) = \frac{4}{36} = \frac{1}{9}$.

On a également :

$$\begin{aligned} \mathbb{P}(A \cap B) &= \frac{6}{36} \neq \frac{1}{2} \frac{1}{2} = \mathbb{P}(A) \mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \frac{1}{36} \neq \frac{1}{2} \frac{1}{9} = \mathbb{P}(A) \mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \frac{3}{36} \neq \frac{1}{2} \frac{1}{9} = \mathbb{P}(B) \mathbb{P}(C). \end{aligned}$$

Toutefois, on a $\mathbb{P}(A \cap B \cap C) = \frac{1}{36} = \mathbb{P}(A) \mathbb{P}(B) \mathbb{P}(C)$.

Les événements A , B et C sont donc trois à trois indépendants mais ne sont pas indépendants deux à deux, donc pas indépendants dans leur ensemble.

2.5.2 Événements indépendants et passage au complémentaire

Dire que A et B sont indépendants équivaut à dire que la réalisation ou la non-réalisation de B n'influe pas sur la probabilité de voir A se réaliser. Il est naturel de penser qu'on peut remplacer dans cette formulation B par l'événement contraire B^c .

Propriété 4 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Si A et B sont deux événements indépendants, les événements A et B^c d'une part, A^c et B d'autre part, A^c et B^c enfin, sont indépendants.

Démonstration. Nous ne faisons la démonstration que pour A et B^c . On a

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$$

et, par indépendance de A et B :

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A) \mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A) \mathbb{P}(B^c).$$

Cette proposition se généralise à toute famille $(A_i)_{i \in I}$ d'événements indépendants. Supposons que l'ensemble I soit partitionné en deux parties I_1 et I_2 . Les événements $(B_i)_{i \in I}$ définis par :

$$B_i = \begin{cases} A_i & \text{si } i \in I_1, \\ A_i^c & \text{si } i \in I_2, \end{cases}$$

forment une famille d'événements indépendants.

2.5.3 Variables aléatoires indépendantes

2.5.3.a Généralités

Définition 25 Toutes les variables aléatoires considérées sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) Soit deux variables aléatoires X_1 et X_2 à valeurs respectivement dans des espaces probabilisables (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) . Les variables aléatoires X_1 et X_2 sont **indépendantes** si, pour tout $A_1 \in \mathcal{E}_1$ et pour tout $A_2 \in \mathcal{E}_2$, les événements $(X_1 \in A_1)$ et $(X_2 \in A_2)$ sont indépendants.
- (ii) Cette notion se généralise à une famille quelconque de variables aléatoires. Soit $(X_i)_{i \in I}$ une famille quelconque de variables aléatoires à valeurs dans des espaces probabilisables respectifs (E_i, \mathcal{E}_i) . Les variables aléatoires X_i sont **indépendantes** si pour toute famille $(A_i)_{i \in I}$ d'ensembles telle que, pour tout $i \in I$, $A_i \in \mathcal{E}_i$, les événements $(X_i \in A_i)$, $i \in I$, sont indépendants.

Propriété 5 Soit I un ensemble fini non vide. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires à valeurs respectivement dans des espaces probabilisables (E_i, \mathcal{E}_i) . Les variables aléatoires X_i sont indépendantes si et seulement si pour toute famille $(A_i)_{i \in I}$ d'ensembles telle que, pour tout $i \in I$, $A_i \in \mathcal{E}_i$, on a la relation :

$$\mathbb{P} \left[\bigcap_{i \in I} (X_i \in A_i) \right] = \prod_{i \in I} \mathbb{P}(X_i \in A_i).$$

2.5.3.b Fonctions de variables aléatoires

Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans un espace probabilisable (E, \mathcal{E}) . Soit f une variable aléatoire définie sur (E, \mathcal{E}) , à valeurs dans un espace (F, \mathcal{F}) . Alors l'application composée $f \circ X$ est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans (F, \mathcal{F}) : en effet, pour tout $B \in \mathcal{F}$, l'ensemble

$$(f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B))$$

appartient à \mathcal{A} . Plutôt que $f \circ X$, cette variable aléatoire est classiquement notée $f(X)$. On parle à son propos de **fonction de la variable aléatoire** X .

Le résultat ci-dessous affirme que "des fonctions de variables aléatoires indépendantes sont indépendantes". Il est très utile car il permet souvent de démontrer sans calculs que des variables aléatoires sont indépendantes.

Propriété 6 Soit $(X_i)_{i \in I}$ une famille quelconque de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs respectivement dans des espaces probabilisables (E_i, \mathcal{E}_i) . On suppose que ces variables aléatoires sont indépendantes. Soit, pour tout $i \in I$, une variable aléatoire f_i définie sur l'espace probabilisable (E_i, \mathcal{E}_i) , à valeurs dans un espace probabilisable (F_i, \mathcal{F}_i) . Alors les variables aléatoires $f_i(X_i)$ sont indépendantes.

Exemple 34 On peut affirmer sans calcul que si X_1 et X_2 sont deux variables aléatoires indépendantes à valeurs dans \mathbb{R} , les variables aléatoires X_1^2 et $\exp(X_2)$ sont indépendantes.

On peut également considérer des fonctions de plusieurs variables aléatoires. Soit X_1, \dots, X_n des variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs respectivement dans $((E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n))$. Soit F un ensemble quelconque et soit $f : E_1 \times \dots \times E_n \mapsto F$ une application quelconque. Alors la composée de l'application $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$ de Ω dans $E_1 \times \dots \times E_n$ et de l'application f se note $f(X_1, \dots, X_n)$, et c'est une variable aléatoire à valeurs dans F .

Propriété 7 Soit $f_1 : E_1 \times \dots \times E_k \mapsto F_1$ et $f_2 : E_{k+1} \times \dots \times E_n \mapsto F_2$ des applications quelconques. Si les variables aléatoires X_1, \dots, X_n sont indépendantes, alors les variables aléatoires $f_1(X_1, \dots, X_k)$ et $f_2(X_{k+1}, \dots, X_n)$ sont indépendantes.

Exemple 35 On peut affirmer sans calcul que si X_1, \dots, X_4 sont quatre variables aléatoires indépendantes à valeurs dans \mathbb{R} , les variables aléatoires $X_1^2 + X_2$ et $\exp(X_3 + X_4)$ sont indépendantes.

2.6 Probabilités conditionnelles

Dans la suite, \mathbb{P} est une probabilité quelconque sur l'espace probabilisable (Ω, \mathcal{A}) .

2.6.1 Définitions. Formule des probabilités totales

Définition 26 Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. Soit $A \in \mathcal{A}$ un événement ; on appelle **probabilité de A conditionnée par B (ou sachant B)** le nombre réel, noté $\mathbb{P}(A|B)$ ou $\mathbb{P}^B(A)$, défini par :

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

Propriété 8 Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. L'application \mathbb{P}^B de \mathcal{A} dans \mathbb{R}^+ qui à tout $A \in \mathcal{A}$ fait correspondre $\mathbb{P}^B(A)$ est une probabilité sur l'espace probablisable (Ω, \mathcal{A}) . Elle est appelée **probabilité conditionnée à B** ou **probabilité conditionnelle sachant B**.

Démonstration. On a d'abord

$$\mathbb{P}^B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

On montre ensuite que l'application \mathbb{P}^B est σ -additive : en effet, pour toute suite $(A_i)_{i \in \mathbb{N}}$ d'événements disjoints deux à deux on a

$$\left(\bigcup_{i \in \mathbb{N}} A_i \right) \cap B = \bigcup_{i \in \mathbb{N}} (A_i \cap B),$$

les $A_i \cap B$ étant disjoints deux à deux, et donc

$$\mathbb{P}^B\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \frac{\sum_{i \in \mathbb{N}} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i \in \mathbb{N}} \mathbb{P}^B(A_i).$$

Remarque 8 1) Cette proposition est très importante puisqu'elle dit que toutes les propriétés établies jusqu'à maintenant pour une probabilité quelconque sont vraies pour la probabilité conditionnelle \mathbb{P}^B .

2) Pour que les événements A et B soient **indépendants**, il faut et il suffit que : $\mathbb{P}^B(A) = \mathbb{P}(A)$.

L'exemple ci-dessous montre comment une probabilité conditionnelle rend compte de l'information apportée par la réalisation d'un événement sur la réalisation d'un autre événement.

Exemple 36 On jette deux fois un dé à 6 faces. Soit A l'événement "on obtient un 6 au premier jet", et soit $B_k, 2 \leq k \leq 12$, les événements "la somme des deux résultats est k ". On modélise les deux lancers de dé par l'espace probablisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, où $\Omega = \{1, 2, \dots, 6\}^2$ et \mathbb{P} est la probabilité uniforme. On a :

$$A = \{6\} \times \{1, 2, \dots, 6\}$$

et

$$B_k = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = k\}.$$

En particulier, on a $B_{12} \subset A$ et $B_{11} = \{(5, 6), (6, 5)\}$ donc

$$\mathbb{P}(A) = \frac{1}{6}, \quad \mathbb{P}(A|B_{11}) = \frac{1}{2}, \quad \mathbb{P}(A|B_{12}) = 1.$$

La proposition suivante résulte immédiatement de la définition.

Propriété 9 Si $\mathbb{P}(B) > 0$, on a :

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

Elle se généralise de la façon suivante.

Propriété 10 (Probabilités conditionnelles en cascade) Soit $(A_i)_{1 \leq i \leq n}$ une suite finie d'événements tels que $\mathbb{P}\left(\bigcap_{1 \leq i \leq n-1} A_i\right) > 0$. On a :

$$\mathbb{P}\left(\bigcap_{1 \leq i \leq n} A_i\right) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Démonstration. On remarque que toutes les probabilités conditionnelles introduites sont bien définies puisque, pour tout $j \in \{1, \dots, n-1\}$, on a :

$$\mathbb{P}\left(\bigcap_{1 \leq i \leq j} A_i\right) \geq \mathbb{P}\left(\bigcap_{1 \leq i \leq n-1} A_i\right) > 0.$$

La formule s'obtient ensuite par simple récurrence.

