



**Haute Ecole de Bruxelles-Brabant
Ecole Supérieure d'Informatique**

Rue Royale, 67 – 1000 Bruxelles
02/219.15.46 – esi@he2b.be

Statistique – Théorie

Bachelor en Informatique

Cours enseigné par :

Patrick Bishop, Laurent Beeckmans, Christine Leignel (coordinatrice)

AVERTISSEMENT

Ce syllabus, qui résulte d'une collaboration entre E. Fontaine et L. Beeckmans est publié et distribué à la Haute Ecole de Bruxelles – Ecole Supérieure d'Informatique, avec l'accord des titulaires du cours.

Les notes de ce syllabus constituent une aide pour l'étudiant et servent de support au cours : elles ne couvrent pas nécessairement l'entièreté de la matière et certaines parties non vues au cours peuvent être présentées à titre d'informations supplémentaires. Inversement, certaines parties du syllabus ne seront pas vues en cours, mais restent dans le syllabus à titre informatif. La prise de notes individuelles et complémentaires reste donc vivement conseillée. De plus des solutions détaillées et expliquées des exercices ne sont données que dans le cadre du cours où théorie et exercices sont répartis de la manière la plus adéquate possible.

Les professeurs titulaires du cours ont toujours la possibilité de distribuer au cours ou de mettre à disposition sur le réseau (poESI) des notes complémentaires.

Les examens faisant intervenir théorie et exercices (cf. organisation de l'année) porteront sur la matière donnée par le professeur et pas nécessairement sur le contenu exclusif de ce syllabus.

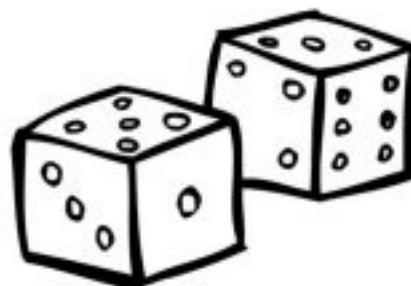
TABLE DES MATIERES

Chapitre 1 – Introduction au calcul des probabilités	3
Chapitre 2 – Variables aléatoires et grandes distributions théoriques	15
Chapitre 3 – Distributions à deux dimensions	27
Chapitre 4 – Inférence statistique	32
Chapitre 5 – Analyse en composantes principales	41
Bibliographie	42

Statistique

Chapitre 1

Introduction au calcul des probabilités



« La loterie est une taxe sur les personnes
qui ne comprennent rien aux probabilités »
Anon.

1.1. Introduction

Calculer une probabilité, c'est essayer de chiffrer la possibilité qu'un événement soumis aux lois du hasard se réalise. L'origine du calcul des probabilités réside dans les jeux de hasard, les joueurs voulant naturellement estimer leurs chances de gain. L'exemple type de cette théorie est le lancer d'une pièce de monnaie : tout le monde sait qu'il y a une chance sur deux d'obtenir *pile* (ou *face*) lors d'un lancer – et c'est sans doute la façon la plus courante et répandue dans le monde de prendre une décision lorsque deux personnes n'arrivent pas à trancher entre deux possibilités ! Le théoricien dira plutôt que la *probabilité* d'obtenir *pile* (ou *face*) est de $1/2$. Nous verrons plus loin la terminologie propre à la théorie des probabilités.

Outre le domaine des jeux où les exemples de probabilités sont nombreux, le hasard intervient encore dans bien des domaines de la vie : tirage au sort dans certains systèmes de sélection, hasard de rencontrer quelqu'un qu'on connaît dans la rue, chance (ou plutôt malchance) de contracter un certain type de maladie...

La notion de hasard elle-même est difficile à définir et peut être le point de départ de cogitations philosophiques ! Pensons au fait de passer un examen, de réussir une année d'études. Certains diront qu'il s'agit d'un processus tout à fait déterministe qui dépend de la seule volonté de l'étudiant ; pourtant, on peut parler de la *probabilité de réussite* d'un étudiant pour une année d'études donnée.

Certains processus très complexes qui semblent déterminés d'un point de vue physique mais qui dépendent de l'action d'un nombre énorme de paramètres se comportent de façon *quasi-aléatoire*. Ceci s'observe par exemple dans le domaine des prévisions météorologiques, où la connaissance et la mise en équation de millions de données (température, pression, direction du vent mesurées en de très nombreux points) ne suffit jamais à assurer des prévisions fiables à très long terme, de sorte qu'on peut aussi parler de la *probabilité qu'il fasse beau* le week-end prochain...

1.2. Terminologie

Expérience aléatoire

On appelle **expérience aléatoire** un processus dont l'issue est aléatoire, c'est-à-dire soumise au hasard, de telle sorte que le résultat de l'expérience ne puisse être connu à l'avance. Quelques exemples :

- a) jeter une pièce de monnaie et noter sur quelle face elle tombe
- b) jeter un dé et noter le point obtenu
- c) jeter un dé et noter la parité du point obtenu
- d) jeter deux dés et noter la somme des points obtenus

- e) jeter deux dés et observer si les points sont identiques
- f) choisir un étudiant au hasard et observer le grade qu'il obtiendra en fin d'année
- g) exécuter un saut en longueur et noter la distance parcourue

On voit que pour être complète, une expérience aléatoire doit décrire le déroulement d'une certaine action ainsi que le type de résultat attendu, une action pouvant être associée à plusieurs résultats différents, cf. ex. b) et c), d) et e) ci-dessus.

Catégorie d'épreuve

La **catégorie d'épreuve** est le nom donné à l'ensemble des issues possibles d'une expérience aléatoire. Cet ensemble est couramment noté Ω . Les catégories d'épreuve respectives des exemples ci-dessus sont les suivantes :

- a) {pile, face}
- b) {1, 2, 3, 4, 5, 6}
- c) {pair, impair}
- d) {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}
- e) {vrai, faux}
- f) {la plus grande distinction, grande distinction, distinction, satisfaction, ajournement, refus}
- g) [0, 10]

De façon similaire aux familles d'observations statistiques, on retrouve plusieurs types de résultats : quantitatifs discrets (b, d), quantitatifs continus (g), qualitatifs (a, c, f) et même booléen (e).

Événement aléatoire

Un **événement aléatoire** est une des issues possibles d'une expérience aléatoire, plus techniquement un sous-ensemble de l'ensemble des possibilités Ω . Donnons pour l'exemple un événement aléatoire associé à chacune des expériences aléatoires ci-dessus :

- a) la pièce tombe sur *pile*
- b) le point du dé est au moins 5
- c) le point du dé est pair
- d) la somme des points des deux dés est au moins 10
- e) les points des dés sont identiques
- f) l'étudiant obtiendra *au moins* une distinction
- g) la longueur du saut sera comprise entre 5 et 7 mètres

Probabilité

La **probabilité** d'un événement aléatoire est un nombre réel compris entre 0 et 1 inclus et donnant une mesure de la chance qu'a cet événement à se réaliser. Au plus la probabilité est proche de 1, au plus l'événement a des chances de se réaliser ; au plus la probabilité est proche de 0, au moins l'événement a des chances de se réaliser. Il y a deux cas extrêmes : si la probabilité est nulle, on dit que l'événement est **impossible**, et lorsque la probabilité vaut 1, on dit que l'événement est **certain**.

Notation : la probabilité de l'événement aléatoire A est notée $P(A)$.

Exemples de probabilités simples pour le jet de la pièce de monnaie :

- $P(\text{la pièce tombe sur } \textit{pile}) = 1/2$
- $P(\text{la pièce tombe sur } \textit{face}) = 1/2$
- $P(\text{la pièce tombe soit sur } \textit{pile} \text{ soit sur } \textit{face}) = 1$ (événement certain)
- $P(\text{la pièce ne retombe pas}) = 0$ (événement impossible)

On peut voir la probabilité comme une fonction de Ω dans l'intervalle $[0, 1]$ associant à tout événement A (assimilé à un sous-ensemble de Ω) la probabilité $P(A)$:

$$P : \Omega \rightarrow [0, 1] : A \mapsto P(A)$$

Evénements équiprobables

Deux événements sont dits **équiprobables** s'ils ont la même probabilité de se réaliser. Dans l'exemple de la pièce de monnaie, les événements *la pièce tombe sur pile* et *la pièce tombe sur face* sont équiprobables (chacun a une probabilité de 1/2), sauf si la pièce est déséquilibrée ! De même pour un dé à six faces, on peut supposer qu'il est équiprobable d'obtenir chacun des points du dé (avec une probabilité 1/6 pour chacun), sauf bien entendu dans le cas d'un dé « pipé » où la fréquence d'obtention du 6 est plus grande que celles des autres points. Nous allons voir que cette notion d'équiprobabilité est cruciale dans la détermination de certaines probabilités.

1.3. Calcul des probabilités

La théorie des probabilités distingue deux façons de calculer la probabilité d'un événement aléatoire. La première (probabilité *à priori*) est de nature essentiellement mathématique et combinatoire tandis que la seconde (probabilité *à posteriori*) se base sur l'expérience et se rapproche de la théorie des statistiques.

Probabilité a priori

C'est la formule bien connue de la théorie des probabilités, elle définit la probabilité d'un événement aléatoire comme le rapport du nombre de *cas favorables* sur le nombre de *cas possibles* :

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

Quelques précisions importantes doivent être apportées à cette formule de base, qui dans sa forme courante n'est pas entièrement correcte. D'abord, par *nombre de cas favorables*, il faut comprendre *nombre de cas favorables à la réalisation de l'événement A parmi les cas possibles envisagés*. Donnons quelques exemples :

- a) probabilité qu'une pièce tombe sur *pile* : il y a deux cas possibles (*pile* ou *face*) et parmi ceux-ci un seul est favorable (*pile*). La probabilité est donc de 1/2.
- b) probabilité que le point du dé soit au moins 5 : il y a six résultats possibles et parmi ceux-ci, 2 mènent à la réalisation de l'événement (5 ou 6). La probabilité est donc de 2/6, soit 1/3.
- h) [attention, piège !] probabilité que la somme des points de deux dés soit au moins 10 : l'ensemble des possibilités Ω est {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. Il y a donc 11 cas possibles et parmi ceux-ci, 3 cas favorables. La probabilité vaut donc 3/11. [résultat erroné !]

A la lueur de ces exemples, le lecteur perspicace aura compris le point faible de la formule ci-dessus. Par *nombre de cas possibles*, on pense d'abord au nombre d'issues possibles de l'expérience aléatoire, soit le nombre d'éléments de Ω .

Mais ceci n'est vrai que si ces éléments sont obtenus de façon équiprobable, ce qui n'est pas le cas dans le dernier exemple : il est clair qu'il est par exemple moins probable d'obtenir 12 en lançant deux dés (il faut que les deux dés tombent sur 6) que 7 (qui peut s'obtenir par les combinaisons 1+6, 2+5, 3+4 mais aussi 4+3, 5+2 et 6+1 !).

Cette remarque conduit à une amélioration de la formule :

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

Voyons maintenant comment procéder pour trouver la probabilité correcte de l'exemple c), la difficulté étant que les cas élémentaires de l'ensemble Ω ne sont pas équiprobables. La solution est alors de

remonter plus loin dans le déroulement de l'expérience aléatoire et d'identifier un ensemble de cas possibles équiprobables. Dans ce cas, ce sont les *couples* de points (dé 1, dé 2) qu'il faut prendre en considération. Chaque dé pouvant prendre 6 valeurs indépendamment de l'autre, ceci conduit à 36 possibilités de couples équiprobables. Parmi ces possibilités, plusieurs mènent à la même somme de points, comme le montre le tableau ci-dessous :

Somme des points des dés 1 et 2

dé 1 dé 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

La probabilité que la somme des points de deux dés soit au moins 10 vaut donc 6/36 puisque 6 cas sur les 36 possibles mènent à une somme valant 10, 11 ou 12.

Cet exemple illustre toute la complexité du calcul des probabilités et montre que parfois l'énumération des cas possibles ne suffit pas : il a fallu ici aller en amont de l'ensemble des possibilités pour en déterminer d'autres plus élémentaires et qui nous semblent plus fiables quant à leur équiprobabilité.

Notons encore que la formule de probabilité *a priori* reste entièrement théorique et suppose une perfection mathématique des acteurs de l'expérience aléatoire. Même pour les exemples a) et b), on suppose implicitement que la pièce et le dé sont parfaitement équilibrés et la formule ne peut s'appliquer dans le cas d'objets « truqués ».

Probabilité *a posteriori*

La formule des probabilités *à priori* convient très bien dans les domaines où les modèles de mathématique combinatoire peuvent être appliqués, comme pour les jeux de hasard, les tirages au sort... Mais dans la plupart des situations de la vie courante, elle n'est d'aucune utilité : probabilité de réussir un examen, probabilité d'attraper un rhume, probabilité que l'ordinateur « se plante », probabilité que le professeur de statistique soit absent au prochain cours... Tous ces cas ne sont pas modélisables mathématiquement ; or, il s'agit d'événements aléatoires bien définis et donc une probabilité y est forcément associée. Dès lors, comment la déterminer ?

La solution est de remplacer la probabilité par une fréquence statistique résultant de l'observation de la réalisation de l'événement aléatoire au cours d'un grand nombre d'expériences. Ainsi, si n représente le nombre d'expériences, soit $f_n(A)$ la fréquence relative des réalisations de A au cours de ces n expériences, obtenue en divisant le nombre de fois où A s'est réalisé par le nombre d'expériences. Cette fréquence donne alors une approximation de la probabilité $P(A)$. Il est intuitif que cette fréquence sera d'autant plus fiable que n est grand, et donc la définition théorique de la probabilité *à posteriori* est :

$$P(A) = \lim_{n \rightarrow \infty} f_n(A)$$

Remarques :

- La fréquence relative est une notion expérimentale, tandis que la probabilité est une notion théorique. Cela veut dire que la valeur réelle de la probabilité $P(A)$ reste inconnue et « inaccessible » par cette méthode, on ne peut que s'en approcher davantage après un grand nombre d'expériences.
- L'utilisation du symbole *limite* dans la formule ci-dessus laisse sous-entendre que la suite des fréquences relatives $f_1(A), f_2(A), \dots, f_n(A), \dots$ est convergente. Si cela paraît clair intuitivement, la justification mathématique en est plus ardue. Dans le chapitre 7, nous

aborderons un résultat important nommé *loi des grands nombres* qui confirme la convergence de la suite des fréquences : au plus n augmente, au plus petite devient la différence entre la fréquence relative et la probabilité théorique.

