Université Pierre et Marie Curie
DEUST Informatique
Année 2002–2003

Module de mathématiques
C. Gonzales 2002

Introduction à la statistique et aux probabilités

Table des matières

Tables de probabilité

1 Introduction à la statistique descriptive

Le but de la statistique descriptive en général, et de ce cours en particulier, est de proposer une méthodologie permettant la synthèse d'informations contenues dans des amas de données, ainsi que leur exploitation.

Exemple 1 : Afin d'étudier la répartition des salaires mensuels de jeunes cadres dans une entreprise, une enquête a été réalisée, qui a permis de consigner les salaires des cent cadres que compte l'entreprise (60 hommes et 40 femmes). Les tableaux 1 et 2 ci-dessous représentent les données «brutes» récoltées, à savoir la liste des salaires de tous les cadres de l'entreprise.

18	10	12	12	10	14	14	14	12	12	16	18	10	16	16
16	16	12	12	14	14	14	14	12	16	14	14	16	16	16
12	10	18	12	18	18	14	14	14	12	16	14	16	16	14
10	16	16	14	14	14	12	16	16	14	14	12	14	10	10

TAB. 1: Salaires des 60 cadres masculins (exprimés en kF)

16	16	14	14	14	14	10	12	12	08	20	14	14	12	16
14	20	14	12	12	16	10	10	18	10	16	12	12	14	10
14	18	12	12	16	08	14	10	12	14					

TAB. 2: Salaires des 40 cadres féminins (exprimés en kF)

Grâce à ces deux tableaux, il devrait être possible d'estimer, par exemple, quel est le salaire moyen des cadres masculins, celui du personnel féminin, quel est le salaire le plus bas, le plus élevé. Est-ce que l'entreprise paye mieux son personnel masculin que son personnel féminin ? La réponse à ces questions pourrait permettre, entre autres, de situer l'entreprise par rapport à ses concurrents...

Malheureusement celle-ci ne découle pas directement des deux tableaux : il faut d'abord synthétiser les informations qu'ils contiennent avant de pouvoir en extraire une information «aisément utilisable». Par exemple, on ne peut pas dire sans effectuer de calcul si l'entreprise paye équitablement son personnel sans distinction de sexe. Nous verrons dans ce cours des outils permettant de répondre à ce type de questions. ♦

1.1 Un peu de vocabulaire

Définition 1 : *On appelle population (statistique) l'ensemble des objets ou des personnes sur lesquels porte une étude. Chaque élément de la population est appelé un individu. Les critères selon lesquels on étudie une population sont appelés des caractères. Les valeurs que peuvent prendre les caractères sont appelés des modalités. On dit qu'un caractère est quantitatif si son ensemble de modalités est un ensemble de nombres, et si ceux-ci correspondent à une échelle mathématique. Dans le cas contraire, le caractère est dit qualitatif. Par convention, un caractère quantitatif est aussi appelé variable statistique.*

Exemple 2 : Afin de déterminer les taxes d'habitation, une commune procède au recensement de toutes ses habitations. Celles-ci sont évaluées grâce au formulaire suivant :

Type de logement	appartement	☐
	maison individuelle	☐
	autre	☐
Surface habitable		☐
Nombre de pièces d'habitation		☐
Y a-t-il une cuisine ?	oui, privée	☐
	oui, commune	☐
	non	☐
Salle de bains ou douche ?	oui, privée	☐
	oui, commune	☐
	non	☐

L'ensemble de toutes les habitations forme alors la population, et un logement particulier est un individu de la population statistique. Les caractères correspondent aux questions posées. Si une habitation contient 3 pièces, alors 3 est une modalité du caractère «*Nombre de pièces d'habitation*». Le «*type de logement*», l'existence d'une «*cuisine*», d'une «*salle de bain ou douche*» sont des caractères qualitatifs tandis que la «*surface habitable*» et le «*nombre de pièces d'habitation*» sont des caractères quantitatifs. ♦

Notons qu'on peut associer des modalités numériques à un caractère qualitatif. Par exemple, on pourrait attribuer au caractère «*type de logement*» les modalités 1, 2, 3, en prenant la convention : 1 signifie «appartement», 2 signifie «maison individuelle», et 3 équivaut à «autres». Mais cela ne transformerait en aucun cas le caractère en une variable statistique. En effet, les chiffres 1, 2, 3, ne reflètent aucune échelle numérique : on aurait pu prendre n'importe quel nombre, et ce dans n'importe quel ordre. Par exemple, les modalités auraient pu être 3, 10, 50 avec la convention : 50 = «appartement», 3 = «maison individuelle», et 10 = «autres». Donc dire qu'une modalité est supérieure à une autre ou vaut deux fois plus qu'une autre n'a aucune signification : le caractère est alors qualitatif.

Exemple 3 : Dans une compagnie d'aviation, on cherche à connaître par une étude s'il existe une relation entre le nombre de langues parlées couramment et le poste occupé au sein de la compagnie. La population est donc l'ensemble du personnel de la compagnie. Il y a deux caractères : le nombre de langues parlées, et le poste occupé (mécanicien, pilote, hôtesse...). Le premier est quantitatif (c'est donc une variable statistique) ; le second est qualitatif. ♦

Définition 2 : *Une variable statistique est discrète si les valeurs numériques qu'elle peut prendre sont isolées (un ensemble d'entiers par exemple). Une variable statistique est continue si elle peut prendre a priori toutes les valeurs numériques d'un intervalle.*

Exemple 2 (suite) : Le caractère «*surface habitable*» est une variable statistique continue car l'ensemble de ses modalités est $\mathbb{R}^+$. Par contre la variable statistique «*Nombre de pièces d'habitation*» est discrète car ses modalités sont des nombres entiers. ♦

Dans la suite du cours, nous nous intéresserons surtout aux variables statistiques.

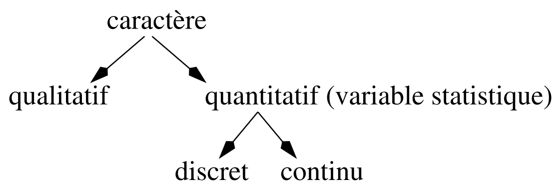


FIG. 1 – Classification des caractères.

1.2 Distribution et représentation d'un caractère qualitatif et d'une variable statistique discrète

Dans la suite du cours, nous noterons les caractères en majuscules : $A, B, C, X, Y \ldots$ Pour chaque caractère X, on notera ses modalités de la manière suivante : $x_1, x_2, x_3 \ldots$

Définition 3 : *Soit X un caractère défini sur une population de N individus et dont l'ensemble des modalités est $\{x_1, \ldots, x_I\}$. L'effectif (on dit encore l'effectif absolu) de la modalité x_i est le nombre d'individus de la population pour lesquels le caractère X prend la valeur x_i. On le note habituellement N_i.*
La fréquence (on dit encore l'effectif relatif) de la modalité x_i, que l'on note f_i, est définie par $f_i = N_i/N$.
La distribution du caractère X est l'ensemble des couples $\{(x_1, f_1), (x_2, f_2), \ldots\}$.

Exemple 1 (suite) : L'entreprise possède 100 cadres, et l'on étudie la répartition des salaires, notamment en fonction du sexe du personnel. Par conséquent, la population étudiée est l'ensemble de tous les cadres de l'entreprise. Un individu de la population correspond à un cadre. Il y a 100 cadres, donc $N = 100$. Considérons maintenant le caractère $X =$ «sexe». Ses modalités sont $\{x_1 = \text{masculin}, x_2 = \text{féminin}\}$. On sait qu'il y a 60 hommes et 40 femmes dans l'entreprise. Donc $N_1 = 60$ et $N_2 = 40$, et $f_1 = N_1/N = 60/100$ et $f_2 = 40/100$. Autrement dit, $f_1 = 60\%$ de la population est masculine, et $f_2 = 40\%$ est féminine. Ces données peuvent être synthétisées dans le tableau suivant :

modalités	masculin	féminin
effectif	60	40
fréquence	$\frac{60}{100}$	$\frac{40}{100}$

TAB. 3: Synthèse du caractère X.

♦

Exemple 3 (suite) : L'enquête menée au sein de la compagnie d'aviation a permis d'établir la liste de données «brutes» suivante :

répondant n°	nb de langues	poste	répondant n°	nb de langues	poste
1	1	mécanicien	13	2	pilote
2	2	hôtesse	14	2	hôtesse
3	1	mécanicien	15	3	hôtesse
4	1	cadre	16	1	cadre
5	2	cadre	17	1	mécanicien
6	3	pilote	18	3	pilote
7	2	cadre	19	2	cadre
8	2	mécanicien	20	2	hôtesse
9	1	pilote	21	1	mécanicien
10	2	pilote	22	2	hôtesse
11	1	mécanicien	23	2	pilote
12	4	hôtesse	24	2	hôtesse

TAB. 4: Nombre de langues parlées par les employés.

La population, c'est-à-dire le nombre d'employés de la compagnie d'aviation, est égal à $N = 24$. Le caractère $X =$ «poste» (qui est un caractère qualitatif) a pour ensemble de modalités $\{x_1 = \text{cadre}, x_2 = \text{hôtesse}, x_3 = \text{mécanicien}, x_4 = \text{pilote}\}$. Dans le tableau ci-dessus, on peut compter 7 hôtesses, donc $N_2 = 7$ et $f_2 = N_2/N = 7/24$. De la même manière, le caractère $Y =$ «nombre de langues»(qui est une variable statistique discrète) a pour ensemble de modalités $\{y_1 = 1, y_2 = 2, y_3 = 3, y_4 = 4\}$. L'effectif et la fréquence des modalités des caractères X et Y se résument au tableau ci-après :

modalité de X	effectif	fréquence	modalité de Y	effectif	fréquence
cadre	5	5/24	1	8	8/24
hôtesse	7	7/24	2	12	12/24
mécanicien	6	6/24	3	3	3/24
pilote	6	6/24	4	1	1/24

TAB. 5: Effectif et fréquence.

♦

Passer des données brutes (tableau 4) au tableau 5 ci-dessus est avantageux car les informations y sont plus facilement exploitables. Il est par exemple plus facile de calculer le nombre moyen de langues parlées par le personnel, ou bien la catégorie de personnel la plus représentée à partir du tableau 5 qu'à partir du tableau 4. Mais on peut encore améliorer la lisibilité des informations en utilisant des représentations graphiques.

Définition 4 : *Soit un caractère X dont la distribution est $\{(x_1, f_1), (x_2, f_2), (x_3, f_3), \ldots\}$. Le diagramme en bâtons est un graphe dans lequel on associe à chacune des modalités x_i (représentées sur l'axe horizontal) un bâton de hauteur f_i. Si l'on élargit les bâtons, on obtient une représentation en colonnes.*

Exemples 3 et 1 (suite) : La partie droite du tableau 5 (qui représente une variable statistique discrète) peut être illustrée par les graphes suivants :

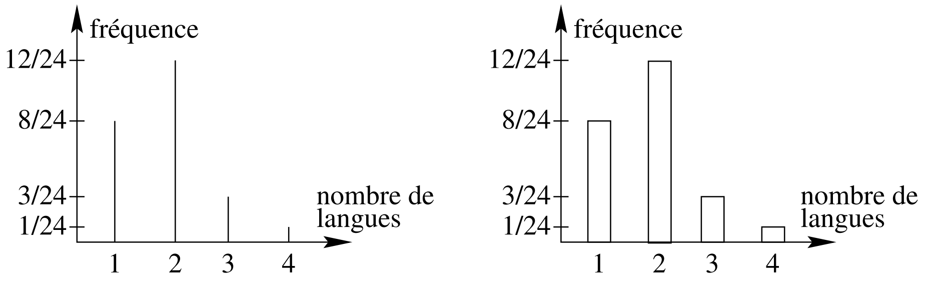


Diagramme en bâtons ***Représentation en colonnes***

FIG. 2: représentation en bâtons et en colonnes.

De manière tout à fait similaire, on peut représenter un caractère qualitatif par un diagramme en bâtons ou par une représentation en colonnes. Par exemple, les caractères «poste» et «sexe» des tableaux 5 et 3 peuvent être illustrés respectivement par :

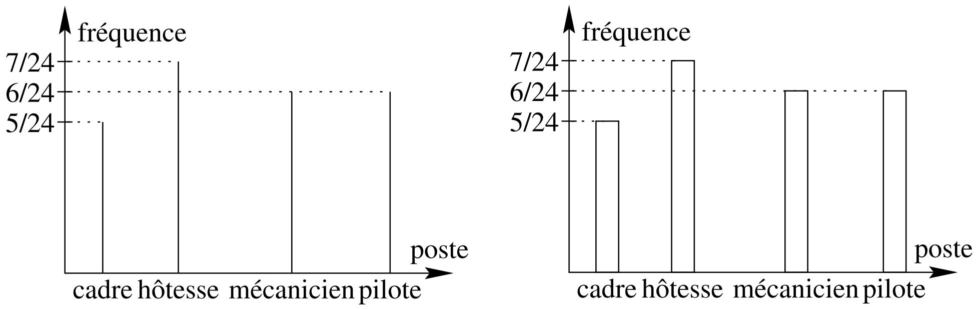


Diagramme en bâtons ***Représentation en colonnes***

FIG. 3: Représentation en bâtons et en colonnes du caractère «poste».

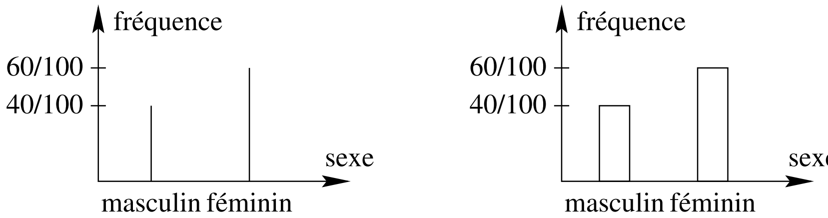


Diagramme en bâtons ***Représentation en colonnes***

FIG. 4: Représentation en bâtons et en colonnes du caractère «sexe».

♦

1.3 Distribution et représentation d'une variable statistique continue

D'une manière générale, on va essayer d'appliquer aux variables statistiques continues les mêmes traitements que pour les variables qualitatives ou bien quantitatives discrètes. Ainsi :

Exemple 1 (suite) : À partir des tableaux 1 et 2 de la page 4 nous pourrions utiliser la définition 3 pour déterminer l'effectif et la fréquence de la variables statistique continue «salaire» :

– Si la population étudiée est limitée à l'ensemble des cadres masculins, alors $N = 60$ car l'entreprise ne compte que 60 hommes. De plus, l'ensemble des modalités de la variable $X =$ «salaire mensuel» est $\{x_1 = 10, x_2 = 12, x_3 = 14, x_4 = 16, x_5 = 18\}$ car ce sont les seuls chiffres que l'on retrouve dans le tableau 1. On peut dénombrer 20 chiffres «14» dans le tableau 1, d'où $N_3 = 20$ et $f_3 = N_3/N = 20/60 = 1/3$. Un tiers du personnel masculin a donc un salaire de 14000F. On peut donc synthétiser le tableau 1 dans le tableau suivant (en particulier, les lignes 1 et 3 donnent la distribution des modalités) :

modalités	10	12	14	16	18
effectifs	7	12	20	16	5
fréquences	$\frac{7}{60}$	$\frac{12}{60}$	$\frac{20}{60}$	$\frac{16}{60}$	$\frac{5}{60}$

TAB. 6: Synthèse des salaires du personnel masculin.

– Si la population étudiée est limitée à l'ensemble des cadres féminins, alors $N = 40$, l'ensemble des modalités de la variable $X =$ «salaire mensuel» est $\{x_1 = 8, x_2 = 10, x_3 = 12, x_4 = 14, x_5 = 16, x_6 = 18, x_7 = 20\}$. Suivant le même principe, on peut résumer le tableau 2 dans le tableau suivant :

modalités	8	10	12	14	16	18	20
effectifs	2	6	10	12	6	2	2
fréquences	$\frac{2}{40}$	$\frac{6}{40}$	$\frac{10}{40}$	$\frac{12}{40}$	$\frac{6}{40}$	$\frac{2}{40}$	$\frac{2}{40}$

TAB. 7: Synthèse des salaires du personnel féminin.

– Si la population étudiée est l'ensemble de tous les cadres de l'entreprise, $N = 100$, l'ensemble des modalités de la variable $X =$ «salaire mensuel» est $\{x_1 = 8, x_2 = 10, x_3 = 12, x_4 = 14, x_5 = 16, x_6 = 18, x_7 = 20\}$ et les tableaux 1 et 2 sont synthétisés dans :

modalités	8	10	12	14	16	18	20
effectifs	2	17	22	32	22	7	2
fréquences	$\frac{2}{100}$	$\frac{17}{100}$	$\frac{22}{100}$	$\frac{32}{100}$	$\frac{22}{100}$	$\frac{7}{100}$	$\frac{2}{100}$

TAB. 8: Synthèse des salaires du personnel.

♦

Cependant la nature continue du caractère peut poser quelques problèmes comme l'illustre l'exemple suivant :

Exemple 4 : On étudie les revenus d'une promotion de 40 ingénieurs trois ans après leur sortie de leur école d'ingénieur. Après avoir demandé à chacun son salaire, on dresse la liste suivante :

19600	12500	17700	18800	19100	14700	21900	22500	21800	20100
16200	20500	12200	25400	20900	21200	15500	21300	17900	25000
21700	14400	18300	16700	23000	17000	24300	27000	27700	22200
18000	17500	23700	21600	13800	19200	20200	15100	19600	17200

TAB. 9: Salaires mensuels des 40 ingénieurs (exprimés en F).

Le problème auquel nous sommes alors confrontés est le suivant : l'effectif de chaque modalité est de 1, autrement dit, on ne rencontre pas deux fois exactement la même valeur dans le tableau ci-dessus. Les effectifs et les fréquences des modalités ne permettent donc pas de synthétiser les données brutes, pas plus que ne le permettent le diagramme en bâtons ou la représentation en colonnes. À titre indicatif, dans la figure ci-dessous se trouve le diagramme en bâtons que l'on aurait obtenu.

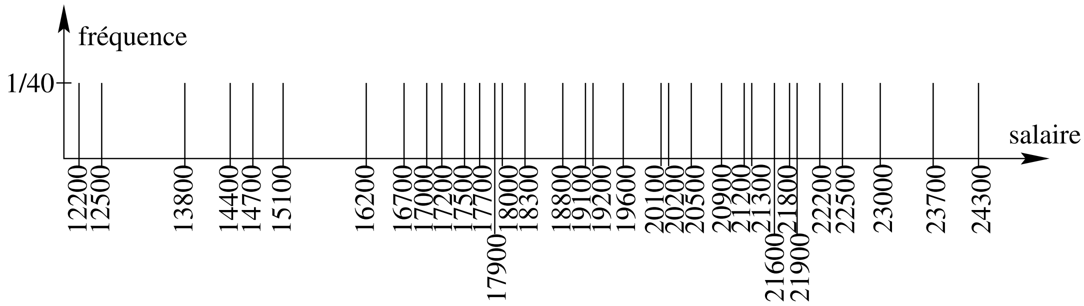


FIG. 5: Diagramme en bâtons.

Ce diagramme n'est manifestement pas plus lisible que le tableau 9. ♦

Pourquoi est-ce que l'effectif et la fréquence résument bien les données brutes lorsque les caractères sont discrets ou qualitatifs, et pas quand ils sont continus ? En fait, dans les exemples de la section 1.2, la taille de la population est bien supérieure au nombre de modalités des caractères (dans l'exemple 1, «sexe» a 2 modalités et «salaire» en a 7 alors que la population compte 100 individus ; dans l'exemple 3, «poste» et «nombre de langues» ont 4 modalités et la population a 24 individus, etc). Par définition, une variable statistique continue peut prendre une infinité de valeurs dans tout intervalle ; on ne peut donc pas raisonnablement espérer que le nombre de modalités d'un caractère continu soit très inférieur à la taille de la population. Comment se sortir de ce piège ? Tout simplement en «discrétisant» le caractère, c'est-à-dire en regroupant toutes les modalités appartenant à certains intervalles dans des $\boxed{\text{classes}}$ de données.

Exemple 4 (suite) : Afin de synthétiser les données du tableau 9, regroupons les modalités en tranches de 2000F. En arrondissant, les salaires sont globalement étalés entre 12000F et 28000F. On aura donc 8 classes : $[12000\text{F}, 14000\text{F}[, [14000\text{F}, 16000\text{F}[, \ldots, [26000\text{F}, 28000\text{F}[$, où $[x, y[$ désigne l'intervalle entre x et y, x compris mais y non compris. En regroupant les valeurs du tableau 9, on obtient le tableau suivant :

Classe de salaire	effectif	fréquence	Classe de salaire	effectif	fréquence
$[12000, 14000[$	3	3/40	$[20000, 22000[$	10	10/40
$[14000, 16000[$	4	4/40	$[22000, 24000[$	4	4/40
$[16000, 18000[$	7	7/40	$[24000, 26000[$	3	3/40
$[18000, 20000[$	7	7/40	$[26000, 28000[$	2	2/40

TAB. 10: Regroupement en classes.

♦

Les classes constituent un ensemble discret. Si l'on remplaçait le caractère «salaire» par un nouveau caractère «classe de salaires», on pourrait donc utiliser les diagrammes en bâtons ou en colonnes vus dans la section 1.2.

Exemple 4 (suite) : À partir des données du tableau 10, on obtiendrait le diagramme en bâtons suivant :

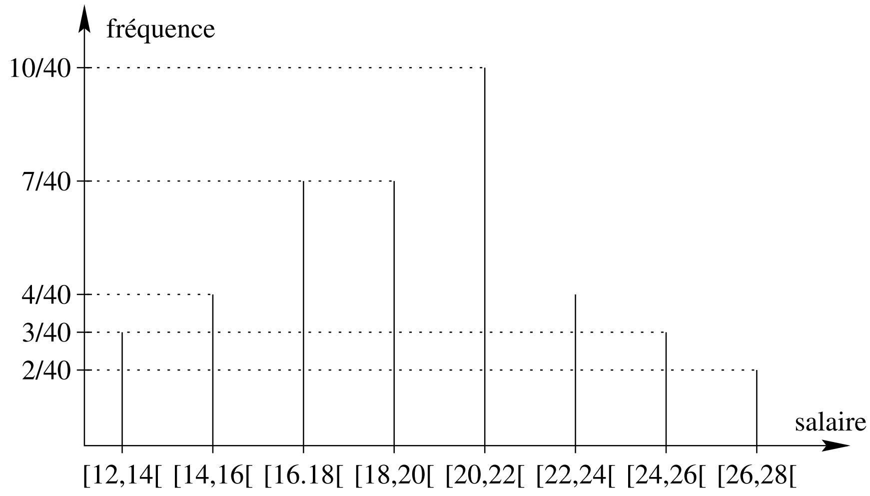


FIG. 6: Répartition du salaire en kF.

En élargissant les bâtons, on obtiendrait une représentation en colonnes. Contrairement au cas discret, dans lequel la largeur des colonnes n'a aucune signification, il serait assez logique d'imposer que la largeur des colonnes corresponde aux bornes des classes. On obtiendrait ainsi la représentation suivante :

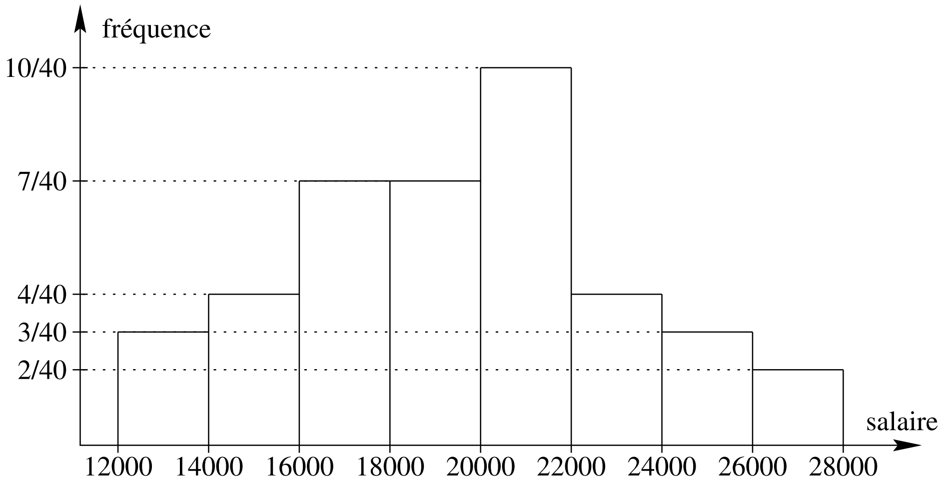


FIG. 7: Répartition du salaire en F.

♦

Malheureusement, si l'on procédait de cette manière, la lecture de la représentation en colonnes pourrait éventuellement nous induire en erreur. En effet, il n'y a aucune raison de ne considérer que des classes de même largeur. Voici ce qui arrive lorsque la largeur des classes n'est pas constante :

Exemple 4 (suite) : Discrétisons la variable «salaire» grâce aux classes suivantes : $\{[12000\text{F}, 17000\text{F}[, [17000\text{F}, 18000\text{F}[, [18000\text{F}, 19000\text{F}[, [19000\text{F}, 20000\text{F}[, [20000\text{F}, 21000\text{F}[, [21000\text{F}, 22000\text{F}[, [22000\text{F}, 27000\text{F}[\}$. On obtient alors le tableau d'effectif et de fréquence ci-après :

Classe de salaire	effectif	fréquence
$[12000, 17000[$	9	$9/40$
$[17000, 18000[$	5	$5/40$
$[18000, 19000[$	3	$3/40$
$[19000, 20000[$	4	$4/40$
$[20000, 21000[$	4	$4/40$
$[21000, 22000[$	6	$6/40$
$[22000, 28000[$	9	$9/40$

TAB. 11: Nouveau regroupement en classes.

ce qui donne lieu à la représentation en colonnes ci-dessous.