Exemple 37 On tire successivement 4 cartes d'un jeu de 52 cartes. Quelle est la probabilité de tirer les 4 as ?

Solution. Notons $A_i, 1 \leq i \leq 4$, l'événement "la $i^{\text{ème}}$ carte tirée est un as". On demande de calculer $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$. Les probabilités immédiatement accessibles, en dehors de $\mathbb{P}(A_1)$, sont les probabilités conditionnelles

$$\mathbb{P}(A_2|A_1) = \frac{3}{51}, \quad \mathbb{P}(A_3|A_1 \cap A_2) = \frac{2}{50}, \quad \mathbb{P}(A_4|A_1 \cap A_2 \cap A_3) = \frac{1}{49}.$$

On a donc

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \mathbb{P}(A_4|A_1 \cap A_2 \cap A_3) = \frac{4 \times 3 \times 2 \times 1}{52 \times 51 \times 50 \times 49} \approx 0,000004.$$

Nous ne sommes pas préoccupés dans cet exemple de la construction d'un espace Ω .

Définition 27 Soit $(A_i)_{i \in I}$ une famille **dénombrable** d'événements disjoints deux à deux telle que

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = 1.$$

Une telle famille est appelée **système complet d'événements**.

Soit N le complémentaire de $\bigcup_{i \in I} A_i$ dans Ω . On a $\mathbb{P}(N) = 0$. Les événements $(A_i)_{i \in I}$ et N forment une partition de Ω . On traduit cette situation en termes probabilistes : avec probabilité 1, l'un des événements A_i et un seul se réalise.

Exemple 38 On lance une pièce de monnaie jusqu'à ce qu'on obtienne un "pile". Soit A_i l'événement "pile apparaît pour la première fois au $i^{\text{ème}}$ lancer". Les A_i sont évidemment deux à deux disjoints. D'autre part, il est possible qu'aucun des A_i ne se réalise mais la probabilité de cet événement est nulle. La famille des événements $(A_i)_{i=1,2,\dots}$ est donc un système complet.

Théorème 1 (Formule des probabilités totales) Soit $(A_i)_{i \in I}$ un système complet d'événements tel que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $A \in \mathcal{A}$:

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|A_i) \mathbb{P}(A_i).$$

Démonstration. Soit N le complémentaire de $\bigcup_{i \in I} A_i$.

$$A = \left[\bigcup_{i \in I} (A \cap A_i) \right] \cup (A \cap N),$$

les deux événements étant disjoints. Comme on a $\mathbb{P}(A \cap N) \leq \mathbb{P}(N) = 0$, on obtient immédiatement

$$\mathbb{P}(A) = \mathbb{P} \left[\bigcup_{i \in I} (A \cap A_i) \right] = \sum_{i \in I} \mathbb{P}[(A \cap A_i)]$$

car les $A \cap A_i$ sont eux aussi disjoints deux à deux. Il ne reste plus alors qu'à utiliser la définition des probabilités conditionnelles.

Un cas particulier de système complet est le système (B, B^c) où B est un événement tel que $0 < \mathbb{P}(B) < 1$. La formule des probabilités totales devient dans ce cas

$$\mathbb{P}(A) = \mathbb{P}(A|B) \mathbb{P}(B) + \mathbb{P}(A|B^c) \mathbb{P}(B^c).$$

Exemple 39 Une urne U_1 (respectivement U_2) contient b_1 (respectivement b_2) boules blanches et n_1 (respectivement n_2) boules noires. On choisit au hasard une urne et on tire ensuite une boule dans cette urne. On cherche la probabilité de tirer une boule noire.

Solution. Notons N , U_1 et U_2 les événements "on tire une boule noire", "le tirage a lieu dans l'urne U_1 " et "le tirage a lieu dans l'urne U_2 ". On a :

$$\mathbb{P}(U_i) = \frac{1}{2} \text{ et } \mathbb{P}(N|U_i) = \frac{n_i}{n_i + b_i}.$$

Les événements U_1 et U_2 forment un système complet donc

$$\mathbb{P}(N) = \mathbb{P}(U_1) \mathbb{P}(N|U_1) + \mathbb{P}(U_2) \mathbb{P}(N|U_2) = \frac{1}{2} \left(\frac{n_1}{n_1 + b_1} + \frac{n_2}{n_2 + b_2} \right).$$

2.6.2 Formule de Bayes

Propriété 11 Soit deux événements $A_1, A_2 \in \mathcal{A}$ de probabilités non nulles. Alors :

$$\mathbb{P}(A_2|A_1) = \mathbb{P}(A_1|A_2) \frac{\mathbb{P}(A_2)}{\mathbb{P}(A_1)}.$$

Cette proposition s'obtient juste en utilisant la définition d'une probabilité conditionnelle.

Théorème 2 (Formule de Bayes) Soit $(A_i)_{i \in I}$ un système complet d'événements tel que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $A \in \mathcal{A}$ de probabilité non nulle et pour tout $i \in I$:

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i) \mathbb{P}(A_i)}{\sum_{j \in I} \mathbb{P}(A|A_j) \mathbb{P}(A_j)}.$$

Démonstration. Il résulte de la proposition 8 que, pour tout $i \in I$:

$$\mathbb{P}(A_i|A) = \mathbb{P}(A|A_i) \frac{\mathbb{P}(A_i)}{\mathbb{P}(A)}.$$

Il suffit alors d'appliquer le théorème 1.

Exemple 40 On reprend l'exemple des deux urnes. On cherche maintenant la probabilité **a posteriori** d'avoir tiré dans l'urne U_i , sachant qu'une boule noire a été tirée.

Solution. On applique le théorème de Bayes avec le système complet formé des événements U_1 et U_2 . Il vient, pour $i = 1, 2$,

$$\mathbb{P}(U_i|N) = \frac{\mathbb{P}(N|U_i) \mathbb{P}(U_i)}{\mathbb{P}(N|U_1) \mathbb{P}(U_1) + \mathbb{P}(N|U_2) \mathbb{P}(U_2)} = \frac{\frac{n_i}{n_i+b_i}}{\frac{n_1}{n_1+b_1} + \frac{n_2}{n_2+b_2}}.$$

On remarque qu'on a bien $\mathbb{P}(U_1|N) + \mathbb{P}(U_2|N) = \mathbb{P}(U_1 \cup U_2|N) = \mathbb{P}(\Omega|N) = 1$.

3 Variables aléatoires discrètes

Mots-clés :

- Variable aléatoire discrète, loi de probabilité.
- Variable aléatoire de loi uniforme, de Bernoulli, binomiale, géométrique, hypergéométrique, de poisson.
- Moment d'ordre p , covariance, variance.
- Variable aléatoire centrée, variable aléatoire réduite.

3.1 Loi de probabilité

3.1.1 Généralités

Définition 28 (Variable aléatoire discrète) Soit (Ω, \mathcal{A}) un espace probabilisable et soit E un ensemble. On dit qu'une application X de Ω dans E est une **variable aléatoire discrète** si les deux conditions sont satisfaites :

- (i) L'ensemble $X(\Omega)$ des valeurs prises par X est dénombrable.
- (ii) Pour tout $x \in E$, on a $X^{-1}(\{x\}) \in \mathcal{A}$.

La plupart du temps, on aura $E = \mathbb{N}$ ou \mathbb{Z} et on munira E de la tribu $\mathcal{E} = \mathcal{P}(E)$.

Propriété 12 Soit X une variable aléatoire définie sur l'espace probabilisable (Ω, \mathcal{A}) à valeurs dans l'espace probabilisable (E, \mathcal{E}) . L'application $A \mapsto \mathbb{P}[X^{-1}(A)]$ de \mathcal{E} dans $[0, 1]$ est une probabilité sur l'espace probabilisable (E, \mathcal{E}) . Elle est notée \mathbb{P}_X . Cette probabilité est appelée **loi (de probabilité) de la variable aléatoire X** .

La loi d'une variable aléatoire est une probabilité sur l'espace des valeurs prises par celle-ci. Si X est une variable aléatoire discrète à valeurs dans (E, \mathcal{E}) , sa loi est déterminée entièrement par la fonction $x \mapsto \mathbb{P}(X = x)$ sur E , désignée comme la **fonction de masse** de la variable X . En effet, pour tout $A \in \mathcal{E}$, on a :

$$\mathbb{P}[X^{-1}(A)] = \mathbb{P}\left[\bigcup_{x \in X(\Omega) \cap A} X^{-1}(\{x\})\right] = \sum_{x \in X(\Omega) \cap A} \mathbb{P}[X^{-1}(\{x\})] = \sum_{x \in A} \mathbb{P}(X = x).$$

3.1.2 Quelques lois usuelles

3.1.2.a Loi uniforme

Comme nous l'avons vu précédemment, cette loi accorde la même probabilité à chaque élément d'un ensemble fini, que nous noterons $\{x_1, \dots, x_n\}$. Si X est l'un des éléments de cet ensemble pris au hasard, c'est une variable aléatoire discrète à valeurs dans $E = \{x_1, \dots, x_n\}$. On dit que X est une variable aléatoire de loi uniforme sur l'ensemble E et on écrit $X \sim \mathcal{U}(\{x_1, \dots, x_n\})$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = \frac{1}{n}$$

pour $x \in E = \{x_1, \dots, x_n\}$. On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{i=1}^n \mathbb{P}(X = x_i) = \sum_{i=1}^n \frac{1}{n} = 1.$$

Exemple 41 1) On effectue un tirage avec un dé à 6 faces équilibré. Soit X le résultat : il suit une loi uniforme sur $\{1, \dots, 6\}$.

2) Une urne est remplie de 10 boules numérotées de 1 à 10. On tire une boule et on note X son numéro : X suit une loi uniforme sur $\{1, \dots, 10\}$.