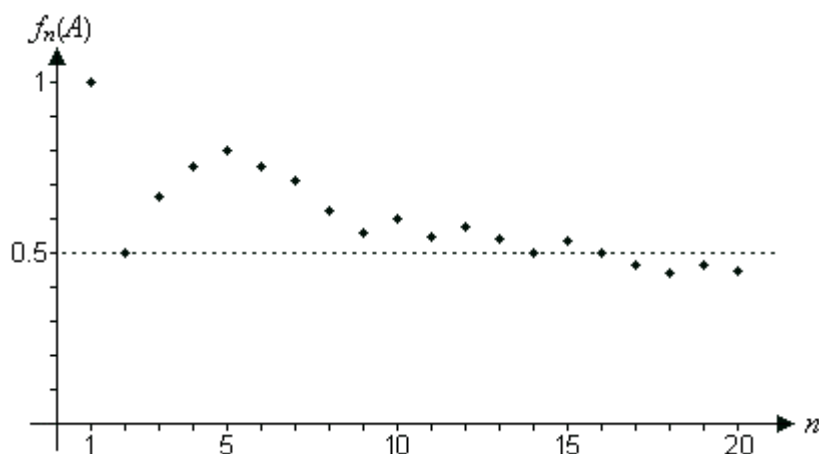
- Il est bien sûr probable que les premières fréquences obtenues subissent de grandes variations, mais après un nombre suffisamment élevé d'expériences, la fréquence tendra à se stabiliser et converger vers la probabilité théorique.

Exemple 1. Imaginons qu'on veuille déterminer la probabilité qu'une pièce tombe sur *pile*. On lance 20 fois cette pièce et on obtient les résultats suivants :

*pile, face, pile, pile, pile, face, pile, face, face, pile,
face, pile, face, face, pile, face, face, face, pile, face*

La suite des fréquences de l'événement « la pièce tombe sur *pile* » est la suivante :

n	$f_n(A)$	n	$f_n(A)$	n	$f_n(A)$	n	$f_n(A)$
1	1/1 = 1	6	4/6 = 0.667	11	6/11 = 0.545	16	8/16 = 0.5
2	1/2 = 0.5	7	5/7 = 0.714	12	7/12 = 0.583	17	8/17 = 0.471
3	2/3 = 0.667	8	5/8 = 0.625	13	7/13 = 0.538	18	8/18 = 0.444
4	3/4 = 0.75	9	5/9 = 0.556	14	7/14 = 0.5	19	9/19 = 0.474
5	4/5 = 0.8	10	6/10 = 0.6	15	8/15 = 0.533	20	9/20 = 0.45



En poursuivant le lancer de la pièce, la suite de fréquences devrait se stabiliser autour de la probabilité attendue, soit 1/2 (ou une autre valeur si la pièce est truquée !). Dans cet exemple, le nombre d'expériences est évidemment très réduit ; nous verrons plus loin que d'après la loi des grands nombres (Chapitre 7), il faudrait en fait 50000 lancers de pièces pour être certain à 95% que la différence entre la fréquence et la probabilité théorique soit inférieure à 1/100 !

Ceci dit, dans le cas de la pièce, nous serons en mesure de calculer à la fin de ce cours qu'en lançant la pièce 20 fois, la probabilité qu'elle tombe sur *pile* entre 9 et 11 fois est de 0.496 ; autrement dit, la fréquence obtenue après 20 lancers sera environ une fois sur deux comprise entre 0.45 et 0.55. Il faut toutefois signaler que ce calcul suppose *à priori* que la probabilité théorique est de 1/2 ! Notez aussi dans l'exemple que certaines fréquences sont égales à la probabilité théorique 1/2 (c'est le cas pour $n = 2, 14$ et 16). Il s'agit bien sûr de cas fortuits, seul le comportement asymptotique de la suite des fréquences est significatif pour déterminer la probabilité théorique (qui est de toute façon supposée inconnue *à priori*).

1.4. Événements aléatoires et ensembles

Nous avons défini un événement aléatoire comme une des issues possibles d'une expérience aléatoire. En fait, en assimilant l'événement aléatoire au sous-ensemble correspondant dans l'ensemble des possibilités Ω , nous commettons un léger abus de notation. Pour être plus précis, un événement aléatoire se définit en fait par une *proposition booléenne* : par exemple « le dé tombe sur

un point pair » est une affirmation booléenne qui sera vraie ou fausse selon le résultat de l'expérience aléatoire. L'ensemble des points du dé $\{2, 4, 6\}$ est quant à lui formé des éléments de Ω pour lesquels l'événement en question est réalisé. Pour faire le lien avec le cours de mathématique (chapitre de logique propositionnelle), un événement aléatoire est similaire à un *prédicat* (ou une *condition*) et le sous-ensemble correspondant de Ω est alors la *classe de vérité* de ce prédicat.

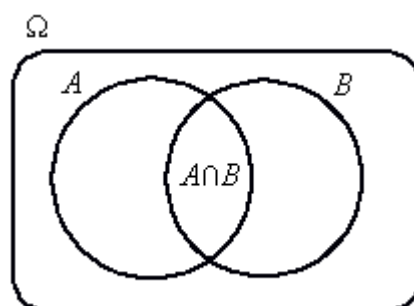
En particulier, l'événement certain (de probabilité 1) est assimilable à l'ensemble Ω tout entier, et l'événement impossible (de probabilité 0) correspond à l'ensemble vide. Dans ce qui suit, nous prendrons la liberté de confondre événements et ensembles, ce qui permet d'unifier les notations et de plagier certains aspects de la théorie sur celle des ensembles. Ainsi les opérateurs logiques ET, OU et NON seront notés respectivement \cap (intersection), \cup (union) et c (ensemble complémentaire).

Théorème d'addition

Le théorème d'addition donne une formule pour la probabilité de l'union de deux événements, cette opération correspondant à l'opérateur logique OU (non exclusif). Soient A et B deux événements liés à la même expérience aléatoire ; alors :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

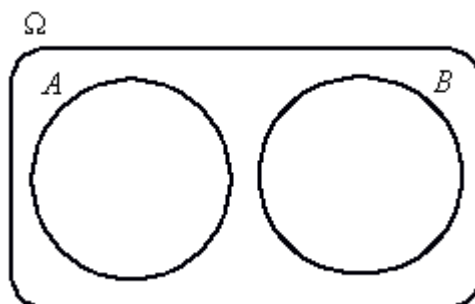
La démonstration de ce théorème est évidente en considérant la vision ensembliste :



En additionnant les probabilités de A et de B , on compte deux fois la probabilité de l'intersection, d'où la soustraction de celle-ci.

Événements incompatibles

Deux événements A et B liés à une même expérience aléatoire sont **incompatibles** s'ils ne peuvent se produire simultanément, c'est-à-dire si les deux sous-ensembles correspondants à A et B dans l'ensemble des possibilités Ω ont une intersection vide, donc si $A \cap B = \emptyset$. Expressions équivalentes : on dit aussi que les événements sont **mutuellement exclusifs** ou encore **disjoints**, par référence à la vision ensembliste.

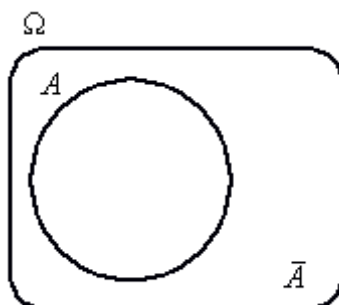


Propriété. Si A et B sont incompatibles, alors $P(A \cap B) = 0$, et la formule du théorème d'addition se simplifie en :

$$P(A \cup B) = P(A) + P(B)$$

Evénement complémentaire

Du point de vue logique, le complémentaire d'un événement A est la négation de celui-ci et est notée \bar{A} . Par exemple, l'événement complémentaire de « le point du dé est au moins 5 » est l'événement « le point du dé est au plus 4 ». Du point de vue ensembliste, c'est l'ensemble complémentaire de A dans Ω tel que définit au cours de mathématiques, c'est-à-dire l'ensemble de tous les éléments de Ω qui ne sont pas dans A . Il est noté $\Omega \setminus A$ ou encore $C_{\Omega}(A)$ ou simplement \bar{A} .

**Propriété.**

$$P(\bar{A}) = 1 - P(A)$$

Démonstration. Il résulte de la définition de l'événement complémentaire que A et \bar{A} sont des événements incompatibles. On a donc par le théorème d'addition :

$$P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

Hors,

$$P(A \cup \bar{A}) = P(\Omega) = 1$$

Et donc :

$$P(\bar{A}) = 1 - P(A)$$

CQFD

1.5. Probabilités conditionnelles

En général, les événements aléatoires liés à une même expérience ont une influence les uns sur les autres, on dit dans ce cas qu'ils sont **dépendants**. La probabilité qu'un événement A se réalise peut être modifiée par le fait de savoir qu'un événement B s'est déjà réalisé précédemment. On notera $P(A/B)$ la **probabilité conditionnelle** de A étant donné que B s'est réalisé (on lit « probabilité de A sachant B » ou encore « probabilité de A si B »).

Prenons l'exemple du jet d'un dé avec les événements A : « le point est 6 » et B : « le point est pair ». On sait que $P(A) = 1/6$. Supposons à présent que B est réalisé, c'est-à-dire que l'on sait déjà que le point est pair. Dans ce cas, il n'y a plus que trois possibilités de résultat et donc $P(A/B) = 1/3$.

Propriété.

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

Démonstration. Supposons la probabilité calculable avec la formule de probabilité *à priori*. Pour calculer $P(A/B)$, on prend pour le calcul des cas possibles uniquement les cas où B se réalise, c'est-à-dire le nombre de cas favorables à B . Au numérateur, on prend parmi ceux du dénominateur les cas où A se réalise aussi, c'est-à-dire le nombre de cas favorables à A ET B :

$$P(A/B) = \frac{\text{nombre de cas favorables à } A \text{ ET } B}{\text{nombre de cas favorables à } B}$$

En divisant le numérateur et le dénominateur par le nombre total de cas possibles, on obtient le résultat. CQFD

Théorème de multiplication

Le théorème de multiplication donne une formule pour la probabilité de l'intersection de deux événements, cette opération correspondant à l'opérateur logique ET. Soient A et B deux événements liés à la même expérience aléatoire ; alors :

$$P(A \cap B) = P(A) * P(B/A)$$

La démonstration découle immédiatement de la propriété précédente.

Evénements indépendants

Lorsque la réalisation d'un événement n'influence pas la probabilité de réalisation d'un autre, on dit alors que ces événements sont **indépendants** :

$$A \text{ et } B \text{ indépendants} \Leftrightarrow P(A/B) = P(A)$$

Propriété. Si A et B sont indépendants, la formule du théorème de multiplication se simplifie en :

$$P(A \cap B) = P(A) * P(B)$$

Exemples. a) On tire deux boules dans une urne contenant 5 boules blanches et 5 boules noires. Quelle est la probabilité de tirer deux boules blanches ? Il faut calculer :

$$\begin{aligned} &P(1^{\text{ère}} \text{ boule blanche ET } 2^{\text{ème}} \text{ boule blanche}) \\ &= P(1^{\text{ère}} \text{ boule blanche}) * P(2^{\text{ème}} \text{ boule blanche} / 1^{\text{ère}} \text{ boule blanche}) \end{aligned}$$

Dans ce cas, le 2^{ème} tirage dépend du 1^{er} : le fait d'avoir tiré une boule blanche en premier influence la probabilité d'en tirer une au second tirage : il ne reste alors plus que 9 boules dans l'urne et parmi celles-ci 4 boules blanches. La probabilité cherchée est donc $5/10 * 4/9 = 2/9 = 0.2222...$

b) même expérience, mais avec remise : on remet la boule tirée dans l'urne après le 1^{er} tirage. Cette fois-ci, les deux tirages sont indépendants. On a donc :

$$\begin{aligned} &P(1^{\text{ère}} \text{ boule blanche ET } 2^{\text{ème}} \text{ boule blanche}) \\ &= P(\text{tirer boule blanche}) * P(\text{tirer boule blanche}) \\ &= 5/10 * 5/10 = 1/4 = 0.25 \end{aligned}$$

1.6. Probabilités totales

La formule des **probabilités totales** s'applique lorsque l'on considère un événement aléatoire se réalisant avec des probabilités différentes suivant différents cas.