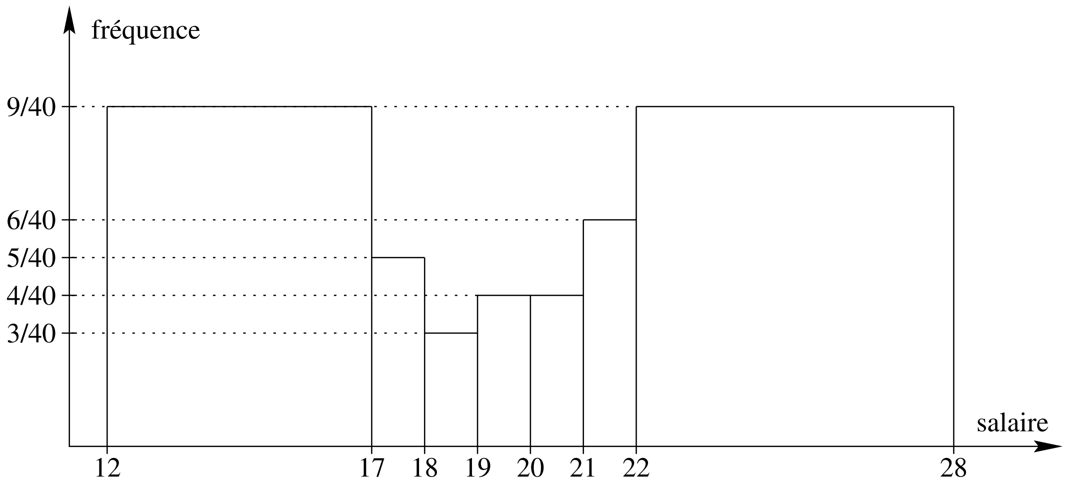
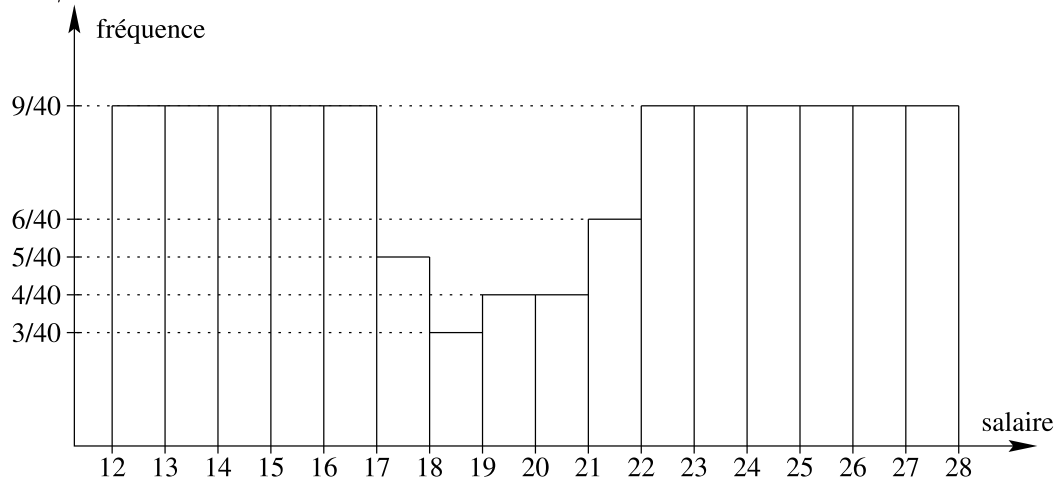


FIG. 8: Représentation en colonnes pour des classes de différentes largeurs.

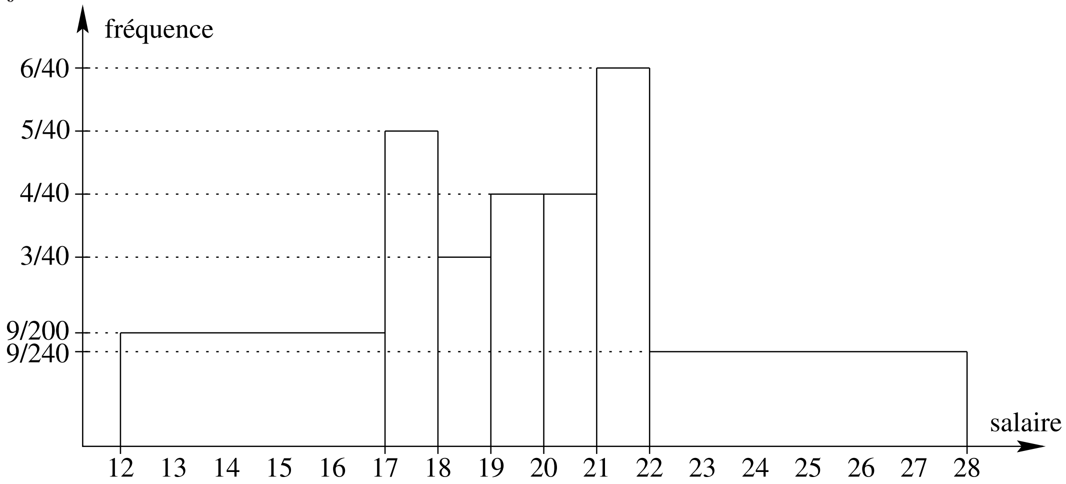
Lorsque l'on regarde la figure 7, on a l'impression qu'il y a plus de salaires «moyens», c'est-à-dire tournant aux alentours de 20000F, que de salaires bas (12000F) ou élevés (28000F). Au contraire, lorsque l'on regarde la figure 8, on a l'impression que la majorité des salaires est soit plutôt élevé, soit plutôt bas, mais qu'il y a peu de salaires moyens. ♦

Cet exemple montre que l'interprétation de la figure dépend des classes que l'on a constituées. Cette propriété est gênante car nous avions créé les représentations graphiques ci-dessus afin de synthétiser les données brutes pour pouvoir les analyser rapidement. Si l'on doit encore analyser la figure avant de comprendre les informations qu'elle contient, on n'a pas gagné grand chose par rapport aux données brutes.

Les figures 7 et 8 paraissent contradictoires parce que cette dernière est un «faux ami». En fait, lorsqu'on regarde la figure 8, on l'interprète plutôt comme si elle correspondait à la figure ci-après, qui n'a, malheureusement, pas du tout la même signification : d'après elle, chaque plage de 1000F entre 12000F et 17000F a une fréquence de 9/40.



Pour ne pas tomber dans ce piège, il suffit de diviser la fréquence en question, à savoir 9/40, par le nombre de tranches de 1000F entre 12000F et 17000F, autrement dit 5. Donc, en moyenne, chaque tranche de 1000F entre 12000F et 17000F a une fréquence de $(9/40)/5 = 9/200$. De même, on devrait associer la fréquence $(9/40)/6$ à chacune des tranches de 1000F entre 22000F et 28000F. Ceci nous donnerait la figure ci-après, qui montre bien que la majorité des salaires tourne autour de 19KF.



En fait, cela revient à créer ce que l'on appelle un histogramme :

Définition 5 : *Un histogramme est un graphe dont l'axe des abscisses représente les modalités et qui associe à chaque classe de modalités un rectangle dont la base correspond aux bornes de cette classe et dont la hauteur est égale à*

$$hauteur = \frac{fréquence}{base}.$$

Autrement dit, dans un histogramme, la fréquence d'une classe est égale au produit de sa base par sa hauteur, donc à la surface du rectangle qui lui est associé.

Exemple 5 : À l'issue d'une enquête statistique, on a récupéré les données suivantes concernant une variable statistique X :

9,4	6,1	3,3	3,5	5,2	8,1	5,7	5,3	3,9	5,9
5,5	3,9	8,8	0,2	4,2	7,3	6,3	6,4	6,4	5,7
5,8	1,9	7,5	5,9	5,8	1,2	3,2	5,9	6,8	4,6
8,3	5,1	6,2	6,7	4,8	5,2	6,3	2,1	0,8	1,6

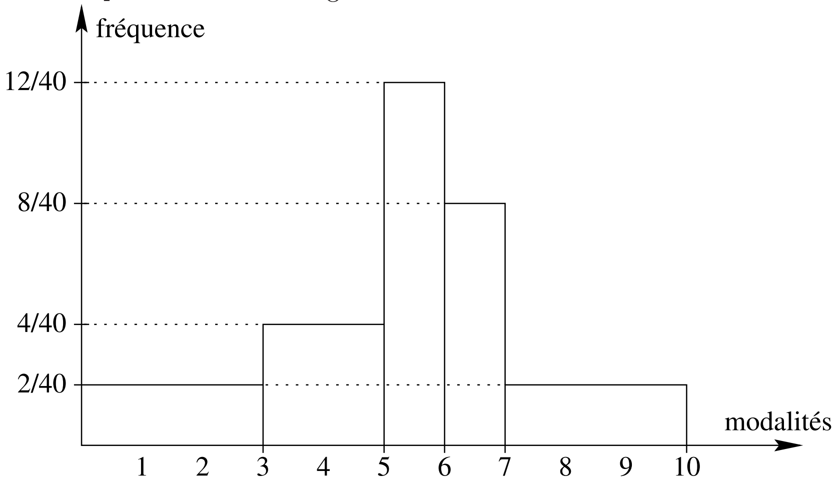
On répartit les données ci-dessus dans les classes de modalités suivantes :
$[0,3[$, $[3,5[$, $[5,6[$, $[6,7[$ et $[7,10[$.
On veut dessiner l'histogramme correspondant à ces classes de modalités.
La première étape consiste à associer à chaque classe les données lui appartenant. On obtient alors :

classe	données de la classe										
$[0,3[$	0,2	0,8	1,2	1,6	1,9	2,1					
$[3,5[$	3,2	3,3	3,5	3,9	3,9	4,2	4,6	4,8			
$[5,6[$	5,1	5,2	5,2	5,3	5,5	5,7	5,7	5,8	5,8	5,9	5,9	5,9
$[6,7[$	6,1	6,2	6,3	6,3	6,4	6,4	6,7	6,8			
$[7,10[$	7,3	7,5	8,1	8,3	8,8	9,4					

On peut maintenant calculer les fréquences et les hauteurs associées à chacune des classes :

classe	fréquence	base	hauteur
$[0,3[$	6/40	3	$(6/40)/3 = 2/40$
$[3,5[$	8/40	2	$(8/40)/2 = 4/40$
$[5,6[$	12/40	1	$(12/40)/1 = 12/40$
$[6,7[$	8/40	1	$(8/40)/1 = 8/40$
$[7,10[$	6/40	3	$(6/40)/3 = 2/40$

Il ne reste plus alors qu'à dessiner l'histogramme :



♦

Définition 6 : *Soit X une variable statistique continue. Supposons que ses modalités aient été regroupées en classes $\{C_1, C_2, \ldots, C_I\}$. Pour chaque classe $C_i = [a_i, b_i[$, le centre de la classe C_i est égal à $(a_i + b_i)/2$.*

D'une manière générale, il n'existe pas de règle précise régissant le choix des classes. Il existe cependant quelques principes qu'il est préférable de respecter :

- Le nombre de classes ne doit être ni trop grand, ni trop petit. Dans le premier cas, on n'obtient pas de synthèse d'information (voir figure 8) ; dans le second cas, il n'y a pas assez de classes pour que l'on puisse en déduire la répartition des modalités dans la population.
- Les classes doivent être adjacentes les unes aux autres et doivent être définies de telle sorte qu'aucun individu n'appartient à deux classes (c'est pour cela que l'on a utilisé des bornes du type $[x, y[$).
- Les classes contiennent toutes au moins une observation.

1.4 Exercices

Exercice 1 L'équipe de recherche d'un groupe pharmaceutique décide de lancer une étude pour déterminer si l'absorption régulière d'aspirine réduit les risques d'infarctus. Pour cela, elle distribue à 22000 volontaires des comprimés, qui peuvent être soit de l'aspirine, soit un placébo. L'expérience est conduite pendant 5 ans.

Quels sont les populations, individus, caractères, modalités de l'étude. Déterminer si le ou les caractère(s) en question est(sont) quantitatif(s) ou qualitatif(s) et, le cas échéant, s'il(s) est(sont) discret(s) ou continu(s).

Exercice 2 Une enquête parue dans le numéro du 29 juin 1993 de USA Today montre les préférences des adolescents américains en matière de fast-foods. Pour cela, on a demandé à 603 personnes de cocher dans le questionnaire ci-après la case qui remporte leur préférence gastronomique(?).

McDonald's	☐
Burger King	☐
Taco Bell	☐
Wendy's food	☐
Pizza Hut	☐
Autres	☐

1/ Déterminez quelle est la population de cette enquête, son caractère, les modalités de ce dernier.

2/ On a pris les questionnaires un par un et on a noté dans le tableau ci-dessous les cases qui étaient cochées (pour simplifier l'exercice, on a juste noté les 50 premiers questionnaires) :

Wendy	autre	autre	McDo	Hut	King	McDo	McDo	McDo	Taco
Taco	autre	autre	McDo	McDo	King	autre	autre	autre	Hut
McDo	McDo	McDo	autre	McDo	Taco	Taco	King	McDo	McDo
autre	autre	Taco	McDo	autre	King	McDo	McDo	autre	Taco
Taco	King	McDo	McDo	McDo	Taco	McDo	autre	McDo	McDo

Déterminer l'effectif et la fréquence des modalités du caractère «*fast food préféré*». Tracer le diagramme en bâtons correspondant.

Exercice 3 En 1993, l'association des chercheurs de la chambre de commerce américaine a recensé le prix moyen des maisons d'environ 150m^2 dans diverses villes. Voici les résultats obtenus :

Huntsville	105111$	Anchorage	155913$	Phoenix	100748$
Little Rock	92750$	San Diego	223600$	Denver	125200$
Orlando	104112$	Bloomington	105613$	Ames	110325$
La Nouvelle Orléans	88000$	Minneapolis	128112$	Lincoln	92925$
Manchester	125750$	Albuquerque	130550$	Réno-Sparks	142300$
Albany	116363$	Charlotte	115800$	Cincinnati	121329$
Salem	108656$	Sioux Falls	99890$	Memphis	91297$
Houston	96900$	Salt Lake City	92281$	Charleston	126000$
Green Bay	114500$				

1/ Regrouper les modalités de la variable statistique «prix» en 5 classes de tailles identiques et déterminer l'effectif et les fréquences de chaque classe. Déterminer l'histogramme correspondant.

2/ Regrouper les 3 dernières classes en une seule. Déterminer l'histogramme correspondant.

Exercice 4 Aux États Unis, la taxe sur les carburants est fixée par chaque état. En 1992, la Federal Highway Administration a établi le tableau suivant, qui indique pour chacun des 50 états sa taxe (exprimée en cents par gallon) :

18.0	8.0	18.0	18.7	17.0	22.0	28.0	19.0	11.8	7.5	16.0	21.0	19.0
18.0	15.4	20.0	19.0	23.5	21.0	15.0	20.0	18.2	13.0	21.4	24.6	24.0
17.0	22.9	22.3	17.0	21.0	17.0	24.0	22.4	26.0	16.0	18.0	20.0	20.0
17.5	23.0	20.4	22.2	9.0	15.0	20.0	18.6	10.5	19.0	16.0		

1/ Doit-on regrouper les modalités du caractère «taxe». Justifier votre réponse.

2/ Déterminer les fréquences ainsi qu'une représentation graphique de la variable statistique «taxe».

2 Moyenne, médiane et dispersion

Problème : existe-t-il quelques indicateurs bien choisis qui nous permettraient de synthétiser les données encore plus que les représentations graphiques vues dans la section précédente ? Autrement dit, peut-on différencier les histogrammes ci-dessous grâce à quelques variables caractéristiques ?

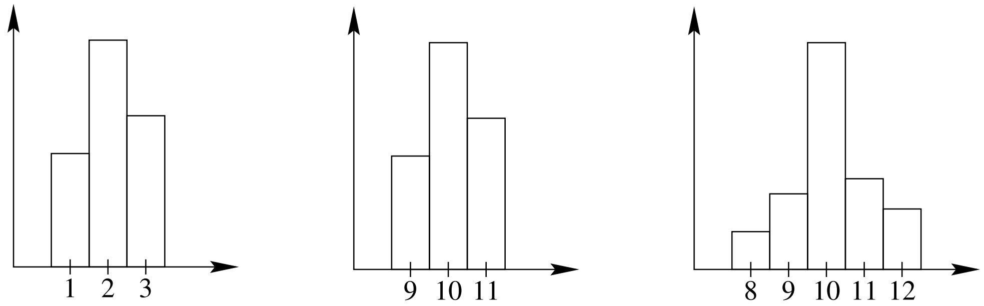


FIG. 9: Quelques histogrammes.

Les deux histogrammes de gauche sont identiques, exceptés qu'ils sont décalés sur l'axe des abscisses. Les deux histogrammes de droites sont situés aux mêmes endroits sur l'axe des abscisses mais celui de droite est visiblement plus étendu en largeur, on dit aussi plus dispersé, que celui du milieu. Ceci va nous amener à introduire les notions de «*moyenne*» et «*d'écart type*».

2.1 La moyenne

Définition 7 : *Soit X une variable statistique discrète définie sur une population de N individus, et dont les modalités sont $\{x_1, x_2, \ldots, x_I\}$. La moyenne de X, notée μ_X, est définie par :*

$$\mu_X = \sum_{i=1}^{I} f_i x_i = \frac{1}{N} \sum_{i=1}^{I} N_i x_i.$$

Soit Y une variable statistique continue définie sur une population de N individus. Supposons que ses modalités aient été regroupées en I classes dont, par analogie avec les variables discrètes, nous noterons les centres $\{y_1, y_2, \ldots, y_I\}$. Alors la moyenne de Y, notée μ_Y, est définie par :

$$\mu_Y = \sum_{i=1}^{I} f_i y_i = \frac{1}{N} \sum_{i=1}^{I} N_i y_i.$$

Exemple 3 (suite) : Illustrons le calcul de la moyenne d'une variable statistique discrète sur l'exemple de la compagnie d'aviation. Le tableau 5, page 6, nous indique les fréquences f_i des modalités de la variable $Y =$ «*nombre de langues parlées*». D'après ce tableau, on a :

$$\mu_Y = \frac{8}{24} \times 1 + \frac{12}{24} \times 2 + \frac{3}{24} \times 3 + \frac{1}{24} \times 4 = \frac{45}{24} = 1,875.$$

En moyenne, le personnel de la compagnie d'aviation parle 1,875 langues. ♦

Dans le cas d'une variable statistique continue, il est pratiquement impossible de connaître la valeur théorique réelle de la moyenne car les valeurs observées de la variable ne sont en général que des approximations des valeurs qu'elle prend. Par exemple, si l'on tirait des nombres réels au hasard, on pourrait très bien tirer le nombre π, mais nous ne pourrions noter qu'un nombre fini de décimales et donc seulement une approximation de sa valeur réelle. C'est pourquoi, pour calculer (approximer) la moyenne d'une variable continue, on peut se contenter de calculer la moyenne des centres des classes de modalités.

Exemple 4 (suite) : Illustrons le calcul de la moyenne d'une variable continue grâce à l'exemple des revenus des 40 ingénieurs. La variable en question est X = «revenu des ingénieurs». D'après le tableau 10, page 9, on aurait alors

$$\begin{aligned}\mu_X &= \frac{3}{40}\times 13000+\frac{4}{40}\times 15000+\frac{7}{40}\times 17000+\frac{7}{40}\times 19000+\frac{10}{40}\times 21000+\\ &\quad \frac{4}{40}\times 23000+\frac{3}{40}\times 25000+\frac{2}{40}\times 27000 = 19550.\end{aligned}$$

Si l'on avait pris les classes définies dans le tableau 11, on aurait obtenu :

$$\begin{aligned}\mu_X &= \frac{9}{40}\times 14500+\frac{5}{40}\times 17500+\frac{3}{40}\times 18500+\frac{4}{40}\times 19500+\frac{4}{40}\times 20500+\\ &\quad \frac{6}{40}\times 21500+\frac{9}{40}\times 25000 = 19687,5.\end{aligned}$$

On remarque que, quelles que soient les classes choisies, les valeurs obtenues pour μ_X qui, rappelons-le, ne sont qu'une approximation, sont assez proches l'une de l'autre. ♦

Il peut arriver que l'on ait à modifier l'échelle d'une variable. Par exemple, les salaires actuels sont exprimés en francs et on passe à l'euro, ou bien une variable représente des températures exprimées en degrés celsius et l'on désire les convertir en farenheit. Problème : comment mettre à jour la moyenne ?

Propriété : *Soit X une variable statistique et soit $Y = aX + b$, où a et b sont des nombres réels. Alors $\mu_Y = a\mu_X + b$.*

Pour conclure au sujet de la moyenne, notons qu'une de ses caractéristiques les plus intéressantes est qu'elle se prête aisément à des manipulations algébriques. Ainsi, si X est une variable statistique définie sur n sous-populations distinctes $S_1, S_2, \ldots, S_n$ comprenant respectivement $N_1, N_2, \ldots, N_n$ individus, et si μ_i, $i = 1, \ldots, n$, représente la moyenne de X sur la i$^{\text{ème}}$ sous-population, alors la moyenne μ_X de X sur la population $S = S_1 \cup S_2 \cup \ldots \cup S_n$ est définie par :

$$\mu_X = \frac{\sum_{i=1}^{n} N_i\mu_i}{\sum_{i=1}^{n} N_i}.$$

Exemple 6 : Le secrétariat d'État chargé de la santé a recensé le nombre de médecins pour 100000 habitants en France au 1[er] janvier 1997, ce qui a permis d'établir le tableau suivant :

région	nb de médecins/100000hab	population de la région
Alsace	309,4	1689722
Aquitaine	289,4	2914306
Auvergne	258,1	1312670
Bourgogne	237,7	1626420
Bretagne	264,8	2884441
Centre	233,8	2480325
Champagne-Ardenne	240,3	1344985
Corse	273,7	262696
Franche-Comte	252,6	1108868
Ile de France	380,8	11100578
Languedoc-Roussillon	320,8	2290212
Limousin	298,9	711989
Lorraine	275,0	2284364
Midi-Pyrenées	329,7	2511677
Nord - Pas-de-Calais	259,7	3975741
Basse-Normandie	240,7	1422933
Haute-Normandie	241,1	1802986
Pays de la Loire	242,7	3165636
Picardie	222,3	1875394
Poitou-Charentes	262,1	1613506
P.A.C.A	372,6	4567364
Rhone-Alpes	282,0	5657092

Question : à partir de ce tableau, peut-on déterminer le nombre moyen de médecins pour 100000 habitants en France métropolitaine ? En fait, tout se passe comme si la variable X = «nombre moyen de médecins pour 100000 habitants» avait été étudiée indépendamment sur chaque région (chaque sous-population), fournissant ainsi les μ_i apparaissant dans la deuxième colonne du tableau ci-dessus. Les N_i correspondent évidemment à la troisième colonne du tableau. Ainsi, μ_X est défini par :

$$\mu_X = \frac{309,4 \times 1689722 + 289,4 \times 2914306 + \cdots + 282,0 \times 5657092}{1689722 + 2914306 + \cdots + 5657092} \approx 297,9$$

♦

2.2 Médiane et quantiles

La moyenne représente la valeur centrale des modalités, mais il existe une autre tendance centrale, la médiane, qui est une valeur de la variable pour laquelle la moitié de la population a une modalité inférieure à cette valeur et l'autre moitié en a une supérieure :

Définition 8 : *Soit X une variable statistique discrète définie sur une population de N individus, et dont les modalités sont $\{x_1, \ldots, x_I\}$. Tout nombre δ tel que*

$$\sum_{i \in \{j : x_j < \delta\}} N_i \leq N/2 \quad \text{et} \quad \sum_{i \in \{j : x_j > \delta\}} N_i \leq N/2$$

est une <u>médiane</u> de X.

Exemple 7 : Considérons une variable statistique discrète X définie sur une population de $N = 19$ individus, et dont l'effectif est le suivant :

modalité	0	1	2	3	4	5
effectif	2	3	4	5	3	2

Si l'on écrit les valeurs prises par la variable pour chacun des individus, on obtient :

$$\begin{array}{c}0,\ 0,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 2,\ \underset{\uparrow}{3},\ 3,\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 5\end{array}$$

La flèche indique la médiane : il y a autant de chiffres à gauche de cette flèche qu'à sa droite.