3.1.2.b Loi de Bernoulli

Cette loi de probabilité sert à coder de manière numérique la réalisation ou non d'un événement. Soit A un événement aléatoire de probabilité $\mathbb{P}(A) = p \in]0, 1[$. On définit la variable aléatoire

$$X = \begin{cases} 1 & \text{si } A \text{ se réalise,} \\ 0 & \text{si } A^c \text{ se réalise.} \end{cases}$$

On dit que X suit la loi de Bernoulli de paramètre p et on écrit : $X \sim \mathcal{B}(1, p)$. X est une variable aléatoire discrète à valeurs dans $E = \{0, 1\}$ et sa fonction de masse est :

$$\mathbb{P}(X = x) = p^x q^{1-x}$$

pour $x \in E = \{0, 1\}$, avec $q = 1 - p$. On vérifie que

$$\mathbb{P}_X(E) = \mathbb{P}(X \in E) = \sum_{x \in E} \mathbb{P}(X = x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = p + q = 1.$$

Exemple 42 1) On lance une pièce de monnaie équilibrée. Soit A l'événement "face". Ici $X = 1$ si la pièce tombe sur face et 0 sinon. On a $p = \frac{1}{2}$.

2) On lance un dé équilibré. Soit A l'événement "résultat < 3". Ici $X = 1$ si le résultat est 1 ou 2 et $X = 0$ sinon. On a $p = \frac{1}{3}$.

3) Observation du sexe du prochain bébé qui naîtra dans une maternité : $A = \text{"fille"}$. Ici $X = 1$ si c'est une fille et 0 sinon. Selon des statistiques de 2008 (<http://www.bartleby.com/151/fields/31.htm>) en France, $p = \frac{1}{2.05} \approx 0.4878$. Noter que p varie selon les pays (voir le site internet).

3.1.2.c Loi binomiale

Cette loi de probabilité sert à modéliser le nombre de réalisations d'un événement. Soit A un événement aléatoire de probabilité $\mathbb{P}(A) = p \in]0, 1[$, susceptible de se produire lors d'une expérience aléatoire. On se propose de répéter l'expérience n fois, de façon indépendante. À la $i^{\text{ème}}$ répétition, on observera si A s'est réalisé ou non et on y associera la variable aléatoire $Y_i \sim \mathcal{B}(1, p)$.

Soit X le nombre de réalisations de A au cours des n répétitions : $X = Y_1 + Y_2 + \dots + Y_n$ est la somme de n variables aléatoires de loi de Bernoulli indépendantes. C'est donc une variable aléatoire discrète à valeurs dans $E = \{0, 1, \dots, n\}$. On dit que X est une variable aléatoire de loi binomiale de paramètres n et p et on écrit $X \sim \mathcal{B}(n, p)$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = C_n^x p^x q^{n-x}$$

pour $x \in E = \{0, 1, \dots, n\}$, où $C_n^x = \frac{n!}{x!(n-x)!}$ est le nombre de combinaisons de x éléments parmi n . On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{x=0}^n C_n^x p^x q^{n-x} = (p + q)^n = 1.$$

Exemple 43 1) On effectue des tirages avec remise dans une urne remplie de boules, dont $\frac{1}{3}$ de blanches et $\frac{2}{3}$ de rouges. Soit X le nombre de boules blanches tirées après 10 lancers : on a $X \sim \mathcal{B}(10, \frac{1}{3})$.

2) On lance une pièce de monnaie équilibrée 6 fois de suite. On s'intéresse à la probabilité d'obtenir, au cours de ces 6 lancers, 2 fois l'événement $A = \text{"face"}$. Si on note X le nombre de "face" sur 6 lancers, on a $X \sim \mathcal{B}(6, \frac{1}{2})$ de sorte que :

$$\mathbb{P}(X = 2) = C_6^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2} = \frac{6!}{2!4!} \left(\frac{1}{2}\right)^6 = \frac{15}{64}.$$

3) On estime qu'en France à l'instant t , la probabilité qu'un individu soit un fumeur est de 0.2. Un bus contient 6 personnes. Quelle est la probabilité que 4 ou plus de ces passagers soient fumeurs ? On suppose leur indépendance. Soit X le nombre de fumeurs parmi les 6 personnes : on a $X \sim \mathcal{B}(6, 0.2)$, $E = \{0, 1, \dots, 6\}$ et

$$\begin{aligned} \mathbb{P}(X \geq 4) &= \mathbb{P}(\{X = 4\} \cup \{X = 5\} \cup \{X = 6\}) \\ &= \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) \\ &= C_6^4 (0.2)^4 (0.8)^2 + C_6^5 (0.2)^5 (0.8)^1 + C_6^6 (0.2)^6 \\ &= 0.01536 + 0.001536 + 0.000064 \\ &= 0.01696 \end{aligned}$$

3.1.2.d Loi géométrique

Cette loi de probabilité sert à modéliser le nombre d'essais nécessaires à la réalisation d'un événement. Soit A un événement aléatoire de probabilité $\mathbb{P}(A) = p \in]0, 1[$, susceptible de se produire lors d'une expérience aléatoire. On se propose de répéter l'expérience, de façon indépendante, tant que l'événement A n'est pas réalisé.

Soit X le nombre de répétitions de l'expérience. C'est une variable aléatoire discrète à valeurs dans $E = \mathbb{N}^*$. On dit que X est une variable aléatoire de loi géométrique de paramètre p et on écrit $X \sim \mathcal{G}(p)$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = pq^{x-1}$$

pour $x \in E = \mathbb{N}^*$. On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{x=1}^{+\infty} pq^{x-1} = p \frac{1}{1-q} = 1.$$

Exemple 44 1) On lance une pièce de monnaie équilibrée jusqu'à la réalisation de l'événement $A = \text{"face"}$. Si on note X le nombre de lancers nécessaires, on a $X \sim \mathcal{G}(\frac{1}{2})$ de sorte que :

$$\mathbb{P}(X = x) = \frac{1}{2^x}.$$

2) On lance un dé à 6 faces jusqu'à l'obtention d'un 6. Si on note X le nombre de lancers nécessaires, on a $X \sim \mathcal{G}(\frac{1}{6})$ de sorte que :

$$\mathbb{P}(X = x) = \frac{1}{6} \times \frac{5^{x-1}}{6^{x-1}}.$$

3.1.2.e Loi hypergéométrique

Cette loi de probabilité sert à modéliser le résultat d'un tirage **sans remise**. Soit U un ensemble de cardinal r partitionné en deux sous-ensembles U_1 et U_2 , de cardinaux respectifs r_1 et $r_2 = r - r_1$. On se propose de choisir "au hasard" n éléments de U ($n \leq r$) et on note X le nombre d'éléments de U_1 parmi ces n . C'est une variable aléatoire discrète à valeurs dans $E = \{\max(0, n - r_2), \dots, \min(n, r_1)\}$. On dit que X est une variable aléatoire de loi hypergéométrique de paramètres n , r et r_1 et on écrit $X \sim \mathcal{H}(n, r, r_1)$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = \frac{C_{r_1}^x C_{r_2}^{n-x}}{C_r^n}$$

pour $x \in E = \{\max(0, n - r_2), \dots, \min(n, r_1)\}$. On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{x=\max(0, n-r_2)}^{\min(n, r_1)} \frac{C_{r_1}^x C_{r_2}^{n-x}}{C_r^n} = \frac{1}{C_r^n} \sum_{x=\max(0, n-r_2)}^{\min(n, r_1)} C_{r_1}^x C_{r_2}^{n-x} = \frac{1}{C_{r_1+r_2}^n} \sum_{x=0}^{r_1} C_{r_1}^x C_{r_2}^{n-x} = 1$$

d'après l'identité de Vandermonde.

Exemple 45 1) Un lac contient r poissons, dont r_1 sont d'une espèce intéressante. On pêche n poissons, en supposant que toutes les espèces se laissent aussi facilement attraper, et on note X le nombre de poissons de l'espèce intéressante parmi les poissons attrapés : X suit une loi hypergéométrique.

2) Ce modèle hypergéométrique est à la base de la théorie des sondages.

3.1.2.f Loi de Poisson

Cette loi de probabilité décrit le comportement du nombre d'évènements se produisant dans un laps de temps fixé, si ces évènements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l'évènement précédent. Soit X ce nombre d'occurrences. C'est une variable aléatoire discrète à valeurs dans $E = \mathbb{N}$. On dit que X est une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$ et on écrit $X \sim \mathcal{P}(\lambda)$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = \exp(-\lambda) \frac{\lambda^x}{x!}$$

pour $x \in E = \mathbb{N}$. On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{x=0}^{+\infty} \exp(-\lambda) \frac{\lambda^x}{x!} = \exp(-\lambda) \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} = 1.$$

Exemple 46 Actuellement, on utilise beaucoup cette loi dans les télécommunications (pour compter le nombre de communications dans un intervalle de temps donné), le contrôle de qualité statistique (nombre de défauts), la description de certains phénomènes liés à la désintégration radioactive, la biologie (mutations), la météorologie, la finance pour modéliser la probabilité de défaut d'un crédit, le Yield Management (pour estimer la demande de passagers).

3.2 Moments

Les moments d'une variable aléatoire sont des paramètres numériques qui donnent des renseignements sur la loi de cette variable aléatoire (sans toutefois, en général, la déterminer complètement). Les plus couramment utilisés sont la moyenne, ou espérance mathématique, et la variance.

3.2.1 Moyenne ou espérance mathématique

Définition 29 (Moyenne ou espérance mathématique) Soit X une variable aléatoire réelle discrète à valeurs dans E . Si la somme $\sum_{x \in E} |x| \mathbb{P}(X = x)$ est finie, on dit que la variable aléatoire X possède une **moyenne** ou **espérance mathématique**

$$\mathbb{E}(X) = \sum_{x \in E} x \mathbb{P}(X = x).$$

En langage courant, $\mathbb{E}(X)$ n'est autre que la **moyenne pondérée** des valeurs x prises par la variable aléatoire X , pondérées par la probabilité que X vaille x .

3.2.1.a Premières propriétés de l'espérance

Propriété 13 On peut montrer les propriétés suivantes :

(i) Soit un réel a . Si la variable aléatoire X est telle que $\mathbb{P}(X = a) = 1$, elle admet une moyenne égale à a et on écrit

$$\mathbb{E}(a) = a.$$

(ii) Toute variable aléatoire discrète bornée admet une moyenne.