Exemple

Soit un service de transport en bus pour lequel on connaît la probabilité de ponctualité suivant les différents jours de la semaine. Le week-end, le trafic étant calme, la probabilité que le bus arrive à l'heure est de 0.9. En semaine, la probabilité est de 0.8, sauf le lundi et le vendredi où à cause du trafic important la probabilité descend à 0.6. On s'interroge alors sur la *probabilité totale* qu'un bus arrive à l'heure, tous jours confondus. Les formules d'addition et de multiplication peuvent nous aider à trouver la réponse à ce problème. Décomposons l'événement « total » *le bus arrive à l'heure* suivant les 3 cas de probabilité :

$$\begin{aligned} P(\text{bus à l'heure}) &= P(\text{bus à l'heure le week-end} \\ &\quad \text{OU bus à l'heure lundi ou vendredi} \\ &\quad \text{OU bus à l'heure mardi, mercredi ou jeudi}) \end{aligned}$$

Les événements séparés par les OU sont clairement incompatibles puisqu'ils concernent des jours différents. On a donc :

$$\begin{aligned}
 P(\text{bus à l'heure}) &= P(\text{bus à l'heure le week-end}) \\
 &\quad + P(\text{bus à l'heure lundi ou vendredi}) \\
 &\quad + P(\text{bus à l'heure mardi, mercredi ou jeudi})
 \end{aligned}$$

A présent, chaque terme de cette somme est la probabilité d'une intersection d'événements :

$$\begin{aligned}
 P(\text{bus à l'heure}) &= P(\text{week-end}) \cdot P(\text{bus à l'heure/week-end}) \\
 &\quad + P(\text{lundi ou vendredi}) \cdot P(\text{bus à l'heure/lundi ou vendredi}) \\
 &\quad + P(\text{mercredi, jeudi ou vendredi}) \cdot P(\text{bus à l'heure/mardi, mercredi ou jeudi})
 \end{aligned}$$

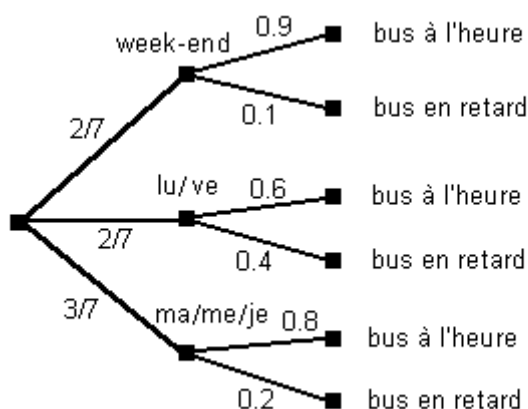
Et finalement : $P(\text{bus à l'heure}) = 2/7 \cdot 0.9 + 2/7 \cdot 0.6 + 3/7 \cdot 0.8 = 0.7714\dots$

Au *total*, il y a donc environ 77% de chances que le bus arrive à l'heure un jour quelconque. Remarquez que le calcul peut aussi se voir comme une *moyenne* des différentes probabilités *pondérée* par les fréquences des différents groupes de jours de la semaine.

Arbre de probabilités

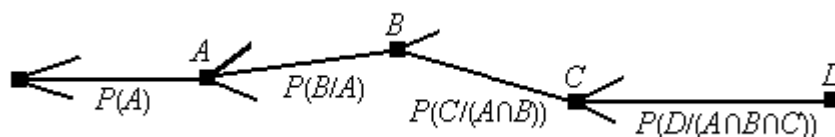
La situation de l'exemple précédent se visualise commodément sur un **arbre de probabilités** : Les caractéristiques d'un arbre de probabilités sont les suivantes :

- les *sommets* de l'arbre (sauf le premier, appelé *racine*) représentent des événements aléatoires
- à chaque *arête* (ou *branche*) est associée une probabilité : la probabilité associée à une arête joignant la racine de l'arbre à un événement A est la probabilité de réalisation de cet événement ; la probabilité associée à une arête joignant un événement A à un événement B est la probabilité conditionnelle $P(B/A)$
- deux événements issus d'un même sommet sont toujours incompatibles
- l'ensemble des événements issus d'un même sommet couvre toutes les possibilités relatives à l'expérience concernée. Dans l'exemple, tous les jours apparaissent au premier niveau de l'arbre ; au second niveau, on a à chaque fois une branche pour l'événement *bus à l'heure* et une autre pour l'événement complémentaire *bus en retard*.
- la somme des probabilités associées aux arêtes issues d'un même sommet vaut toujours 1, ce qui découle des deux points précédents



On visualise facilement sur l'arbre la probabilité que nous avons calculée : on additionne les probabilités des différents chemins menant à l'événement qui nous intéresse ; la probabilité de prendre un chemin donné est obtenue par la multiplication des probabilités associées aux arêtes formant ce chemin.

Le nombre de niveaux de l'arbre peut bien sûr être supérieur à 2, il est théoriquement illimité et peut même être infini (cf. ex. 4 de la série 10). Sur un arbre « multi-niveaux », des branches consécutives forment un enchaînement de probabilités conditionnelles :



Nous allons énoncer à présent de façon formelle la formule des probabilités totales. Mais il faut pour cela définir d'abord le concept de

Système exhaustif d'événements aléatoires

Lorsqu'on partitionne l'ensemble des possibilités Ω en une série de sous-ensembles disjoints, on est en présence d'un **système exhaustif d'événements aléatoires**. La définition est la suivante :

Soit un ensemble d'événements aléatoires $\{E_1, E_2, \dots, E_n\}$ relatifs à une même expérience aléatoire. Cet ensemble forme un système exhaustif si

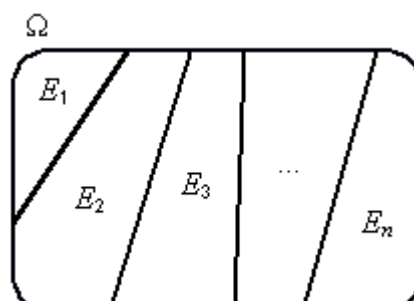
- aucun événement n'est impossible, c'est-à-dire que $P(E_i) > 0 \forall i$
- les événements sont incompatibles deux à deux, c'est-à-dire que $P(E_i \cap E_j) = 0 \forall i, j$
- l'ensemble des événements recouvre la totalité de l'ensemble des résultats possibles, c'est-à-dire que $E_1 \cup E_2 \cup \dots \cup E_n = \Omega$.

On peut visualiser un système exhaustif de la façon suivante :

Exemples et contre-exemples :

- Un événement et son complémentaire constituent toujours un système exhaustif, par exemple $\{\text{le bus arrive à l'heure}\}, \{\text{le bus n'arrive pas à l'heure}\}$
- $\{\text{week-end}\}, \{\text{lundi, vendredi}\}, \{\text{mardi, mercredi, jeudi}\}$ est un système exhaustif pour le choix aléatoire d'un jour de la semaine
- $\{\text{cœurs}\}, \{\text{carreaux}\}, \{\text{cartes noires}\}$ est un système exhaustif pour le tirage d'une carte
- $\{\text{cœurs}\}, \{\text{cartes noires}\}$ n'est pas un système exhaustif pour le tirage d'une carte
- $\{\text{images}\}, \{\text{rouges}\}, \{\text{noires}\}$ n'est pas un système exhaustif pour le tirage d'une carte.
- $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ est un système exhaustif pour le jet d'un dé
- $\{\text{impairs}\}, \{\text{pairs}\}$ est un système exhaustif pour le jet d'un dé

En particulier, tout ensemble d'événements associés aux arêtes issues d'un même sommet d'un arbre de probabilités constitue un système exhaustif.



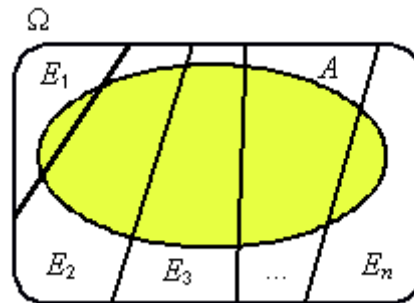
Théorème des probabilités totales

Soit un système exhaustif d'événements aléatoires $\{E_1, E_2, \dots, E_n\}$ et un événement A associé à la

même expérience aléatoire. Alors :
$$P(A) = \sum_{i=1}^n P(E_i) * P(A/E_i)$$

Démonstration. Comme les événements E_i partitionnent l'ensemble des possibilités Ω , on a :

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \dots \cup (A \cap E_n)$$



Comme les E_i sont disjoints :

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_n) = \sum_{i=1}^n P(A \cap E_i)$$

Par le théorème de multiplication, $P(A \cap E_i) = P(E_i) * P(A/E_i)$ et donc :

$$P(A) = \sum_{i=1}^n P(E_i) * P(A/E_i)$$

CQFD

Théorème de Bayes

Dans le type de problèmes où interviennent les probabilités totales, on s'intéresse à la probabilité de réalisation d'un événement A suivant les différents cas du contexte dans lequel cet événement se réalise (cas représentés par les événements E_i formant un système exhaustif). Le théorème de Bayes nous permet de changer l'angle de vue de la situation, en retournant l'ordre chronologique des événements : *sachant que* l'événement A s'est produit, quelle est la probabilité qu'il se soit produit dans le contexte E_i ?

Thèse. Soit un système exhaustif d'événements aléatoires $\{E_1, E_2, \dots, E_n\}$ et un événement A associé à la même expérience aléatoire. Alors :

$$P(E_k/A) = \frac{P(E_k) * P(A/E_k)}{\sum_{i=1}^n P(E_i) * P(A/E_i)}$$

Démonstration. Par le théorème de multiplication, on a $P(A \cap E_k) = P(A) * P(E_k/A)$. Par commutativité de l'intersection, cette expression est aussi égale à $P(E_k) * P(A/E_k)$, d'où :

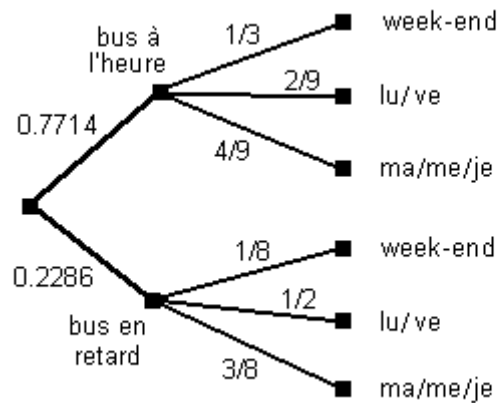
$$P(E_k/A) = \frac{P(E_k) * P(A/E_k)}{P(A)}$$

Ensuite, le dénominateur s'obtient en appliquant le théorème des probabilités totales. CQFD

Exemple. Reprenons les données du problème du bus. Nous sommes à présent en mesure de répondre à la question : *sachant que le bus est arrivé à l'heure, quelle est la probabilité que cela se soit passé durant le week-end ?* Par le théorème de Bayes :

$$\begin{aligned} P(\text{week-end}/\text{bus à l'heure}) &= P(\text{week-end}) * P(\text{bus à l'heure}/\text{week-end}) / P(\text{bus à l'heure}) \\ &= 2/7 * 0.9 / 0.7714... = 1/3 \end{aligned}$$

Il y a donc une chance sur 3 que cela se soit passé durant le week-end. En calculant les autres probabilités de ce type, on peut ainsi établir l'*arbre inverse*, échangeant les causes et conséquences dans le déroulement des événements :



Chapitre 2

Variables aléatoires et grandes distributions théoriques.



2.1. Définition

Une **variable aléatoire** associe une valeur numérique au résultat d'une expérience aléatoire. Si l'expérience est de type quantitatif, il peut s'agir de la valeur obtenue à l'issue de l'expérience ou une fonction de celle-ci. Dans le cas d'un jeu de hasard, il s'agit souvent du gain (qui est plus souvent une perte !) associé aux différents résultats du jeu. Quelques exemples :

- on jette un dé et on note le point obtenu. On définit ainsi une variable aléatoire dont la valeur est le point du dé
- on jette un dé avec les règles suivantes : si on obtient 6, on gagne 3 € ; si on obtient 2 ou 4 on perd 2 €, pour les autres faces, il ne se passe rien. On définit ainsi une variable aléatoire dont les valeurs possibles sont : -2, 0 et 3
- on jette 3 fois un dé et on compte le nombre de 6 obtenu. Ceci définit une variable aléatoire dont les valeurs sont 0, 1, 2 et 3
- on jette un dé plusieurs fois de suite, jusqu'à ce que le point obtenu soit 6. Ceci définit une variable aléatoire dont les valeurs possibles sont tous les naturels non nuls
- on mesure la taille d'un individu choisi au hasard dans la rue. Ceci définit une variable aléatoire dont les valeurs sont comprises dans l'intervalle] 0m, 2.5m]

Nous distinguons dans ces exemples deux types de variables aléatoires : la variable aléatoire est **discrète** si elle possède un ensemble fini (exemples a, b, c) ou infini dénombrable de valeurs (exemple d) ; elle est **continue** si ses valeurs forment un ensemble infini non dénombrable de valeurs, le plus souvent un intervalle de valeurs réelles (exemple e). Dans ce chapitre, nous étudierons d'abord les variables discrètes, les variables continues feront l'objet du chapitre suivant.

La notation habituelle d'une variable aléatoire est X . Une variable aléatoire X se définit donc comme une fonction de Ω dans \mathbb{R} , qui associe une valeur réelle à un résultat d'une expérience aléatoire.

2.2. Distribution de probabilité d'une variable aléatoire discrète

La **distribution de probabilité** d'une variable aléatoire se définit par la donnée des différentes valeurs possibles de la variable (notées x_i) accompagnées de leur probabilité respective (notée p_i). Ainsi, la distribution de la variable aléatoire X = « gain du jeu » de l'exemple b) serait la suivante :

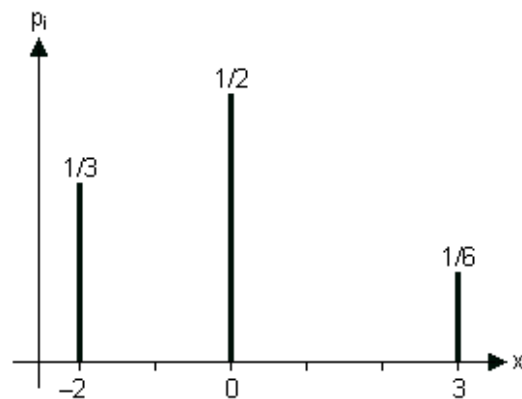
x_i	p_i
-2	1/3
0	1/2
3	1/6

Par définition, on a donc $p_i = P[X = x_i]$.