Considérons maintenant une variable statistique discrète Y définie sur une population de $N = 18$ individus, et dont l'effectif est le suivant :

modalité	0	1	2	3	4	5
effectif	2	3	4	4	3	2

Si l'on écrit les valeurs prises par la variable pour chacun des individus, on obtient :

$$\begin{array}{c}0,\ 0,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ \underset{\uparrow}{2},\ \underset{\uparrow}{3},\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5,\ 5\end{array}$$

Il y a deux médianes car non seulement $\displaystyle\sum_{i\in\{j:x_j<2\}} N_i = 5 \leq 18/2$ et $\displaystyle\sum_{i\in\{j:x_j>2\}} N_i = 9 \leq 18/2$, mais aussi $\displaystyle\sum_{i\in\{j:x_j<3\}} N_i = 9 \leq 18/2$ et $\displaystyle\sum_{i\in\{j:x_j>3\}} N_i = 5 \leq 18/2$. ♦

Définition 9 : *Soit X une variable statistique continue. La médiane est le nombre δ tel que les aires situées de part et d'autre de ce nombre dans l'histogramme représentant X sont égales.*

Exemple 4 (suite) : Si l'on applique la définition ci-dessus à l'histogramme de la figure 7, page 9, on obtient alors la médiane :

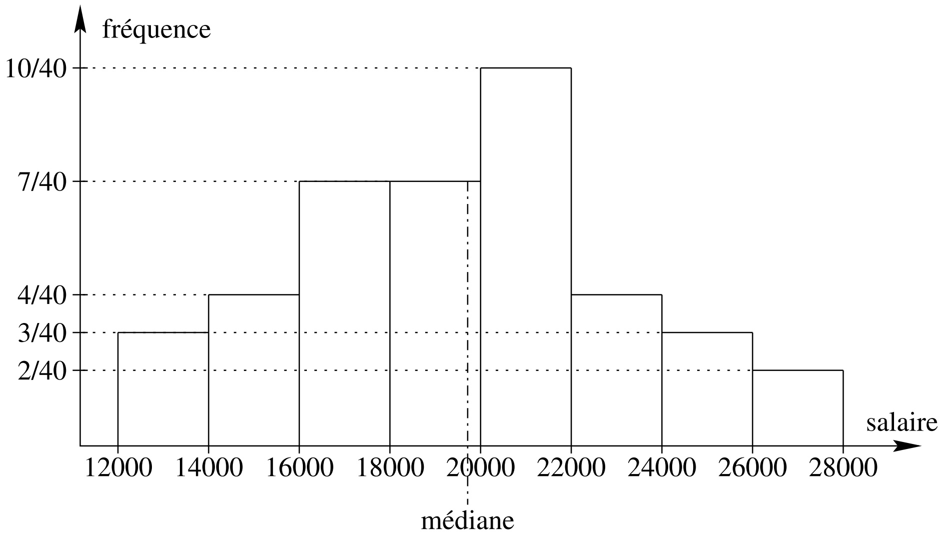


FIG. 10: Répartition du salaire en F.

D'une manière générale, il n'est pas très pratique de travailler directement sur l'histogramme. Voici une méthode simple pour déterminer la médiane : reprenons le tableau des fréquences des modalités (celui-ci avait été calculé page 9) :

Classe de salaire	effectif	fréquence	fréquence cumulée
$[12000, 14000[$	3	3/40	3/40
$[14000, 16000[$	4	4/40	7/40
$[16000, 18000[$	7	7/40	14/40
$[18000, 20000[$	7	7/40	21/40
$[20000, 22000[$	10	10/40	31/40
$[22000, 24000[$	4	4/40	35/40
$[24000, 26000[$	3	3/40	38/40
$[26000, 28000[$	2	2/40	40/40

La $\boxed{\text{fréquence cumulée}}$ pour la i$^{\text{ème}}$ classe de modalité, que nous noterons F_i, représente la somme des fréquences des classes de modalités x_j, $j \leq i$, autrement dit, $F_i = f_1 + f_2 + \cdots + f_i$.

Remarquons que la fréquence cumulée pour la classe $[16000, 18000[$ est $14/40$, ce qui signifie que $14/40^{\text{ème}}$ de la population, c'est-à-dire moins de la moitié de la population, a un revenu inférieur ou égal à 18000F. De plus, $21/40^{\text{ème}}$, soit plus de la moitié de la population, a un revenu inférieur à 20000F. Dans le rectangle de base $[18000, 2000[$, le coin gauche représente donc $14/40^{\text{ème}}$ de la population et le coin droit $21/40^{\text{ème}}$. Pour obtenir la moitié de la population, on fait une règle de trois :

$$\frac{1}{2} = \frac{20}{40} = \frac{14}{40} + \frac{6}{7} \times \frac{21-14}{40} \quad \Rightarrow \quad \text{médiane} = 18000 + \frac{6}{7} \times 20000 = 19714,286\text{F}.$$

♦

Les quantiles généralisent la notion de médiane :

Définition 10 : *Soit X une variable statistique discrète définie sur une population de N individus, et dont les modalités sont $\{x_1, \ldots, x_I\}$. Tout nombre δ tel que*

$$\sum_{i \in \{j : x_j < \delta\}} N_i \leq \alpha N \quad \text{et} \quad \sum_{i \in \{j : x_j > \delta\}} N_i \leq (1-\alpha) N$$

est un quantile d'ordre α de X.
Soit X une variable statistique continue. Un quantile d'ordre α est le nombre δ tel que les aires situées de part et d'autre de ce nombre dans l'histogramme représentant X sont égales respectivement à $\alpha \times$ aire totale et $(1-\alpha) \times$ aire totale.
Les principaux quantiles sont :

valeur de α	*nom du quantile d'ordre α*
$1/2$	*médiane*
$i/4$ $(i = 1, 2, 3)$	*$i^{\text{ème}}$ quartile*
$i/5$ $(i = 1, 2, 3, 4)$	*$i^{\text{ème}}$ quintile*
$i/10$ $(i = 1, 2, \ldots, 9)$	*$i^{\text{ème}}$ décile*
$i/100$ $(i = 1, \ldots, 99)$	*$i^{\text{ème}}$ centile*

Propriété : *Soit X une variable statistique et soit $Y = aX + b$, où a et b sont des nombres réels. Alors médiane$(Y) = a \times$ médiane$(X) + b$.*

2.3 Variance et écart-type

Maintenant que l'on sait comment caractériser les tendances «centrales» de l'amas de données, il faut aussi caractériser sa dispersion. La première idée qui vient à l'esprit est la suivante :

Définition 11 : *Soit X une variable statistique discrète. L'étendue de X est la différence entre la plus grande modalité de X et la plus petite modalité.*
Soit X une variable statistique continue. L'étendue de X est la différence entre la borne supérieure de la classe associée à la plus grande valeur observée et la borne inférieure de la classe associée à la plus petite valeur observée.

Exemple 4 (suite) : D'après le tableau 10, page 9, les classes de X = «revenu» vont de $[12000, 14000[$ à $[26000, 28000[$. Donc l'étendue de X est égale à $28000 - 12000 = 16000$. ♦

Propriété : *Soit X une variable statistique et soit $Y = aX + b$, où a et b sont des nombres réels. Alors étendue$(Y) = |a| \times$ étendue(X).*

Cependant l'étendue n'est pas une mesure de dispersion très utilisée. En effet, si elle a le mérite de montrer l'intervalle sur lequel s'étend la variable statistique, elle présente l'inconvénient majeur de ne donner aucune information sur la répartition de celle-ci sur cet intervalle. Ainsi, dans une population P contenant 100 individus, si X et Y sont deux variables statistiques telles que

- pour la variable X, les effectifs des modalités i, $i \in \{1, 2, \ldots, 100\}$, sont tous égaux à 1,
- pour la variable Y, les effectifs des modalités i, $i \in \{1, 2, 3, 4, 5\}$, sont respectivement égaux à 20, 20, 20, 20 et 19 ; ceux des modalités i, $i \in \{6, 7, \ldots, 99\}$ sont égaux à 0 ; et enfin, l'effectif de la modalité 100 est égal à 1,

alors les variables X et Y ont la même étendue. Pourtant, ces deux variables paraîssent très différentes. C'est pourquoi les statisticiens préfèrent utiliser une autre mesure de dispersion, à savoir l'écart-type, une mesure qui se définit à partir de la notion de variance. L'idée est d'analyser globalement les déviations observées entre les valeurs prises par une variable sur un individu et la moyenne de celle-ci.

Définition 12 : *Soit X une variable statistique discrète définie sur une population de N individus et dont l'ensemble des modalités est $x_1, \ldots, x_I$. La variance de X, notée σ_X^2, est égale à :*

$$\sigma_X^2 = \sum_{i=1}^{I} f_i (x_i - \mu_X)^2 = \frac{1}{N} \sum_{i=1}^{I} N_i (x_i - \mu_X)^2.$$

Soit Y une variable statistique continue définie sur une population de N individus. Supposons que ses modalités aient été regroupées en I classes dont, par analogie avec les variables discrètes, nous noterons les centres $\{y_1, y_2, \ldots, y_I\}$. Alors la variance de Y, notée σ_Y^2, est définie par :

$$\sigma_Y^2 = \sum_{i=1}^{I} f_i (y_i - \mu_Y)^2 = \frac{1}{N} \sum_{i=1}^{I} N_i (y_i - \mu_Y)^2.$$

Ainsi, plus σ_X^2 est grand, plus les valeurs de la variable sont dispersées. Pourquoi prendre des carrés ? Parce que si l'on avait défini la variance comme étant égale à $\sum_{i=1}^{I} f_i (x_i - \mu_X)$, on aurait systématiquement obtenu la valeur 0.

Exemple 8 : Soit X une variable statistique discrète sur une population de 18 individus. Les valeurs de la variable observées dans la population sont :

2 0 3 3 4 1 2 5 2 3 0 4 2 1 1 5 3 4

En ordonnant les données, nous obtenons les effectifs suivants :

modalité	0	1	2	3	4	5
effectif	2	3	4	4	3	2

On peut ainsi calculer la moyenne μ_X :

$$\mu_X = \frac{1}{18}(2 \times 0 + 3 \times 1 + 4 \times 2 + 4 \times 3 + 3 \times 4 + 2 \times 5) = \frac{45}{18} = \frac{5}{2}.$$

Par conséquent, la variance de X est :

$$\sigma_X^2 = \frac{1}{18} \Bigg[2 \times \left(0 - \frac{5}{2}\right)^2 + 3 \times \left(1 - \frac{5}{2}\right)^2 + 4 \times \left(2 - \frac{5}{2}\right)^2 + 4 \times \left(3 - \frac{5}{2}\right)^2 + 3 \times \left(4 - \frac{5}{2}\right)^2 + 2 \times \left(5 - \frac{5}{2}\right)^2 \Bigg] = \frac{81}{2}.$$

♦

Exemple 4 (suite) : Pour illustrer la variance d'une variable statistique continue, reprenons le tableau 10 :

Classe de salaire	centre de la classe	effectif	fréquence
$[12000, 14000[$	13000	3	3/40
$[14000, 16000[$	15000	4	4/40
$[16000, 18000[$	17000	7	7/40
$[18000, 20000[$	19000	7	7/40
$[20000, 22000[$	21000	10	10/40
$[22000, 24000[$	23000	4	4/40
$[24000, 26000[$	25000	3	3/40
$[26000, 28000[$	27000	2	2/40

D'après ce tableau, la moyenne, déjà calculée page 16, est de 19550. La variance est donc :

$$\begin{aligned}\sigma_X^2 = \frac{1}{40}\Big[&3\times(13000-19550)^2+4\times(15000-19550)^2+7\times(17000-19550)^2\\&+7\times(19000-19550)^2+10\times(21000-19550)^2+4\times(23000-19550)^2\\&+3\times(25000-19550)^2+2\times(27000-19550)^2\Big] = 13197500\mathrm{F}^2.\end{aligned}$$

♦

En pratique, on ne calcule pas la variance directement à partir de la définition 12. Il est en effet souvent plus simple d'utiliser la formule suivante :

$$\boxed{\sigma_X^2 = \frac{1}{N}\left(\sum_{i=1}^{I} N_i x_i^2\right) - \mu_X^2.}$$

Démonstration :

$$\begin{aligned}\sigma_X^2 &= \frac{1}{N}\sum_{i=1}^{I} N_i(x_i-\mu_X)^2 = \frac{1}{N}\sum_{i=1}^{I} N_i(x_i^2 - 2\mu_X x_i + \mu_X^2)\\&= \frac{1}{N}\sum_{i=1}^{I} N_i x_i^2 - \frac{2\mu_X}{N}\sum_{i=1}^{I} N_i x_i + \frac{\mu_X^2}{N}\sum_{i=1}^{I} N_i = \frac{1}{N}\sum_{i=1}^{I} N_i x_i^2 - \frac{2\mu_X}{N}N\mu_X + \frac{\mu_X^2}{N}N\\&= \frac{1}{N}\left(\sum_{i=1}^{I} N_i x_i^2\right) - \mu_X^2.\end{aligned}$$

♦

Propriété : *Soit X une variable statistique et soit $Y = aX + b$, où a et b sont des nombres réels. Alors $\sigma_Y^2 = a^2\sigma_X^2$.*

Le problème de la variance, qu'on peut constater sur l'exemple ci-dessus, réside dans le fait qu'on a des données exprimées dans une unité (ici des francs), et que la variance est exprimée dans le carré de cette unité (F^2). Pour pallier cela, on introduit, en statistique, l'écart-type :

Définition 13 : *Soit X une variable statistique (discrète ou continue). L'écart-type de X, noté σ_X, est la racine carrée de la variance de X.*

Propriété : *Soit X une variable statistique et soit $Y = aX + b$, où a et b sont des nombres réels. Alors $\sigma_Y = |a|\sigma_X$.*

L'écart-type est assurément la mesure de dispersion la plus utilisée en statistique. Le théorème suivant le justifie :

> **Théorème 1 (Inégalité de Bienaymé-Tchébicheff) :** *Soit X une variable statistique discrète de moyenne μ_X et d'écart-type σ_X. Alors l'ensemble des valeurs de X observées entre $\mu_X - k\sigma_X$ et $\mu_X + k\sigma_X$ est strictement supérieur à $1 - \frac{1}{k^2}$.*
> *Cette propriété s'étend aux variables statistiques continues : l'ensemble des individus pour lesquels la valeur de X se situe entre $\mu_X - k\sigma_X$ et $\mu_X + k\sigma_X$ est strictement supérieure à $1 - \frac{1}{k^2}$ ème de la population.*

Cette propriété est importante car elle montre que plus de 88,89% de la population se situe entre $\mu_X - 3\sigma_X$ et $\mu_X + 3\sigma_X$, et même plus de 96% de la population se situe entre $\mu_X - 5\sigma_X$ et $\mu_X + 5\sigma_X$. L'écart-type, associé au théorème 1, est donc un indicateur de dispersion très puissant.

Démonstration : Soit k un nombre réel strictement positif et soit X une variable statistique discrète dont l'ensemble des modalités est $\{x_1, \ldots, x_I\}$ et dont la moyenne et l'écart-type valent respectivement μ_X et σ_X. Soient $I_1 = \{i : \mu_X - k\sigma_X \leq x_i \leq \mu_X + k\sigma_X\}$ et $I_2 = \{i : x_i < \mu_X - k\sigma_X \text{ ou } \mu_X + k\sigma_X < x_i\}$. Alors :

$$\sigma_X^2 = \sum_{i=1}^{I} f_i(x_i - \mu_X)^2 = \sum_{i \in I_1} f_i(x_i - \mu_X)^2 + \sum_{i \in I_2} f_i(x_i - \mu_X)^2.$$

Lorsque $i \in I_2$, $(x_i - \mu_X)^2 > k^2\sigma_X^2$. Donc :

$$\sigma_X^2 > \sum_{i \in I_1} f_i(x_i - \mu_X)^2 + \sum_{i \in I_2} f_i k^2 \sigma_X^2 \geq \sum_{i \in I_2} f_i k^2 \sigma_X^2 = k^2\sigma_X^2 \sum_{i \in I_2} f_i.$$

En remarquant que $\sum_{i \in I_2} f_i = 1 - \sum_{i \in I_1} f_i$, on obtient $\sigma_X^2 > k^2\sigma_X^2 \left(1 - \sum_{i \in I_1} f_i\right)$, d'où :

$$\frac{1}{k^2} > 1 - \sum_{i \in I_1} f_i, \text{ ce qui équivaut à :}$$
$$\sum_{i \in I_1} f_i > 1 - \frac{1}{k^2}.$$

♦

2.4 Exercices

Exercice 1 La chambre de commerce américaine a effectué une étude sur le prix des vins dans différentes villes des États Unis. Voici les données récoltées (ci-dessous les prix de Chablis Paul Masson, bouteille de 1,5 litre) :

Huntsville	5,95$	Phoenix	4,40$	San Diego	5,41$
Denver	4,39$	Bloomington	4,67$	La Nouvelle Orléans	4,52$
Manchester	4,24$	Albuquerque	4,93$	Albany	5,91$
Charlotte	4,69$	Salem	4,51$	Charleston	6,05$

Calculer la moyenne de la variable «prix des vins».

Exercice 2 Le service administratif d'une compagnie de jeux vidéos de 40 personnes a comptabilisé le nombre de jour d'absentéisme au cours de l'année dernière. Il a ainsi obtenu le tableau suivant :

Nombre de jours	nombre d'employés
de 0 à 2	13
de 3 à 5	14
de 6 à 8	6
de 9 à 11	4
de 12 à 14	3

En moyenne, combien de jours un employé est-il absent ?

Exercice 3 D'après le Business Week daté du 3 aout 1992, le gouvernement américain a accordé en 1991 aux 12 firmes automobiles listées ci-dessous le nombre de brevets suivant :

General Motors	506	Ford	234	Chrysler	97
Nissan	385	Porsche	50	Honda	249
Mazda	210	Volvo	23	Daimler-Benz	275
Toyota	257	Mitsubishi	36	BMW	13

Déterminer la médiane de la variable «nombre de brevets», ainsi que les premiers et deuxièmes quartiles.

Exercice 4 Aux États Unis, la taxe sur les carburants est fixée par chaque état. En 1992, la Federal Highway Administration a établi le tableau suivant, qui indique pour chacun des 50 états sa taxe (exprimée en cents par gallon) :

18.0	8.0	18.0	18.7	17.0	22.0	28.0	19.0	11.8	7.5	16.0	21.0	19.0
18.0	15.4	20.0	19.0	23.5	21.0	15.0	20.0	18.2	13.0	21.4	24.6	24.0
17.0	22.9	22.3	17.0	21.0	17.0	24.0	22.4	26.0	16.0	18.0	20.0	20.0
17.5	23.0	20.4	22.2	9.0	15.0	20.0	18.6	10.5	19.0	16.0		

1/ Regrouper la variable statistique «taxe» en 5 classes de tailles identiques.
2/ Calculer le troisième quintile de la variable «taxe».

Exercice 5 Reprenez l'exercice 1 et calculez l'écart-type de la variable «prix des vins». Recalculez la moyenne et l'écart-type en francs, et non plus en dollars.
Faites de même avec l'exercice 4.

Exercice 6 Une société d'assurances veut proposer à ses clients une nouvelle police d'assurance. Pour déterminer son tarif, elle a sélectionné les dossiers de 40 de ses clients, et a estimé ce qu'elle aurait dû rembourser aux dits clients l'année dernière. Voici les remboursements qu'elle a calculés :

500F	150F	0F	0F	2500F	250F	0F	100000F	1000F	500F
0F	0F	5500F	560F	0F	230F	150F	1000F	0F	1250F
60F	0F	1500F	0F	750F	1500F	2000F	0F	750F	0F
700F	1800F	200F	0F	600F	0F	20000F	300F	0F	0F

1/ Calculer la moyenne et l'écart-type de la variable «remboursement».
2/ Calculer le premier et le neuvième décile.
3/ Quelle est la mesure de dispersion la plus appropriée ?

3 Introduction aux probabilités

Dans tout ce que l'on a vu jusqu'à maintenant intervient implicitement la notion de probabilité. Par exemple, le théorème de Bienaymé-Tchebicheff peut se résumer à $\text{Prob}(\mu_X - k\sigma_X \leq X \leq \mu_X + k\sigma_X) > 1 - 1/k^2$. Mais qu'est ce, au juste, qu'une probabilité ? C'est une question épineuse à laquelle il est difficile de répondre : plusieurs réponses ont déjà été apportées, mais aucune ne satisfait entièrement la communauté scientifique travaillant sur le sujet. Grossièrement, les théories qui s'affrontent pour le moment sont réparties en deux classes :

- les «*probabilités subjectives*», prônées en particulier par des chercheurs comme DeFinetti[1] ou Lindley[2], qui considèrent que les probabilités expriment le degré d'incertitude que les <u>individus</u> accordent à des événements incertains. Par exemple si l'on lit dans un journal «certains observateurs de la vie politique pensent qu'il y a une chance sur quatre pour que la rencontre au sommet ait lieu avant novembre», alors Prob(rencontre avant novembre) $= 0,25$. Cet exemple peut paraître un peu simpliste, mais dans la pratique, cette interprétation des probabilités est souvent utilisée : dans la gestion des catastrophes naturelles, il est assez rare d'avoir beaucoup de données «objectives» ; si l'on veut utiliser des outils «probabilistes» (eh oui, ça existe), on demande alors à des experts es catastrophes d'estimer la probabilité que tel ou tel événement se produise.
- les «*probabilités objectives*», qui suggèrent que les probabilités ne dépendent absolument pas d'appréciations personnelles et doivent donc être fondées uniquement sur des données *objectives*. Historiquement c'est ainsi que les premiers mathématiciens pensèrent les probabilités aux XVIème, XVIIème et XVIIIème siècles (Cardano, De Moivre, Pascal, Bernoulli, Laplace) : en étudiant les jeux de hasard, ils cherchèrent à découvrir les harmonies naturelles du hasard. L'approche des probabilités objectives est encore d'actualité de nos jours puisqu'elle a donné lieu dans les années 30 à l'axiomatique de Kolmogoroff. En pratique, elle est aussi très utile : par exemple, c'est un logiciel à base de probabilités «objectives» qui gère les incidents du système de propulsion de la navette spatiale[3]. Un autre logiciel à base de probabilités gère le diagnostic de pannes des imprimantes sous Windows 95, etc.

Quoi qu'il en soit, toutes ces théories s'accordent pour dire que les probabilités sont une des nombreuses manières de représenter de l'incertitude. Richard T. Cox , un physicien, a même montré mathématiquement que, si l'on est «rationnel», c'est la seule représentation qu'on puisse choisir, toutes les autres aboutissent à des inconsistances[4]. Par rationnel, Cox entend le respect de 5 desideratas :

1. Si vous évaluez l'incertitude d'un événement de plusieurs manières, vous devez toujours obtenir le même résultat.
2. L'incertitude d'un événement ne doit pas changer brutalement parce que l'incertitude sur un autre événement subit une très légère modification. Par exemple, ce n'est pas parce que vous pipez légèrement un dé que le «1», qui sortait avant une fois sur six, va maintenant sortir une fois sur deux.
3. La représentation de l'incertitude doit être universelle, c'est-à-dire ne pas s'appliquer seulement à quelques problèmes. Par exemple s'appliquer à l'incertitude sur le jet d'un dé, mais pas pour le numéro qui va sortir à la roulette.
4. Vous ne voulez évaluer l'incertitude que d'événements définis sans ambiguïté. Par exemple, prenez une personne au hasard dans la rue, combien de chances a-t-elle d'être des DOM-TOM ? Que veut dire «être des DOM-TOM» ? Est-ce qu'on parle des personnes nées dans les DOM-TOM ? recensées dans les DOM-TOM ? habitant actuellement les DOM-TOM ? L'énoncé est ambigu.
5. La représentation doit pouvoir utiliser toutes les informations dont vous disposez. Autrement dit, si vous ne fournissez pas à votre représentation toutes les informations que vous avez, ne contestez pas ses évaluations d'incertitude, sous pretexte que votre processus mental vous a conduit à estimer différemment l'incertitude.

Puisque c'est la seule manière d'être rationnel, adoptons-la. Dans le reste du cours, nous suivrons l'approche objectiviste parce qu'elle correspond bien à l'esprit des statistiques. Étudions maintenant une définition des probabilités.

[1] cf. "Theory of Probability", Vol 1 & 2, *Bruno DeFinetti*, Wiley & Sons

[2] cf. "Scoring Rules and the Inevitability of Probability", *Dennis V. Lindley*, *International Statistical Review*, vol 50 (1982), pp. 1-26

[3] cf. le site internet : `http ://www.research.microsoft.com/research/dtg/horvitz/vista.htm`

[4] cf. "The Algebra of Probable Inference", *Richard T. Cox*, Johns Hopkins University Press

3.1 Les événements

Définition 14 : *Tout ce qui peut se réaliser ou ne pas se réaliser à la suite d'une expérience est appelé événement. Lorsque l'on est assuré qu'un événement va se produire, on dit que l'événement est certain. Lorsqu'on est assuré qu'il ne pourra jamais se produire, il est dit impossible. Un événement est élémentaire si seulement un seul résultat de l'expérience permet de le réaliser.*

Exemple 9 : Supposons qu'on prenne une pièce pour jouer à pile ou face. Des événements possibles sont alors «pile», «face», «tranche» (si l'on considère que la pièce peut retomber sur la tranche). On peut aussi composer des événements : «pile ou face», «pile ou tranche», «pile et face», etc. Notons qu'un événement n'est pas forcément réalisable, par exemple on n'obtiendra jamais «pile et face» en même temps. «pile et face» est donc un événement impossible. On le note habituellement $\emptyset$.