(iii) Si $X \geq 0$ alors $\mathbb{E}(X) \geq 0$. Si $X \geq 0$ et $\mathbb{E}(X) = 0$ alors $X = 0$ avec probabilité 1.

(iv) Si X et Y possèdent une moyenne et vérifient $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

(v) Si X possède une moyenne, on a :

$$|\mathbb{E}(X)| \leq \mathbb{E}|X|.$$

Exemple 47 1) Si $X \sim \mathcal{U}(\{x_1, \dots, x_n\})$, sa moyenne est la moyenne arithmétique des x_i :

$$\mathbb{E}(X) = \frac{1}{n} \sum_{i=1}^n x_i.$$

2) Si $X \sim \mathcal{B}(1, p)$, sa moyenne est

$$\mathbb{E}(X) = 1 \times p + 0 \times q = p.$$

3.2.1.b Espérance d'une fonction de variable aléatoire

Théorème 3 (Théorème de transfert) Soit X une variable aléatoire discrète à valeurs dans E . Soit f une application de E dans \mathbb{R} . L'application composée $Y = f \circ X = f(X)$ est une variable aléatoire réelle discrète. Pour qu'elle admette une moyenne, il faut et il suffit que la somme

$$\sum_{x \in E} |f(x)| \mathbb{P}(X = x)$$

soit finie. Dans ce cas, on a :

$$\mathbb{E}(f(X)) = \sum_{x \in E} f(x) \mathbb{P}(X = x).$$

Propriété 14 Soit X et Y deux variables aléatoires réelles discrètes et a et b deux réels. On a :

$$\mathbb{E}(aX + bY) = a \mathbb{E}(X) + b \mathbb{E}(Y).$$

Cette égalité s'obtient en appliquant le théorème de transfert avec $Z = (X, Y)$, variable aléatoire discrète, et $f : (x, y) \mapsto ax + by$.

3.2.1.c Moyennes des lois discrètes classiques

Loi binomiale Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$. On rappelle que X peut s'écrire comme la somme de n variables aléatoires indépendantes Y_1, \dots, Y_n suivant la loi de Bernoulli $\mathcal{B}(1, p)$:

$$X = Y_1 + \dots + Y_n \Rightarrow \mathbb{E}(X) = \mathbb{E}(Y_1) + \dots + \mathbb{E}(Y_n) = np.$$

Loi géométrique Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$. On a :

$$\mathbb{E}(X) = \sum_{i=1}^{+\infty} ipq^{i-1} = p \sum_{i=0}^{+\infty} iq^{i-1} = p \frac{\partial}{\partial q} \left[\sum_{i=0}^{+\infty} q^i \right] = p \frac{\partial}{\partial q} \left[\frac{1}{1-q} \right] = p(1-q)^{-2} = \frac{1}{p}.$$

Loi de Poisson Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$. On a :

$$\mathbb{E}(X) = \sum_{i=0}^{+\infty} i \exp(-\lambda) \frac{\lambda^i}{i!} = \exp(-\lambda) \sum_{i=1}^{+\infty} \frac{\lambda^i}{(i-1)!} = \exp(-\lambda) \lambda \sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} = \lambda.$$

On remarque que, si l'on sait qu'une variable aléatoire suit une loi géométrique ou une loi de Poisson, pour identifier entièrement sa loi, il suffit de connaître sa moyenne, ce qui n'est pas le cas pour la loi binomiale.

3.2.2 Moments d'ordre deux et variance

3.2.2.a Généralités

Définition 30 (Moment d'ordre p) Soit X une variable aléatoire réelle discrète et p un entier supérieur ou égal à 1. Si $\mathbb{E}(|X|^p) < +\infty$, on dit que X admet un **moment d'ordre p** , qui n'est autre que le nombre réel $\mathbb{E}(X^p)$.

Propriété 15 (Inégalité de Cauchy-Schwarz).

Soit X et Y deux variables aléatoires réelles discrètes admettant un moment d'ordre deux. On a :

$$[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

Démonstration. Définissons le polynôme Q du second degré à coefficients réels

$$Q(\lambda) = \mathbb{E}([X + \lambda Y]^2) = \mathbb{E}(X^2) + 2\lambda\mathbb{E}(XY) + \lambda^2\mathbb{E}(Y^2).$$

Puisque $Q(\lambda) \geq 0$ pour tout $\lambda \in \mathbb{R}$, le discriminant est négatif :

$$\Delta = 4[\mathbb{E}(XY)]^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0 \Rightarrow [\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

Définition 31 (Variable aléatoire centrée) (i) Si X admet un moment d'ordre 1, la variable aléatoire $\dot{X} = X - \mathbb{E}(X)$ est appelée **variable aléatoire centrée** associée à X . On a $\mathbb{E}(\dot{X}) = 0$.

(ii) Si X admet un moment d'ordre 2, X admet une moyenne et le réel $\mathbb{E}[(X - \mathbb{E}(X))^2]$ est appelé **variance** de X et noté $\mathbb{V}(X)$. Sa racine carrée est appelée **écart-type** de X et notée σ_X . On a donc $\mathbb{V}(X) = \sigma_X^2$.

(iii) Si X admet un moment d'ordre 2 et que $\sigma_X \neq 0$, la variable aléatoire $\tilde{X} = \frac{X - \mathbb{E}(X)}{\sigma_X}$ est appelée **variable aléatoire centrée réduite** associée à X . On a $\mathbb{E}(\tilde{X}) = 0$ et $\mathbb{V}(\tilde{X}) = 1$.

D'après le théorème de transfert, la variance de X s'obtient de la manière suivante :

$$\mathbb{V}(X) = \sum_{x \in E} [x - \mathbb{E}(X)]^2 \mathbb{P}(X = x).$$

Propriété 16 Si X admet un moment d'ordre deux, on a :

(i)

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

(ii) Pour tous réels a et b , on a :

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X).$$

Cela découle de la linéarité de l'espérance.

3.2.2.b Covariance de deux variables aléatoires

Définition 32 (Covariance) Si X et Y admettent un moment d'ordre 2, il résulte de l'inégalité de Cauchy-Schwarz que la variable aléatoire $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$ admet une moyenne, que nous appelons **covariance** de X et Y :

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

La proposition suivante permet en général un calcul plus simple de la covariance de deux variables aléatoires et permet également de calculer la variance d'une somme de variables aléatoires.

Propriété 17 Si X et Y admettent un moment d'ordre 2, on a :

(i)

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \mathbb{E}(Y).$$

(ii)

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2 \text{cov}(X, Y).$$

Cela découle de la linéarité de l'espérance.

Propriété 18 Si X et Y admettent un moment d'ordre 1 et sont **indépendantes**, on a :

$$\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y).$$

Démonstration. On considère que X est à valeurs dans E_X et Y à valeurs dans E_Y . La variable aléatoire XY est donc discrète, à valeurs dans $E_Z = \{z = xy \in \mathbb{R} : x \in E_X, y \in E_Y\}$. On a :

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{z \in E_Z} z \mathbb{P}(XY = z) \\ &= \sum_{z \in E_Z} z \sum_{x \in E_X, y \in E_Y : xy=z} \mathbb{P}(X = x, Y = y) \\ &= \sum_{z \in E_Z} z \sum_{x \in E_X, y \in E_Y : xy=z} \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &= \sum_{z \in E_Z} \sum_{x \in E_X, y \in E_Y : xy=z} x y \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &= \sum_{x \in E_X, y \in E_Y} x y \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &= \sum_{x \in E_X} x \mathbb{P}(X = x) \sum_{y \in E_Y} y \mathbb{P}(Y = y) \\ &= \mathbb{E}(X) \mathbb{E}(Y). \end{aligned}$$

Ainsi, si X et Y sont **indépendantes**, on a :

$$\text{cov}(X, Y) = 0 \text{ et } \mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

Par extension, la variance d'une somme de variables aléatoires indépendantes est la somme de leurs variances. Attention ! L'égalité $\text{cov}(X, Y) = 0$ **n'entraîne pas** l'indépendance entre X et Y (cf TD3).

3.2.2.c Variance des lois discrètes classiques

Loi de Bernoulli Soit X une variable aléatoire suivant une loi de Bernoulli $\mathcal{B}(1, p)$. On a :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 1^2 \times p + 0^2 \times q - p^2 = p - p^2 = pq.$$

Loi binomiale Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$. On rappelle que X peut s'écrire comme la somme de n variables aléatoires indépendantes Y_1, \dots, Y_n suivant la loi de Bernoulli $\mathcal{B}(1, p)$ donc

$$\mathbb{V}(X) = \mathbb{V}(Y_1) + \dots + \mathbb{V}(Y_n) = npq.$$

Loi géométrique Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$. Calculons d'abord :

$$\mathbb{E}(X(X-1)) = \sum_{i=1}^{+\infty} i(i-1)pq^{i-1} = pq \sum_{i=0}^{+\infty} i(i-1)q^{i-2} = pq \frac{\partial^2}{\partial q^2} \left[\sum_{i=0}^{+\infty} q^i \right] = pq \frac{\partial^2}{\partial q^2} \left[\frac{1}{1-q} \right] = 2pq(1-q)^{-3} = \frac{2}{p}$$

ce qui nous permet d'obtenir :

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

Loi de Poisson Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$. Calculons d'abord :

$$\mathbb{E}(X(X-1)) = \sum_{i=0}^{+\infty} i(i-1) \exp(-\lambda) \frac{\lambda^i}{i!} = \exp(-\lambda) \sum_{i=2}^{+\infty} \frac{\lambda^i}{(i-2)!} = \exp(-\lambda) \lambda^2 \sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} = \lambda^2,$$

ce qui nous permet d'obtenir :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - [\mathbb{E}(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

On remarque que, pour une variable aléatoire de Poisson, le paramètre représente à la fois la moyenne et la variance.

3.2.3 Fonctions génératrices

3.2.3.a Généralités

Définition 33 (Fonction génératrice) Soit X une variable aléatoire à valeurs dans \mathbb{N} . La **fonction génératrice** de X est la fonction G_X qui, à tout $s \in \mathbb{R}$, associe

$$G_X(s) = \mathbb{E}(s^X)$$

lorsque cette quantité existe.