Propriété : La somme des probabilités des différentes valeurs d'une variable aléatoire est toujours

égale à 1.
$$\sum_{i=1}^n p_i = 1$$

Comme en statistique descriptive, on peut représenter cette distribution par un diagramme en bâtonnets :



2.3. Fonction de répartition d'une variable aléatoire discrète

Pour une variable aléatoire discrète, la **fonction de répartition** est une fonction en escalier définie par :

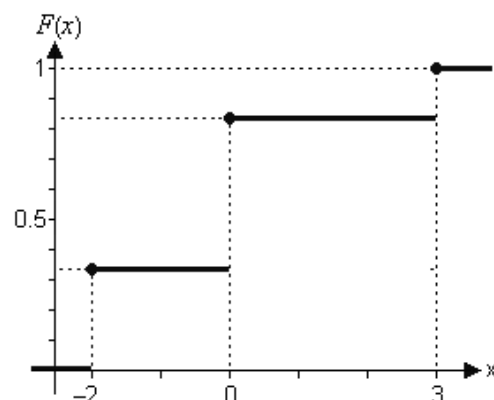
$$F(x) = P[X \leq x]$$

Elle possède des valeurs pour tout réel x . Ses valeurs extrêmes sont 0 lorsque x tend vers $-\infty$ et 1 lorsque x tend vers $+\infty$. En particulier, pour les valeurs x_i prises par la variable aléatoire, $F(x_i)$ est égal à la somme des probabilités correspondant aux valeurs inférieures ou égales à x_i .

Pour l'exemple b), ajoutons dans le tableau précédent une colonne donnant les valeurs de la fonction de répartition en chaque x_i :

x_i	p_i	$F(x_i)$
-2	1/3	1/3
0	1/2	5/6
3	1/6	1

Le graphe de $F(x)$ a alors l'allure suivante :



Notez que $F(x)$ est définie pour tout x réel. Par exemple $F(1) = 5/6$. La probabilité de gagner 1 € au jeu du dé est nulle, mais pas la probabilité de gagner *au plus* 1 € (c'est-à-dire 1 € ou moins).

2.4. Lien avec la statistique descriptive

L'analogie avec la statistique descriptive est évidente : aux fréquences relatives (f_i) et fréquences relatives cumulées (F_i) de la statistique descriptive correspondent maintenant les probabilités (p_i) et les valeurs de la fonction de répartition ($F(x_i)$). Les graphiques (en bâtonnets et en escalier) sont similaires. La différence réside dans la signification des valeurs : en statistique elles résultent d'une série d'observations existantes, tandis qu'ici les valeurs sont des probabilités, c'est-à-dire des valeurs théoriques liées à une expérience dont le résultat n'est pas encore connu.

L'analogie se poursuit dans l'étude des différents paramètres liés à une variable aléatoire : ainsi, à la notion de moyenne va correspondre celle d'espérance mathématique.

2.5. Paramètres d'une variable aléatoire discrète

2.5.1. L'espérance mathématique

L'**espérance mathématique** d'une variable aléatoire X , notée $E[X]$ (ou plus simplement μ) est la valeur moyenne que peut prendre cette variable aléatoire. Par exemple, si X représente le gain associé à un jeu de hasard, $E[X]$ est le gain moyen de ce jeu, c'est-à-dire la valeur que l'on peut espérer gagner en jouant à ce jeu. L'espérance mathématique est donc l'équivalent de la moyenne arithmétique de la statistique descriptive, et elle se calcule par une formule similaire :

$$E[X] = \sum_{i=1}^n p_i x_i$$

Exemple : pour le jet du dé (exemple b) l'espérance mathématique donne $E[X] = (-2) \cdot 1/3 + 0 \cdot 1/2 + 3 \cdot 1/6 = -1/6$ €. Cela veut dire qu'en moyenne, on perd à chaque partie environ 0.17 €.

2.5.2. La variance et l'écart-type

Comme en statistique, l'**écart-type** (noté σ) mesure la dispersion des valeurs de la variable aléatoire autour de sa moyenne (c'est-à-dire μ ou $E[X]$). La formule de l'écart-type est analogue à celle de la statistique descriptive :

$$\sigma^2 = \sum_{i=1}^n p_i x_i^2 - \mu^2$$

La formule précédente donne le carré de l'écart-type, aussi appelé la **variance**. L'écart-type est donc la racine carrée de la variance.

Formule de calcul pratique : on retiendra la formule donnée par la propriété suivante. Elle est équivalente pour le calcul de σ , mais nécessite un nombre moins élevé d'opérations.

Propriété :

$$\sigma^2 = \sum_{i=1}^n p_i x_i^2 - \mu^2$$

Signification intuitive de l'écart-type : on se rappellera surtout cette propriété, similaire à ce qui a été vu en statistique : connaissant l'espérance μ et l'écart-type σ d'une distribution de probabilité, on peut s'attendre qu'environ 68% des valeurs prises par la variable aléatoire se trouveront à au plus un écart-type de distance de la moyenne, c'est-à-dire dans l'intervalle $[\mu - \sigma, \mu + \sigma]$. Environ 95% des valeurs seront éloignées au plus de 2 écart-types de distance, c'est-à-dire qu'elles se trouveront dans

l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$. A peine 1% des valeurs se trouveront à plus de 3 écart-types d'écart de la moyenne.

2.6. Variable aléatoire continue

De façon analogue aux observations statistiques de type quantitatif continu, une variable aléatoire continue peut prendre des valeurs réelles dans un intervalle donné (qui peut être à priori non borné).

Exemples :

- a) on choisit une personne au hasard et on note son poids
- b) on lance une pierre dans une direction donnée et on mesure la distance parcourue
- c) on se rend à un arrêt de bus et on note le temps d'attente

2.6.1. Distribution par classes

De la même manière qu'en statistique descriptive pour les observations continues, on donne les probabilités pour des classes (intervalles) de valeurs de la variable aléatoire :

$$p_i = P[x_{i-1} \leq X < x_i]$$

Prenons comme exemple la taille d'un individu choisi au hasard dans une grande population où les probabilités connues seraient les suivantes (issues par ex. d'observations statistiques)

Classes $[x_{i-1}, x_i[$ (données en cm)	p_i	dp_i	$F(x_i)$
[150, 170[0.3	0.015	0.3
[170, 180[0.45	0.045	0.75
[180, 200[0.25	0.0125	1

Notez que dans ce type de définition, on suppose l'**uniformité** des probabilités dans chaque classe. Par exemple, le tableau indique que $P[150 \leq X < 170] = 0.3$, mais ne permet pas de savoir s'il y a plus d'individu dans la 1^{ère} moitié de cette classe (de 150 à 160) ou dans la seconde (de 160 à 170). On suppose donc dans ce cas que la probabilité est uniforme : il y a une probabilité de 0.15 dans chaque moitié.

C'est cette idée qui conduit à définir la **densité de probabilité** (notée dp_i) : elle est égale à la probabilité p_i d'une classe divisée par la largeur de la classe (que nous notons l_i) :

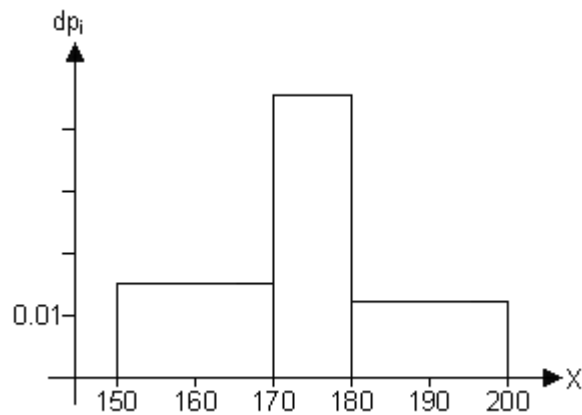
$$dp_i = \frac{p_i}{l_i}$$

Cette densité donne dans une classe, la probabilité par intervalle d'une unité. Par exemple, pour la première classe, on obtient $dp_i = 0.015$, ce qui signifie que la probabilité de trouver quelqu'un mesurant entre 150 et 151 cm vaut 0.015, et de même pour toutes les autres tranches d'un centimètre de cette classe (toujours par le principe d'uniformité).

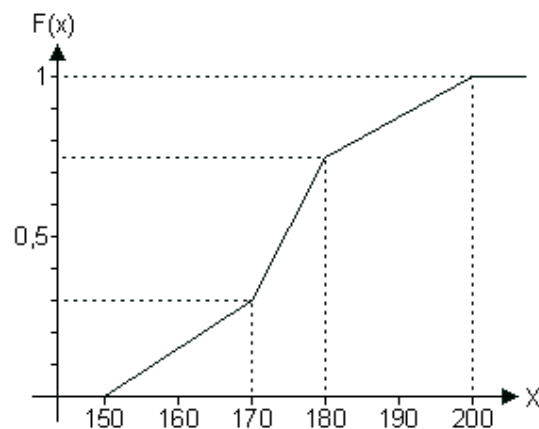
La distribution de probabilité peut être visualisée par l'histogramme des densités de probabilités. Notez que sur ce graphique :

- a) l'aire de chaque rectangle est égale à p_i
- b) l'aire totale vaut 1

la probabilité de se trouver dans un intervalle donné est égale à l'aire des rectangles et portions de rectangles correspondant à cet intervalle.



La **fonction de répartition** $F(x)$ est (encore et toujours !) analogue à celle vue dans le chapitre de statistique. Son graphique est une ligne continue constituée de différents segments :



Nous en rappelons les principales propriétés :

- a) $F(x)$ est la probabilité que X soit inférieure ou égale à x , quel que soit le réel x
- b) $F(x_i)$ (où x_i est la borne droite d'une classe) est la somme des probabilités de toutes les classes à gauche de x_i
- c) la pente d'un segment équivaut à la densité de probabilité dp_i de la classe correspondante
- d) l'équation du segment de la classe $[x_{i-1}, x_i[$ est $y = dp_i (x - x_{i-1}) + F(x_{i-1})$
- e) $P[a \leq X \leq b] = F(b) - F(a)$
- f) La probabilité en un point donné est toujours nulle : $P[X = a] = 0$

Pour calculer les paramètres d'une telle distribution, on travaille avec les **centre-classes** c_i ; pour l'espérance mathématique :

$$E[X] = \sum_{i=1}^n p_i c_i \quad (\text{où } c_i = \frac{x_{i-1} + x_i}{2})$$

Pour la variance et l'écart-type :

$$\sigma^2 = \sum_{i=1}^n \frac{c_i^2 p_i}{c_i}$$

2.6.2. Fonction de densité de probabilité

Imaginons dans l'exemple précédent que nous donnions la probabilité pour des classes de largeur de 1 cm : au lieu de 3 classes, nous aurions un tableau avec 50 classes dont les probabilités seront très petites. Imaginons encore des classes plus petites : les probabilités se rapprocheraient de 0, tandis que les densités vont se stabiliser ponctuellement.

Dans le cas limite où la largeur des classes tend vers 0, la probabilité ponctuelle disparaît, et seule la densité de probabilité a encore une signification. On peut dès lors définir une variable aléatoire par la donnée d'une **fonction de densité de probabilité**. C'est une fonction $f(x)$ positive en tout point et dont l'intégrale définie sur son domaine est égale à 1 :

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

La probabilité que X se trouve dans un intervalle $[a, b]$ est égale à l'aire sous le graphique entre les points a et b , ce qu'on obtient par un calcul d'intégrale :

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

La **fonction de répartition** est alors définie par :

$$F(x) = P[X \leq x] = \int_{-\infty}^x f(t) dt$$

$F(x)$ étant la primitive de $f(x)$, la connaissance de la fonction de répartition permet donc aussi de calculer les probabilités par une simple soustraction :

$$P[a \leq X \leq b] = F(b) - F(a)$$

2.7. Paramètres d'une variable aléatoire continue

Les paramètres de la variable aléatoire se définissent par des formules analogues aux précédentes, les signes de sommation devenant ici des intégrales :

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

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2$$

En général, chaque expérience aléatoire possède sa distribution de probabilité particulière, mais il existe aussi beaucoup de situations qui ont en commun un même modèle de distribution théorique, différant seulement par la valeur de quelques paramètres.

Nous étudions deux modèles importants : la **distribution binomiale** et la **distribution normale** (aussi appelée **distribution de Gauss**).

2.8. La distribution binomiale

2.8.1. Définition

La **loi binomiale** fait partie des plus anciennes lois de probabilités étudiées. Elle a été introduite par le mathématicien suisse Jacques Bernoulli (1654-1705). C'est un modèle théorique important de la théorie des probabilités qui s'applique dans le cas où une même expérience aléatoire est répétée plusieurs fois dans des conditions identiques et indépendantes.

Soit n le nombre de répétitions de cette expérience. On considère alors la variable aléatoire discrète qui compte le nombre de réalisations d'un certain événement aléatoire A au cours de ces n expériences. Chaque fois que A se réalise, on parle de **succès**, dans le cas contraire (c'est le complémentaire de A qui se réalise) on parle d'un **échec**. La probabilité de succès est notée p et celle d'échec est notée q (avec bien entendu $p + q = 1$)

Si X est la variable aléatoire comptabilisant le nombre de succès réalisés au cours des n expériences, on dit alors que X admet une **distribution binomiale** de paramètres n et p , ce qu'on note $X \sim B(n, p)$. Ses valeurs possibles sont les entiers de 0 à n et les probabilités associées à ces valeurs sont données par la formule :

$$P[X=k] = C_n^k p^k q^{n-k} \quad (0 \leq k \leq n)$$

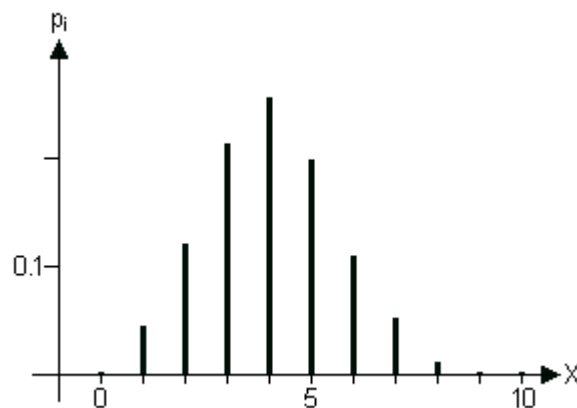
Pour prouver qu'il s'agit d'une distribution valide, il faut vérifier que la somme des probabilités est bien égale à 1. La preuve fait intervenir la formule du binôme de Newton :

$$p + q = 1 \implies (p + q)^n = 1^n = 1$$

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

2.8.2. Profil de la distribution binomiale

Le graphe en bâtonnets ci-dessous représente les valeurs de la distribution binomiale de paramètres $n = 10$ et $p = 0,4$



La probabilité de succès p étant inférieure à $1/2$, le graphique présente une dissymétrie vers la gauche. Dans le cas où p est supérieur à $1/2$, les valeurs seraient plutôt concentrées du côté droit. Le graphique n'est parfaitement symétrique que lorsque $p = q = 1/2$, on a alors $P[X = k] = P[X = n - k]$ pour toute valeur de k .