Si l'on veut jouer aux dés, des événements possibles sont : «obtenir un six», «ne pas obtenir un six», «obtenir un chiffre inférieur à trois». «obtenir un chiffre inférieur à 7» est un événement certain. «obtenir un chiffre pair» n'est pas un événement élémentaire car trois résultats d'expérience permettent de le réaliser : lorsque le dé tombe sur «2», lorsqu'il tombe sur «4» et lorsqu'il tombe sur «6». Par contre, l'événement «obtenir 4» est élémentaire. ♦

Définition 15 : *L'univers associé à une expérience, que l'on note habituellement Ω, est l'ensemble des événements élémentaires. Ainsi, tout événement est un sous-ensemble de Ω.*

Exemple 10 : L'univers associé à l'expérience (on parle aussi d'une $\boxed{\text{expérience aléatoire}}$) «jet d'un dé à six faces» est l'ensemble : $\Omega = \{$«obtenir 1», «obtenir 2», «obtenir 3», «obtenir 4», «obtenir 5», «obtenir 6»$\}$. Notons que l'on n'aurait pas pu rajouter dans l'ensemble Ω des événements élémentaires tels que «obtenir 7» ou «obtenir 8» car aucun résultat ne permet de les réaliser. ♦

Définition 16 : *Soient A et B deux événements (c'est-à-dire des sous-ensembles de Ω). Alors :*
- *$A \cup B$ désigne un événement qui est réalisé dès lors qu'au moins un des événements A et B est réalisé.*
- *$A \cap B$ désigne un événement qui est réalisé si à la fois A et B sont réalisés.*
- *$\overline{A}$ (complémentaire de A dans Ω) est l'événement qui est réalisé si et seulement si A ne l'est pas.*
- *$A \cap B = \emptyset$: il s'agit de deux événements qui ne peuvent se réaliser simultanément. On dit qu'ils sont incompatibles ou encore disjoints.*

$\cup$, $\cap$ et le complémentaire correspondent aux opérations usuelles sur les ensembles (ce qui est normal puisque les événements sont en fait des ensembles).

Exemple 11 : Dans l'exemple du jet de dés, l'événement «obtenir un nombre inférieur à 10» est égal à Ω. En effet, il peut se mettre sous la forme : «obtenir 1» $\cup$ «obtenir 2» $\cup$ «obtenir 3» $\cup$ «obtenir 4» $\cup$ «obtenir 5» $\cup$ «obtenir 6».
«obtenir un nombre pair» $\cap$ «obtenir un nombre inférieur ou égal à 4» = [«obtenir 2» $\cup$ «obtenir 4» $\cup$ «obtenir 6»] $\cap$ [«obtenir 1» $\cup$ «obtenir 2» $\cup$ «obtenir 3» $\cup$ «obtenir 4»] = «obtenir 2 ou 4».
«obtenir un nombre pair» $\cap$ «obtenir un nombre impair» = $\emptyset$. ♦

3.2 Définition des probabilités

Dans toute cette section, ainsi que dans les deux suivantes, on supposera toujours que l'univers Ω est fini.

À partir de la notion d'événement, Kolmogoroff a défini les probabilités de la manière suivante :

Définition 17 : *Soit Ω un ensemble fini. Une probabilité* $\mathrm{Prob}(\cdot)$ *est une fonction de l'ensemble des parties de Ω, c'est-à-dire l'ensemble des sous-ensembles de Ω, que l'on note $\mathcal{P}(\Omega)$, dans $[0, 1]$, qui vérifie les trois propriétés suivantes :*

1. $\forall A \in \mathcal{P}(\Omega)$, $\mathrm{Prob}(A) \geq 0$.
2. $\mathrm{Prob}(\Omega) = 1$ *et* $\mathrm{Prob}(\emptyset) = 0$.
3. $\forall A \in \mathcal{P}(\Omega)$, $\forall B \in \mathcal{P}(\Omega)$ *tels que* $A \cap B = \emptyset$, $\mathrm{Prob}(A \cup B) = \mathrm{Prob}(A) + \mathrm{Prob}(B)$.

Grâce à cette définition, on voit qu'étant données les probabilités des événements élémentaires, la propriété 3 nous permet de calculer la probabilité de n'importe quel événement (puisque ce dernier est en fait une réunion finie d'événements élémentaires disjoints).

Corollaire 1 : *Soient A et B deux événements quelconques. Alors :*

$$\mathrm{Prob}(A \cup B) = \mathrm{Prob}(A) + \mathrm{Prob}(B) - \mathrm{Prob}(A \cap B).$$

3.3 Cas des probabilités uniformes

Considérons l'expérience aléatoire «on lance un dé à six faces non pipé». L'univers de l'expérience est donc l'ensemble de tous les résultats possibles du jet, c'est-à-dire $\{$«1», «2», «3», «4», «5», «6»$\}$. Les événements élémentaires sont «i», $i = 1, \ldots, 6$. On suppose le dé non pipé, donc toutes les faces ont la même chance de sortir. Si l'on désigne par $\mathrm{Prob}(«i»)$ la probabilité que la face «i» sorte, alors on doit avoir :

$$\mathrm{Prob}(«1») = \mathrm{Prob}(«2») = \mathrm{Prob}(«3») = \mathrm{Prob}(«4») = \mathrm{Prob}(«5») = \mathrm{Prob}(«6»).$$

$\mathrm{Prob}(\Omega) = 1$, autrement dit :

$$\mathrm{Prob}(\Omega) = \mathrm{Prob}(«1») + \mathrm{Prob}(«2») + \mathrm{Prob}(«3») + \mathrm{Prob}(«4») + \mathrm{Prob}(«5») + \mathrm{Prob}(«6») = 1.$$

Des deux égalités ci-dessus, on déduit que :

$$\mathrm{Prob}(«1») = \mathrm{Prob}(«2») = \mathrm{Prob}(«3») = \mathrm{Prob}(«4») = \mathrm{Prob}(«5») = \mathrm{Prob}(«6») = \frac{1}{|\Omega|} = \frac{1}{6}.$$

On dit alors que la probabilité est $\boxed{\text{uniforme}}$ et que les événements élémentaires sont $\boxed{\text{équiprobables}}$.

Prenons maintenant un événement composé : «1 ou 2». Il y a une chance sur six d'obtenir un «1» et une chance sur six d'obtenir un «2». On ne peut obtenir ces résultats en même temps, autrement dit les événements sont disjoints, donc, d'après la propriété 3 de la définition 17 ci-dessus, il y a deux chances sur six d'avoir soit un «1» soit un «2». Autrement dit, $\mathrm{Prob}(«1» \text{ ou } «2») = 2/6$. En généralisant, soit E un événement. Alors

$$\mathrm{Prob}(E) = \frac{|E|}{|\Omega|}. \tag{1}$$

Définition 18 : *Dans le cas d'une probabilité uniforme, la probabilité d'un événement est égale au nombre de cas favorables à la réalisation de cet événement ($|E|$) divisé par le nombre d'événements dans l'univers ($|\Omega|$).*

On remarque que cette définition est totalement en accord avec la définition 17 de Kolmogoroff : le cardinal d'un ensemble étant toujours positif ou nul, l'équation (1) entraîne la propriété 1 de la définition 17. $\mathrm{Prob}(\Omega) = |\Omega|/|\Omega| = 1$ et $\mathrm{Prob}(\emptyset) = |\emptyset|/|\Omega| = 0$. La propriété 3 de la définition 17, quant à elle, est vérifiée puisque si $A \cap B = \emptyset$, $|A \cup B| = |A| + |B|$.

3.4 Probabilités et fréquences

Considérons maintenant une expérience aléatoire quelconque. Par exemple le jet d'un dé à six faces. Réalisons n fois l'expérience et notons les fréquences relatives de chacun des événements. Ainsi $f_n(A)$ dénote le nombre de fois où l'événement A a été réalisé parmi ces n expériences. Il est évident que $f_n(A)$ dépend des jets de dés que l'on a effectués. Deux séries de jets peuvent donc nous donner deux valeurs différentes pour $f_n(A)$. Cependant, on observe expérimentalement (et on peut le démontrer mathématiquement) que lorsque n augmente, $f_n(A)$ fluctue de moins en moins. Lorsque n tend vers $+\infty$, la limite vers laquelle tend $f_n(A)$, nous dirons que c'est la probabilité de A : c'est assez logique puisque cela correspond en gros au nombre de chances que A a de se réaliser.

Définition 19 (loi des grands nombres) : *Soit $f_n(A)$ la fréquence de réalisation d'un événement A parmi n expériences aléatoires. Alors :*

$$\text{Prob}(A) = \lim_{n \to +\infty} f_n(A).$$

Exemple 12 : Vous jetez un dé à six faces un millier de fois et vous notez les fréquences des résultats obtenus :

modalité	1	2	3	4	5	6
fréquence	0,165	0,169	0,166	0,167	0,166	0,167

$n = 1000$ étant un «grand» nombre, on peut approximer les probabilités des événements par les fréquences ci-dessus. Ainsi, Prob(«nombre < 4») $=$ Prob(«1») $+$ Prob(«2») $+$ Prob(«3») $= 0,165 + 0,169 + 0,166 = 0.5$. ♦

Remarque : insistons sur le fait que les fréquences ne nous fournissent qu'une approximation des probabilités, et que cette approximation n'est bonne que pour n «très grand».

3.5 Probabilités sur un univers continu

On verra précisément dans la section 6 comment les choses se passent lorsque l'univers est continu. Toutefois, on peut déjà en donner une esquisse. Repartons du cas où l'univers est fini, prenons l'exemple 4 concernant les salaires de 40 ingénieurs : à partir de données recueillies dans la population, on a déterminé des classes de salaire, leur fréquence ainsi que l'histogramme ci-dessous.

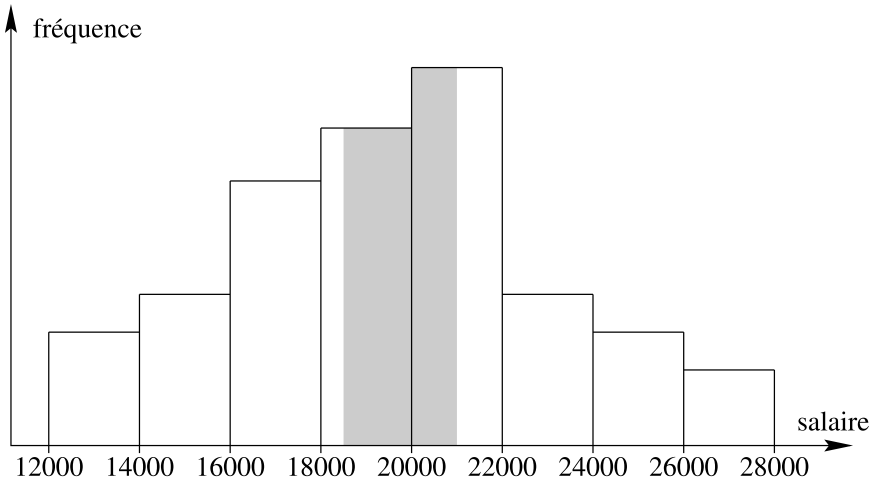


FIG. 11: Histogramme des salaires.

D'après la section 3.4, la probabilité d'appartenir à une classe devrait correspondre au nombre de cas favorables / nombre d'événements dans l'univers, c'est-à-dire à la fréquence de la classe. Or, nous avons vu page 12 que, dans un histogramme, les fréquences correspondent aux surfaces des rectangles. Par analogie, si l'on veut calculer la probabilité que le salaire soit compris entre 18500F et 21000F, on va calculer la surface grisée sur la figure 11.

On peut tout de même se dire que c'est une technique assez grossière : l'univers est sensé être continu, et on ne dessine que des rectangles. Si l'on affinait un peu les données (en supposant que la population est aussi grande que l'on veut), on pourrait peut-être séparer l'univers en plus de classes. Multiplions par 2 le nombre de classes et voyons ce que nous obtenons :

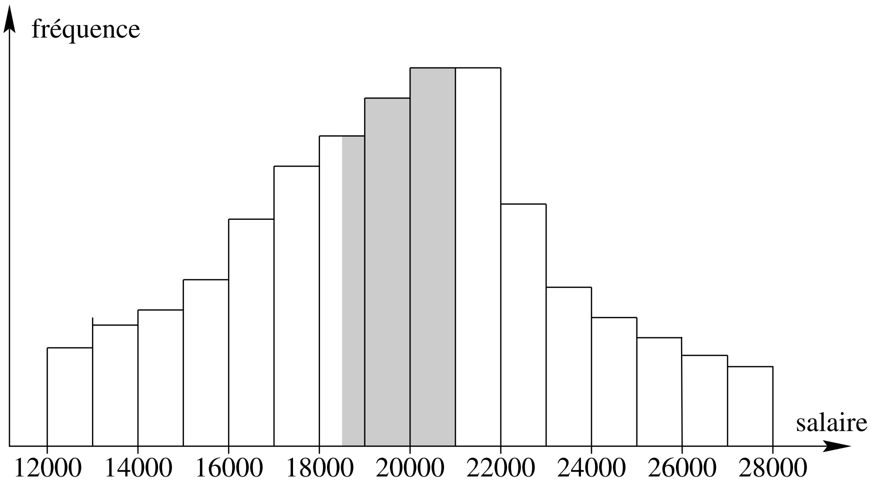


FIG. 12: Multiplication du nombre de classes.

Le calcul de Prob(salaire entre 18500F et 21000F) s'effectue de la même manière. Cependant, on voit que la surface est calculée à partir d'un polygone dont les contours sont beaucoup plus précis que dans la figure 11. En multipliant par deux le nombre de classes une infinité de fois, on s'aperçoit que les rectangles finissent par former une sorte de courbe assez lisse (voir figure ci-dessous). La probabilité d'avoir un salaire entre 18500F et 21000F se calcule par analogie en prenant la surface délimitée par le morceau de courbe entre 18500F et 21000F. En termes plus mathématiques, on calcule l'intégrale entre 18500F et 21000F de la fonction dont la courbe est représentée sur la figure 13. Cette fonction s'appelle une $\boxed{\text{fonction de densité}}$.

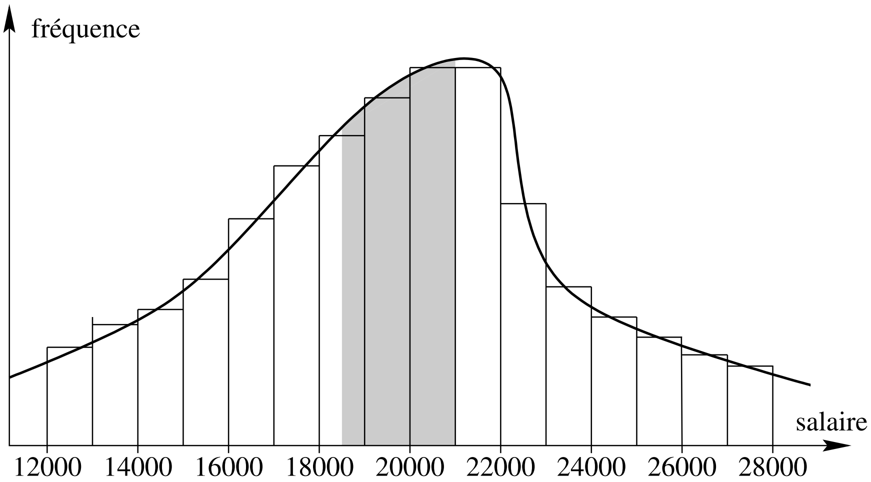


FIG. 13: Fonction de densité.

En généralisant, voilà comment on définit et comment on utilise les probabilités sur un univers continu : on commence par définir une fonction de densité, qui nous donne la répartition (on dit aussi la <u>distribution</u>) des événements. Ensuite, lorsque l'on veut calculer Prob(A), il suffit de calculer la surface délimitée par la fonction de densité dans la zone où les événements sont inclus dans A.

Ceci nous amène aux remarques suivantes :

- Lorsqu'une variable est continue, la probabilité d'un événement élémentaire est forcément nulle. En effet, un événement élémentaire ne doit être le résultat que d'une seule expérience. Sur la figure 13, un événement élémentaire correspond ainsi à un point sur l'axe des abscisses, et donc la surface à calculer est forcément nulle. Exemple : calculer la probabilité qu'un salaire soit égal exactement à 20000F revient à calculer la surface de la fonction de densité entre 20000F et 20000F. C'est donc bien une surface nulle.
- La surface délimitée par la fonction de densité sur tout l'univers Ω est forcément égale à 1 (puisqu'elle correspond à $\text{Prob}(\Omega)$). On avait bien évidemment la même propriété pour les histogrammes : la somme des surfaces des rectangles est égale à 1 car elle correspond à la somme des fréquences.
- Les valeurs prises par la fonction de densité sont forcément positives ou nulles partout. Autrement, en calculant les surfaces par des intégrales, on pourrait obtenir des surfaces négatives, et donc des probabilités négatives.

3.6 Probabilités conditionnelles et indépendance

La notion de probabilité conditionnelle est l'une des fonctionnalités les plus importantes et les plus simples des probas :

> **Définition 20 :** *la probabilité d'un événement A conditionnellement à un événement B, que l'on note* $\text{Prob}(A|B)$*, est la probabilité que A se produise sachant que B s'est ou va se produire.*

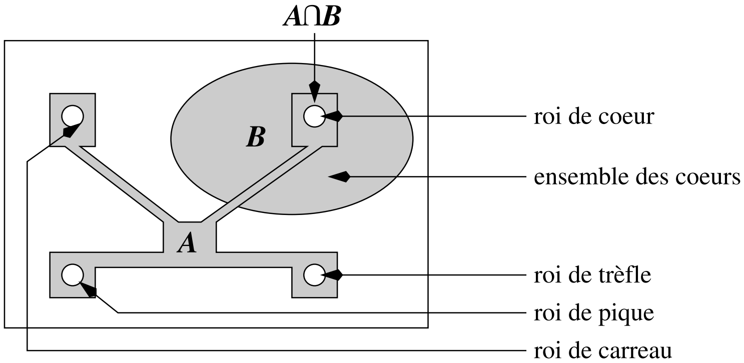
Exemple 13 : On tire au hasard deux cartes d'un jeu qui en comporte 32. Problème : quelle est la probabilité de tirer deux rois ? Si l'on s'en réfère à la section 3.3, Prob(2 rois) = nombre de cas favorables / nombre d'événements possibles dans l'univers. Le dénominateur est simple à calculer : il y a 32 possiblités de tirer la première carte. Lorsque celle-ci est tirée, il ne reste que 31 cartes, et on tire une carte parmi celles-ci. Donc le dénominateur est égal à $32 \times 31 = 992$. Calculons maintenant le numérateur : au départ, il y a 4 rois dans le jeu et on tire un de ces rois. Il y a donc 4 possibilités. Ensuite, parmi les cartes restantes, il ne reste que 3 rois. Donc le numérateur est égal à $4 \times 3 = 12$. Par conséquent, Prob(2 rois) $= 12/992 = 3/248$.

On a donc 3 chances sur 248 d'obtenir 2 rois lorsque l'on tire au hasard deux cartes. Mais supposons que la première carte que l'on ait tirée soit un roi, que peut-on dire sur la probabilité d'avoir encore un roi au deuxième tirage ? Le raisonnement du paragraphe précédent nous dit alors que le nombre de cas favorables est 3 et que le nombre d'événements possibles dans l'univers est 31. Donc la probabilité de tirer un deuxième roi conditionellement au fait que la première carte tirée était un roi, c'est-à-dire Prob($2^{\text{ème}}$ roi | première carte est un roi) est égale à $3/31$. ♦

Voyons comment calculer $\text{Prob}(A|B)$. D'abord, remarquons que $\text{Prob}(A|\Omega) = \text{Prob}(A)$. En effet, $\text{Prob}(A|\Omega)$ représente la probabilité que A soit réalisé conditionnellement au fait qu'au moins un événement élémentaire sera réalisé. Or on sait que cette condition est toujours vérifiée puisque $\text{Prob}(\Omega) = 1$. Donc conditionner par Ω ne diminue en rien ou n'augmente en rien les chances que A a de se réaliser. Donc $\text{Prob}(A|\Omega)$ doit nécessairement être égal à $\text{Prob}(A)$. L'exemple suivant va nous permettre de comprendre comment $\text{Prob}(A|B)$ se calcule pour un événement B quelconque :

Exemple 14 : Tirons au hasard une carte parmi un jeu de 32 cartes. $\Omega = \{32 \text{ cartes}\}$. Considérons les événements A = tirer un roi et B = tirer un cœur. On veut calculer $\text{Prob}(A|B)$, c'est-à-dire la probabilité de tirer un roi sachant que l'on a tiré un cœur.

Si A se produit, sachant que B s'est aussi produit, cela signifie que $A \cap B$ se produit. Autrement dit, $\text{Prob}(A|B) = \text{Prob}(A \cap B|B)$, ou encore la probabilité de tirer un roi sachant que l'on a tiré un cœur est égale à la probabilité de tirer le roi de cœur sachant que l'on a tiré un cœur. Cela se voit bien sur la figure ci-après : on cherche la probabilité que l'événement A (c'est-à-dire l'un des quatre cercles) soit réalisé, sachant que l'événement B l'est (c'est-à-dire que l'on est dans l'ellipse). Il n'y a donc qu'un seul cercle qui peut être réalisé, et celui-ci correspond à $A \cap B$.



Cette petite manipulation est très intéressante. En effet, on cherche la probabilité que le cercle $A \cap B$, qui est inclus dans B, soit réalisé, sachant que l'événement B est réalisé, c'est-à-dire que $\mathrm{Prob}(B) = 1$. Cela suggère de remplacer l'univers Ω par l'univers B, et de calculer dans cet univers $\mathrm{Prob}(A \cap B|B) = \mathrm{Prob}(A \cap B)$. La section 3.3 nous dit alors que :

$$\mathrm{Prob}(A \cap B) = \frac{|A \cap B|}{|B|}.$$

On peut alors en déduire :

$$\mathrm{Prob}(A|B) = \frac{|A \cap B|}{|\Omega|} \times \frac{|\Omega|}{|B|} = \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(B)}.$$

♦

Il s'avère que la formule obtenue ci-dessus est toujours vérifiée :

Théorème 2 : *Soient A et B deux événements. Si* $\mathrm{Prob}(B) > 0$, *alors :*

$$\mathrm{Prob}(A|B) = \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(B)}. \tag{2}$$

Pourquoi imposer $\mathrm{Prob}(B) > 0$? Outre le fait que diviser $\mathrm{Prob}(A \cap B)$ par zéro risque de donner un résultat pour le moins curieux, il y a aussi une question sémantique : $\mathrm{Prob}(A|B)$ correspond à la probabilité que l'événement A soit réalisé sachant que B l'est. Or, justement, si $\mathrm{Prob}(B) = 0$, B ne peut être réalisé, d'où un petit problème au niveau de la signification du $\mathrm{Prob}(A|B)$ dans un tel cas.

L'équation (2) paraît toute bête mais elle est globalement à la base de tous les travaux actuels sur les probabilités. Par exemple, elle permet de démontrer le théorème de Bayes :

Théorème 3 (Bayes) : *Soient A et B deux événements tels que* $\mathrm{Prob}(A) \neq 0$ *et* $\mathrm{Prob}(B) \neq 0$. *Alors :*

$$\mathrm{Prob}(B|A) = \mathrm{Prob}(A|B) \times \frac{\mathrm{Prob}(B)}{\mathrm{Prob}(A)}. \tag{3}$$

Démonstration : Si $\mathrm{Prob}(A) \neq 0$ et $\mathrm{Prob}(B) \neq 0$ alors :

$$\mathrm{Prob}(B|A) = \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(A)} = \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(B)} \times \frac{\mathrm{Prob}(B)}{\mathrm{Prob}(A)} = \mathrm{Prob}(A|B) \times \frac{\mathrm{Prob}(B)}{\mathrm{Prob}(A)}.$$

♦

FIG. 14: Le révérend Thomas Bayes (1702 – 1761).

Le théorème de Bayes nous sera utile dans les tests d'hypothèses, section 9. En effet, si A représente l'observation d'un événement lors d'une expérience et B une hypothèse, on a alors :

$$\text{Prob}(\text{hypothèse}|\text{observation}) = \text{Prob}(\text{observation}|\text{hypothèse}) \times \frac{\text{Prob}(\text{hypothèse})}{\text{Prob}(\text{observation})}.$$

Cette équation stipule que la croyance que l'on a en une hypothèse H (ou encore la probabilité de l'hypothèse) après avoir observé un événement e peut se calculer grâce à la croyance que l'on avait en H avant toute observation (Prob(H)), la probabilité que e se réalise, et la probabilité que l'on observe e sachant que H s'est réalisée. On peut abandonner Prob(e) en remarquant que :

$$\text{Prob}(e) = \text{Prob}(e|H)\text{Prob}(H) + \text{Prob}(e|\overline{H})\text{Prob}(\overline{H}).$$

Démonstration : $\text{Prob}(e|H)\text{Prob}(H) + \text{Prob}(e|\overline{H})\text{Prob}(\overline{H}) = \text{Prob}(e,H) + \text{Prob}(e,\overline{H}) = \text{Prob}(e).$ ♦

Autrement dit, la probabilité qu'une hypothèse H soit vraie sachant qu'on observe e peut se calculer à partir de la probabilité que l'hypothèse H soit vraie quand on n'a aucune information, de la probabilité d'obtenir e quand on sait que H est vraie, et de la probabilité d'obtenir e quand on sait que H est fausse.

Exemple 15 : Vous êtes réveillés la nuit par la délicate sonnerie du signal d'alarme de votre maison. Problème : quel degré de confiance devez-vous accorder dans la possibilité d'une effraction de votre domicile ?

Vous avez mené votre enquête auprès du fabricant du signal d'alarme et, d'après les statistiques, il ressort que Prob(alarme|effraction) $= 0,95$. Autrement dit, le signal a 95% de chances de se déclencher lors d'une tentative d'effraction. De plus, le fabricant vous a assuré qu'en moyenne le signal ne se déclenche pour rien qu'une fois sur cent. Traduisez $\text{Prob}(\text{alarme}|\overline{\text{effraction}}) = 0,01$. Enfin, le commissariat de votre quartier vous a communiqué des chiffres sur la délinquance, qui laissent apparaître que Prob(effraction) $= 10^{-4}$: vous avez une chance sur 10000 d'être cambriolés.