Propriété 19 Soit X une variable aléatoire à valeurs dans \mathbb{N} , de fonction génératrice G_X et de loi déterminée par :

$$\mathbb{P}(X = n) = p_n, \quad \forall n \in \mathbb{N}.$$

(i) Le domaine de définition de G_X contient l'intervalle $[-1, 1]$. On a :

$$\forall s \in [-1, 1], \quad |G_X(s)| \leq 1 \text{ et } G_X(1) = 1.$$

(ii) Pour tout $s \in [-1, 1]$, on a :

$$G_X(s) = \sum_{n=0}^{+\infty} p_n s^n.$$

(iii) La fonction G_X est continue sur $[-1, 1]$, et C^∞ sur $] - 1, 1[$.

(iv) La fonction génératrice G_X caractérise la loi de X puisque, pour tout $k \in \mathbb{N}$, on a :

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

3.2.3.b Fonctions génératrices des lois discrètes classiques

Loi binomiale Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$. On a, $\forall s \in [-1, 1]$:

$$G_X(s) = \sum_{i=0}^n C_n^i p^i q^{n-i} s^i = (ps + q)^n.$$

Loi géométrique Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$. On a, $\forall s \in [-1, 1]$:

$$G_X(s) = \sum_{i=0}^{+\infty} pq^{i-1} s^i = \frac{ps}{1 - qs}.$$

Loi de Poisson Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$. On a, $\forall s \in [-1, 1]$:

$$G_X(s) = \sum_{i=0}^{+\infty} \exp(-\lambda) \frac{\lambda^i}{i!} s^i = \exp(\lambda s - \lambda).$$

3.2.3.c Fonctions génératrices et moments

La fonction génératrice d'une variable aléatoire déterminant sa loi, il est naturel qu'elle donne ses moments lorsque ceux-ci existent. Nous noterons, pour $r \in \mathbb{N}^*$, $G_X^{(r)}(1^-)$ la $r^{\text{ième}}$ dérivée à gauche de G_X en 1 lorsqu'elle existe.

Propriété 20 Soit X une variable aléatoire à valeurs dans \mathbb{N} . Pour que X admette un moment d'ordre $r \in \mathbb{N}^*$, il faut et il suffit que sa fonction génératrice G_X soit r fois dérivable à gauche en 1 et dans ce cas, on a

$$G_X^{(r)}(1^-) = \sum_{k=r}^{+\infty} k(k-1) \cdots (k-r+1) p_k = \mathbb{E}[X(X-1) \cdots (X-r+1)].$$

4 Variables aléatoires continues et lois usuelles continues

Mots-clés :

- variable aléatoire continue, loi de probabilité, densité de probabilité, fonction de répartition. Interprétation géométrique d'une probabilité (aire sous la courbe de densité).
- Espérance et variance d'une variable aléatoire.
- Variables aléatoires uniformes, exponentielles, normales : domaine d'applications. Lecture de tables.
- Approximation de la loi de probabilité de la moyenne aléatoire d'un échantillon : Théorème Central Limite.

4.1 Introduction

Rappelons qu'on appelle **variable aléatoire réelle** toute application X de Ω dans \mathbb{R} où Ω est l'univers associé à une expérience aléatoire.

On appelle **univers-image**, noté \mathcal{F} (noté aussi parfois \mathcal{E} dans les chapitres précédents), l'ensemble de toutes les valeurs que peut potentiellement prendre la v.a. X .

Exemple 48 X est la v.a qui à chaque prélèvement d'eau effectué aléatoirement dans une rivière associe la concentration en ion nitrate.

$$\begin{array}{ccc} X : & \Omega & \longrightarrow & \mathcal{F} = \mathbb{R}^+ \\ & \omega & \mapsto & X(\omega) \\ & \text{prélèvement} & & \text{concentration} \end{array}$$

4.2 Variables Aléatoires Continues - Loi de probabilité : densité et fonction de répartition

Dans toute la suite, on considère que \mathcal{F} est un ensemble **continu** c'est-à-dire : \mathbb{R} ou un intervalle I de \mathbb{R} (borné ou non borné).

Définition 34 (Variable aléatoire continue) Une v.a. X est dite **continue** si elle peut prendre *a priori*, toutes les valeurs d'un intervalle I de \mathbb{R} .

Exemple 49 Les variables aléatoires suivantes sont continues :

- La durée de vie d'un organisme $\mathcal{F} = \mathbb{R}^+$
- La taille d'un individu dans une population donnée $\mathcal{F} = \mathbb{R}^+$.
- La concentration d'un polluant dans une rivière $\mathcal{F} = \mathbb{R}^+$.
- La taille d'une espèce de truites $\mathcal{F} =]0; 35]$ (si l'on sait que la taille ne peut dépasser 35 cm).

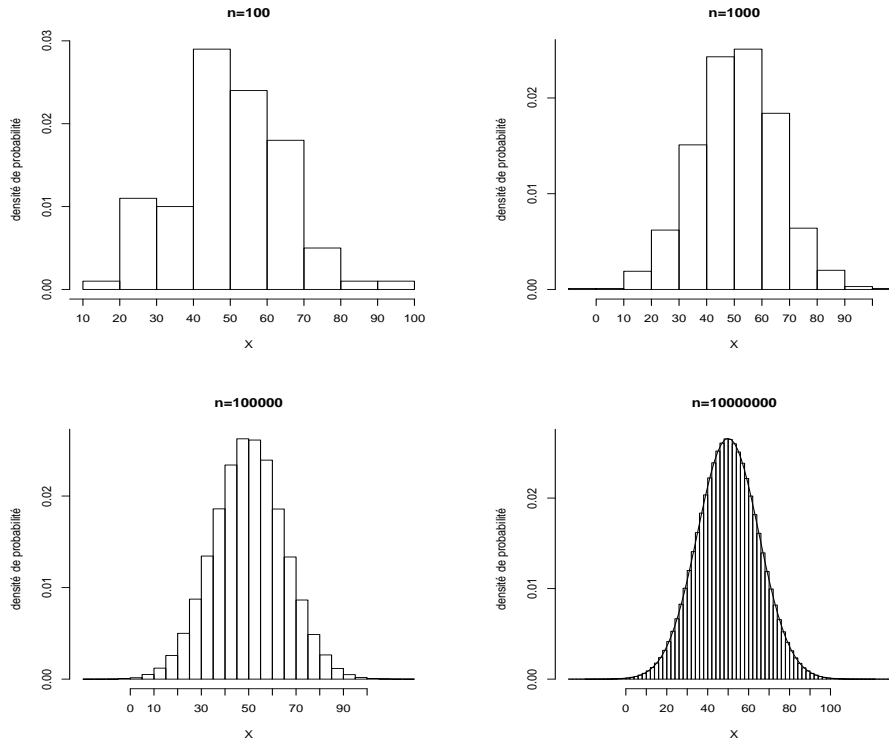
Comme on ne peut pas savoir avec certitude quelles seront les valeurs prises par une variable aléatoire X , on utilise sa **loi de probabilité** qui permet de calculer les probabilités que la v.a. X prenne certaines valeurs. On s'intéresse en général à des **événements** du type :

$$X \in [a, b] \text{ ou } X \geq a \text{ ou } X \leq b.$$

La **loi de probabilité** d'une v.a. continue est entièrement connue si l'on connaît sa **densité de probabilité** ou sa **fonction de répartition**. C'est ce que nous allons voir dans ce qui suit.

- **Notion de densité de probabilité** Observons différents histogrammes qui représentent la

distribution observée de la concentration en nitrate en mg/ℓ de n prélèvements effectués dans une rivière : il apparait clairement que plus le nombre d'observations n devient grand, plus l'on augmente le nombre de classes, et plus l'histogramme se rapproche d'une courbe "théorique" continue :



Dans toute la suite, l'ensemble des valeurs prises par la v.a. continue X est un intervalle I .

Définition 35 (Densité de probabilité) Si X est une v.a. **continue** qui prend ses valeurs dans un intervalle I de \mathbb{R} , on appelle **densité (de probabilité)** de X , une fonction f positive ou nulle telle que pour tout intervalle $[a, b]$ inclus dans I , on a :

$$P(X \in [a, b]) = \int_a^b f(x) dx$$

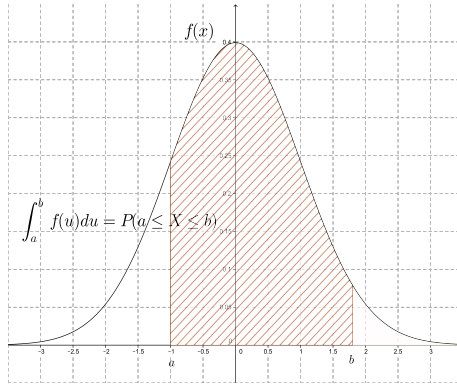
en particulier, $P(X \in I) = \int_I f(x) dx = 1$.

- Si $I = [a, b]$ alors la quantité $\int_I f(x) dx$ désigne simplement $\int_a^b f(x) dx$
- Si I est un intervalle non borné, par exemple $I = [a, +\infty[$ alors, la quantité $\int_I f(x) dx$ désigne si elle existe, la limite suivante :

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

De même, si $I =]-\infty, b]$.