2.8.3. Paramètres de la distribution binomiale

Propriété. Si $X \sim B(n, p)$, alors $\mu = E[X] = np$ et $\sigma = \sqrt{npq}$

Nous voyons d'après la formule de l'écart-type que si p est proche de 0 ou de 1, la dispersion sera proche de 0. La dispersion maximale est atteinte pour $p = q = 1/2$.

2.9. La distribution normale

2.9.1. Définition

La **distribution normale**, aussi appelée **distribution de Gauss**, du nom de son principal instigateur Johann Carl Friedrich Gauss (1777-1855) est une distribution continue. C'est certainement la loi statistique la plus utile et la plus répandue, puisqu'on la retrouve dans une multitude de situations où le hasard apparaît de façon « naturelle », par exemple :

- la taille des individus d'une population donnée
- le poids des œufs produits par un élevage de volaille
- la température observée à un endroit donné
- le quotient intellectuel d'une population d'étudiants
- la durée de vie des individus d'une population donnée

Les exemples témoignant de l'importance de cette distribution sont innombrables. De plus, la distribution normale est l'acteur central des théories des échantillons et de l'estimation que nous étudierons plus loin. Nous verrons en outre comment elle se relie aux autres distributions discrètes étudiées précédemment.

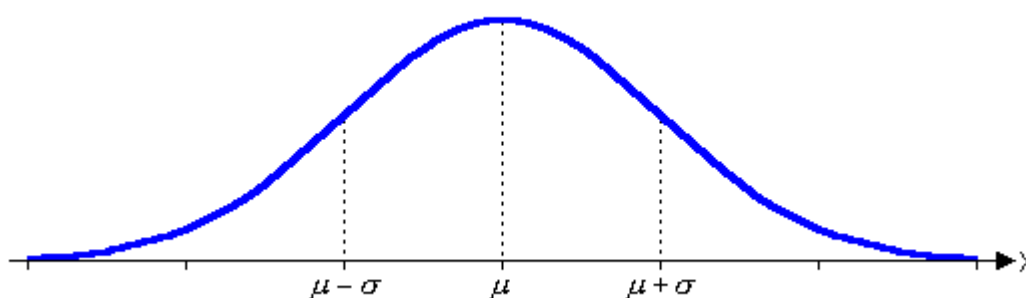
La fonction de densité de probabilité définissant la loi normale est la suivante :

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

Deux paramètres y apparaissent : μ et σ qui sont respectivement l'espérance (moyenne) et l'écart-type de la distribution. Nous ne développerons pas ici le calcul de ces paramètres, qui fait appel à des techniques d'intégration ardues. Si X est une variable aléatoire admettant cette distribution, on dira que X est une normale de paramètre μ et σ , ce qu'on note : $X \sim N(\mu, \sigma)$.

2.9.2. Profil de la distribution normale

Le graphique de la fonction de densité de probabilité $f(x)$ de la distribution normale est la fameuse *courbe en cloche*, encore appelée *courbe de Gauss* ou tout simplement *gaussienne*.



On retiendra ses principales caractéristiques :

- l'axe des x est une asymptote horizontale au graphique : lorsque x est très grand (ou très petit), l'exponentielle tend vers 0
- symétrie en μ : la courbe a la même allure de part et d'autre de la moyenne
- sommet en μ : le maximum de $f(x)$ est atteint lorsque $x = \mu$ (on peut le montrer facilement en dérivant la fonction)
- points d'inflexion en $\mu - \sigma$ et $\mu + \sigma$: la courbe change de concavité à un écart-type de distance de part et d'autre de la moyenne (on peut le montrer en vérifiant que ces valeurs annulent la dérivée seconde de $f(x)$)
- l'aire sous la courbe vaut 1, quel que soient les valeurs des deux paramètres ; c'est obligatoire d'ailleurs pour une fonction de densité, mais le calcul le démontrant n'est pas aisé, on admettra donc cette propriété sans démonstration.

Nous rappelons que le calcul de probabilité d'une variable aléatoire continue résulte d'un calcul d'intégrale : la probabilité que X prenne des valeurs comprises entre a et b est égale à l'aire sous la courbe de $f(x)$ entre ces deux points :

$$f(x) dx = F(b) - F(a)$$

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

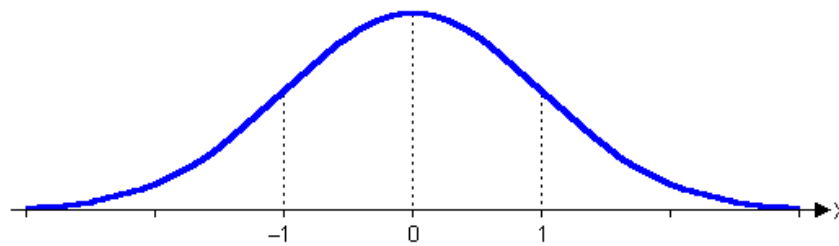
où $F(x)$ est la fonction de répartition, primitive de $f(x)$. Ce calcul pose un problème pratique dans le cas de la distribution normale. En effet, il n'est pas possible d'intégrer $f(x)$ par les techniques traditionnelles, car la fonction e^{-x^2} ne possède pas de primitive élémentaire. Pour calculer une telle probabilité, il faudra consulter une table de valeurs de la fonction de répartition qui est établie pour un cas particulier de distribution normale, appelée **distribution normale réduite**. Nous verrons ensuite comment on passe d'une normale quelconque à cette distribution particulière.

2.9.3. La distribution normale réduite

C'est le nom donné à la distribution normale de paramètres $\mu = 0$ et $\sigma = 1$. Pour la distinguer d'une normale « quelconque », elle est habituellement notée Z plutôt que X : $Z \sim N(0, 1)$. L'équation de sa fonction de densité de probabilité se simplifie en :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Son graphique est donc centré en 0 et les points d'inflexions sont en -1 et 1 . L'aire sous la courbe vaut toujours 1 :



2.9.4. Fonction de répartition de la distribution normale réduite

Pour rappel, la fonction de répartition $F(x)$ donne la probabilité que la variable aléatoire soit inférieure à une valeur x donnée :

$$F(x) = P[Z \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$$

Ce calcul d'intégrale est toujours inaccessible par les méthodes classiques d'intégration, c'est pourquoi des tables ont été développées donnant des valeurs de $F(x)$ obtenues par des techniques d'approximations. La façon la plus simple d'approcher $F(x)$ est d'effectuer un développement de l'intégrande en série de Mac Laurin. Par exemple, en se limitant à l'ordre 6, on obtient :

$$F(x) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \left(1 - \frac{1}{2}t^2 + \frac{1}{8}t^4 - \frac{1}{48}t^6 \right) dt$$

Après intégration, on obtient la formule suivante, qui donne des valeurs satisfaisantes pour $|x| < 1$:

$$F(x) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(x - \frac{1}{6}x^3 + \frac{1}{40}x^5 - \frac{1}{336}x^7 \right)$$

2.9.5. Table de la fonction de répartition de la distribution normale réduite $Z \sim N(0, 1)$

La table suivante donne les valeurs arrondies à la 4^{ème} décimale de $F(x) = P[-\infty \leq Z \leq x]$ pour toutes les valeurs positives de x multiples de 0,01 entre 0 et 3.

x	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	0,7518	0,7549
0,7	0,7580	0,7612	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
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,5	0,9998									
4,0	1									

Pour des valeurs négatives de x , on se rappellera que le graphique de la normale réduite est symétrique par rapport à 0. Comme l'aire totale sous la courbe est égale à 1, on a la formule :

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

2.9.6. Passage d'une normale quelconque à la normale réduite

La table ci-dessus ne donne que les valeurs de la distribution normale réduite $Z \sim N(0, 1)$. Heureusement, les valeurs pour toute autre normale $X \sim N(\mu, \sigma)$ peuvent être déterminées par la même table, via une transformation simple de la variable aléatoire. Il s'agit d'un changement d'origine et d'unité :

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

Nous avons vu dans le chapitre 2 les propriétés d'un tel changement linéaire : la moyenne de la distribution se transforme de la même façon, tandis que l'écart-type ne subit que le changement d'unité. On voit ainsi que Z possède à présent les paramètres de la normale réduite : $Z \sim N(0, 1)$.

Il est à présent simple de calculer la probabilité sur un intervalle $[a, b]$ d'une normale $X \sim N(\mu, \sigma)$ quelconque :

$$P[a \leq X \leq b] = P\left[\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right] = F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right)$$

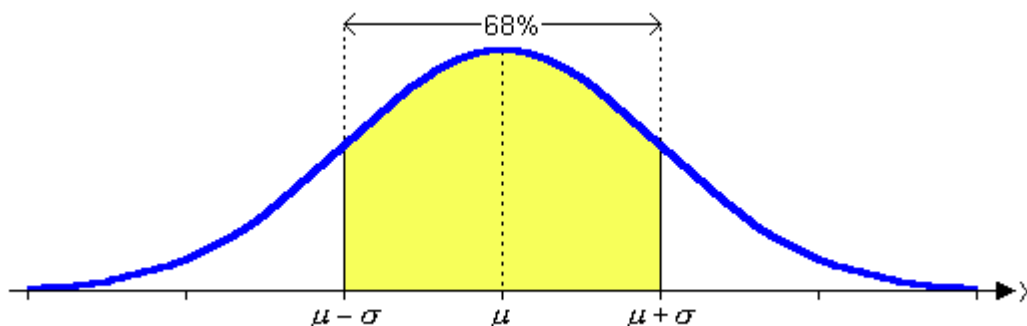
où $F(x)$ désigne toujours la fonction de répartition de la normale réduite.

Remarquer que le passage d'une normale quelconque à la normale réduite est un simple changement d'échelle. A la moyenne d'une distribution normale correspond le 0 de la normale réduite, et à un écart-type sur le graphique d'une normale quelconque correspond une unité sur le graphique de la normale réduite.

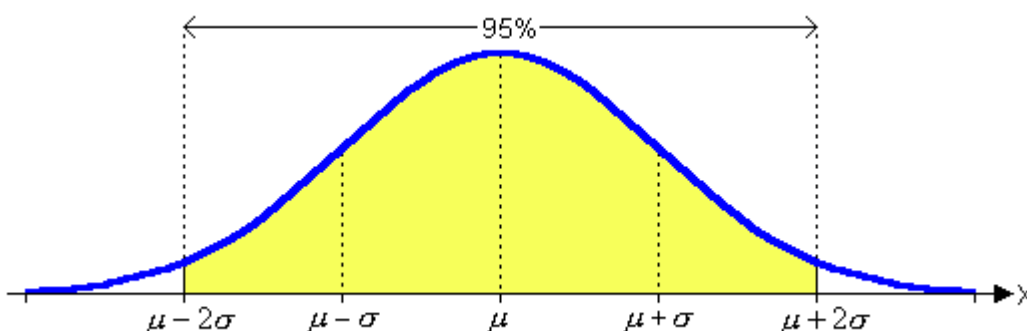
2.9.7. Quelques valeurs typiques

On essaiera de se rappeler ces valeurs remarquables :

- $P[\mu - \sigma \leq X \leq \mu + \sigma] = P[-1 \leq Z \leq 1] = 0,6826$; autrement dit, la probabilité de se trouver à moins d'un écart-type de distance de la moyenne est d'environ 68%



- $P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = P[-2 \leq Z \leq 2] = 0,9544$; la probabilité de se trouver à moins de deux écart-types de distance de la moyenne est d'environ 95%, et donc la probabilité de se trouver au-delà de deux écart-type est très faible, elle ne vaut que 0,0456.



Si on considère une distance de 3 écart-types par rapport à la moyenne, la probabilité de se trouver dans l'intervalle vaut alors 0,9974. La probabilité de se trouver au-delà de 3 écart-types est donc quasi nulle.

2.10. Approximation de la binomiale par la normale

Lorsque n est grand, le calcul de probabilité d'une variable distribuée par la loi binomiale s'avère long et fastidieux. Ce calcul peut être facilité et remplacé par une approximation à l'aide de la distribution normale. Ce fait s'appuie sur le théorème de Moivre-Laplace qui affirme que *la distribution binomiale est asymptotiquement normale*.

En pratique, si on superpose le graphique en bâtonnets d'une distribution binomiale $X_B \sim B(n, p)$ avec un n « grand » (on peut le considérer comme tel à partir de 30) et celui de la normale de paramètres correspondants, $X_N \sim N(np, \sqrt{npq})$, on observe que le tracé de la courbe de Gauss se confond avec les extrémités des bâtonnets.

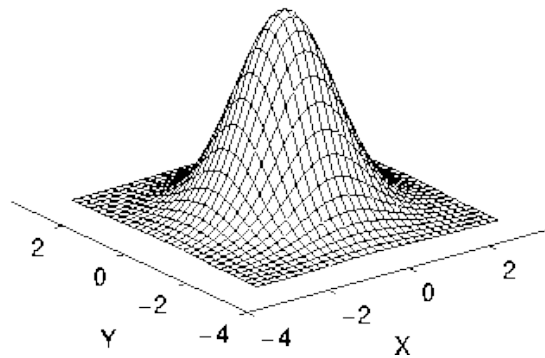
Dès lors, on peut approcher chaque probabilité $P[X_B = k]$ par l'aire sous la courbe de Gauss entre les valeurs $k - 1/2$ et $k + 1/2$ sans grande perte de précision. En général on aura donc :

$$P[a \leq X_B \leq b] \approx P\left[a - \frac{1}{2} \leq X_N \leq b + \frac{1}{2}\right]$$

L'ajout des $1/2$ de part et d'autre de l'intervalle s'appelle la correction de continuité. Elle peut être négligée pour des très grandes valeurs de n .