Vous êtes donc réveillés la nuit et vous vous interrogez sur Prob(effraction|alarme). D'après Bayes,

$$\text{Prob}(\text{effraction}|\text{alarme}) = \text{Prob}(\text{alarme}|\text{effraction}) \times \frac{\text{Prob}(\text{effraction})}{\text{Prob}(\text{alarme})}.$$

Or, il est facile de calculer Prob(alarme) :

$$\begin{aligned}\text{Prob}(\text{alarme}) &= \text{Prob}(\text{alarme}|\text{effraction})\text{Prob}(\text{effraction}) + \text{Prob}(\text{alarme}|\overline{\text{effraction}})\text{Prob}(\overline{\text{effraction}})\\ &= 0,95 \times 10^{-4} + 0,01 \times (1 - 10^{-4}) = 0,010094.\end{aligned}$$

Par conséquent,

$$\text{Prob}(\text{effraction}|\text{alarme}) = 0,95 \times \frac{10^{-4}}{0,010094} \approx 0,00941.$$

Donc la probabilité qu'il y ait effectivement un cambriolage en cours dans votre maison est seulement de 0,9%. Ce chiffre peut paraître surprenant au premier abord, mais en fait, il se justifie assez bien : vous avez une chance sur 10000 d'être cambriolés, on peut donc raisonnablement penser que c'est une défaillance de votre signal d'alarme qui a déclenché la sonnerie. Conclusion : rendormez-vous. ♦

Exemple 16 : On a 28 dominos. Chaque domino contient 2 chiffres numérotés de 0 à 6. Quelle est la probabilité de sélectionner un couple de dominos compatibles, c'est-à-dire ayant au moins un chiffre en commun, lorsque l'on prend deux dominos au hasard ?

Ici, l'univers Ω est égal à l'ensemble de tous les couples de dominos que l'on peut choisir. Soient deux dominos, notons les X et Y. D'après le théorème 2 sur la page 30,

$$\mathrm{Prob}(X, Y) = \mathrm{Prob}(X) \times P(Y|X).$$

Autrement dit, la probabilité d'avoir choisi le couple de dominos (X, Y) est égale à la probabilité d'avoir choisi le domino X parmi l'ensemble des dominos, multipliée par la probabilité de choisir le domino Y sachant que l'on a déjà choisi X. Puisqu'il y a 28 dominos, $\mathrm{Prob}(X) = \frac{1}{28}$. Lorsque l'on a choisi X, il ne reste plus que 27 dominos. Donc $\mathrm{Prob}(Y|X) = \frac{1}{27}$. Par conséquent, $\mathrm{Prob}(X, Y) = \frac{1}{28\times 27} = \frac{1}{756}$, ou encore $|\Omega| = 756$.

Soit E l'événement «on a choisi des dominos compatibles». On a déjà vu que

$$\mathrm{Prob}(E) = \frac{|E|}{|\Omega|}.$$

Il ne nous reste donc plus qu'à calculer $|E|$. $E = E_1 \cup E_2 \cup E_3$, où E_1 = «le domino X possède deux nombres identiques», E_2 = «le domino Y possède deux nombres identiques» et E_3 = «aucun des deux dominos n'a deux nombres identiques». Bien entendu, $E_1 \cap E_2 = \emptyset$, $E_1 \cap E_3 = \emptyset$ et $E_2 \cap E_3 = \emptyset$. Donc

$$\mathrm{Prob}(E) = \frac{|E_1| + |E_2| + |E_3|}{|\Omega|}.$$

Calculons $|E_1|$. Les numéros sur les dominos vont de 0 à 6. Il y a donc 7 possibilités de choisir un domino ayant deux numéros identiques. Lorsque X est choisi, il faut, pour Y, choisir un domino compatible, c'est-à-dire ayant un chiffre en commun. Il ne reste donc plus que 6 possibilités parmi les 27 dominos restants. Autrement dit, $\mathrm{Prob}(Y|X) = \frac{6}{27}$. Par conséquent, $|E_1| = 7 \times 6 = 42$. Par symétrie, $|E_2| = 42$.

Calculons maintenant $|E_3|$. Il y a 7 dominos ayant deux nombres identiques. Par conséquent, il y a $28-7 = 21$ dominos n'ayant pas deux nombres identiques. Donc $\mathrm{Prob}(X) = \frac{21}{28}$. Sachant que X a été tiré, il ne reste que 27 dominos. Parmi ceux-là, il n'y en a que 5 qui soient compatibles avec le premier numéro de X sans avoir pour autant deux numéros identiques. De même, il y en a 5 compatibles avec le deuxième numéro de X mais n'ayant pas deux numéros identiques. Par conséquent, $\mathrm{Prob}(Y|X) = \frac{5+5}{27} = \frac{10}{27}$. Donc $|E_3| = 21 \times 10 = 210$. Ainsi,

$$\mathrm{Prob}(E) = \frac{|E_1| + |E_2| + |E_3|}{|\Omega|} = \frac{42 + 42 + 210}{756} = \frac{294}{756} = \frac{7}{18}.$$

Il y a donc 7 chances sur 18 pour que deux dominos choisis au hasard soient compatibles. Attention : ici, les dominos sont choisis de manière ordonnée, c'est-à-dire que choisir (X, Y) est différent de choisir (Y, X). ♦

Exemple 17 : Un cable vidéo peut avoir à chacune de ses extrémités une prise VGA mâle, VGA femelle, S-vidéo mâle, S-vidéo femelle, Péritel mâle, Péritel femelle, composite mâle, et composite femelle. Les deux extrémités peuvent avoir des prises identiques ou bien différentes. Par exemple, les cables suivants sont valides : cable VGA mâle-composite mâle, cable Péritel femelle-Péritel mâle, cable Péritel femelle-Péritel femelle, etc. Un panier renferme un exemplaire et un seul de tous les cables possibles (un cable Péritel femelle-Péritel mâle est supposé identique au cable Péritel mâle-Péritel femelle). On veut savoir la probabilité qu'en choisissant deux cables au hasard dans le panier on puisse les raccorder entre eux (autrement dit, que l'un ait une prise mâle et l'autre une prise femelle du même type).

Soit Ω l'ensemble des cables possibles. Numérotons les prises de 1 à 8. Pour déterminer $|\Omega|$, afin d'éviter de compter deux fois le même cable, on remarque que si l'on a un cable (x, y), où x (resp. y) représente la première prise (resp. la deuxième), alors (y, x) représente le même cable. Par conséquent, dans notre comptage de Ω, il suffit de ne compter que les cables pour lesquels $y \geq x$. Pour un x donné, les valeurs de y possibles sont donc $x, x+1, \ldots, 8$. Par conséquent,

$$|\Omega| = \sum_{x=1}^{8}(9-x) = \sum_{x=1}^{8} 9 - \sum_{x=1}^{8} x = 9 \times 8 - \frac{9 \times 8}{2} = 36.$$

Soit E l'ensemble des couples de cables répondant à la question. Soit E_1 le sous-ensemble de E tel que les deux cables ont les mêmes prises à leurs extrémités. Soit E_2 le sous-ensemble de E tel que seulement l'un des cables a les mêmes prises à ses deux extrémités. Enfin, soit E_3 le sous-ensemble tel qu'aucun des deux cables n'a les mêmes prises à ses extrémités. Bien évidemment, $|E| = |E_1| + |E_2| + |E_3|$. Or $|E_1| = 4$ car, si les deux cables ont leurs extrémités identiques, une fois qu'on a sélectionné le cable mâle, le deuxième cable est déterminé de manière unique. Puisqu'il y a 4 prises mâles, $|E_1| = 4$. Pour calculer $|E_2|$, commençons par choisir le cable ayant ses deux extrémités identiques : il y a 8 choix possibles. Pour le deuxième cable, l'une des extrémités est déjà déterminée pour correspondre avec le premier cable. Pour la deuxième extrémité, il reste 7 choix. Donc $|E_2| = 8 \times 7 = 56$. Calculons maintenant $|E_3|$: celui-ci se partitionne en deux sous-ensembles : E_{31} et E_{32} qui représentent respectivement les cables qui peuvent être raccordés seulement par un bout, et les cables qui peuvent être raccordés simultanément par les deux bouts. Dans E_{31}, on commence par choisir la prise mâle de raccordement : il y a 4 choix. Pour le premier cable, on peut encore choisir 7 prises pour la deuxième extrémité. Lorsque c'est fait, une des extrémités du deuxième cable est choisie (une femelle pour se raccorder à la prise mâle) et il y a 6 possibilités pour la deuxième extrémité. Donc $|E_{31}| = 4 \times 7 \times 6 = 168$. Il nous reste maintenant les prises se raccordant par les deux bouts, E_{32}. Bon, là, il y a une petite subtilité : il faut considérer le cas où un cable a pour extrémité une prise mâle et à l'autre extrémité la même prise, mais femelle, dans ce cas, on ne peut pas trouver d'autre cable qui puisse se raccorder. C'est pourquoi, pour calculer E_{32}, nous allons commencer par choisir un cable dont les deux extrémités sont des prises mâles différentes : il y a 6 possibilités. Le deuxième cable est alors déterminé de manière unique. Si, par contre le premier cable a une prise mâle (4 cas possibles) et une prise femelle (3 cas possibles d'après ce qui précède), le deuxième cable est déterminé encore de manière unique mais on compte ces couples de cables deux fois. Donc $|E_{32}| = 6 + 4 \times 3/2 = 12$. Par conséquent, $|E| = 4 + 56 + 168 + 12 = 240$.

Il nous faut maintenant calculer le nombre de couples de cables possibles. On a 36 choix possibles pour le premier cable. Lorsqu'on a choisi celui-ci, il reste 35 cables possibles. Il y a donc $35 \times 36 = 1260$ paires ordonnées de cables. Pour éliminer la notion d'ordre, on divise par 2. Par conséquent, il y a 630 couples de cables. Conclusion : la probabilité que l'on cherchait est $240/630 \approx 0,38$. ♦

Les probabilités conditionnelles permettent, entre autres choses, de définir la notion d'indépendance probabiliste :

Définition 21 : *Deux événements A et B sont <u>indépendants</u> si* $\text{Prob}(A \cap B) = \text{Prob}(A) \times \text{Prob}(B)$.

Grâce à cette définition, on peut déduire assez aisément que si A et B sont des événements tels que $\text{Prob}(B) \neq 0$, alors $\text{Prob}(A|B) = \text{Prob}(A)$.

Démonstration : Si $\text{Prob}(B) \neq 0$, alors $\text{Prob}(A|B) = \text{Prob}(A \cap B)/\text{Prob}(B)$. De plus, puisque les événements A et B sont indépendants, $\text{Prob}(A \cap B) = \text{Prob}(A) \times \text{Prob}(B)$. Donc $\text{Prob}(A|B) = \dfrac{\text{Prob}(A) \times \text{Prob}(B)}{\text{Prob}(B)} = \text{Prob}(A)$. ♦

Cette propriété nous montre que lorsque deux événements sont indépendants, alors savoir qu'un des deux événements est réalisé (ici B) n'apporte aucune information sur la réalisation de l'autre événement. C'est en ce sens que l'on peut dire que les deux événements sont indépendants. On peut aussi dire que la réalisation d'un événement A ne dépend pas de la réalisation de l'autre événement.

Cette propriété, quoiqu'anodine au premier abord, revêt une importance primordiale en informatique : c'est elle qui permet de stocker en mémoire vive des probabilités sur des univers de tailles importantes.

Exemple 18 : Soit un univers représentant le jet de 100 dés à 6 faces (indépendants les uns des autres). Les événements élémentaires correspondent à toutes les combinaisons possibles des dés. Il y a donc 6^{100} possibilités, soit environ $6,53 \times 10^{77}$ possibilités.

Quand on sait que pour effectuer des calculs probabilistes, il faut absolument connaître les probabilités des événements élémentaires, on peut se dire qu'aucun ordinateur actuel n'a suffisamment de mémoire pour calculer des probabilités sur notre univers de 100 dés. En fait, l'utilisation de la définition 21 nous indique qu'il suffit de rentrer pour chaque dé les probabilités que chacune des faces sorte, pour que l'on ait toutes les informations pour retrouver les probabilités des événements élémentaires de l'univers. Par exemple, si $\text{Prob}_i(j)$ représente la probabilité que le i$^{\text{ème}}$ dé tombe sur la face j, et si $\text{Prob}(1, 1, \ldots, 1)$ représente la probabilité de l'événement élémentaire «tous les dés sont retombés sur la face 1», alors :

$$\text{Prob}(1, 1, \ldots, 1) = \prod_{i=1}^{100} \text{Prob}_i(1).$$

Retrouver les probabilités des événements de l'univers demande un peu de calcul, certes, mais cela permet de manipuler des probabilités sur des univers qui sont *a priori* trop gros pour être manipulés sur ordinateur. ♦

3.7 Introduction aux variables aléatoires

Jusqu'à maintenant, nous avons vu comment l'on pouvait définir l'univers Ω sur lequel portent les probabilités et comment, à partir des événements élémentaires (en nombre supposé fini), l'on pouvait calculer la probabilité de n'importe quel événement de Ω. Cependant, en pratique, ce sont rarement les événements eux-mêmes qui nous intéressent, mais plutôt les conséquences de ces événements.

Exemple 19 : Supposons que vous ayez 1000F à votre disposition. Jouons au jeu suivant : je tire 3 cartes parmi un jeu de 32 cartes et, suivant le tirage, je vous donne ou vous prend de l'argent :

- si 3 rois ont été tirés, vous gagnez 2000F ;
- si 2 rois ont été tirés, vous gagnez 500F ;
- si 1 roi a été tiré, vous ne gagnez ni ne perdez rien ;
- si aucun roi n'a été tiré, vous perdez vos 1000F.

En fait, les événements de ce jeu correspondent à l'ensemble des triplets de cartes que l'on peut tirer. Maintenant, songez à ce qui vous intéresse dans ce jeu. Sont-ce les 3 cartes tirées ou plutôt le gain qu'elles vont vous rapporter ? Il y a fort à parier que la deuxième option remporte votre intérêt. Dans ce cas, ce ne sont pas les événements qui vous intéressent, mais plutôt leurs conséquences (économiques). ♦

Il faudrait donc disposer d'un outil qui nous permette de travailler directement sur ces gains, et non plus sur les triplets de cartes. Cet outil répond au doux nom de *variable aléatoire*.

Définition 22 : *Soit Ω un univers muni d'une loi de probabilité* $\text{Prob}(\cdot)$, *et soit Ω' un autre ensemble. Notons $\mathcal{P}(\Omega)$ et $\mathcal{P}(\Omega')$ l'ensemble des sous-ensembles de Ω et de Ω'. Une variable aléatoire est une fonction Γ de $\mathcal{P}(\Omega)$ dans $\mathcal{P}(\Omega')$ telle que :*

$$\Gamma^{-1}(A') \in \mathcal{P}(\Omega) \quad \forall A' \in \mathcal{P}(\Omega').$$

On peut alors définir une loi de probabilité $\text{Prob}'(\cdot)$ *sur Ω' de la façon suivante :*

$$\text{Prob}'(A') = \text{Prob}\left(\Gamma^{-1}(A')\right) \quad \forall A' \in \mathcal{P}(\Omega').$$

Exemple 19 (suite) : Dans ce jeu, Ω est relativement aisé à déterminer : il s'agit de l'ensemble de tous les triplets de cartes que l'on peut tirer du jeu. Mathématiquement, nous pourrions noter cela de la manière suivante :

$$\Omega = \{(x, y, z) : y \neq x,\ z \neq x, y, \text{ et } x, y, z \in \{\text{jeu de 32 cartes}\}\}.$$

Ce qui nous intéresse dans cet exemple, ce sont les gains ou pertes d'argent occasionnés par le jeu. On a donc :

$$\Omega' = \{2000\text{F}, 500\text{F}, 0\text{F}, \text{-}1000\text{F}\}.$$

La fonction $\Gamma(\cdot)$ qui nous permet de passer de Ω à Ω' est elle-aussi relativement simple à définir :

$$\Gamma(A) = \begin{cases} \text{2000F} & \text{si } A = \text{trois rois,} \\ \text{500F} & \text{si } A \supset \text{exactement deux rois,} \\ \text{0F} & \text{si } A \supset \text{exactement un roi,} \\ \text{-1000F} & \text{si } A \cap \text{rois} = \emptyset. \end{cases}$$

Ainsi, il est relativement facile de savoir si l'on va gagner de l'argent ou en perdre en jouant à ce jeu :

$$\begin{aligned}
\text{Prob}'(\text{2000F}) &= \text{Prob(tirer 3 rois)} = \tfrac{4}{32} \times \tfrac{3}{31} \times \tfrac{2}{30} = \tfrac{1}{1240} \\
\text{Prob}'(\text{500F}) &= \text{Prob(tirer 2 rois)} \\
&= \text{Prob(roi,roi,autre)} + \text{Prob(roi,autre,roi)} + \text{Prob(autre,roi,roi)} \\
&= \tfrac{4}{32} \times \tfrac{3}{31} \times \tfrac{28}{30} \quad + \quad \tfrac{4}{32} \times \tfrac{28}{31} \times \tfrac{3}{30} \quad + \quad \tfrac{28}{32} \times \tfrac{4}{31} \times \tfrac{3}{30} = \tfrac{42}{1240} \\
\text{Prob}'(\text{0F}) &= \text{Prob(tirer 1 roi)} \\
&= \text{Prob(roi,autre,autre)} + \text{Prob(autre,roi,autre)} + \text{Prob(autre,autre,roi)} \\
&= \tfrac{4}{32} \times \tfrac{28}{31} \times \tfrac{27}{30} \quad + \quad \tfrac{28}{32} \times \tfrac{4}{31} \times \tfrac{27}{30} \quad + \quad \tfrac{28}{32} \times \tfrac{27}{31} \times \tfrac{4}{30} = \tfrac{378}{1240} \\
\text{Prob}'(\text{-1000F}) &= \text{Prob(tirer aucun roi)} \\
&= \tfrac{28}{32} \times \tfrac{27}{31} \times \tfrac{26}{30} = \tfrac{819}{1240}
\end{aligned}$$

♦

Résumons : Nous partons d'un univers Ω et, grâce à une fonction $\Gamma : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega')$, nous pouvons travailler dans un autre univers Ω'. Il est bien entendu possible de définir simultanément plusieurs variables aléatoires $\Gamma : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega')$, $\Phi : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega'')$, $\Psi : \mathcal{P}(\Omega') \to \mathcal{P}(\Omega'')$... Par définition,

$$\begin{aligned}
\text{Prob}'(A') &= \text{Prob}\left(\Gamma^{-1}(A')\right) & \forall A' \in \mathcal{P}(\Omega'), \\
&= \text{Prob}\left(A : A \in \Omega \text{ et } \Gamma(A) = A'\right) & \forall A' \in \mathcal{P}(\Omega'), \\
\text{Prob}''(A'') &= \text{Prob}\left(\Phi^{-1}(A'')\right) & \forall A'' \in \mathcal{P}(\Omega''), \\
&= \text{Prob}\left(A : A \in \Omega \text{ et } \Phi(A) = A''\right) & \forall A'' \in \mathcal{P}(\Omega''), \\
\text{Prob}''(A'') &= \text{Prob}'\left(\Psi^{-1}(A'')\right) & \forall A'' \in \mathcal{P}(\Omega''), \\
&= \text{Prob}'\left(A' : A' \in \Omega' \text{ et } \Psi(A') = A''\right) & \forall A'' \in \mathcal{P}(\Omega'').
\end{aligned}$$

Lorsque l'on travaille avec plusieurs variables aléatoires, cette notation devient vite fastidieuse car : 1) chaque univers doit avoir sa propre loi de probabilité $\text{Prob}(\cdot)$, $\text{Prob}'(\cdot)$, $\text{Prob}''(\cdot)$; 2) dans le calcul de $\text{Prob}''(A'')$, il faut impérativement préciser la variable aléatoire qu'on utilise (Γ, Φ, Ψ) ainsi que la loi de probabilité dans l'univers de départ de la variable aléatoire (cf. les lignes 3 à 6 parmi les égalités ci-dessus).

> Afin d'éviter toutes ces complications relativement inutiles, on écrit par abus de notation $\text{Prob}(\Gamma = A')$ à la place de $\text{Prob}'(A')$ et de $\text{Prob}\left(\Gamma^{-1}(A')\right)$.

Ainsi, dans l'exemple 19, pourrait-on écrire :

$$\begin{aligned}
\text{Prob}(\Gamma = 2000F) &= \frac{1}{1240}, \\
\text{Prob}(\Gamma = 500F) &= \frac{42}{1240}, \\
\text{Prob}(\Gamma = 0F) &= \frac{378}{1240}, \\
\text{Prob}(\Gamma = -1000F) &= \frac{819}{1240}.
\end{aligned}$$

Exemple 20 : La roulette est incontestablement l'un des jeux les plus connus au casino. Mais existe-t-il une stratégie assurant une chance raisonnable de gagner à ce jeu ? Si l'on n'est pas trop «gourmand», la réponse est oui.

Dans ce jeu, les joueurs peuvent parier sur le numéro (inférieur à 36) sur lequel la boule va s'arrêter, mais ils peuvent aussi miser sur la couleur de ce numéro (Noir ou Rouge), sur le fait qu'il est Pair ou

Impair, ou encore s'il est Passe ou Manque (supérieur ou inférieur à 18). Ces six dernières possibilités ont chacune 18 chances sur 37 de s'avérer vraies et permettent de remporter une fois sa mise. On peut donc dire que l'univers Ω du jeu auquel nous allons jouer est :

$$\begin{aligned}\Omega = \{&(\text{Noir,Pair,Passe}), (\text{Noir,Pair,Manque}), (\text{Noir,Impair,Passe}), (\text{Noir,Impair,Manque}),\\ &(\text{Rouge,Pair,Passe}), (\text{Rouge,Pair,Manque}), (\text{Rouge,Impair,Passe}), (\text{Rouge,Impair,Manque})\}.\end{aligned}$$

De plus, comme indiqué ci-dessus, on a :

$$\begin{aligned}\text{Prob(Noir)} &= \frac{18}{37} & \text{Prob(Rouge)} &= \frac{18}{37},\\ \text{Prob(Pair)} &= \frac{18}{37} & \text{Prob(Impair)} &= \frac{18}{37},\\ \text{Prob(Passe)} &= \frac{18}{37} & \text{Prob(Manque)} &= \frac{18}{37}.\end{aligned}$$

Notons à ce propos que, d'après Kolmogoroff, on a aussi Prob(Noir) = Prob(Noir,Pair,Passe) + Prob(Noir,Pair,Manque) + Prob(Noir,Impair,Passe) + Prob(Noir,Impair,Manque), mais que les probabilités des événements élémentaires comme Prob(Noir,Pair,Passe) ne sont pas égales, comme on pourrait le croire au premier abord, à $\left(\frac{18}{37}\right)^3$. En effet, une fois que l'on sait que la boule est tombée sur le noir, par exemple, on sait que le numéro «0» n'est pas sorti, et donc il y a une chance sur deux pour que le numéro sortant soit pair ou impair, passe ou manque. Donc Prob(Noir,Pair,Passe) $= \frac{18}{37} \times \frac{1}{2} \times \frac{1}{2} = \frac{9}{74}$.

Maintenant, ce qui nous intéresse n'est pas vraiment le fait que la bille sortante soit noire, rouge, paire, impaire, passe ou manque, mais plutôt l'argent qu'elle va nous rapporter ou nous faire perdre. On aurait donc intérêt à modéliser notre jeu grâce à une variable aléatoire associant aux billes sortantes les gains qu'elles procurent. Supposons donc que l'on possède au départ M francs et que l'on décide de jouer une séquence de k mises, $S_k = (\{m_1, c_1\}, \{m_2, c_2\}, \ldots, \{m_k, c_k\})$. Par exemple, $\{m_1, c_1\}$ pourrait consister à miser m_1 francs sur c_1 = rouge. On peut alors créer la variable aléatoire $\Gamma(M, S_k)$, qui représente l'état de nos finances après une séquence S_k de mises. Ainsi, si l'on note s_i l'état (dans Ω) de la bille sortante,

$$\Gamma(M, \{(m_1, c_1)\}) = \begin{cases} M + m_1 \text{ si } s_1 \in c_1, \\ M - m_1 \text{ si } s_1 \notin c_1. \end{cases}$$

$$\Gamma(M, \{(m_1, c_1), (m_2, c_2)\}) = \begin{cases} M + m_1 + m_2 \text{ si } s_1 \in c_1 \text{ et } s_2 \in c_2, \\ M - m_1 + m_2 \text{ si } s_1 \notin c_1 \text{ et } s_2 \in c_2, \\ M + m_1 - m_2 \text{ si } s_1 \in c_1 \text{ et } s_2 \notin c_2, \\ M - m_1 - m_2 \text{ si } s_1 \notin c_1 \text{ et } s_2 \notin c_2. \end{cases}$$

Par conséquent, comme les résultats des différentes billes sont indépendants les uns des autres, on a :

$$\text{Prob}(\Gamma(M, \{(m_1, c_1)\}) = M + m_1) = \frac{18}{37} \qquad \text{Prob}(\Gamma(M, \{(m_1, c_1)\}) = M - m_1) = \frac{19}{37}$$

$$\text{Prob}(\Gamma(M, \{(m_1, c_1), (m_2, c_2)\}) = M + m_1 + m_2) = \frac{18}{37} \times \frac{18}{37}$$

$$\text{Prob}(\Gamma(M, \{(m_1, c_1), (m_2, c_2)\}) = M - m_1 + m_2) = \frac{19}{37} \times \frac{18}{37}$$

$$\text{Prob}(\Gamma(M, \{(m_1, c_1), (m_2, c_2)\}) = M + m_1 - m_2) = \frac{18}{37} \times \frac{19}{37}$$

$$\text{Prob}(\Gamma(M, \{(m_1, c_1), (m_2, c_2)\}) = M - m_1 - m_2) = \frac{19}{37} \times \frac{19}{37}.$$

Par récurrence, il est facile de montrer que

$$\text{Prob}\left(\Gamma(M, \{m_1, c_1\}, \{m_2, c_2\}, \ldots, \{m_k, c_k\}) = M - \sum_{i=1}^{k} m_k\right) = \left(\frac{19}{37}\right)^k.$$

Autrement dit, il y a $\left(\frac{19}{37}\right)^k$ chances de définir une séquence S_k de mises et de perdre à chaque fois. Par exemple, il y a $\frac{19}{37}^5 \approx 0,0357$ chances que 5 lancés de billes successifs tombent, le premier sur un numéro noir, le deuxième sur passe, le troisième sur rouge, le quatrième sur rouge et le cinquième sur impair et que l'on ait misé successivement sur rouge, puis sur manque, puis sur noir, noir, pair. On voit donc que l'on perd tout son argent lorsque la séquence opposée à la mise, c'est-à-dire (noir, passe, rouge, rouge, impair) sort. Or nous venons de voir que cette séquence n'a que 3,57% de chances de sortir. Donc, même avec une séquence de 5 mises, on a vraiment très peu de chances de **tout** perdre.