Interprétation de l'aire comme une probabilité : L'aire sous la courbe de densité délimitée par l'axe des abscisses et les droites $x = a$ et $x = b$, s'interprète comme une probabilité : $\int_a^b f(x) dx = P(a \leq X \leq b)$



Exemple 50 On appelle X la v.a. qui représente la direction du vent en degrés. L'ensemble de toutes les valeurs possibles prises par X est donc l'intervalle $I = [0, 360]$. Considérons la fonction f donnée par :

$$f(x) = \begin{cases} 0.001 & \text{si } 0 \leq x < 100 \\ 0.01 & \text{si } 100 \leq x < 160 \\ 0.0015 & \text{si } 160 \leq x \leq 360 \end{cases}$$

On définit la probabilité que la direction du vent soit comprise entre a et b par :

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$$

Calculer $\mathbb{P}(X \in [0, 45])$ et $\mathbb{P}(X \in [90, 135])$.

$$\mathbb{P}(X \in [0, 45]) = \int_0^{45} 0.001 dx = 0.001 \times 45 = 0.045$$

et

$$\begin{aligned} \mathbb{P}(X \in [90, 135]) &= \int_{90}^{135} f(x) dx = \int_{90}^{100} 0.001 dx + \int_{100}^{135} 0.01 dx \\ &= 0.001 \times (100 - 90) + 0.01 \times (135 - 100) = 0.01 + 0.35 = 0.36 \end{aligned}$$

On observe que même si les intervalles $[0, 45]$ et $[90, 135]$ ont une taille identique, leurs probabilités sont différentes : le vent a plus de chance de souffler entre 90 et 135 degrés qu'entre 0 et 45 degrés.

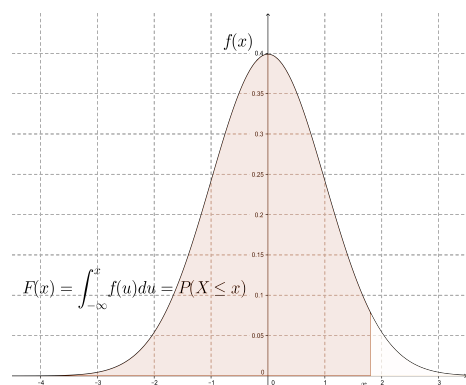
• Fonction de répartition d'une v.a. continue

Définition 36 (Fonction de répartition) La fonction de répartition F (f.d.r) d'une v.a. X est définie pour tout $x \in \mathbb{R}$ par :

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

$$\text{Et on a } F(x) = \int_{-\infty}^x f(u) du = \lim_{a \rightarrow -\infty} \int_a^x f(u) du.$$

aire sous la courbe de densité ci-contre.



Propriété 21 • F est croissante sur \mathbb{R} .

• $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Propriété 22 Pour tout $a, b \in \mathbb{R}$ avec $a < b$,

$$\begin{aligned}
 P(a < X \leq b) &= F(b) - F(a) = \int_a^b f(u) du \\
 &\text{et aussi puisque } P(X = a) = P(X = b) = 0 : \\
 &= P(a \leq X \leq b) \\
 &= P(a \leq X < b) \\
 &= P(a < X < b).
 \end{aligned}$$

Exemple 51 On donne la fonction de répartition d'une v.a. $X : F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x/2 & \text{si } 0 \leq x < 2 \\ 1 & \text{si } x \geq 2. \end{cases}$

- 1) Représenter graphiquement F .
- 2) Déterminer $P(X < 0)$ et $P(X \leq 0)$
 $P(X \leq 1/2)$ et $P(X < 1/2)$
 $P(X \leq 1)$ et $P(X < 1)$
 $P(1/2 \leq X \leq 1)$ et $P(1/2 \leq X < 1)$.

Exemple 52 Soit X une v.a. qui prend ses valeurs dans $I = [0, 2]$ de densité $f(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t < 1 \\ 2 - t & \text{si } 1 \leq t < 2 \\ 0 & \text{si } t \geq 2 \end{cases}$

- 1) Vérifier que la fonction f est positive ou nulle sur \mathbb{R} et que $\int_0^2 f(t) dt = 1$.
- 2) Calculer $P(1 \leq X \leq 2)$ si X est une v.a. de densité f .

Remarque 9 Pour faire des calculs de probabilité relatifs à la v.a. X , il est équivalent de connaître la f.d.r F ou la densité f de X .

Propriété 23 La fonction de répartition F d'une v.a. continue X est continue sur \mathbb{R} . De plus, elle est dérivable sur I et on a $F' = f$.

• **Espérance et variance d'une v.a. continue**

Définition 37 (Espérance mathématique ou moyenne théorique) Si la v.a. continue X prend ses valeurs dans un intervalle I et admet pour densité la fonction f alors, on définit son **espérance** par :

$$\mu = E(X) = \int_I x f(x) dx$$

L'espérance correspond à la valeur moyenne "théorique" prise par la v.a. X .

Définition 38 (Variance)

$$V(X) = \int_I (x - \mu)^2 f(x) dx$$

où μ est l'espérance mathématique de X .

La variance mesure la dispersion de la v.a. X autour de son espérance.

Propriété 24 (Formule de Kœnig-Huyghens)

$$V(X) = \int_I x^2 f(x) dx - \mu^2$$

Cette formule peut être utile pour le calcul de la variance (souvent plus simple).

Exemple 53 Calculer l'espérance et la variance de la v.a. X de l'exemple ?? et de l'exemple ?. Dans l'exemple ??, on donne la f.d.r de la v.a. X , nous allons tout d'abord déterminer la densité f associée. Pour cela, on utilise la Propriété ?? :

$$f(x) = 1/2 \text{ si } 0 \leq x < 2 \text{ et } f(x) = 0 \text{ si } x \notin [0, 2[.$$

Puis,

$$\mu = E(X) = \int_0^2 x f(x) dx = \int_0^2 x \times \frac{1}{2} dx = \left[\frac{x^2}{4} \right]_0^2 = 1.$$

$$V(X) = \int_0^2 x^2 f(x) dx - \mu^2 = \int_0^2 x^2 \times \frac{1}{2} dx - 1^2 = \left[\frac{x^3}{6} \right]_0^2 - 1 = \frac{8}{6} - 1 = \frac{1}{3}.$$

Dans l'exemple ??, on a directement la densité f de la v.a. X , donc :

$$\begin{aligned} E(X) &= \int_0^2 x f(x) dx = \int_0^1 x f(x) dx + \int_1^2 x f(x) dx \\ &= \int_0^1 x^2 dx + \int_1^2 x(2-x) dx = \left[\frac{x^3}{3} \right]_0^1 + \left[x^2 - \frac{x^3}{3} \right]_1^2 \\ &= \frac{1}{3} + 2^2 - \frac{2^3}{3} - 1^2 + \frac{1^3}{3} = 1 \end{aligned}$$

Et

$$\begin{aligned} V(X) &= \int_0^2 x^2 f(x) dx - 1^2 = \int_0^1 x^2 f(x) dx + \int_1^2 x^2 f(x) dx - 1 \\ &= \int_0^1 x^3 dx + \int_1^2 x^2(2-x) dx - 1 = \left[\frac{x^4}{4} \right]_0^1 + \left[\frac{2}{3}x^3 - \frac{x^4}{4} \right]_1^2 - 1 \\ &= \frac{1}{4} + \frac{2}{3}2^3 - \frac{2^4}{4} - \frac{2}{3}1^3 + \frac{1^4}{4} - 1 = \frac{3}{2}. \end{aligned}$$

Propriété 25 Soient $a, b \in \mathbb{R}$, ($a \neq 0$),

- $E(aX + b) = aE(X) + b$
- $V(aX + b) = a^2V(X)$

• **Centrer et Réduire une v.a.** : C'est une transformation de la v.a. X qui permet d'obtenir une nouvelle v.a. Z telle que $E(Z) = 0$ (on dit que Z est **centrée**) et $V(Z) = 1$ (on dit que Z est **réduite**).

Propriété 26 Soit X une v.a. d'espérance μ et d'écart-type σ . La v.a.

$$Z = \frac{X - \mu}{\sigma}$$

est centrée et réduite. Il suffit d'appliquer la Propriété ?? avec $a = 1/\sigma$ et $b = -\mu/\sigma$.

4.3 Variables aléatoires usuelles continues

• **v.a. uniforme**

Définition 39 Une v.a. X suit une loi uniforme si sa densité f est de la forme :

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

Notation : On écrit $X \sim \mathcal{U}([a, b])$. La v.a. X prend ses valeurs dans $[a, b]$.

Exemple 54 – Calculer l'espérance et la variance d'une v.a. X qui suit une loi uniforme sur $[0, 1]$.

– De façon générale, montrer que l'espérance et la variance d'une v.a. uniforme sur $[a, b]$ sont égales à :

$$E(X) = \frac{a+b}{2} \quad V(X) = \frac{(b-a)^2}{12}$$

Exemple 55 On remplit une éprouvette de volume maximum 20 cl d'une quantité aléatoire Q d'eau de façon uniforme entre 0 et 20 cl. Quelle est la probabilité d'obtenir moins de 5 cl d'eau dans l'éprouvette ?

• **v.a. exponentielle**

Définition 40 Une v.a. X positive suit une loi exponentielle de paramètre λ , avec $\lambda > 0$ si sa densité f est de la forme :

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

On note $X \sim \text{Exp}(\lambda)$ et X est à valeurs dans \mathbb{R}^+ .

La v.a. X est souvent utilisée pour représenter une **durée de vie** : durée de vie d'un matériel, durée d'un organisme, d'un patient, ...

Propriété 27 La f.d.r F d'une loi exponentielle est de la forme :

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - \exp(-\lambda x) & \text{si } x \geq 0 \end{cases}$$

Propriété 28

$$E(X) = \frac{1}{\lambda} \quad V(X) = \frac{1}{\lambda^2}$$

Exemple 56 Un laboratoire de physique dispose d'un parc d'oscilloscopes identiques. La durée de vie en années d'un oscilloscope est une variable aléatoire notée X qui suit la loi de durée de vie sans vieillissement ou encore loi exponentielle de paramètre $\lambda > 0$.

1. Sachant que $P(X > 10) = 0.286$, montrer que $\lambda = 0.125$ au centième près. Dans la suite, on prend $\lambda = 0.125$.
2. Calculer la probabilité qu'un oscilloscope du modèle étudié ait une durée de vie inférieure à 6 mois.
3. Sachant qu'un appareil a déjà fonctionné 8 années, quelle est la probabilité qu'il ait une durée de vie supérieure à 10 ans.