Chapitre 3

Distributions à deux dimensions



3.1. Introduction

Toutes les distributions étudiées jusqu'ici étaient unidimensionnelles, c'est-à-dire qu'un seul aspect d'une population était abordé, par exemple le poids ou la taille d'individus. Dans une distribution à deux dimensions, on considérera plutôt un **couple** de variables aléatoires, par exemple la taille X en cm **et** le poids Y en kg.

Le tableau suivant donne un exemple de la distribution du couple (X, Y) dans une population donnée. Dans cet exemple, les valeurs possibles de X et Y ont été groupées par classes:

Y	$[50,65[$	$[65,75[$	$[75,85[$	$[85,100[$	$p_{i.}$
X					
$[140,160[$	0.05	0.10	0.05	0	0.20
$[160,170[$	0.05	0.12	0.08	0.05	0.30
$[170,180[$	0	0.08	0.17	0.08	0.33
$[180,200[$	0	0.05	0.10	0.02	0.17
$p_{.j}$	0.10	0.35	0.40	0.15	1

Pour fixer les idées, ce tableau nous renseigne que dans la population étudiée, la probabilité qu'un individu choisi au hasard mesure entre 170 et 180 cm **et** pèse entre 75 et 85 kg vaut 0.17. Bien sûr, pour avoir une distribution correcte, il faut que la somme totale des probabilités soit égale à 1. Remarquez aussi que X et Y possèdent chacune leur propre distribution de probabilité dont il est facile de retrouver les valeurs: il suffit pour cela de calculer respectivement les totaux par ligne et par colonne; ces probabilités sont appelées **probabilités marginales**.

3.2. Définitions

Soient deux variables aléatoires X et Y , prenant respectivement les valeurs x_1, \dots, x_m et y_1, \dots, y_n . La distribution de probabilité à deux dimensions du couple (X, Y) est définie par l'ensemble des valeurs

$$X = x_i \quad \text{et} \quad Y = y_j$$

$$p_{ij} = P(X = x_i, Y = y_j)$$

La distribution est valide si :

$$\sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1$$

On définit encore les probabilités marginales:

$$p_{i.} = P[X = x_i] = \sum_{j=1}^n p_{ij} \quad \text{et} \quad p_{.j} = P[Y = y_j] = \sum_{i=1}^m p_{ij}$$

Remarquez que ces définitions ne sont pas étrangères au théorème des probabilités totales vu au chapitre 2.

3.3. Covariance et corrélation

Dans une distribution à deux dimensions, les variables X et Y possèdent chacune leurs valeurs caractéristiques (moyennes μ_x et μ_y et écarts-type σ_x et σ_y), qu'on calcule cette fois à l'aide des probabilités marginales $p_{i.}$ et $p_{.j}$ par les formules vues pour les distributions à une dimension.

Un nouveau paramètre important fait son apparition, il s'agit de la **covariance**. Ce paramètre mesure la dépendance des variables X et Y et est défini par:

$$\sigma_{xy} = \sum_{i=1}^m \sum_{j=1}^n p_{ij} (x_i - \mu_x)(y_j - \mu_y)$$

On retiendra pour le calcul pratique la formule équivalente suivante :

Propriété :

$$\sigma_{xy} = \sum_{i=1}^m \sum_{j=1}^n p_{ij} x_i y_j - \mu_x \mu_y$$

La covariance conduit au **coefficient de corrélation** défini par :

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

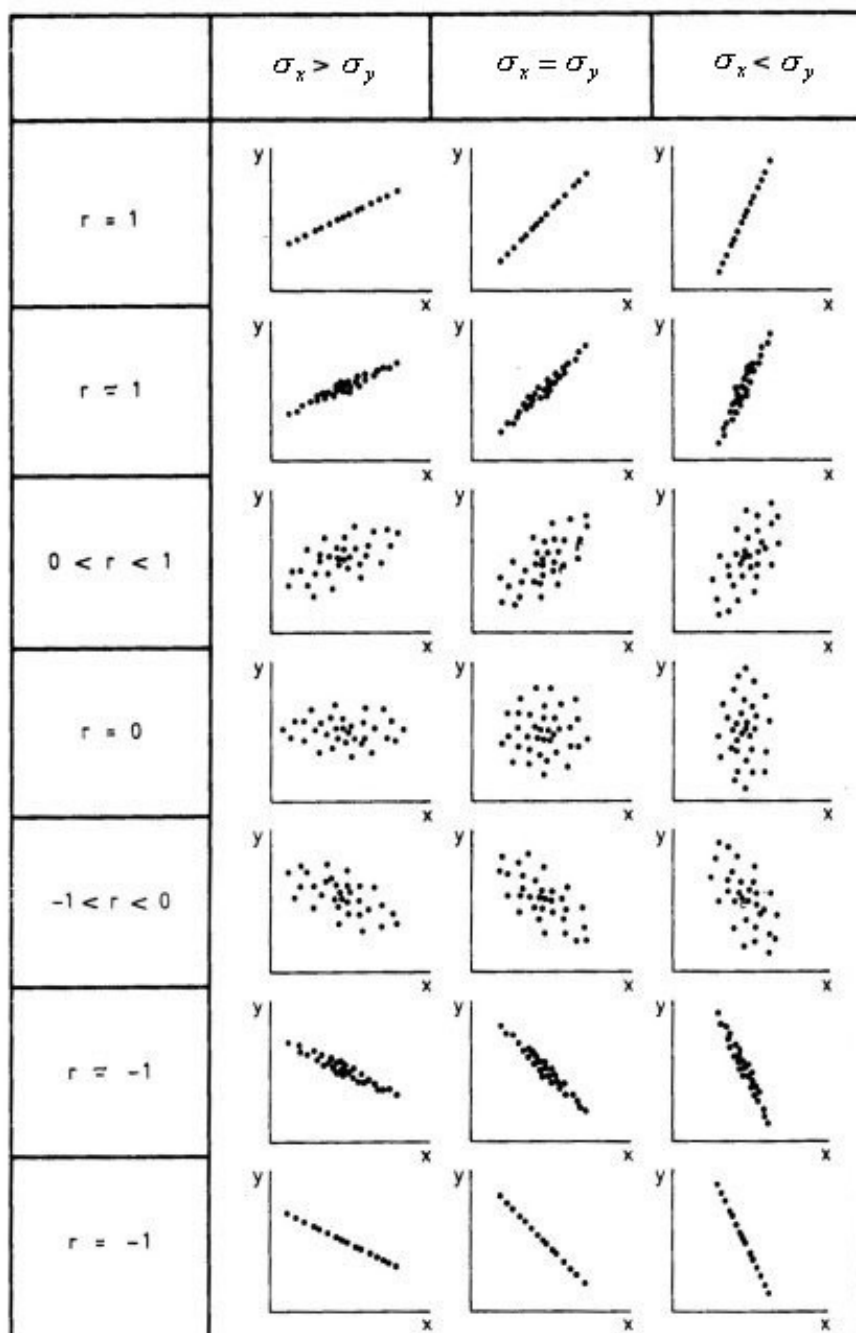
On peut montrer que ce coefficient est toujours compris entre -1 et 1 . Selon que sa valeur absolue est plus proche de 0 ou de 1, la dépendance (ou corrélation) entre X et Y est respectivement plus faible ou plus forte.

Pour l'exemple introduit dans ce chapitre, on trouve les résultats suivants:

$\mu_x = 169.55 \text{ cm}$; $\mu_y = 76.125 \text{ kg}$; $\sigma_x = 12.789 \text{ cm}$; $\sigma_y = 9.698 \text{ kg}$; $\sigma_{xy} = 45.25625$ et $r = 0.365$.

Les dessins qui suivent montrent l'allure générale du nuage de points pour différentes valeurs de r entre -1 et 1 . Les points tendent à s'aligner lorsque r est proche de 1 en valeur absolue. Lorsque r est nul, le nuage de points n'accuse aucun profil linéaire (nous verrons plus loin que deux variables indépendantes ont toujours un coefficient de corrélation nul). Notez que lorsque r est positif, les valeurs de Y ont tendance à croître avec celles de X . Lorsque r est négatif, les croissances sont opposées : lorsque X augmente, Y diminue et vice-versa.

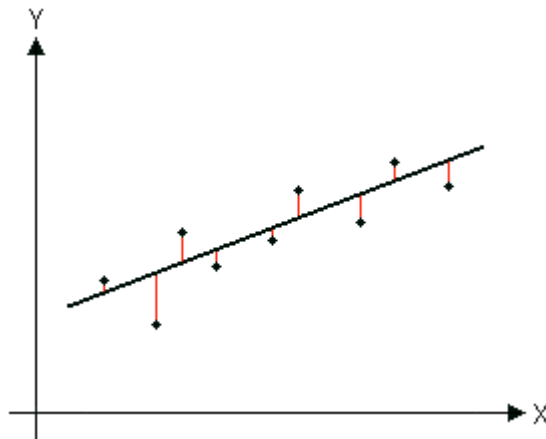
En colonne, on compare les situations déterminées par la comparaison entre les écart-types de X et de Y . Par exemple dans la 1^{ère} colonne, c'est σ_x qui est le plus grand, c'est-à-dire que le nuage de points à une dispersion plus grande dans le sens des X que dans le sens des Y , et donc sa forme est plus écrasée dans le sens vertical que dans le sens horizontal.



3.4. Droite de régression

Dans le cas d'une corrélation forte entre X et Y , nous avons vu que les points formés par les valeurs du couple (X, Y) forment un « nuage » d'allure linéaire, autrement dit, la donnée de x_i permet de « deviner » à peu près la valeur y_i . D'où l'idée d'essayer de trouver l'équation d'une droite qui approcherait au mieux l'ensemble de ces points quasi alignés.

Une solution est donnée par la **droite de régression** (encore nommée **droite des moindres carrés**). C'est la droite qui minimise la somme des carrés des écarts verticaux entre les points (x_i, y_i) et son graphe.



Propriété : l'équation de la droite de régression $y = ax + b$ a comme coefficients :

$$a = \frac{\sigma_{xy}}{\sigma_x^2} \quad \text{et} \quad b = \mu_y - a\mu_x$$

Notez que le signe du coefficient angulaire de la droite de régression est également celui de la covariance. En particulier, lorsque la covariance est nulle, la droite est horizontale et son équation est $y = \mu_y$.

3.5. Indépendance des deux variables

La définition des probabilités p_{ij} ci-dessus peut nous faire penser au théorème de multiplication et nous savons alors que la probabilité sera plus facile à calculer en cas d'indépendance. En effet, si X et Y sont indépendantes, on aura

$$X = x_i \quad \text{et} \quad Y = y_j \quad p_{ij} = P[X = x_i] * P[Y = y_j] = p_{i.} * p_{.j}$$

Ainsi, si X et Y sont indépendantes, chaque probabilité p_{ij} du tableau est égale au produit des probabilités marginales en fin de ligne et de colonne.

Propriété : la covariance de deux variables indépendantes est nulle

3.6. Généralisation à plusieurs dimensions

Dans la suite du cours, nous serons amenés à considérer l'action de plusieurs variables aléatoires X_1, X_2, \dots, X_n simultanément. Nous donnons ici seulement quelques définitions et résultats utiles pour la suite.

Définition : les variables aléatoires X_1, X_2, \dots, X_n sont indépendantes ssi

$$P[X_1 = a_1 \text{ et } X_2 = a_2 \text{ et } \dots \text{ et } X_n = a_n] = P[X_1 = a_1] * P[X_2 = a_2] * \dots * P[X_n = a_n] \quad \forall a_1, a_2, \dots, a_n$$

Nous aurons aussi à considérer des variables aléatoires résultant d'une somme de variables aléatoires. Notons alors les propriétés suivantes:

Propriété : l'espérance mathématique d'une somme de variables aléatoires (indépendantes ou non) est égale à la somme des espérances mathématiques de ces variables aléatoires :

$$E[X+Y] = \sum_{i=1}^m \sum_{j=1}^n p_{ij}(x_i + y_j) = E[X] + E[Y]$$

Exemple : on considère la variable aléatoire égale à la somme des points obtenus en lançant deux dés. Sa moyenne est égale à la somme des moyennes des points obtenus pour chacun des dés (soit 3,5 pour chaque dé). La moyenne de la somme est donc 7.

Propriété :
$$\sigma^2[X+Y] = \sigma_x^2 + 2\sigma_{xy} + \sigma_y^2$$

Corrolaire : Si X et Y sont deux variables indépendantes, alors la variance de leur somme est égale à la somme de leurs variances.

Les résultats énoncés sont valables pour un nombre quelconque de variables aléatoires.

Chapitre 4

Inférence statistique



Ce chapitre regroupe les théories de l'échantillonnage, de l'estimation et la technique des tests d'hypothèse. Ces notions appartiennent au domaine de l'**inférence statistique**, qui consiste à fournir un certain type d'information sur un ensemble d'objets ou d'individus (la population) à partir d'un sous-ensemble de celle-ci (l'échantillon), l'idée de base étant que les caractéristiques du sous-ensemble reflètent – avec une certaine marge d'erreur – celles de la population. C'est donc un domaine qui fait le lien entre les probabilités et les statistiques, en proposant un ensemble de méthodes permettant de tirer des conclusions fiables à partir des données d'échantillons statistiques.

4.1. Echantillonnage

Un **échantillon** est le sous-ensemble de la population que l'on observe ; la **taille** de l'échantillon est le nombre d'individus observés (rappelons que le terme *individu* recouvre aussi bien des êtres humains que des animaux, des objets ou des caractéristiques abstraites telles qu'une mesure, la taille, le poids, etc...)