Ainsi, si l'on peut miser k fois, quelque soit la séquence choisie, on perd son argent seulement si la séquence opposée sort, c'est-à-dire dans $\left(\frac{19}{37}\right)^k$ des cas. Donc, si l'on veut s'assurer p chances de succès, il suffit de trouver k tel que $1-p = \left(\frac{19}{37}\right)^k$. Autrement dit,

$$k = \left\lceil \frac{\ln(1-p)}{\ln 19 - \ln 37} \right\rceil .$$

Par exemple, si l'on veut se garantir 99% de succès, il faut pouvoir assurer $\lceil 6,91 \rceil = 7$ mises. Pour se garantir 999 chances de succès sur 1000, il faut $\lceil 10,36 \rceil = 11$ mises.

Ce que nous venons de voir nous garantit de ne pas rentrer ruinés du casino, mais ne nous garantit absolument pas que nous n'allons pas perdre de l'argent, c'est-à-dire revenir du casino avec moins de M francs. Voyons maintenant comment miser. Supposons que notre stratégie nous indique la séquence de mises suivante : $S_k = (\{m_1, c_1\}, \{m_2, c_2\}, \ldots, \{m_k, c_k\})$. Si l'on gagne sur la première boule, on remporte m_1 francs. Si l'on perd la première mise, mais que l'on gagne sur la deuxième, on gagne globalement $m_2 - m_1$ francs. Si l'on a perdu les deux premières mises, mais que l'on gagne la troisième, on a globalement gagné $m_3 - m_2 - m_1$. Par récurrence, il est facile de montrer que si l'on a perdu les $i-1$ premières mises, mais que l'on gagne sur la $i^{\text{ème}}$mise, on remporte globalement $m_i - \sum_{j<i} m_j$. Si l'on veut s'assurer que, lorsqu'on gagne sur une mise, globalement cela rapporte de l'argent, il faut donc que :

$$m_i \geq \sum_{j<i} m_j \qquad \text{pour tout } i.$$

Au pire, par récurrence, il suffit de miser $m_i = i \times m_1$. Mais cela ne nous rapportera de l'argent que si l'on gagne à l'étape 1. Essayons maintenant de gagner la même somme d'argent à chaque étape. Pour cela, il suffit de résoudre l'équation :

$$m_i - \sum_{j<i} m_j = m_1 \qquad \text{pour tout } i.$$

Par récurrence sur i, on montre que $m_i = 2^i m_1$.

Conclusion : si l'on veut être assuré de gagner m_1 francs avec une probabilité p de succès, il suffit de choisir une séquence arbitraire de longueur $k = \left\lceil \frac{\ln(1-p)}{\ln 19 - \ln 37} \right\rceil$ de Pair/Impair/Noir/Rouge/Passe/Manque, et de miser $m_i = 2^i m_1$ francs sur la $i^{\text{ème}}$boule. Par exemple, si l'on veut se garantir un gain 50 francs avec une probabilité de succès de 99%, il faut constituer la séquence de mises : (50F, 100F, 200F, 400F, 800F, 1600F, 3200F). Il faut donc avoir un capital de départ de 6350F. Par récurrence, on peut montrer que le capital de départ doit être de $M = (2^{k+1} - 1)m_1$ francs. ♦

3.8 Exercices

Exercice 1 Prenons un jeu de 32 cartes. Soit une expérience consistant à tirer aléatoirement 3 cartes au hasard. Quelle est la probabilité de l'événement «obtenir un roi + une dame + un valet de pique» ?

Exercice 2 Dans une entreprise, 150 personnes ont passé un test de personnalité. Voici les résultats obtenus :

	type A	type B
hommes	78	42
femmes	19	11

Quelle est la probabilité qu'un individu tiré au hasard

1. présente une personnalité de type A ;
2. soit une femme ;
3. soit un homme de type A ;
4. soit de type B sachant que c'est une femme ;
5. soit un homme ou soit de type B.

Exercice 3 À Saint-Michel, un voyageur peut utiliser soit le RER B, soit le RER C pour se rendre à Massy-Palaiseau. Ce voyageur est extrêmement pressé et prend toujours le premier train en partance pour Massy. Quel que soit le RER, la fréquence de passage des trains est toujours d'exactement 10 minutes. Expliquez pourquoi le voyageur prend 90% du temps le RER C.

Exercice 4 Les mots de passe sous Unix peuvent comporter de 1 à 8 caractères (choisis dans un jeu de 127 caractères). Quelle est la probabilité de trouver un mot de passe donné en 1 seul essai ?

Exercice 5 Il était une fois, dans un pays lointain, trois prisonniers A, B, C, qui étaient tous trois jugés pour un meurtre. Devant la cruauté du meurtre, les jurés décidèrent que la personne reconnue coupable serait exécutée et que les deux autres seraient relachées. En attendant que les jurés rendent leur verdict, les prisonniers furent conduits dans trois cellules séparées. Pendant le procès, il était difficile de savoir ce que les jurés pensaient des trois accusés, aussi chaque prisonnier avait-il la même chance d'être reconnu coupable.

Le jugement fut rendu, mais les prisonniers ne furent pas informés tout de suite de leur sort. Seul le geôlier le fut. Dans l'éventualité où il serait reconnu coupable, A écrivit une lettre pour sa femme. Il demanda alors au geôlier de donner cette lettre à l'un des prisonniers B ou C qui serait relâché. Celui-ci donna la lettre à B. Quelle est maintenant la probabilité que A soit exécuté ? Quelle est celle de C ?

Exercice 6 Un QCM contient 20 questions. Chaque question contient 3 réponses, dont une seule est correcte. Quelle est la probabilité d'obtenir la moyenne à ce QCM si l'on considère les règles de calcul suivantes :

1. les réponses correctes valent 1 point ; les réponses incorrectes et les questions non répondues valent 0 point.
2. les réponses correctes valent 1 point ; les questions non répondues valent 0 point ; les réponses incorrectes valent -1 point.

4 Calcul pratique des probabilités

Dans de nombreuses situations, des décideurs — que ce soient des chefs d'entreprise, des hommes politiques ou tout simplement des ingénieurs — doivent prendre des décisions dans des contextes incertains. Par exemple, lorsque vous allez chez le médecin, ce dernier doit diagnostiquer la maladie dont vous souffrez alors même qu'il lui manque un certain nombre de données pour le faire (les mêmes symptômes peuvent avoir des causes différentes). C'est pourquoi la gestion d'incertitudes a reçu de nombreuses contributions de la part de la communauté scientifique.

Il n'existe pas qu'une seule manière de gérer les incertitudes. On peut citer par exemple la logique floue, les fonctions de croyance (une extension des probabilités), les facteurs de certitude, etc. Dans les années 70-80, la mode était aux systèmes experts. Leur idée fondamentale était de stipuler des règles logiques du type «*Si telle chose est vraie alors telle autre chose est vraie*», par exemple «*Si fièvre ET douleurs ET toux alors grippe*». Mais assez rapidement, on s'est aperçu que ce type de règles manquait de souplesse. Par exemple, il n'est pas abherrant d'avoir dans son système les règles «tous les oiseaux volent», «les pingouins sont des oiseaux». Une inférence logique nous apprend donc que les pingouins volent. Pour avoir donc une base cohérente, il faut énormément de règles, en particulier pour gérer toutes les exceptions. Or cela s'avère assez délicat à réaliser en pratique. C'est pourquoi, actuellement, beaucoup de chercheurs en informatique travaillent sur des systèmes experts probabilistes. L'idée des systèmes experts probabilistes est de formuler les mêmes règles mais en leur donnant une probabilité d'être vraie : «*Si fièvre ET douleurs ET toux alors on a une probabilité de 90% d'avoir la grippe*». Ainsi, la plupart du temps, on peut inférer une grippe, mais on se laisse quand même une marge d'erreur.

Qui dit système expert probabiliste dit calcul de probabilité. Dans la section précédente, nous avons effectué «à la main» quelques calculs probabilistes simples. Cependant, en pratique, les espaces de probabilité sur lesquels on doit travailler sont extrêmement complexes et nécessitent des algorithmes de calcul adaptés. La méthode la plus en vogue actuellement dans la communauté «Intelligence Artificielle» pour les réaliser s'appelle un réseau bayésien. Dans la littérature, on l'appelle aussi réseau probabiliste, ou encore réseau de croyance bayésien.

Ce type de méthode est utilisée, en autres, pour le dépistage des problèmes d'imprimantes réseaux par HP, pour aider les utilisateurs à trouver le plus rapidement possible les informations qui les intéressent dans l'aide d'Office 98, et par la NASA pour diagnostiquer les pannes du système de propulsion de la navette spatiale. Pour en comprendre l'intérêt, nous allons examiner le problème auquel était confronté la NASA.

4.1 La NASA découvre l'intérêt des probabilités

Lors des débuts de la conquête spatiale, moulte fusées se sont écrasées à cause de problèmes de propulsion. C'était bien évidemment embêtant, mais, bon, sans trop de gravité : le contribuable américain pouvait se permettre de gaspiller quelques millions de dollars. Mais arriva un moment où la NASA voulut envoyer des hommes dans l'espace et, un peu plus tard, où elle décida de créer un engin qui, contrairement aux fusées conventionnelles, pourrait effectuer plusieurs vols spatiaux : la navette spatiale. Et là, un problème de propulsion signifiait non seulement une perte de milliards de dollars, mais aussi une perte humaine, ce qui n'était plus du tout admissible.

Aussi la NASA réfléchit-elle à un moyen de gérer *rapidement* tous les incidents des propulseurs de la navette. Impossible de laisser des experts traiter ces problèmes en temp réel : les temps de réaction humains sont beaucoup trop lents. Aussi la NASA décida-t-elle de confier cette tâche à un système informatique. Celui-ci devait prendre en entrée les données fournies par divers capteurs installés en divers endroits du système de propulsion, et retourner en sortie la ou les pannes qui s'étaient produites. Or, cela s'avéra relativement ardu car nombreuses sont les pannes qui ont à peu près les mêmes conséquences. Plutôt qu'un système expert conventionnel, qui n'aurait pas été capable de différencier les multiples conséquences possibles, il était plus intéressant d'avoir un outil donnant les probabilités que l'ensemble des pannes apparaissent.

L'idée est donc de travailler sur un univers Ω représentant toutes les valeurs possibles des données fournies par les capteurs. On introduit alors des variables aléatoires $X_1, \ldots, X_n$ représentant les pannes. Comme nous avons vu dans la section 3, calculer la probabilité qu'une variable aléatoire X_i prenne telle ou telle valeur x_i (on parle aussi de modalité) revient à calculer la probabilité d'apparition des événements de Ω engendrant

$X_i = x_i$. Bien entendu, pour pouvoir faire tous ces calculs, il suffit de connaître les probabilités des événements élémentaires sur Ω (d'après la définition 17).

Or c'est là que tout se complique. En effet, le nombre d'événements élémentaires de Ω peut être prohibitif. Pour le comprendre, prenons un exemple simple : je veux utiliser des probabilités pour décrire l'incertitude sur le résultat des jets de 2 dés à 6 faces. Dans ce cas, Ω est égal à l'ensemble de tous les couples (i, j), où i représente le résultat de premier dé, et j le résultat du second. Il faut bien comprendre que ce qui arrive après le jet des 2 dés doit forcément pouvoir être décrit à l'aide **d'un seul** événement élémentaire. Donc décrire l'incertitude sur le résultat des jets de 2 dés requiert la connaissance des probabilités de 36 événements élémentaires. Par récurrence, si l'on veut décrire l'incertitude sur le résultat des jets de 100 dés, il nous faut des probabilités de $6^{100} \approx 6,53\ 10^{77}$ événements élémentaires. Autrement dit, on ne peut pas stocker ces probabilités élémentaires sur ordinateur car elles demandent une capacité mémoire gigantesque, et pourtant le problème ne paraissait pas trop complexe à l'origine.

Revenons au problème de la NASA. La première partie du travail à réaliser consistait à déterminer tous les éléments de Ω pour le problème de navette. Et là, il fallut bien se rendre à l'évidence qu'il y en avait beaucoup trop pour les déterminer tous. En effet, les pannes suivantes peuvent se produire simultanément sur le système de propulsion : fuites des réservoirs d'oxygène, d'hélium, d'azote, de fuel, problèmes de 2 valves d'admission, de pressions dans 4 réservoirs, de trop pleins... Chaque panne énoncée possède un capteur qui lui est spécifique. Autrement dit, Ω devait être constitué de l'ensemble des n-uplets de toutes les pannes possibles. Supposons que ces pannes soient évaluées sur une échelle de 0 à 9 selon leur gravité. Dans ce cas, Ω doit avoir 10^{11} =100 milliards d'éléments. Non seulement la consommation mémoire était prohibitive, mais encore les calculs risquaient de s'éterniser. Pour que le système informatique soit viable, il fallait donc impérativement trouver un outil permettant de limiter les probabilités à fournir au modèle de gestion des incertitudes : celui-ci fut un réseau bayésien.

4.2 La NASA découvre les réseaux bayésiens

Pour expliquer comment les réseaux bayésiens permettent de réduire la consommation mémoire, reprenons l'exemple du jet des 2 dés. Selon Kolmogoroff, $\Omega = \{(x, y), 1 \leq x, y \leq 6\}$ a 36 éléments, et on doit affecter à chacun de ces éléments une probabilité. Appelons X la variable aléatoire représentant le résultat du premier jet et Y celui du deuxième jet. X et Y peuvent prendre chacun 6 valeurs. Remarquez que les résultats des deux jets de dés ne dépendent pas l'un de l'autre. Dans ce cas, on dit que les variables aléatoires X et Y sont indépendantes. Elles vérifient alors la propriété suivante :

Définition 23 (Indépendance (marginale) de variables aléatoires) : *Deux variables aléatoires X et Y sont* ***indépendantes*** *si* $\text{Prob}(X = x, Y = y) = \text{Prob}(X = x) \times \text{Prob}(Y = y)\ \forall x, y$.

Autrement dit, il est inutile de stocker les probabilités des 36 événements élémentaires, il suffit de stocker les probabilités suivantes :

$$\begin{array}{cccccc} \text{Prob}(X=1) & \text{Prob}(X=2) & \text{Prob}(X=3) & \text{Prob}(X=4) & \text{Prob}(X=5) & \text{Prob}(X=6) \\ \text{Prob}(Y=1) & \text{Prob}(Y=2) & \text{Prob}(Y=3) & \text{Prob}(Y=4) & \text{Prob}(Y=5) & \text{Prob}(Y=6) \end{array}$$

et d'utiliser la formule de la définition 23 pour recalculer toutes les probabilités $\text{Prob}(X = x, Y = y)$ qui vous intéressent. Ainsi, même le problème du résultat des jets des 100 dés peut être aisément stocké sur ordinateur (il suffit de stocker 6×100 probabilités).

Oui, mais dans les situations pratiques, et dans le problème de la navette en particulier, les variables aléatoires auxquelles on s'intéresse sont rarement indépendantes. Par exemple, les variables aléatoires «fuites des réservoirs d'oxygène» et «pressions dans les réservoirs» ont peu de chance d'être indépendantes. C'est pourquoi une propriété moins forte a été introduite par les mathématiciens : l'indépendance conditionnelle, qui s'appuie sur la notion de probabilité conditionnelle que nous avions vue page 29.

Définition 24 (indépendance conditionnelle) : *Soient X_1, X_2, X_3 des ensembles de variables aléatoires. X_1 et X_2 sont* ***indépendantes conditionnellement*** *à X_3 si les propriétés suivantes (qui sont équivalentes) sont vérifiées :*

1. $\text{Prob}(X_1 = x_1 | X_2 = x_2, X_3 = x_3) \ = \ \text{Prob}(X_1 = x_1 | X_3 = x_3) \ \ \forall x_1, x_2, x_3,$
2. $\text{Prob}(X_1 = x_1, X_2 = x_2 | X_3 = x_3) \ = \ \text{Prob}(X_1 = x_1 | X_3 = x_3) \times \text{Prob}(X_2 = x_2 | X_3 = x_3) \ \ \forall x_1, x_2, x_3.$

La barre verticale sépare les variables conditionnées des variables conditionnantes. Ainsi $\text{Prob}(X_1 = x_1 | X_2 = x_2, X_3 = x_3) = \text{Prob}(X_1 = x_1 | (X_2 = x_2, X_3 = x_3))$ *et* $\text{Prob}(X_1 = x_1, X_2 = x_2 | X_3 = x_3) = \text{Prob}((X_1 = x_1, X_2 = x_2) | X_3 = x_3)$.

Traduction de la propriété 1 : $\text{Prob}(X_1 = x_1 | X_3 = x_3)$ représente le nombre de chances que X_1 prenne la valeur x_1 sachant l'information que X_3 a pris la valeur x_3. $\text{Prob}(X_1 = x_1 | X_2 = x_2, X_3 = x_3)$ représente le nombre de chances que X_1 prenne la valeur x_1 connaissant les informations X_3 a pris la valeur x_3 et X_2 a pris la valeur x_2. Autrement dit, la propriété 1 traduit simplement le fait que la connaissance de la valeur prise par X_2 n'apporte aucune information sur la valeur de X_1 quand on connaît déjà la valeur de X_3.
La propriété 2 traduit, elle, le fait que, connaissant la valeur de X_3, X_1 et X_2 sont indépendantes.
Notez que deux variables X_1, X_2 peuvent très bien être indépendantes conditionnellement à une troisième, et ne pas être indépendantes marginalement. Exemple : en général, tousser provoque des maux de tête. Donc X_1 = «tousser» et X_2 = «mal de tête» ne sont pas indépendantes marginalement. Par contre, sachant que vous avez attrapé la grippe (X_3), les variables X_1 et X_2 deviennent indépendantes car toux et céphalées sont deux symptômes du virus, qui peuvent apparaître séparément.
Autre illustration de l'indépendance conditionnelle : certains immeubles sont dotés d'alarmes incendie. Vous admettrez sans trop de difficulté qu'il existe un lien entre le déclenchement de ces alarmes et le fait qu'il y ait réellement un incendie dans l'immeuble. Donc les variables aléatoires «incendie» et «alarme» ne sont pas indépendantes marginalement. De même, vous conviendrez que les variables «incendie» et «fumée» ne sont pas non plus indépendantes, de même que «fumée» et «alarme» puisqu'en principe ces dernières se déclenchent lorsqu'il y a un certain taux de fumée dans l'atmosphère. Par contre, à partir du moment où l'appareil a détecté de la fumée, il se déclenche, qu'il y ait ou non un incendie. Donc «incendie» et «alarme» sont indépendants conditionnellement à la variable «fumée».
Maintenant, quel est l'intérêt de l'indépendance conditionnelle ? Eh bien, elle permet de réduire la place nécessaire au stockage des probabilités. En effet, si $X_1, \ldots, X_n$ sont des variables aléatoires, alors on peut démontrer aisément par récurrence grâce au théorème 2 ($\text{Prob}(A, B) = \text{Prob}(A|B) \times \text{Prob}(B)$) que :

$$\text{Prob}(X_1, \ldots, X_n) = \text{Prob}(X_1) \times \text{Prob}(X_2|X_1) \times \text{Prob}(X_3|X_1, X_2) \times \cdots \times \text{Prob}(X_n|X_1, \ldots, X_{n-1}). \quad (4)$$

L'équation ci-dessus devient particulièrement intéressante lorsqu'on l'utilise conjointement avec la propriété 1 de la définition 24 : si $\{i_1, \ldots, i_k\}$ et $\{i_{k+1}, \ldots, i_{j-1}\}$ forment une partition de $\{1, \ldots, j-1\}$ et si X_j est indépendant de $X_{i_{k+1}}, \ldots, X_{i_{j-1}}$ conditionnellement à $X_{i_1}, \ldots, X_{i_k}$, alors $\text{Prob}(X_j|X_1, \ldots, X_{j-1}) = \text{Prob}(X_j|X_{i_1}, \ldots, X_{i_k})$. L'équation (4) peut alors se simplifier énormément, et, d'une manière générale, le nombre de probabilités conditionnelles à rentrer dans l'ordinateur (et à stocker en mémoire) est souvent très inférieur à celui qu'on aurait dû rentrer si l'on avait utilisé les événements élémentaires. Cela permet de traiter des problèmes qui, *a priori*, devraient être impossibles à traiter.
L'idée des réseaux bayésiens consiste donc :

1. à recenser toutes les variables aléatoires. Prenons un exemple : la dyspnée, une maladie respiratoire, peut être engendrée par une tuberculose, un cancer des poumons, une bronchite, par plusieurs de ces maladies, ou bien par aucune. Un séjour récent en Asie augmente les chances de tuberculose, tandis que fumer augmente les risques de cancer des poumons. Un patient se rend à l'hopital parce qu'il éprouve des difficultés à respirer. Dans quelle mesure peut-on dire qu'il est atteint de dyspnée ? Les variables aléatoires de ce problème sont donc «dyspnée» (le patient est-il atteint de dyspnée ? oui/non), «tuberculose» (a-t-il la tuberculose ? oui/non), «cancer des poumons» (oui/non), «bronchite» (oui/non), «séjour en Asie» (oui/non), «fumer» (oui/non).
2. à choisir un ordre pour ces variables. Prenons par exemple X_1 = «séjour en Asie», X_2 = «fumer», X_3 = «tuberculose», X_4 = «cancer des poumons», X_5 = «bronchite», X_6 = «dyspnée». En fait, on peut ranger

les variables dans n'importe quel ordre. Cependant, certains ordres sont plus avantageux que d'autres parce qu'ils vont permettre plus de simplifications des probabilités conditionnelles. En règle générale, on obtient plus de simplifications lorsque l'on place les variables correspondant à des causes avant celles correspondant aux conséquences (par exemple, on a intérêt à placer «fumer» avant «cancer des poumons»).

3. à simplifier les probabilités conditionnelles. On ne peut guère simplifier $\text{Prob}(X_1)$. *A priori*, on peut considérer que le fait d'avoir séjourné en Asie est indépendant du fait que l'on fume (quoique...). Donc $\text{Prob}(X_2|X_1) = \text{Prob}(X_2)$. Il n'y a aucun lien entre le fait de fumer et celui d'avoir la tuberculose ; par contre il y en a un avec un éventuel séjour en Asie. Donc $\text{Prob}(X_3|X_1, X_2) = \text{Prob}(X_3|X_1)$. Un cancer du poumon n'a pas de rapport avec un séjour en Asie, ni avec la tuberculose. Donc $\text{Prob}(X_4|X_1, X_2, X_3) = \text{Prob}(X_4|X_2)$.

   De même avec la bronchite, et comme «fumer» ne provoque pas de bronchite, $\text{Prob}(X_5|X_1, X_2, X_3, X_4) = \text{Prob}(X_5)$. Enfin la dyspnée n'est provoquée que par la tuberculose, la bronchite et le cancer des poumons, donc $\text{Prob}(X_6|X_1, X_2, X_3, X_4, X_5) = \text{Prob}(X_6|X_3, X_4, X_5)$. On peut objecter que X_6 devrait aussi dépendre de X_2 puisque fumer augmente les risques de cancer du poumon. Mais sachant si l'on a (ou non) ce cancer, le fait de fumer ne modifie pas les chances d'avoir la dyspnée. Donc, connaissant X_3, X_4, et X_5, X_6 est indépendante de X_1 et de X_2. Ainsi, l'équation (4) se réduit-elle à :

$$\begin{aligned}\text{Prob}(X_1, X_2, X_3, X_4, X_5, X_6) = {} & \text{Prob}(X_1) \times \text{Prob}(X_2) \times \text{Prob}(X_3|X_1) \times \text{Prob}(X_4|X_2) \\ & \times \text{Prob}(X_5) \times \text{Prob}(X_6|X_3, X_4, X_5).\end{aligned} \tag{5}$$

   Notez que si l'on avait saisi les probabilités des événements élémentaires comme le suggérait la définition de Kolmogoroff, on aurait dû saisir $2^6 = 64$ nombres, tandis qu'avec l'équation (5), on a besoin de saisir uniquement $2 + 2 + 4 + 4 + 2 + 16 = 30$ nombres.