4.4 Variable aléatoire de Loi Gaussienne

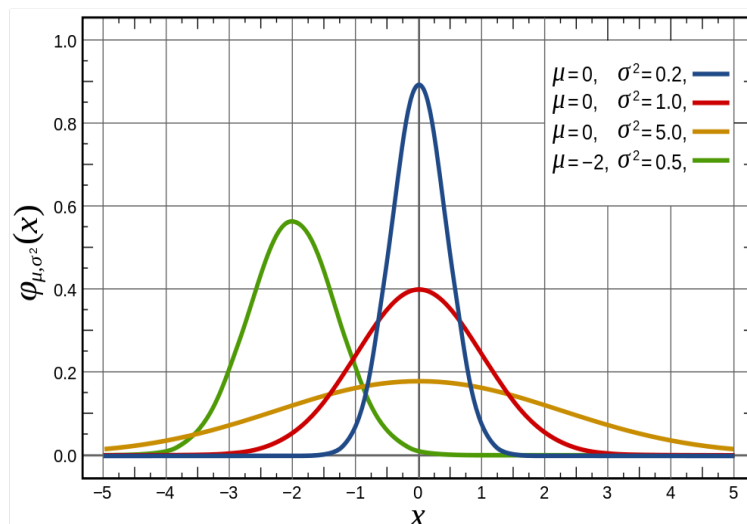
La **loi normale** est très importante car elle gouverne de nombreux phénomènes aléatoires. La loi de nombreuses variables rencontrées dans les sciences de la nature peut être modélisée par une loi normale. La loi normale est souvent associée à des variables que l'on mesure : c'est la loi des erreurs. De plus, le théorème de la **limite centrale** montre qu'elle est la **loi limite** de la somme de n v.a. indépendantes de même loi.

Définition 41 Soit $\mu \in \mathbb{R}$ et $\sigma > 0$, une v.a. X suit une loi normale si sa densité f est de la forme :

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),$$

Une v.a. X normale est donc entièrement définie par la connaissance des deux **paramètres** μ et σ et on note $X \sim \mathcal{N}(\mu, \sigma)$.

- Exemples de densités normales pour différentes valeurs de μ et σ^2 (d'après Wikipedia) :



La courbe représentative de la densité d'une v.a. normale est **en forme de cloche**. C'est une courbe symétrique par rapport à l'axe vertical d'équation $x = \mu$. Plus le paramètre σ est grand, plus la courbe en cloche est aplatie. Au contraire, lorsque σ est petit, la courbe est très "piquée".

- Propriété 29**
- 1) Si $X \sim \mathcal{N}(\mu, \sigma)$ alors $E(X) = \mu$ et $V(X) = \sigma^2$.
 - 2) Si la v.a. Z suit une loi normale $\mathcal{N}(0, 1)$, on dit que Z est **centrée** et **réduite**.
 - 3) De plus, toute v.a. normale de paramètres μ et σ peut être transformée en v.a. normale **centrée** et **réduite** de la façon suivante :

$$X \sim \mathcal{N}(\mu, \sigma) \iff Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

La f.d.r d'une v.a. normale centrée réduite est donnée par

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

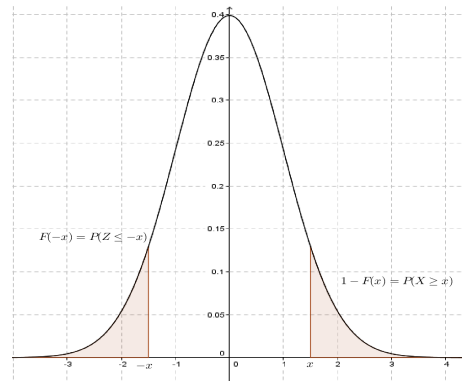
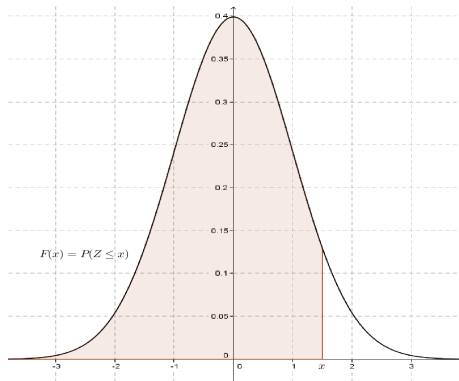
Elle n'a pas d'expression explicite : c'est-à-dire que l'on ne peut pas la calculer "à la main" mais ses valeurs ont été tabulées.

• **Lecture de tables de la loi normale centrée réduite**

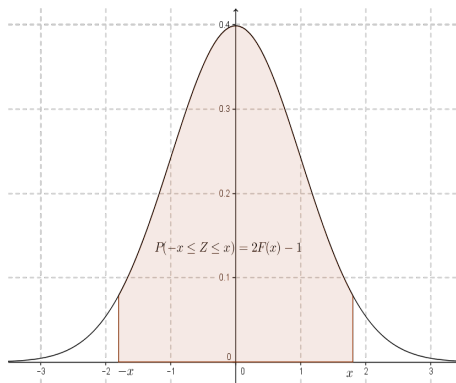
Propriété 30 Soit $Z \sim \mathcal{N}(0, 1)$ alors,

$$F(x) = P(Z \leq x) \text{ et } 1 - F(x) = P(Z > x)$$

$$F(-x) = 1 - F(x)$$



$$P(-x \leq Z \leq x) = F(x) - F(-x) = 2F(x) - 1$$



• Table de la loi normale $\mathcal{N}(0, 1)$

$$F(t) = P(X \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \text{et} \quad F(-t) = 1 - F(t).$$

t	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	0.7704	0.7734	0.7764	0.7794	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	0.8907	0.8925	0.8944	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

Exemple 57 Soit $Z \sim \mathcal{N}(0, 1)$. Lire la probabilité $P(Z < 1.45)$:

$$P(Z < 1.45) = F(1.45) = 0.9265.$$

t	0.00	0.01	...	0.05	...
0.0	0.5	0.5040	...	0.5199	...
0.1	0.5398	0.5438	...	0.5596	...
0.2	0.5793	0.5832	...	0.5987	...
⋮	⋮	⋮		⋮	
⋮	⋮	⋮		⋮	
1.4	0.9192	0.9207	...	0.9265	...
⋮	⋮	⋮		⋮	
⋮	⋮	⋮		⋮	

Exemple 58 Soit $X \sim \mathcal{N}(70, 5)$. Calculer la probabilité $P(X < 77.25)$:

$$\begin{aligned}
 P(X < 77.25) &= P(X - 70 < 77.25 - 70) \\
 &= P\left(\frac{X - 70}{5} < \frac{77.25 - 70}{5}\right) \\
 &= P(Z < 1.45) \quad \text{où } Z \sim \mathcal{N}(0, 1) \\
 &= F(1.45) \\
 &= 0.9265
 \end{aligned}$$

4.5 Théorème Central-Limite (T.C.L)

$\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n$ désignent n expériences aléatoires identiques, et réalisées de façon indépendante les unes des autres. A chaque expérience \mathcal{E}_i , on associe une v.a. X_i . Les v.a. X_1, X_2, \dots, X_n sont dites (mutuellement) **indépendantes**.

Exemple 59 On répète n fois l'expérience aléatoire qui consiste à faire un prélèvement d'eau dans une rivière. Les prélèvements sont réalisés dans les mêmes conditions et sont considérés indépendants. Ici, une expérience aléatoire \mathcal{E}_i est un prélèvement d'eau \mathcal{E}_i : on associe à chaque \mathcal{E}_i la variable aléatoire X_i qui donne la mesure de la concentration en nitrates (en mg/l). Les v.a. X_1, X_2, \dots, X_n des n mesures de concentrations sont mutuellement indépendantes.

Théorème 1 ("central-limite" T.C.L.) Soit X_1, X_2, \dots, X_n des variables aléatoires **indépendantes et de même loi de probabilité**. Notons μ et σ^2 l'espérance et la variance de ces v.a. Si on considère la moyenne aléatoire

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

alors, pour n est assez grand (en pratique on considère $n \geq 30$), la loi de probabilité de \bar{X}_n peut être approchée par une loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. On note $\bar{X}_n \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

La loi normale est donc loi de la v.a. \bar{X}_n quelle que soit la loi des X_i (il faut juste que μ et σ^2 existent). Ceci montre le caractère universel des lois normales. Le théorème central-limite est fondamental dans la théorie de l'échantillonnage.

Exemple 60 On fait 100 mesures indépendantes de la concentration en nitrates dans une rivière. On ne connaît pas la loi de la variable concentration mais on a pu évaluer son espérance à 50 mg/l et son écart-type à 50 mg/l. Quelle est la probabilité que la concentration moyenne dépasse le seuil de 55 mg/l, de 60 mg/l ?

Si X_i représente la concentration du prélèvement i , alors la concentration moyenne sur les 100 prélèvements effectués est

$$\bar{X}_{100} = \frac{1}{100} \sum_{i=1}^{100} X_i$$

On sait que $E(X_i) = 50$ et $V(X_i) = 50^2$. Par le TCL, comme $n \geq 30$ on sait que l'on peut approcher la loi de la variable \bar{X}_n par une loi normale $\mathcal{N}(50, 50/\sqrt{100})$.

$$\begin{aligned} P(\bar{X}_n \geq 55) &= P\left(\frac{\bar{X}_{100} - 50}{50/\sqrt{100}} \geq \frac{55 - 50}{50/\sqrt{100}}\right) \\ &\simeq P(Z \geq 1) \text{ où } Z \sim \mathcal{N}(0, 1) \\ &\simeq 1 - 0.8413 = 0.1587 \text{ soit } 15.87\% \end{aligned}$$

Et de même, $P(\bar{X}_n \geq 60) \simeq P(Z \geq 2) = 1 - 0.9772 = 0.0228$ soit 2.28%.

5 Échantillonnage et Estimation

Mots-clés :

- Population-mère, échantillon, fluctuation d'échantillonnage.
- Échantillon aléatoire. Variables aléatoires indépendantes.
- Estimateur, Estimateur sans biais, Estimation ponctuelle, Estimation par intervalles, niveau de confiance et risque d'erreur.
- Test d'hypothèses, hypothèse nulle et alternative, tests unilatéral et bilatéral, niveau ou seuil de risque d'un test.
- Test du χ^2 d'ajustement à une loi.