Echantillons représentatifs

La pertinence des conclusions que l'on va tirer d'un échantillon dépend de sa nature, celle-ci étant caractérisée par sa taille et par la manière dont les individus de l'échantillon sont choisis. Pour qu'un échantillon soit représentatif, il est clair que sa taille ne peut être trop petite, sans qu'il existe pour autant des règles strictes fixant la taille d'un échantillon. Ensuite, les individus faisant partie de l'échantillon doivent être choisis aléatoirement de sorte que chaque individu de la population doit avoir la même probabilité de faire partie de l'échantillon.

On distingue en théorie deux types d'échantillons :

1. l'échantillon *avec remise* (ou échantillon simple) : les individus sont choisis dans la population indépendamment les uns des autres, c'est-à-dire qu'à chaque tirage d'un individu, les chances d'être choisis sont identiques pour tous. Il est donc possible qu'un même individu soit choisi plusieurs fois
2. l'échantillon *sans remise* : à chaque tirage, on choisit un individu non encore choisi précédemment ; un même individu ne peut donc faire partie plusieurs fois de l'échantillon.

En pratique, si la taille de la population est grande par rapport à celle de l'échantillon (ce qui est souvent le cas), les deux types d'échantillons peuvent être considérés comme identiques. Si la population est grande, la probabilité de choisir deux fois le même individu est alors proche de 0.

4.2. Distribution échantillonnée de la moyenne

Nous allons étudier à présent les relations qui existent entre les paramètres d'une population et ceux de tous ses échantillons **de même taille n** . Dans ce qui suit, N représentera la taille de la population. Notons que dans le cas d'une grande population, N est le plus souvent inconnu, et on considère d'un point de vue mathématique que sa valeur est infinie.

Supposons que l'on étudie une valeur caractéristique X de cette population, et soient μ et σ la moyenne et l'écart-type de X calculés sur l'ensemble de cette population. Au niveau de l'échantillon, ces paramètres seront notés respectivement \bar{x} et s .

Considérons à présent l'opération qui consiste à associer à un échantillon aléatoire de taille n la moyenne \bar{x} de cet échantillon. Cette opération définit une variable aléatoire appelée **moyenne échantillonnée** que nous noterons \bar{X} . Cette variable aléatoire possède sa propre distribution de probabilité appelée **distribution échantillonnée de la moyenne**.

Prenons à titre d'exemple le poids (en kg) observé dans un groupe de 5 personnes, dont on extrait des échantillons de 3 individus. Les poids observés sont : 57, 63, 72, 75, 81 (ce qui donne une moyenne $\mu = 69,6$ kg et un écart-type $\sigma = 8,6$ kg).

Enumérons tous les échantillons de taille 3 que l'on peut extraire de ce groupe (on considère ici les échantillons « sans remise »), et calculons leur moyenne.

échantillon	moyenne \bar{x}
{57, 63, 72}	64
{57, 63, 75}	65
{57, 63, 81}	67
{57, 72, 75}	68
{57, 72, 81}	70
{57, 75, 81}	71
{63, 72, 75}	70
{63, 72, 81}	72
{63, 75, 81}	73
{72, 75, 81}	76

On observe que les valeurs de la moyenne échantillonnée sont moins dispersées que les valeurs observées. Nous verrons ci-dessous qu'au plus n est grand, au plus la dispersion est faible. Notez que dans le cas limite où n est égal à la taille de la population, le seul échantillon sans remise possible est la population elle-même, et dans ce cas la dispersion est nulle.

Dans ce qui suit, $\mu_{\bar{x}}$ et $\sigma_{\bar{x}}$ désignent respectivement la moyenne et l'écart-type de la moyenne échantillonnée \bar{X} .

Propriété. Pour un échantillon de taille n établi avec remise, la moyenne et l'écart-type de la moyenne échantillonnée \bar{X} ont pour valeur :

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Nous admettrons sans démonstrations les formules suivantes :

Propriété. Pour un échantillon de taille n établi sans remise, la moyenne et l'écart-type de la moyenne échantillonnée \bar{X} ont pour valeur :

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Remarque : lorsque N est grand par rapport à la taille de l'échantillon, le second facteur dans l'expression de $\sigma_{\bar{x}}$ est proche de 1. Si N est inconnu, on considère sa valeur infinie, et dans ce cas la valeur limite de ce facteur est aussi égale à 1. On voit donc que pour les « grandes » population, les formules trouvées dans le cas « avec » et « sans remise » peuvent être considérées comme identiques, et que le second facteur peut être négligé. En pratique, on utilisera la formule complète uniquement dans le cas d'échantillons établis sans remise pour une population de taille connue et telle que $N < 20n$.

Les formules confirment qu'en moyenne, la moyenne de l'échantillon \bar{x} coïncide avec la moyenne de la population. Etant donné le hasard des tirages, elle sera plus ou moins proche de μ et en fournira donc une estimation. Quant à $\sigma_{\bar{x}}$, on voit qu'il est inversement proportionnel à la racine carrée de n , et donc au plus n est grand, au plus sa valeur est petite, en d'autres termes cela veut dire que meilleure sera l'estimation de μ donnée par \bar{x} .

En vertu du théorème Central-Limite, la distribution échantillonnée de la moyenne tend vers une distribution normale, et ce quelle que soit la distribution initiale de X pour la population. En pratique, on considère que c'est le cas dès que $n \geq 30$. En outre, si la population est de distribution normale, la distribution de \bar{X} est alors d'office normale, et ce quelle que soit alors la taille de l'échantillon.

4.3. Théorie de l'estimation

Inversément à la démarche précédente, nous allons voir à présent comment **estimer** la moyenne μ (en général inconnue) de la population dont est issu un échantillon de taille n , de moyenne \bar{x} et d'écart-type s .

Estimation ponctuelle

Elle consiste simplement à prendre comme estimation de μ la valeur de la moyenne \bar{x} d'un échantillon. Cette estimation n'étant basée que sur un seul échantillon, on parle d'**estimation ponctuelle**. Il est évident qu'il y a dans cette démarche un facteur approximatif non négligeable. Imaginons que l'on veut estimer la taille moyenne d'un groupe de 20 personnes en prenant un échantillon de 3 personnes : le hasard pourrait faire que les trois plus petits (ou les trois plus grands) aient été choisis, ce qui conduirait à des conclusions faussées.

C'est pourquoi l'usage privilégie plutôt la seconde démarche, qui consiste à donner une estimation par intervalle.

4.4. Intervalle de confiance

Plutôt qu'une valeur ponctuelle, on va donner un intervalle centré sur la moyenne \bar{x} de l'échantillon, et dans lequel la moyenne réelle μ **a une certaine probabilité de se trouver**. Un tel intervalle s'appelle **intervalle de confiance**, et la probabilité qui lui est associée est le **niveau de confiance**. Les niveaux habituels sont 90%, 95% ou plus rarement 99%. La probabilité complémentaire est le **seuil d'erreur** (ou encore **risque d'erreur**).

Par exemple, si on donne l'intervalle [1,68 ; 1,72] au niveau de confiance 0,95 pour estimer la taille moyenne (en mètre) d'une population, cela veut dire que la moyenne réelle de la taille de la population a 95% de chance de se trouver dans cet intervalle. On commet donc un risque d'erreur de 5% qui est la probabilité que la moyenne réelle soit en dehors de l'intervalle.

Notez qu'il n'y a pas moyen de contourner la notion de risque d'erreur qui est inhérente à ce type de résultat. La taille de l'intervalle étant d'autant plus grande que le risque d'erreur est petit, le statisticien est amené à choisir entre une précision faible ou un risque d'erreur important. S'il veut un intervalle encadrant la moyenne réelle avec trop de précision, il y aura un risque important que celle-

ci ne s'y trouve pas ! Et inversement, s'il opte pour plus d'assurance pour un intervalle dans lequel la moyenne réelle a une grande chance de se trouver, l'intervalle risque d'être trop grand que pour donner une information significative. Dans notre exemple de la taille, ceci se traduirait par exemple par donner l'intervalle $[1, 2]$ qui possède sans risque d'erreur une probabilité de 100% de contenir la taille moyenne de la population !

Pour calculer les bornes d'un intervalle de confiance, nous allons utiliser la moyenne échantillonnée introduite au chapitre précédent. Nous envisagerons plusieurs cas, selon que l'écart-type σ de la population est connu ou non. De le cas où il est inconnu, nous serons amenés à distinguer les petits ($n < 30$) et les grands échantillons ($n \geq 30$).

4.4. 1. Intervalle de confiance avec σ connu

Le but étant d'estimer la moyenne μ d'une population, il peut paraître paradoxal que soit connu l'écart-type σ , puisque le calcul de ce dernier fait précisément intervenir μ ! Néanmoins, certaines situations peuvent faire supposer la stabilité de σ (penser à un changement d'origine des observations sans changement d'unité). On peut aussi imaginer que la valeur proposée pour σ résulte elle-même d'une estimation ou est issue de statistiques antérieures...

Nous avons vu précédemment que la moyenne échantillonnée \bar{X} admet une distribution normale (ou approximativement normale lorsque n est petit) de moyenne $\mu_{\bar{X}} (= \mu)$ et d'écart-type $\sigma_{\bar{X}} = \sigma / \sqrt{n}$.

Ceci nous permet de calculer facilement la probabilité que la moyenne \bar{X} d'un échantillon se trouve dans un intervalle centré sur la moyenne de la population μ :

$$P[\mu - z \sigma_{\bar{X}} \leq \bar{X} \leq \mu + z \sigma_{\bar{X}}] = 2F(z) - 1$$

où z est un réel positif quelconque et $F(z)$ la fonction de répartition de la normale réduite $N(0, 1)$. Une simple réécriture des inégalités montre que cette probabilité est aussi égale à :

$$P[\bar{X} - z \sigma_{\bar{X}} \leq \mu \leq \bar{X} + z \sigma_{\bar{X}}]$$

Les rôles de \bar{X} et de μ sont donc interchangeables. On voit donc que la probabilité que μ appartienne à l'intervalle $[\bar{X} - z \sigma_{\bar{X}}, \bar{X} + z \sigma_{\bar{X}}]$ est égale à $2F(z) - 1$, ce qui nous donne l'expression de l'intervalle de confiance recherché, et du niveau de confiance correspondant.

Quelques valeurs caractéristiques

Niveau de confiance $2F(z) - 1$	z	Intervalle de confiance $[\bar{X} - z \sigma_{\bar{X}}, \bar{X} + z \sigma_{\bar{X}}]$
68,28%	1	$[\bar{X} - \sigma_{\bar{X}}, \bar{X} + \sigma_{\bar{X}}]$
90%	1,65	$[\bar{X} - 1.65 \sigma_{\bar{X}}, \bar{X} + 1.65 \sigma_{\bar{X}}]$
95%	1,96	$[\bar{X} - 1.96 \sigma_{\bar{X}}, \bar{X} + 1.96 \sigma_{\bar{X}}]$
95,44%	2	$[\bar{X} - 2 \sigma_{\bar{X}}, \bar{X} + 2 \sigma_{\bar{X}}]$
99%	2,58	$[\bar{X} - 2.58 \sigma_{\bar{X}}, \bar{X} + 2.58 \sigma_{\bar{X}}]$

		$2.58 \sigma_{\bar{x}}]$
99,74%	3	$[\bar{x} - 3 \sigma_{\bar{x}} , \bar{x} + 3 \sigma_{\bar{x}}]$

4.4. 2. Intervalle de confiance avec σ inconnu – grands échantillons

On considère qu'un échantillon est « grand » dès que n est supérieur ou égal à 30. Si σ est inconnu, il faut travailler avec une estimation de celui-ci. Nous admettrons sans démonstration que l'estimateur de σ est s' donné par la formule suivante :

$$s' = \sqrt{\frac{n}{n-1}} s$$

où s est l'écart-type de l'échantillon (s' s'appelle **estimateur non biaisé** de σ). La distribution de la moyenne échantillonnée \bar{X} admet encore une distribution normale de moyenne $\mu_{\bar{x}} (= \mu)$ et d'écart-type $\sigma_{\bar{x}}$, à la différence que celui-ci sera calculé avec la valeur de s' :

$$\sigma_{\bar{x}} = \frac{s'}{\sqrt{n}}$$

Les intervalles de confiance sont donnés par les mêmes formules que dans le tableau ci-dessus.

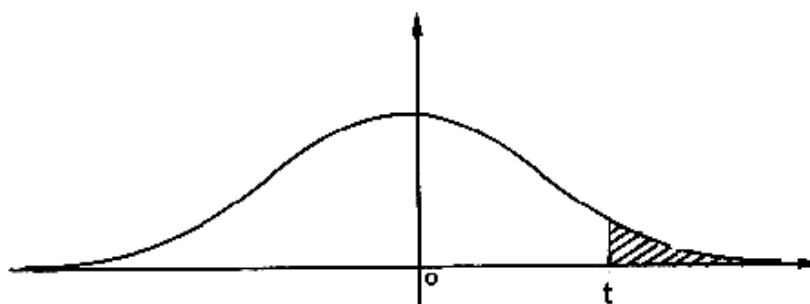
4.4. 3. Intervalle de confiance avec σ inconnu – petits échantillons

Lorsque σ est inconnu et $n < 30$, on montre que la moyenne échantillonnée \bar{X} est distribuée suivant une **loi de Student à $n - 1$ degrés de libertés**. La loi de Student est une distribution semblable à la distribution normale réduite, également symétrique et de moyenne 0, mais plus étendue et « aplatie ». Asymptotiquement, elle tend vers la normale réduite, et lorsque $n = 30$, on peut considérer que les distributions coïncident.

La moyenne et l'écart-type de la distribution échantillonnée sont donnés par les mêmes formules que dans le cas des grands échantillons. Le changement réside dans la formule de l'intervalle de confiance : il faut remplacer le z de la normale réduite par le t de la distribution de Student à $n - 1$ degrés de liberté :

$$[\bar{x} - t \sigma_{\bar{x}} , \bar{x} + t \sigma_{\bar{x}}]$$

Les valeurs de t sont données par le tableau suivant ; pour chaque degré de liberté, la table donne la valeur réduite t pour laquelle la fréquence des valeurs supérieures à t est reprise en première ligne.