4. à utiliser les formules de probabilités ci-dessus pour calculer les probabilités cherchées.

Afin de mieux visualiser les simplifications des probabilités conditionnelles décrites dans l'équation (5), on construit un graphe dans lequel les noeuds représentent les variables aléatoires et dans lequel on crée des arcs allant des variables conditionnantes vers les variables conditionnées : $\text{Prob}(X_6|X_3, X_4, X_5)$ sera ainsi représentée grâce aux arcs (X_3, X_6), (X_4, X_6) et (X_5, X_6). On obtient alors le graphe de la figure 15.

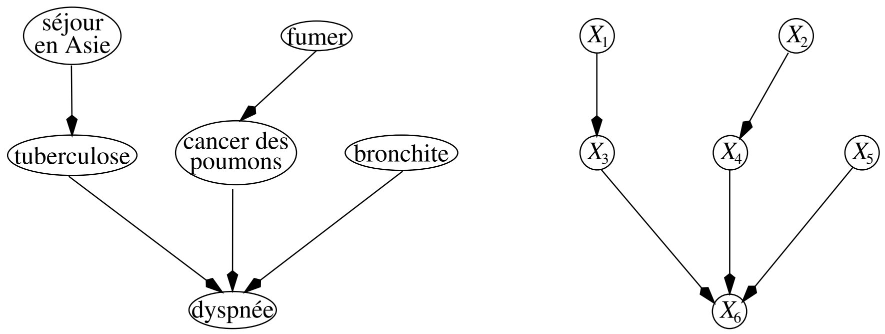


FIG. 15 – Le réseau bayésien de la dyspnée.

Conclusion : un réseau bayésien est constitué d'une part d'un graphe tel que celui ci-dessus et, d'autre part d'un ensemble de probabilités conditionnelles. Il est d'usage de considérer que chaque noeud stocke sa probabilité conditionnellement à ses parents.

À titre d'exemple, voici sur la page suivante le réseau utilisé par la NASA.

4.3 Comment calculer une probabilité avec un réseau bayésien ?

Il existe deux types de probabilités calculables par réseaux bayésiens, à savoir les probabilités *a priori* et les probabilités *a posteriori*. Les premières représentent les probabilités de chacune des variables aléatoires lorsque l'on n'a aucune information particulière à sa disposition. Ce sont par exemple les probabilités qu'un

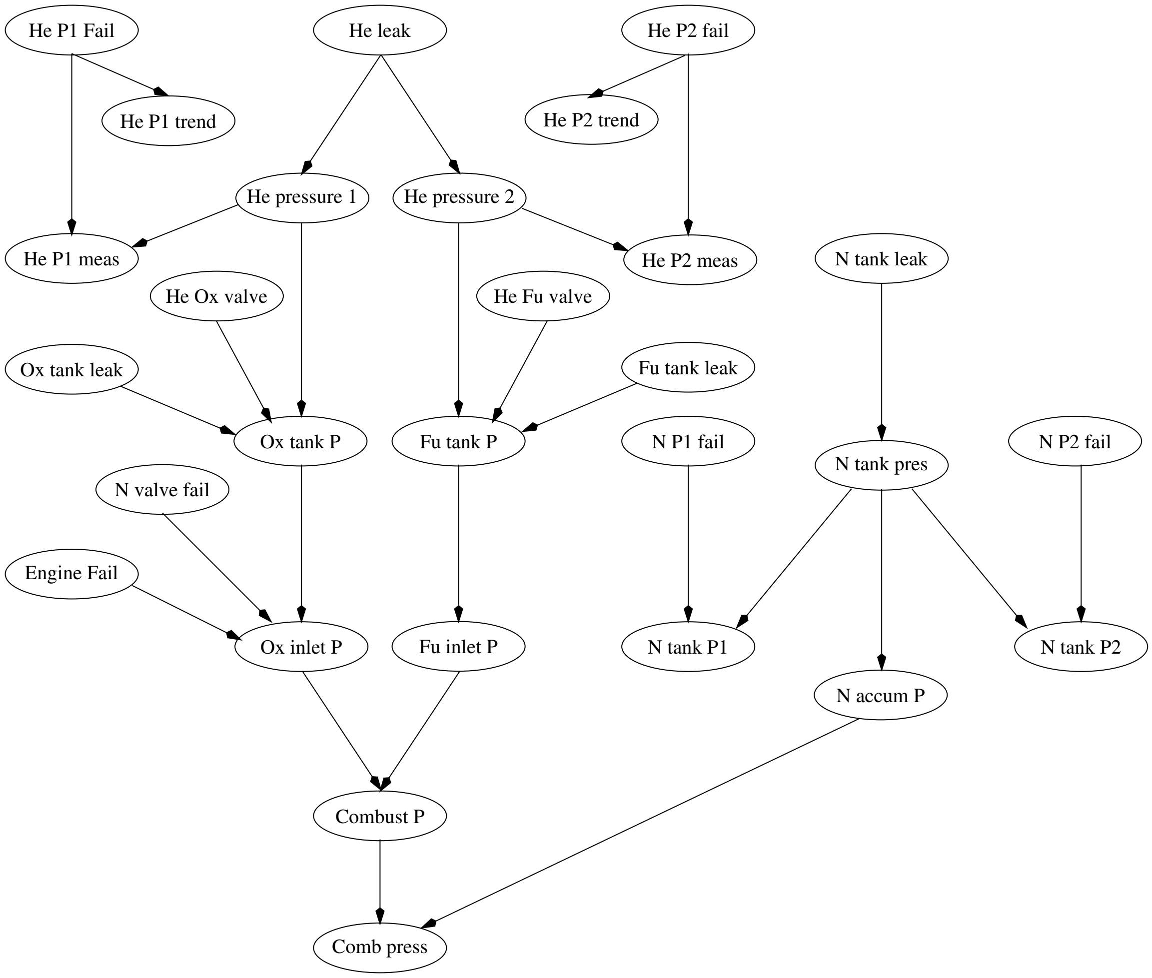


FIG. 16 – Le réseau bayésien de la NASA.

dé tombe sur la face trois, qu'un individu quelconque soit atteint de dyspnée, ou encore qu'il y ait une fuite dans le réservoir d'oxygène de la navette spatiale. Les probabilités *a posteriori* sont les probabilités des variables aléatoires sachant certaines informations que l'on n'avait pas lorsque l'on a réalisé l'étape 3 de l'algorithme ci-dessus. Ce sont par exemple la probabilité qu'un patient ait une dyspnée sachant qu'il fume et qu'il a voyagé en Asie, qu'une des valves d'admission ait un problème sachant que la pression du réservoir d'oxygène baisse et qu'il y a une fuite dans ce même réservoir. Autrement dit, les probabilités *a priori* sont des probabilités *dans l'absolu* et les probabilités *a posteriori* sont des probabilités après avoir reçu de nouvelles observations/informations.

Pour donner une idée des calculs, nous allons voir en détail le calcul des probabilités *a priori* et *a posteriori*.

4.3.1 Calcul des probabilités *a priori*

Les probabilités marginales *a priori*, que l'on va calculer dans cette sous-section, sont les $\text{Prob}(X_i)$ de l'exemple de la dyspnée. $\text{Prob}(X_1)$, $\text{Prob}(X_2)$ et $\text{Prob}(X_5)$ ont déjà été saisies dans l'ordinateur, donc pas besoin de calcul pour les déterminer. En généralisant, on peut dire que cela correspond aux probabilités des noeuds sans parents dans le graphe (les noeuds orphelins). Rappelons que dans les noeuds sont stockées les probabilités de ces mêmes noeuds conditionnellement à leurs parents. En appliquant la propriété 3 de la définition 17,

$$\text{Prob}(X_3 = x_3) = \sum_{i=1}^{|X_1|} \text{Prob}(X_3 = x_3, X_1 = x_1^i),$$

et, en appliquant le théorème 2, on obtient :

$$\text{Prob}(X_3 = x_3) = \sum_{i=1}^{|X_1|} \text{Prob}(X_3 = x_3 | X_1 = x_1^i) \times \text{Prob}(X_1 = x_1^i).$$

De la même manière,

$$\text{Prob}(X_4 = x_4) = \sum_{i=1}^{|X_2|} \text{Prob}(X_4 = x_4 | X_2 = x_2^i) \times \text{Prob}(X_2 = x_2^i).$$

Enfin, pour calculer $\text{Prob}(X_6)$, on remarque que :

$$\begin{aligned}\text{Prob}(X_6 = x_6) &= \sum_{i=1}^{|X_3|}\sum_{j=1}^{|X_4|}\sum_{k=1}^{|X_5|} \text{Prob}(X_6 = x_6, X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k),\\ &= \sum_{i=1}^{|X_3|}\sum_{j=1}^{|X_4|}\sum_{k=1}^{|X_5|} \text{Prob}(X_6 = x_6 | X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k) \times \text{Prob}(X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k).\end{aligned} \tag{6}$$

Il reste maintenant à calculer $\text{Prob}(X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k)$. Pour cela, on reprend la propriété 3 de la définition 17 :

$$\begin{aligned}\text{Prob}(X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k) &= \sum_{a=1}^{|X_1|}\sum_{b=1}^{|X_2|}\sum_{c=1}^{|X_6|} \text{Prob}(X_1 = x_1^a, X_2 = x_2^b, X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k, X_6 = x_6^c),\\ &= \sum_{a=1}^{|X_1|}\sum_{b=1}^{|X_2|}\sum_{c=1}^{|X_6|} \text{Prob}(X_1 = x_1^a) \times \text{Prob}(X_2 = x_2^b) \times \text{Prob}(X_3 = x_3^i | X_1 = x_1^a)\\ &\qquad \times \text{Prob}(X_4 = x_4^j | X_2 = x_2^b) \times \text{Prob}(X_5 = x_5^k)\\ &\qquad \times \text{Prob}(X_6 = x_6^c | X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k)\end{aligned}$$

et on lui applique l'équation (5) :

$$\begin{aligned}\text{Prob}(X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k) &= \left[\sum_{a=1}^{|X_1|} \text{Prob}(X_3 = x_3^i | X_1 = x_1^a) \times \text{Prob}(X_1 = x_1^a)\right]\\ &\quad \times \left[\sum_{b=1}^{|X_2|} \text{Prob}(X_4 = x_4^j | X_2 = x_2^b) \times \text{Prob}(X_2 = x_2^b)\right]\\ &\quad \times \left[\sum_{c=1}^{|X_6|} \text{Prob}(X_6 = x_6^c | X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k)\right] \times \text{Prob}(X_5 = x_5^k)\\ &= \text{Prob}(X_3 = x_3^i) \times \text{Prob}(X_4 = x_4^j) \times \text{Prob}(X_5 = x_5^k).\end{aligned}$$

Par conséquent, d'après ce qui précède et l'équation (6), on peut calculer $\text{Prob}(X_6 = x_6)$ de la manière suivante :

$$\begin{aligned}\text{Prob}(X_6 = x_6) = \sum_{i=1}^{|X_3|}\sum_{j=1}^{|X_4|}\sum_{k=1}^{|X_5|} &\text{Prob}(X_6 = x_6 | X_3 = x_3^i, X_4 = x_4^j, X_5 = x_5^k) \times \text{Prob}(X_3 = x_3^i)\\ &\times \text{Prob}(X_4 = x_4^j) \times \text{Prob}(X_5 = x_5^k).\end{aligned}$$

Les calculs que nous venons de mener nous permettent d'en déduire l'algorithme de calcul des probabilités a priori pour des graphes quelconques[5] :

[5] En fait, si l'on pousse les calculs jusqu'au bout, cet algorithme ne fonctionne que pour des graphes **sans cycles**, c'est-à-dire dans lesquels il n'existe pas d'ensemble de noeuds $\{X_{i_1}, \ldots, X_{i_k}\}$ tel que le graphe contienne soit l'arc (X_{i_1}, X_{i_k}) soit l'arc (X_{i_k}, X_{i_1}), et, pour tout $j < k$, soit $(X_{i_j}, X_{i_{j+1}})$ soit l'arc $(X_{i_{j+1}}, X_{i_j})$.

Les probabilités des noeuds orphelins (sans parents dans le graphe) sont déjà calculées. Pour tous les noeuds X_i dont tous les parents $Y_1, \ldots, Y_n$ ont été calculés, $\mathrm{Prob}(X_i = x_i)$ se calcule de la manière suivante :

$$\mathrm{Prob}(X_i = x_i) = \sum_{j_1=1}^{|Y_1|} \cdots \sum_{j_n=1}^{|Y_n|} \mathrm{Prob}(X_i = x_i | Y_1 = y_1^{j_1}, \ldots, Y_n = y_n^{j_n}) \times \mathrm{Prob}(Y_1 = y_1^{j_1}) \times \cdots \times \mathrm{Prob}(Y_n = y_n^{j_n}).$$

En l'absence de cycle dans le graphe, une récursion sur le nombre de noeuds non calculés montre que soit i) tous les noeuds ont été calculés et l'algorithme est terminé ; soit ii) il existe un noeud dont tous les parents ont déjà été calculés, auquel cas on peut appliquer l'équation ci-dessus.

4.3.2 Les calculs selon une notation matricielle

Nous allons reprendre les calculs ci-dessus, mais en utilisant des opérations matricielles plutôt que des opérations sur des probabilités. Les probabilités marginales sont en effet représentées sur ordinateur par des vecteurs et les probabilités conditionnelles par des hypermatrices (c'est-à-dire des matrices ayant un nombre de dimensions non obligatoirement égal à 2). Par conséquent, à chaque dimension d'une hypermatrice ou à la dimension d'un vecteur correspond un noeud. Pour mieux comprendre ce qui suit, nous allons systématiquement indicer les vecteurs, les matrices et hypermatrices par les noms des noeuds correspondant à chaque dimension. Ainsi V_n désignera un vecteur dont la dimension est relative au noeud n ; par exemple V_n pourra représenter le vecteur de probabilité $\mathrm{Prob}(n)$. La taille du vecteur V_n sera donc égale à $|n|$, c'est-à-dire au nombre de modalités du noeud n. De même, M_{pn} désignera une matrice dont chaque ligne correspond à une modalité du noeud p et dont chaque colonne correspond à une modalité du noeud n, soit une matrice à $|p| \times |n|$ éléments. De manière analogue, $M_{i_1 i_2 \ldots i_p}$ représentera une hypermatrice dont les dimensions correspondent aux noeuds $i_1, i_2, \cdots, i_p$, autrement dit une matrice à $|i_1| \times |i_2| \times \cdots \times |i_p|$ éléments. De plus, nous désignerons dans les calculs les vecteurs, matrices et hypermatrices par leurs termes génériques de la manière suivante : un vecteur $V_n = \begin{pmatrix} v_1 \\ \vdots \\ v_{|n|} \end{pmatrix}$ sera noté $[\![v_i]\!]_n$, où v_i représente l'élément du vecteur sur la $i^{\text{ème}}$ ligne ; une matrice $M_{pn} = \begin{pmatrix} m_{11} & \ldots & m_{1j} & \ldots & m_{1|n|} \\ \vdots & & \vdots & & \vdots \\ m_{i1} & \ldots & m_{ij} & \ldots & m_{i|n|} \\ \vdots & & \vdots & & \vdots \\ m_{|p|1} & \ldots & m_{|p|j} & \ldots & m_{|p||n|} \end{pmatrix}$ sera notée $[\![m_{ij}]\!]_{pn}$, et une hypermatrice $M_{i_1 i_2 \ldots i_p}$ sera notée $[\![m_{j_1 j_2 \ldots j_p}]\!]_{i_1 i_2 \ldots i_p}$. Bien entendu, on supposera que dans les matrices et hypermatrices, toutes les dimensions sont relatives à des noeuds différents (probabilités conditionnelles d'un noeud sachant ses parents dans le réseau). Voyons maintenant quelques opérations de base sur les hypermatrices.

Somme de deux hypermatrices :

Soient $M_{i_1 i_2 \ldots i_p}$ et $N_{i_1 i_2 \ldots i_p}$ deux hypermatrices de mêmes dimensions. Alors la somme de ces deux hypermatrices est l'hypermatrice suivante : $M_{i_1 i_2 \ldots i_p} + N_{i_1 i_2 \ldots i_p} = [\![m_{j_1 j_2 \ldots j_p} + m_{j_1 j_2 \ldots j_p}]\!]_{i_1 i_2 \ldots i_p}$.

Produit d'une hypermatrice avec un scalaire :

Soient λ un nombre réel quelconque et $M_{np} = [\![m_{ij}]\!]_{np}$ une matrice quelconque. On sait que le produit de M_{np} par λ est égal à la matrice $\lambda M_{np} = [\![\lambda\, m_{ij}]\!]_{np}$. On généralise très simplement ce produit aux hypermatrices : soit $M_{i_1 i_2 \ldots i_p} = [\![m_{j_1 j_2 \ldots j_p}]\!]_{i_1 i_2 \ldots i_p}$, alors $\lambda M_{i_1 i_2 \ldots i_p} = [\![\lambda\, m_{j_1 j_2 \ldots j_p}]\!]_{i_1 i_2 \ldots i_p}$.

Produit d'une hypermatrice avec un vecteur :

Soient $V_n = [\![v_j]\!]_n$ un vecteur de taille $|n|$, et $M_{pn} = [\![m_{ij}]\!]_{pn}$ une matrice. Notons $\otimes$ le produit en question. On sait que le produit de la matrice M_{pn} par le vecteur V_n est égal au vecteur :

$$N_p = M_{pn} \otimes V_n = \begin{pmatrix} \sum_{i=1}^{|n|} v_j \times m_{1j} \\ \vdots \\ \sum_{i=1}^{|n|} v_j \times m_{|p|j} \end{pmatrix} = \left[\!\!\left[\sum_{j=1}^{|n|} v_j \times m_{ij} \right]\!\!\right]_p .$$

De même, le produit de la matrice M_{pn} par un vecteur V_p est égal au vecteur :

$$N_n = M_{pn} \otimes V_p = \left(\sum_{i=1}^{|p|} v_j \times m_{j1}, \cdots, \sum_{i=1}^{|p|} v_j \times m_{j|n|} \right) = \left[\!\!\left[\sum_{j=1}^{|p|} v_j \times m_{ji} \right]\!\!\right]_n .$$

On peut généraliser ce produit à des hypermatrices de la manière suivante : Soit $M_{i_1 i_2 \ldots i_p}$ une hypermatrice et V_{i_k} un vecteur. Alors le produit de l'hypermatrice avec le vecteur est égal à :

$$M_{i_1 i_2 \ldots i_p} \otimes V_{i_k} = \left[\!\!\left[\sum_{i=1}^{|i_k|} v_i \times m_{j_1 j_2 \ldots j_{k-1} i j_{k+1} \ldots j_p} \right]\!\!\right]_{i_1 i_2 \ldots i_{k-1} i_{k+1} \ldots i_p} .$$

Produit tensoriel (ou terme à terme) entre deux vecteurs :

Soient deux vecteurs $V_k = [\![v_i]\!]_k$ et $W_k = [\![w_j]\!]_k$. Alors le produit tensoriel de V_k et de W_k est égal à $V_k \odot W_k = [\![v_i \times w_i]\!]_k$.

4.3.3 Calcul des probabilités *a priori* en informatique

Revenons maintenant aux calculs qui nous intéressent. Considérons un noeud X dont on connaît les probabilités *a priori* de ses parents, $Y_1, \ldots, Y_n$. On a vu que $\text{Prob}(X = x_j)$ se calcule en théorie de la manière suivante :

$$\begin{aligned} \text{Prob}(X = x_j) = \sum_{j_1=1}^{|Y_1|} \sum_{j_2=1}^{|Y_2|} \cdots \sum_{j_n=1}^{|Y_n|} & \text{Prob}(X = x_j | Y_1 = y_1^{j_1}, Y_2 = y_2^{j_2}, \ldots, Y_n = y_n^{j_n}) \\ & \times \text{Prob}(Y_1 = y_1^{j_1}) \times \cdots \times \text{Prob}(Y_n = y_n^{j_n}). \end{aligned}$$

En pratique, nous allons stocker dans l'ordinateur la probabilité conditionnelle $\text{Prob}(X = x_i | Y_1 = y_1^{j_1}, \ldots, Y_n = y_n^{j_n})$ pour l'ensemble des valeurs $x_i, y_1^{j_1}, \ldots, y_n^{j_n}$, grâce à une hypermatrice : $[\![\text{Prob}(x_i | y_1^{j_1}, \ldots, y_n^{j_n})]\!]_{XY_1 \ldots Y_n}$. De la même manière, nous allons stocker les probabilités des parents sous forme de vecteurs $[\![\text{Prob}(y_k^{j_k})]\!]_{Y_k}$. D'après les formules qui précèdent, il est relativement évident que l'équation ci-dessus correspond à la ligne x_j du vecteur $[\![\text{Prob}(x_i)]\!]_X$ défini par :

$$[\![\text{Prob}(x_i)]\!]_{X_i} = \left(\cdots \left(\left([\![\text{Prob}(x_i | y_1^{j_1}, \ldots, y_n^{j_n})]\!]_{XY_1 \ldots Y_n} \otimes [\![\text{Prob}(y_1^{j_1})]\!]_{Y_1} \right) \otimes [\![\text{Prob}(y_2^{j_2})]\!]_{Y_2} \right) \cdots \right) \otimes [\![\text{Prob}(y_n^{j_n})]\!]_{Y_n} .$$

Autrement dit, le calcul des probabilités *a priori* se ramène à l'algorithme suivant :

Algorithme 1 (Calcul des probabilités *a priori*) *Les probabilités des noeuds orphelins (sans parents dans le graphe) sont déjà calculées. Pour tous les noeuds X dont tous les parents $Y_1, \ldots, Y_n$ ont été calculés, $[\![\text{Prob}(X = x_i)]\!]_X$ se calcule de la manière suivante :*

$$[\![\text{Prob}(x_i)]\!]_X = \left(\cdots \left(\left([\![\text{Prob}(x_i | y_1^{j_1}, \ldots, y_n^{j_n})]\!]_{XY_1 \ldots Y_n} \otimes [\![\text{Prob}(y_1^{j_1})]\!]_{Y_1} \right) \otimes [\![\text{Prob}(y_2^{j_2})]\!]_{Y_2} \right) \cdots \right) \otimes [\![\text{Prob}(y_n^{j_n})]\!]_{Y_n} .$$

En l'absence de cycle dans le graphe, soit i) tous les noeuds ont été calculés et l'algorithme est fini ; soit ii) il existe un noeud dont tous les parents ont déjà été calculés, auquel cas on peut appliquer l'équation ci-dessus.

En fait, on peut interpréter l'algorithme ci-dessus de la manière suivante : les noeuds orphelins (sans parents) envoient à leurs enfants les messages correspondant à leurs probabilités marginales, à savoir les $[\![\mathrm{Prob}(Y_j = y_j^i)]\!]_{Y_j}$. Lorsqu'un noeud a reçu des messages de tous ses parents, il peut alors calculer sa probabilité marginale grâce à l'équation ci-dessus. Celle-ci peut se comprendre grâce à des équations aux dimensions : Avant l'arrivée des messages, le noeud X ne connaît que la probabilité conditionnelle $\mathrm{Prob}(X|Y_1,\ldots,Y_n)$. Autrement dit, avant l'arrivée des messages, le noeud X ne connaît qu'une table de probabilité de dimensions $|X|\times|Y_1|\times|Y_2|\times\cdots\times|Y_n|$. Le noeud X voit l'arrivée des messages de taille $|Y_1|$, $|Y_2|$, etc, et veut calculer $\mathrm{Prob}(X)$, soit une matrice de probabilité de taille $|X|$. Il lui faut donc partir de la matrice de taille $|X|\times|Y_1|\times|Y_2|\times\cdots\times|Y_n|$, et éliminer les dimensions $|Y_1|$, $|Y_2|$, etc, jusqu'à $|Y_n|$. Comment faire ? Eh bien tout simplement en faisant des produits matrice-vecteur qui permettent justement l'élimination de dimensions. En multipliant donc successivement l'hypermatrice $\mathrm{Prob}(X|Y_1,\ldots,Y_n)$ par des vecteurs de dimensions $|Y_1|$,$|Y_2|$,. . .,$|Y_n|$, on obtient un vecteur de dimension $|X|$, et c'est exactement ce que l'on cherche.

Lorsqu'un noeud a pu calculer par produits matrice-vecteur sa probabilité marginale, il ne lui reste plus qu'à envoyer à ses enfants des messages (vecteurs) égaux à cette probabilité marginale, et le processus peut recommencer avec les noeuds ayant reçu des messages de tous leurs parents.