Pour étudier une variable dans une population \mathcal{P} , on dispose de 2 méthodes :

- La **méthode exhaustive** qui consiste à recenser TOUS les individus de la population. Cette méthode est fort peu employée.
- La **méthode d'échantillonnage** qui consiste à n'étudier qu'un échantillon E prélevé dans la **population-mère** \mathcal{P} , et à induire, à partir des résultats observés sur l'échantillon des résultats concernant la population-mère entière. Nous admettrons qu'en réalisant un tirage au sort des individus de l'échantillon, on obtient un échantillon représentatif de la population.

5.1 Estimation ponctuelle d'un paramètre - Qualité d'un estimateur - Notion de biais

L'**estimation** consiste à évaluer un paramètre inconnu θ de la population-mère \mathcal{P} à partir d'un échantillon.

Si on veut estimer, des paramètres comme l'espérance μ et l'écart-type σ d'une variable d'intérêt X dans une population \mathcal{P} , on ne dispose le plus souvent que d'un seul échantillon de taille n , qui nous a permis de calculer la moyenne observée m_e et l'écart-type observé σ_e . En estimant μ par m_e , ou σ par σ_e nous commettons une erreur. L'espérance μ de la population est inconnue mais c'est une constante, alors que la valeur m_e dépend de l'échantillon : on aurait une autre valeur de m_e si l'on avait pris un autre échantillon, c'est ce qu'on appelle la **fluctuation d'échantillonnage**. Comment savoir si l'échantillon que nous avons observé fournit une "bonne" valeur pour évaluer les paramètres de la population-mère ?

Définition 42 (Échantillon aléatoire) Considérons X une variable aléatoire qui associe à tout individu de la population-mère \mathcal{P} une valeur réelle. Soit θ un paramètre d'intérêt réel (comme l'espérance μ ou l'écart-type σ) de la v.a. X .

On prélève dans la population-mère, un **échantillon aléatoire** de taille n . On appelle X_i la variable aléatoire qui associe à l'individu i prélevé la valeur observée au i -ème tirage. Les variables X_i sont **indépendantes** et ont toutes la **même loi de probabilité** puisqu'elles sont issues de la même population-mère \mathcal{P} .

Définition 43 (Estimateur) On appelle **estimateur** toute variable aléatoire $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ qui s'exprime en fonction des v.a. X_1, \dots, X_n d'un échantillon aléatoire de taille n prélevé dans la population-mère.

Exemple 61 La v.a. $\hat{\theta}_n = \bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ qui à chaque échantillon prélevé dans \mathcal{P} , associe la moyenne des X_i est un **estimateur** du paramètre $\theta = \mu$. Ici $\theta = \mu$ est l'espérance

inconnue de la v.a. X dans la population-mère \mathcal{P} .

• **Lien entre la statistique descriptive et l'estimation** : On suppose que les valeurs observées x_1, \dots, x_n des tailles des 100 écrevisses sont les réalisations de $n = 100$ v.a. X_1, \dots, X_n indépendantes et de **même espérance μ inconnue**.

Ainsi la valeur observée $m_e = \frac{1}{n} \sum_{i=1}^n x_i = 122 \text{ mm}$ est la réalisation pour l'échantillon prélevé des 100 écrevisses de la v.a.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Et, un autre échantillon de 100 écrevisses donnerait une autre valeur de la variable \bar{X}_n .

En conclusion, il ne faut pas confondre les 3 quantités suivantes :

μ	la valeur inconnue du paramètre dans la population-mère \mathcal{P} .
\bar{X}_n	la variable aléatoire qui à chaque échantillon associe sa valeur moyenne.
m_e	la valeur observée (ou réalisation) de la v.a. \bar{X}_n pour un échantillon donné.

Définition 44 Un estimateur $\hat{\theta}_n$ d'un paramètre θ est dit **sans biais** si $E(\hat{\theta}_n) = \theta$.

Exemple 62

$$E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n}(\mu + \dots + \mu) = \mu.$$

L'estimateur \bar{X}_n est un estimateur sans biais de μ .

Exemple 63 On peut aussi montrer (mais c'est plus délicat !) que si X_1, \dots, X_n est un échantillon de variables aléatoires de même variance σ^2 alors

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

est un estimateur sans biais de σ^2 .

5.2 Estimation ponctuelle

Soit une variable d'intérêt X dans la population-mère \mathcal{P} , on peut donner une **estimation ponctuelle** de :

– l'**espérance** μ de X par la moyenne observée m_e de l'échantillon

– l'**écart-type** σ de X par $s_e = \sqrt{\frac{n}{n-1}} \sigma_e$ où σ_e est l'écart-type observé sur l'échantillon.

Les estimations ponctuelles dépendent beaucoup de l'échantillon prélevé au hasard.

La question qui se pose alors est le degré de confiance que l'on peut accorder à ces estimations ponctuelles. On a aucun renseignement sur la proximité de la valeur donnée par l'échantillon et la valeur inconnue du paramètre.

Exemple 64 Retour sur l'exemple de la Taille des écrevisses.

Question : En quoi la valeur de $m_e = 122$ mm observée sur l'échantillon des 100 écrevisses nous renseigne t'elle sur la valeur de l'espérance μ de la taille de toute la population \mathcal{P} des écrevisses de cette rivière ?

– Si on avait prélevé un autre échantillon de 100 écrevisses, donnerait-il la même valeur ou une valeur différente de 122 mm ?

5.3 Estimation par intervalles

Nous allons introduire la notion d'estimation par intervalles de confiance qui permet de mesurer un certain type d'erreur lorsque l'on extrapole un résultat observé sur un échantillon à toute la population-mère.

D'après le Théorème ??, nous savons que la variable aléatoire \bar{X}_n qui, à chaque échantillon de taille n , associe sa moyenne suit approximativement la loi normale $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

- **Estimation de μ lorsque σ est connu**

On veut construire un intervalle de confiance pour l'espérance inconnue μ de la variable X dans la population-mère.

Or, d'après le Théorème ??, nous savons que la variable aléatoire \bar{X}_n qui, à chaque échantillon de taille n , associe sa moyenne suit approximativement :

$$\bar{X}_n \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

donc si on centre et réduit la v.a. \bar{X}_n

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1)$$

D'après la table de la loi normale $\mathcal{N}(0, 1)$, on sait que $P(-1.96 \leq Z \leq 1.96) = 0.95$.

$$\begin{aligned} P(-1.96 \leq Z \leq 1.96) &= P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) \\ &= P\left(\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95 \end{aligned}$$

L'espérance inconnue μ appartient donc à l'**intervalle aléatoire**

$$\left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

avec une probabilité de 0.95. Cela signifie que parmi toutes les réalisations possibles de la v.a. \bar{X}_n , seulement 5% d'entre elles fournissent un intervalle qui ne contient pas la valeur μ .

Ensuite, pour donner une fourchette numérique on remplace la v.a. \bar{X}_n par sa réalisation ou valeur observée m_e de l'échantillon prélevé,

L'intervalle observé $\left[m_e - 1.96 \frac{\sigma}{\sqrt{n}}; m_e + 1.96 \frac{\sigma}{\sqrt{n}}\right]$ fournit une fourchette pour μ avec un niveau de confiance de 95 % ou un risque d'erreur de 5%.

Plus généralement, pour un risque d'erreur de $\alpha\%$, si $z_{1-\alpha/2}$ est la valeur lue dans la table de la loi normale telle que $P(Z \leq z_{1-\alpha/2}) = 1 - \alpha/2$:

L'intervalle observé $\left[m_e - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; m_e + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$ **fournit une fourchette pour μ avec un niveau de confiance de $1 - \alpha\%$ ou un risque d'erreur de $\alpha\%$.**

• **Estimation de μ lorsque σ est inconnu** Lorsque σ est inconnu, on le remplace par son estimation ponctuelle $s_e = \sigma_e \sqrt{\frac{n}{n-1}}$ et :

L'intervalle observé $\left[m_e - z_{1-\alpha/2} \frac{s_e}{\sqrt{n}}; m_e + z_{1-\alpha/2} \frac{s_e}{\sqrt{n}} \right]$ **donne une fourchette pour μ avec un niveau de confiance de $1 - \alpha\%$ ou un risque d'erreur de $\alpha\%$.**

Exemple 65 On reprend l'exemple de la taille des écrevisses. Une étude précédente, il y a 10 ans, avait indiqué que la taille moyenne observée sur un échantillon de la population des écrevisses de Californie dans cette rivière était de 118 mm. L'étude actuelle fournit les valeurs suivantes :

$$m_e = 122 \text{ mm} \quad \text{et} \quad s_e = 15.5 \text{ mm}.$$

Donner un intervalle de confiance à 95% de la taille moyenne (l'espérance de la taille) de la population des écrevisses aujourd'hui. Peut-on dire que la taille moyenne des écrevisses a augmenté en 10 ans, au vu de cette nouvelle étude ?

$$P \left(\bar{X}_n - z_{0.975} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{0.975} \frac{\sigma}{\sqrt{n}} \right) = 1 - 0.05 = 0.95$$

Ici, l'échantillon des tailles observées donne $n = 100$, $m_e = 122$ et on remplace σ par $s_e = 15.5$, donc l'intervalle de confiance observé pour l'espérance μ de la taille des écrevisses à 95 % est donné par :

$$122 - 1.96 \times \frac{15.5}{\sqrt{100}} \leq \mu \leq 122 + 1.96 \times \frac{15.5}{\sqrt{100}}$$

soit

$$118.9 \leq \mu \leq 125$$

Interprétation : On a 95 % de chance d'avoir observé un échantillon de taille 100 qui donne une fourchette contenant la vraie valeur μ . De plus, on pourrait penser que la taille de cette espèce d'écrevisses a eu tendance à s'accroître (puisque la valeur 118 mm donnée il y a 10 ans, n'est pas dans l'intervalle de confiance calculé aujourd'hui). Nous allons voir dans le paragraphe suivant comment tester plus précisément l'égalité de deux moyennes.