**Table de la distribution de Student**

<i>d.l.</i>	0.10	0.05	0.025	0.01	0.005	<i>d.l.</i>	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.65	18	1.330	1.734	2.101	2.552	2.878
2	1.886	2.920	4.303	6.965	9.925	19	1.328	1.729	2.093	2.539	2.861
3	1.638	2.353	3.182	4.541	5.841	20	1.325	1.725	2.086	2.528	2.845
4	1.533	2.132	2.776	3.747	4.604	21	1.323	1.721	2.080	2.518	2.831
5	1.476	2.015	2.571	3.365	4.032	22	1.321	1.717	2.074	2.508	2.819
6	1.440	1.943	2.447	3.143	3.707	23	1.319	1.714	2.068	2.500	2.807
7	1.415	1.895	2.365	2.998	3.499	24	1.318	1.711	2.064	2.492	2.797
8	1.397	1.860	2.306	2.896	3.355	25	1.316	1.708	2.060	2.485	2.787
9	1.383	1.833	2.262	2.821	3.250	26	1.315	1.706	2.056	2.479	2.779
10	1.372	1.812	2.228	2.764	3.169	27	1.314	1.793	2.052	2.473	2.771
11	1.363	1.796	2.201	2.718	3.106	28	1.313	1.701	2.048	2.467	2.763
12	1.356	1.782	2.179	2.681	3.055	29	1.311	1.699	2.045	2.462	2.756
13	1.350	1.771	2.160	2.650	3.012	30	1.310	1.697	2.042	2.457	2.750
14	1.345	1.761	2.145	2.624	2.977	40	1.303	1.684	2.021	2.423	2.704
15	1.341	1.753	2.131	2.602	2.947	60	1.296	1.671	2.000	2.390	2.660
16	1.337	1.746	2.210	2.583	2.921	120	1.289	1.658	1.980	2.358	2.617
17	1.333	1.740	2.110	2.567	2.898	—	1.282	1.645	1.960	2.326	2.576

La valeur de t est à prendre en fonction du niveau de confiance voulu et du nombre de degré de liberté (égal à la taille de l'échantillon moins 1). La table ci-dessus ne donne que les fréquences les plus utiles.

Par exemple, si on veut un intervalle de confiance à 95% pour un échantillon de taille 16, il faudra prendre la valeur de t dans la ligne 15 pour la fréquence égale à 2.5%, c'est-à-dire 0.025, ce qui donnera $t = 2.131$.

Observez dans ce tableau que lorsque le degré de liberté augmente, les valeurs dans une colonne fixée convergent vers la valeur correspondante de z dans la table de la normale réduite. Les valeurs de t sont aussi toutes supérieures à celles de z , ce qui donne des intervalles de confiance plus grands, ce qui est logique puisque nous sommes dans la situation où on peut s'attendre au plus grand « flou » concernant l'estimation de μ .

4.5. Test d'hypothèse et prise de décision

4.5.1. Introduction

Dans ce qui précède nous avons vu comment estimer la moyenne d'une population sous forme d'un intervalle de confiance construit à partir des données d'un échantillon. Nous allons voir à présent comment cet intervalle de confiance va nous aider à trancher sur la validité d'une hypothèse portant sur la valeur supposée de la moyenne de la population, cette démarche étant appelée *prise de décision*.

Elle apparaît naturellement dans des situations où on s'interroge sur la valeur réelle des paramètres d'une population, bien que cette valeur réelle reste inconnue ou « inaccessible ». Les tests d'hypothèses sont ainsi fréquents dans les usines des contrôles de fabrication sont quasi permanents. Prenons quelques exemples :

a) Un responsable d'un atelier s'interroge si une des machines est réglée correctement. Elle est sensée découper des barres de métal d'un mètre de longueur en moyenne – avec un certain écart-type, dû à l'imprécision de la mesure et à son dérèglement possible. Dans un échantillon, on retrouve des barres légèrement plus petites, mais aussi quelques unes plus grandes. Peut-on dès lors sur base de cet échantillon affirmer que la moyenne des longueurs des barres est toujours d'un mètre ? Le résultat du test confirmera ou rejettera cette hypothèse, et amènera le responsable de la machine à envisager une révision si la réponse est négative.

b) Une marque de vin réputée est vendue dans des bouteilles de 75 cl. Un consommateur alerte vérifie le contenu de 12 bouteilles achetées dans une caisse, et constate qu'il est souvent inférieur de quelques ml. Il se demande dès lors si la moyenne réelle n'est pas inférieure au contenu annoncé sur l'étiquette, mettant en doute l'honnêteté du fabricant. Son hypothèse peut être vraie, mais aussi fausse car il se peut que par malchance, l'échantillon à sa disposition donne une moyenne inférieure à 75 cl, mais que dans d'autres caisses, le contenu pourrait être supérieur. Il y a donc ici un risque d'accuser injustement le fabricant.

4.5.2. Hypothèses et risques

Le but est de répondre à des questions du type : « telle valeur est-elle possible comme moyenne de la population ? », « la valeur réelle de la moyenne de la population n'est-elle pas plus grande (ou plus petite) que la valeur annoncée ? »... L'hypothèse formulée est notée H_0 (appelée **hypothèse nulle**) et la négation de cette hypothèse (l'**alternative**) est notée H_1 .

L'analyse d'un échantillon par un test adéquat va conduire à accepter ou rejeter H_0 . Mais indépendamment du résultat de ce test, l'hypothèse H_0 peut-être vraie ou fausse. Ceci conduit à 4 cas de figure :

- L'hypothèse H_0 est vraie et elle est acceptée par le test : dans ce cas la conclusion est exacte. Dans l'exemple a) ci-dessus, cela veut dire que la machine est bien réglée et que le test confirme ce fait.
- H_0 est vraie mais est rejetée par le test : c'est le **risque de 1ère espèce**. Dans l'exemple, la machine est bien réglée mais les données de l'échantillon laissent penser que ce n'est pas le cas. La machine sera réglée « inutilement ».
- H_0 est fausse (c'est H_1 qui est vraie) mais le test accepte H_0 : c'est le **risque de 2ème espèce** : la machine nécessite bien un réglage mais son dysfonctionnement n'a pas été confirmé sur base de l'échantillon choisi.
- H_0 est fausse et le test rejette également H_0 : dans ce cas, la conclusion est exacte : la machine est mal réglée et le test confirme également ce fait.

4.6. Test de comparaison de la moyenne

4.6.1. Test bilatéral

Il s'agit de tester si une valeur μ_0 est acceptable comme moyenne de la population. Les hypothèses nulle et alternative se formulent donc ainsi :

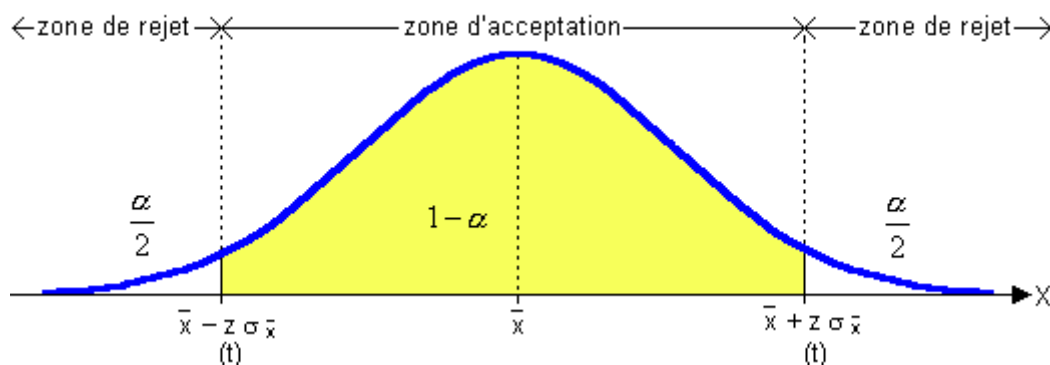
$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Pour décider de la validité de l'hypothèse, nous allons utiliser l'intervalle de confiance centré sur la moyenne \bar{x} de l'échantillon que nous avons introduit au chapitre précédent. Pour rappel, cet intervalle est associé à un risque d'erreur α qui correspond au risque de 1^{ère} espèce du test : la moyenne véritable a une probabilité $1 - \alpha$ d'appartenir à l'intervalle, et une probabilité α de ne pas s'y trouver.

Dès lors, la prise de décision sera la suivante : si μ_0 appartient à l'intervalle de confiance de niveau de confiance $1 - \alpha$, H_0 est acceptée, sinon elle est rejetée. La probabilité de faire une décision erronée est égale au risque d'erreur.

Le graphique ci-dessus montre deux zones de rejet, disposées symétriquement. Par exemple, en effectuant un test au niveau de confiance de 95%, on tolère un risque d'erreur de 5% qui est réparti symétriquement de part et d'autre de l'intervalle : il y a donc une probabilité 0.025 que la moyenne véritable se trouve à gauche de la zone d'acceptation, et de même à droite.



4.6.2. Test unilatéral

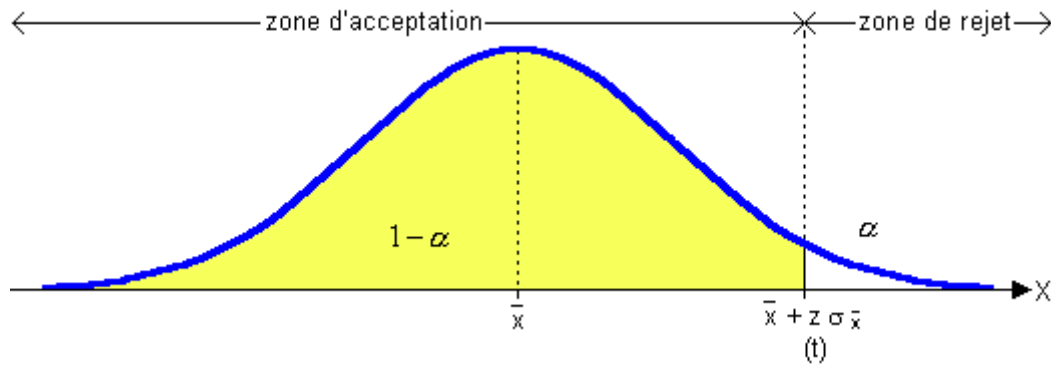
On se demande à présent si une valeur μ_0 présentée comme moyenne d'une population est surestimée ou sous-estimée. En d'autres termes, on répond à la question : μ est-il inférieur (ou supérieur) à la valeur μ_0 ? La différence avec le cas précédent est qu'il n'y aura qu'une seule zone de rejet. L'intervalle de confiance considéré sera décentré et s'étendra jusqu'à l'infini sur un de ses côtés.

1^{er} cas :

$$H_0 : \mu < \mu_0$$

$$H_1 : \mu \geq \mu_0$$

On se demande si la moyenne μ est inférieure à la valeur μ_0 . Logiquement, si nous avons été amenés à nous poser cette question, c'est que la moyenne \bar{x} de l'échantillon est également inférieure à μ_0 . On construit alors à l'aide de la distribution échantillonnée de la moyenne un intervalle $]-\infty, x]$ tel que la moyenne véritable a une probabilité $1 - \alpha$ de s'y trouver (et donc une probabilité α de ne pas s'y trouver). Autrement dit, il y a une probabilité $1 - \alpha$ que la vraie moyenne est inférieure à la borne de droite x , et seulement une probabilité α qu'elle soit supérieure. H_0 sera donc acceptée si μ_0 appartient à cet intervalle, et rejetée sinon.

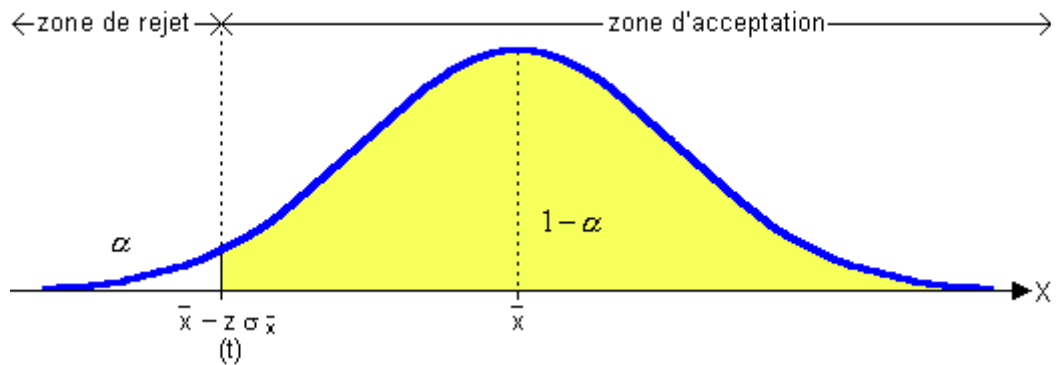


2^{ème} cas :

$$H_0 : \mu > \mu_0$$

$$H_1 : \mu \leq \mu_0$$

C'est le cas symétrique du précédent : on se demande si la moyenne μ est supérieure à la valeur μ_0 . L'intervalle de confiance est maintenant repoussé vers la droite.



N.B. : dans les deux cas unilatéraux, on déduira les bornes des intervalles à l'aide de la distribution normale ou de Student, conformément aux différents cas étudiés au chapitre précédents (dépendant de la connaissance ou non de σ , et de la taille de l'échantillon).



Chapitre 5

Analyse en composantes principales

Bibliographie

Droesbecque J-J, *Eléments de statistique*, Ed. Ellipse

Statistiques, Première approche, Dossier de la commission pédagogique de la SBPMef

Wonnacott T.H & R.J., *Statistique: Economie – Gestion – Sciences – Médecine (avec exercices d'application)*, Ed. Economica

Justens Daniel, *Statistique pour décideurs*, De Boeck Université

Donald H. Sanders & François Allard, *Les statistiques - une approche nouvelle*, Ed. McGraw-Hill

Théorie et applications de la statistique, Collection Schaum