4.3.4 Probabilités *a posteriori* : l'ajout d'informations dans le réseau

L'intérêt des réseaux bayésiens ne réside pas seulement dans le calcul des probabilités a priori ; heureusement, sinon nous aurions développé un outil bien compliqué pour l'utilité qu'il procurerait ! Non, l'intérêt principal des réseaux bayésiens réside dans le calcul de probabilités *a posteriori*. Reprenons l'exemple de la dyspnée. Que nous indiquent les probabilités *a priori* ? Eh bien tout simplement la probabilité qu'une personne choisie au hasard dans la population fume ($\mathrm{Prob}(X_2)$), ou bien qu'elle soit allée en Asie ($\mathrm{Prob}(X_1)$), ou bien encore la probabilité qu'une personne choisie au hasard ait une dyspnée ($\mathrm{Prob}(X_6)$). Cependant, ce n'est pas ce qui intéresse le médecin pour faire son diagnostic : lui sait que son patient fume et qu'il n'a jamais séjourné en Asie. Donc ce qui l'intéresse, c'est la probabilité que le patient ait une dyspnée sachant ces informations, autrement dit $\mathrm{Prob}(X_6 = \text{«oui»}|X_1 = \text{«non»}, X_2 = \text{«oui»})$. Une probabilité conditionnée par des informations (des observations) s'appelle une $\boxed{\text{probabilité } \textit{a posteriori}}$.

Pour qu'un logiciel de réseaux bayésiens soit utile, il faut donc qu'il permette l'insertion, l'ajout, de nouvelles informations, et qu'il permette de recalculer simplement les probabilités des différentes variables aléatoires conditionnellement à celles-ci. Les nouvelles informations rentrées dans le réseau, que l'on appelle aussi observations ou encore $\boxed{\text{évidences}}$, sont dans les logiciels commerciaux (comme HUGIN) de plusieurs natures :

- soit elles représentent, grâce à des valeurs booléennes (1/0), l'observation ou la non observation des valeurs que peuvent prendre les variables aléatoires du réseau ;
- soit elles représentent les nouvelles lois de probabilité marginale observées pour une ou des variables. Ces nouvelles lois étant connues par statistiques.

Dans le cadre de ce cours, nous pencherons pour le premier type d'observation, c'est-à-dire l'utilisation de variables booléennes. L'exemple suivant va illustrer comment l'on s'y prend : supposons qu'une variable X ait 5 modalités et que l'on ait observé que seules la deuxième et la quatrième sont possibles. Alors le vecteur d'évidence que l'on associe au noeud X est le suivant : $(0,1,0,1,0)$. Le médecin de la dyspnée a devant lui un patient qui n'a jamais séjourné en Asie et qui fume, eh bien l'on rentre dans les noeuds «séjour en asie» et «fume», dont les modalités sont («oui»,«non»), respectivement les évidences $(0,1)$ et $(1,0)$.

4.3.5 Calcul des probabilités *a posteriori*

Grâce à un exemple simple et des équations aux dimensions, nous allons déduire comment peuvent se calculer les probabilités *a posteriori*.

Lorsque les observations proviennent d'une racine du réseau

Considérons donc le réseau de la figure 17. On apprend par ailleurs que T, qui *a priori* pouvait prendre les valeurs $t_1, t_2, \ldots, t_{|T|}$ ne peut plus prendre que deux valeurs t_1 et t_2. Appelons e cette information (e comme *evidence* en anglais). Quelle est maintenant la probabilité de chacune des variables aléatoires connaissant e ? Autrement dit, que valent $\mathrm{Prob}(T|e)$, $\mathrm{Prob}(U|e)$, $\mathrm{Prob}(X|e)$, $\mathrm{Prob}(Y|e)$, $\mathrm{Prob}(Z|e)$?

$$T \longrightarrow U \longrightarrow X \longrightarrow Y \longrightarrow Z$$

FIG. 17 – Exemple de réseau bayésien

Tout d'abord, il faut calculer $\text{Prob}(T = t_i|e)$ pour tout i. Bien évidemment, si $i \neq 1, 2$, $\text{Prob}(T = t_i|e) = 0$. Supposons que T représente le couple (âge,sexe) des étudiants dans un groupe de TD. Si t_0, t_1, t_2, t_3 correspondent respectivement aux couples (âge$\leq$21 ans, masculin), (âge>21 ans, masculin), (âge$\leq$21 ans, féminin), (âge>21 ans, féminin), et si en moyenne les groupes contiennent 23 hommes, dont 14 ont plus de 21 ans, et 7 femmes dont 3 ont plus de 21 ans, alors :

$$\text{Prob}(T = t_1) = \frac{9}{30}, \quad \text{Prob}(T = t_2) = \frac{14}{30}, \quad \text{Prob}(T = t_3) = \frac{4}{30}, \quad \text{Prob}(T = t_4) = \frac{3}{30}.$$

On stocke sur ordinateur la probabilité de T grâce au vecteur $[\![\text{Prob}(t_i)]\!]_T = (\frac{9}{30}, \frac{14}{30}, \frac{4}{30}, \frac{3}{30})$. Maintenant, on sélectionne un groupe et l'on s'aperçoit qu'il n'y a aucune femme dans ce groupe. Dans ce cas, on peut dire qu'en sélectionnant un individu au hasard dans le groupe, on a 14 chances sur 23 qu'il ait plus de 21 ans, et 9 chances sur 23 qu'il ait moins de 21 ans. Ainsi,

$$\text{Prob}(T = t_1|e) = \frac{9}{23}, \quad \text{Prob}(T = t_2|e) = \frac{14}{23}, \quad \text{Prob}(T = t_3|e) = 0, \quad \text{Prob}(T = t_4|e) = 0.$$

Notons que $\sum_{i=1}^{4} \text{Prob}(T = t_i|e) = 1$. Créons un vecteur représentant l'information sur T, $E_T = (1, 1, 0, 0)$, où les 1 représentent les valeurs t_i que peut prendre T et les 0 représentent les valeurs que T ne peut plus prendre, et effectuons le produit tensoriel de $[\![\text{Prob}(t_i)]\!]_T$ avec E_T. On obtient alors le vecteur $[\![\text{Prob}(t_i)]\!]_T \odot E_T = (\frac{9}{30}, \frac{14}{30}, 0, 0)$, qui n'est autre que le vecteur représentant $\text{Prob}(T|e)$ à un coefficient multiplicatif près. Ce dernier est visiblement égal à $\frac{30}{23}$, ce qui correspond en fait à $\frac{1}{\text{Prob}(e)}$ puisqu'on a 23 chances sur 30 de sélectionner un homme lorsque l'on choisit un individu au hasard dans un groupe choisi lui aussi au hasard. Cependant, nous calculerons ce coefficient de la manière suivante : on sait que $\sum_{i=1}^{4} \text{Prob}(T = t_i|e) = 1$, donc obligatoirement, le coefficient multiplicatif est l'inverse de la somme des éléments du vecteur $[\![\text{Prob}(t_i)]\!]_T \odot E_T$. Sous forme matricielle, cette somme peut s'écrire de la manière suivante : $([\![\text{Prob}(t_i)]\!]_T \odot E_T) \otimes \mathbb{1}_T$, où $\mathbb{1}_T$ désigne le vecteur de taille $|T|$ composé uniquement de 1. Donc

$$[\![\text{Prob}(t_i|e)]\!]_T = \frac{1}{\text{Prob}(e)} ([\![\text{Prob}(t_i)]\!]_T \odot E_T) \quad \text{et} \quad \text{Prob}(e) = ([\![\text{Prob}(t_i)]\!]_T \odot E_T) \otimes \mathbb{1}_T.$$

En principe, on ne calcule jamais le coefficient multiplicatif $\text{Prob}(e)$ car les calculs restent cohérents quels que soient ces coefficients. Lorsqu'on a besoin de montrer à l'utilisateur les probabilités calculées, il suffit de se rappeler que $\sum_i \text{Prob}(A = a_i|e) = 1$, quelle que soit la variable A.

Calculons maintenant $\text{Prob}(U|e)$.

$$\text{Prob}(U = u|e) = \frac{\text{Prob}(U = u, e)}{\text{Prob}(e)} = \frac{\sum_{i=1}^{4} \text{Prob}(U = u, e|T = t_i) \times \text{Prob}(T = t_i)}{\text{Prob}(e)}.$$

Notons que $\text{Prob}(U = u, e|T = t_3) = \text{Prob}(U = u, e|T = t_4) = 0$ puisque ce sont les probabilités que $U = u$ et $T = t_1$ ou t_2 sachant que $T = t_3$ ou t_4. De même $\text{Prob}(T = t_3, e) = \text{Prob}(T = t_4, e) = 0$. Par contre, $\text{Prob}(U = u, e|T = t_1) = \text{Prob}(U = u, e|T = t_2) = \text{Prob}(U = u|T = t_2)$, $\text{Prob}(T = t_1, e) = \text{Prob}(T = t_1)$, et $\text{Prob}(T = t_2, e) = \text{Prob}(T = t_2)$. Donc

$$\text{Prob}(U = u|e) = \frac{\sum_{i=1}^{4} \text{Prob}(U = u|T = t_i) \times \text{Prob}(T = t_i, e)}{\text{Prob}(e)} = \sum_{i=1}^{4} \text{Prob}(U = u|T = t_i) \times \text{Prob}(T = t_i|e).$$

Autrement dit, pour calculer $\text{Prob}(U = u|e)$, il suffit de faire le produit de la matrice $[\![\text{Prob}(u_i|t_j)]\!]_{UT}$ avec le vecteur $[\![\text{Prob}(t_j|e)]\!]_T$. De la même manière, on voit bien que $[\![\text{Prob}(x_i|e)]\!]_X = [\![\text{Prob}(x_i|u_j)]\!]_{XU} \otimes [\![\text{Prob}(u_j|e)]\!]_U$. Par récurrence,

> Supposons que le réseau bayésien soit une chaîne et que l'information e provienne de la racine de cette chaîne. Soit B un noeud quelconque de la chaîne (excepté la racine), et soit A son père.
> Alors $[\![\mathrm{Prob}(b_i|e)]\!]_B = [\![\mathrm{Prob}(b_i|a_j)]\!]_{BA} \otimes [\![\mathrm{Prob}(a_j|e)]\!]_A$.

D'une manière plus visuelle (cf. figure 18), voilà comment se passe la propagation de l'information $T = t_1$ ou t_2 dans le réseau :

> 1. T reçoit l'information, sous forme d'un vecteur E_T, qu'il ne peut plus prendre que certaines valeurs.
> 2. T met alors à jour sa probabilité : $[\![\mathrm{Prob}(t_i|e)]\!]_T = \frac{1}{\mathrm{Prob}(e)} ([\![\mathrm{Prob}(t_i)]\!]_T \odot E_T)$. Il envoie alors à son fils un message $\pi_{TU} = [\![\mathrm{Prob}(t_i|e)]\!]_T$.
> 3. Lorsque U reçoit le message envoyé par T, il met à jour sa propre probabilité a posteriori : $[\![\mathrm{Prob}(u_i|e)]\!]_U = [\![\mathrm{Prob}(u_i|t_j)]\!]_{UT} \otimes \pi_{TU}$. Il envoie ensuite à son fils un message $\pi_{UX} = [\![\mathrm{Prob}(u_i|e)]\!]_U$.
> 4. lorsque X reçoit le message, il met à jour sa probabilité et émet un message π_{XY}, et ainsi de suite...

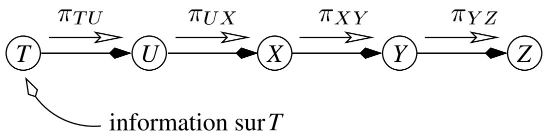


FIG. 18 – Les messages à envoyer pour calculer les probabilités *a posteriori*

Lorsque les observations proviennent d'une feuille

Supposons maintenant que l'information e ne soit plus en T mais en Z : on apprend que Z peut seulement prendre les valeurs z_1 et z_2 au lieu des 4 valeurs z_1, z_2, z_3, z_4. On connaît déjà $\mathrm{Prob}(Z)$, cette probabilité ayant été calculée dans les sous-sections 4.3.1 et 4.3.2. De la même manière que l'on avait calculé $\mathrm{Prob}(T|e)$, on peut calculer $\mathrm{Prob}(Z|e)$:

$$[\![\mathrm{Prob}(z_i|e)]\!]_Z = \frac{1}{\mathrm{Prob}(e)} ([\![\mathrm{Prob}(z_i)]\!]_Z \odot E_Z) \quad \text{et} \quad \mathrm{Prob}(e) = ([\![\mathrm{Prob}(z_i)]\!]_Z \odot E_Z) \otimes \mathbb{1}_Z.$$

Pour calculer $\mathrm{Prob}(Y|e)$, nous allons appliquer le théorème de Bayes (cf. page 30) :

$$\mathrm{Prob}(Y|e) = \mathrm{Prob}(e|Y) \times \frac{\mathrm{Prob}(Y)}{\mathrm{Prob}(e)}.$$

Maintenant $\mathrm{Prob}(e|Y) = \mathrm{Prob}(Z = z_1 \text{ ou } z_2|Y) = \mathrm{Prob}(Z = z_1|Y) + \mathrm{Prob}(Z = z_2|Y)$. En termes matriciels, on obtient :

$$[\![\mathrm{Prob}(e|y_i)]\!]_Y = [\![\mathrm{Prob}(z_j|y_i)]\!]_{ZY} \otimes E_Z.$$

Par conséquent,

$$[\![\mathrm{Prob}(y_i|e)]\!]_Y = \frac{1}{\mathrm{Prob}(e)} (([\![\mathrm{Prob}(z_j|y_i)]\!]_{ZY} \otimes E_Z) \odot [\![\mathrm{Prob}(y_i)]\!]_Y).$$

De la même manière,

$$\mathrm{Prob}(X|e) = \mathrm{Prob}(e|X) \times \frac{\mathrm{Prob}(X)}{\mathrm{Prob}(e)} = \sum_{i=1}^{|Y|} \mathrm{Prob}(e|X, Y = y_i) \times \mathrm{Prob}(Y = y_i|X) \times \frac{\mathrm{Prob}(X)}{\mathrm{Prob}(e)}.$$

Or, d'après la signification des arcs dans le graphe, le fait que l'arc (Y, Z) existe, mais que l'arc (X, Z) n'existe pas, signifie que Z est indépendant de X conditionnellement à Y. Autrement dit, $\mathrm{Prob}(e|X, Y = y_i) = \mathrm{Prob}(e|Y = y_i)$. Donc

$$\mathrm{Prob}(X|e) = \sum_{i=1}^{|Y|} \mathrm{Prob}(e|Y = y_i) \times \mathrm{Prob}(Y = y_i|X) \times \frac{\mathrm{Prob}(X)}{\mathrm{Prob}(e)}.$$

En termes matriciels, on obtient :

$$[\![\mathrm{Prob}(x_i|e)]\!]_X = \frac{1}{\mathrm{Prob}(e)} \left([\![\mathrm{Prob}(e|y_j)]\!]_Y \otimes [\![\mathrm{Prob}(y_j|x_i)]\!]_{YX}\right) \odot [\![\mathrm{Prob}(x_i)]\!]_X.$$

Par récurrence, on obtient la propriété suivante :

> Supposons que le réseau bayésien soit une chaîne et que l'information e provienne de la feuille de cette chaîne. Soit un noeud B quelconque de la chaîne (excepté la feuille), et soit A son fils. Alors $[\![\mathrm{Prob}(b_i|e)]\!]_B = \frac{1}{\mathrm{Prob}(e)} \left([\![\mathrm{Prob}(e|a_j)]\!]_A \otimes [\![\mathrm{Prob}(a_j|b_i)]\!]_{AB}\right) \odot [\![\mathrm{Prob}(b_i)]\!]_B$. Là encore, le terme multiplicatif $\frac{1}{\mathrm{Prob}(e)}$ n'a pas besoin d'être calculé explicitement puisqu'on sait que $[\![\mathrm{Prob}(b_i|e)]\!]_B \otimes \mathbb{1}_B = 1$.

D'une manière plus visuelle (cf. figure 19), voilà comment se passe la propagation de l'information $Z = z_1$ ou z_2 dans le réseau :

1. Z reçoit l'information sous forme d'un vecteur E_Z qu'il ne peut plus prendre que certaines valeurs.
2. Z met alors à jour sa probabilité : $[\![\mathrm{Prob}(z_i|e)]\!]_Z = \frac{1}{\mathrm{Prob}(e)} \left([\![\mathrm{Prob}(z_i)]\!]_Z \odot E_Z\right)$. Il envoie alors à son père un message $\lambda_{ZY} = [\![\mathrm{Prob}(e|y_i)]\!]_Y = [\![\mathrm{Prob}(z_j|y_i)]\!]_{ZY} \otimes E_Z$.
3. Lorsque Y reçoit le message envoyé par Z, il met à jour sa propre probabilité a posteriori : $[\![\mathrm{Prob}(y_i|e)]\!]_Y = \frac{1}{\mathrm{Prob}(e)} \left(\lambda_{ZY} \odot [\![\mathrm{Prob}(y_i)]\!]_Y\right)$. Il envoie ensuite à son père un message $\lambda_{YX} = \lambda_{ZY} \otimes [\![\mathrm{Prob}(y_j|x_i)]\!]_{YX}$.
4. lorsque X reçoit le message, il met à jour sa probabilité et émet un message λ_{XU}, et ainsi de suite...

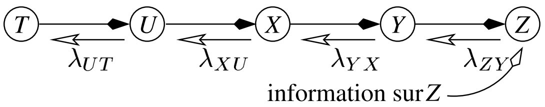


FIG. 19 – Les messages à envoyer pour calculer les probabilités *a posteriori*

Lorsque les observations proviennent d'une racine et d'une feuille

Supposons maintenant qu'on ait eu une information e_T en T et une information e_Z en Z. Pour calculer les probabilités *a posteriori*, dans ce cas, l'algorithme n'est pas compliqué, c'est en fait juste une petite généralisation de ce que nous avons fait précédemment :

1. On calcule les messages π et λ avec les formules données précédemment.
2. Pour les noeuds non observés, U, X, Y, si l'on note $\propto$ l'égalité à un coefficient multiplicatif près, les probabilités a posteriori sont égales à :

$$\begin{aligned}
[\![\mathrm{Prob}(u_i|e_T, e_z)]\!]_U &\propto \left([\![\mathrm{Prob}(u_i|t_j)]\!]_{UT} \otimes \pi_{TU}\right) \odot \lambda_{XU},\\
[\![\mathrm{Prob}(x_i|e_T, e_z)]\!]_X &\propto \left([\![\mathrm{Prob}(x_i|u_j)]\!]_{XU} \otimes \pi_{UX}\right) \odot \lambda_{YX},\\
[\![\mathrm{Prob}(y_i|e_T, e_z)]\!]_Y &\propto \left([\![\mathrm{Prob}(y_i|x_j)]\!]_{YX} \otimes \pi_{XY}\right) \odot \lambda_{ZY}.
\end{aligned}$$

3. Pour les noeuds observés, T et Z :

$$\begin{aligned}
[\![\mathrm{Prob}(t_i|e_T, e_z)]\!]_T &\propto \left([\![\mathrm{Prob}(t_j)]\!]_T \odot \lambda_{UT}\right) \odot E_T,\\
[\![\mathrm{Prob}(z_i|e_T, e_z)]\!]_Z &\propto \left([\![\mathrm{Prob}(z_i|y_j)]\!]_{ZY} \otimes \pi_{YZ}\right) \odot E_Z.
\end{aligned}$$

En fait, dans les calculs, on peut ne pas différencier les variables non observées de celles qui le sont si l'on rajoute à ces derniers un fils fictif qui enverrait un message λ égal au message $\mathbb{1}$. Par exemple, pour U, on rajouterait un fils qui enverrait le message E_U.

> Dans tout ce qui suit, on appellera les messages π lorsqu'ils sont envoyés des parents vers les enfants, et λ lorsqu'ils vont des enfants vers les parents.

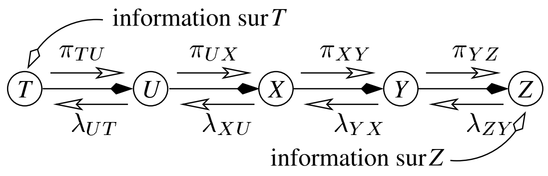


FIG. 20 – Les messages à envoyer pour calculer les probabilités *a posteriori*

4.3.6 Calcul des probabilités *a posteriori* : le cas général

Pour calculer des probabilités *a posteriori*, il nous faut bien comprendre ce que représentent les arcs du réseau : les arcs relient les variables qui sont dépendantes les unes des autres, ou plus exactement l'absence d'arc représente une absence de dépendance. Si la probabilité *a posteriori* d'une variable X est différente de sa probabilité *a priori*, cela signifie que la variable X est influencée par les informations que l'on vient de rentrer dans le réseau, ou encore que d'une manière directe ou indirecte X est dépendante des variables qui ont reçu les nouvelles informations. Comme les dépendances sont représentées par les arcs du réseau, on peut comprendre intuitivement que les informations se propagent dans le réseau grâce aux arcs. L'idée de l'algorithme de calcul des probabilités *a posteriori* est donc tout simplement d'envoyer des messages le long des arcs du réseau, messages qui vont contenir les nouvelles informations reçues. Lorsque les noeuds recevront ces messages, ils pourront mettre à jour leurs probabilités marginales et renvoyer de nouveaux messages vers leurs voisins. Ainsi les informations seront propagées de proche en proche.

Illustration sur l'exemple de la dyspnée : on vient de recevoir une information sur la bronchite et sur un séjour en Asie. Nous noterons $\pi_k(X, Y, Z)$ les messages contenant les informations sur les noeuds X, Y, Z, envoyés lors de la $k^{\text{ème}}$ étape de l'algorithme vers les enfants. Par exemple $\pi_2(A)$ est le message contenant les informations sur A, envoyé à la deuxième étape de l'algorithme. De même nous noterons $\lambda_k(X, Y, Z)$ les messages contenant

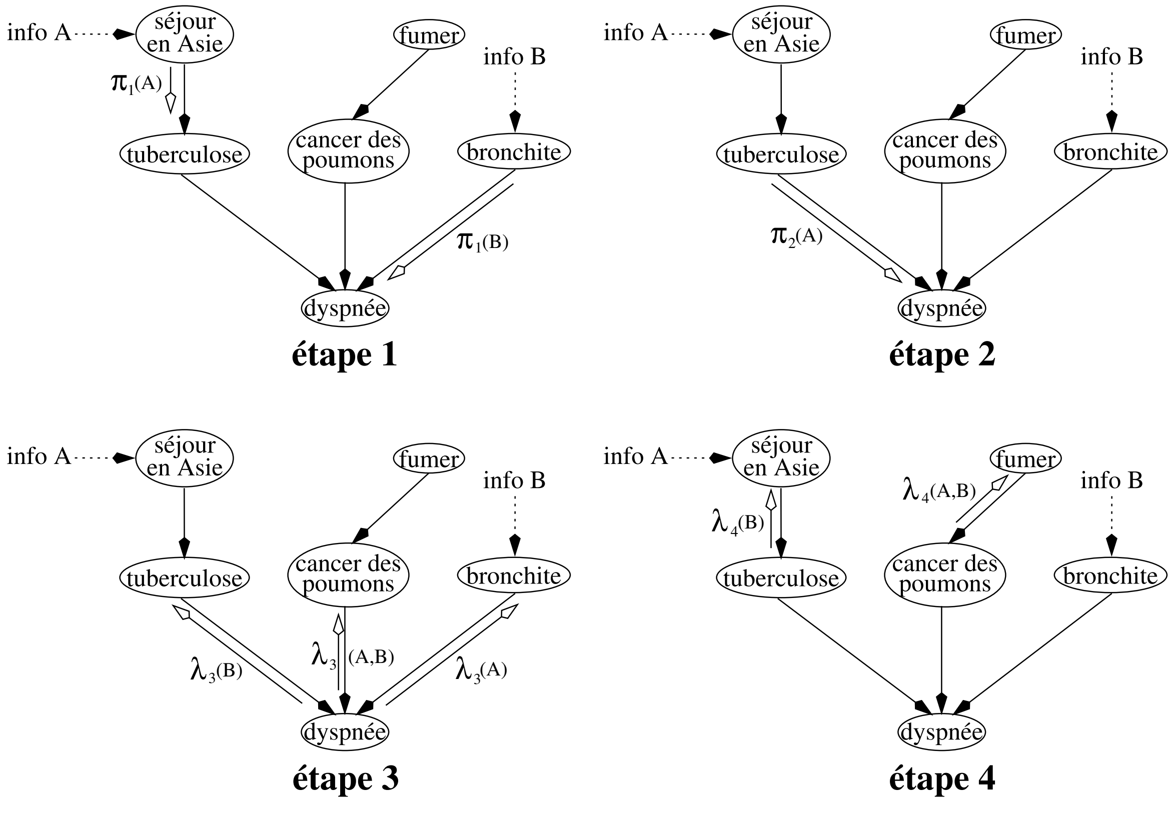


FIG. 21 – Calcul des probabilités *a posteriori*

les informations sur les noeuds X, Y, Z, envoyés vers les parents lors de la $k^{\text{ème}}$ étape de l'algorithme.

Au début de l'algorithme, le noeud «séjour en asie» reçoit l'information A et le noeud «bronchite» reçoit l'information B. Ces noeuds vont produire des messages contenant ces informations, et vont les envoyer à leurs voisins. Ce sont les messages $\pi_1(A)$ et $\pi_1(B)$. Les voisins peuvent à leur tour envoyer des messages contenant les informations sur A et/ou sur B. Ainsi les informations A et B transiteront dans tout le graphe.

Pour que les probabilités *a posteriori* des noeuds soient bien calculées, il suffit que l'ensemble des messages reçus par chacun des noeuds contienne l'ensemble des informations entrées dans le réseau. Ici, il faut donc que chaque noeud reçoive un ou plusieurs messages contenant les informations A et B. On pourra remarquer que c'est bien le cas sur la figure 21. Pour réaliser cela, on adopte en principe l'algorithme suivant : on choisit un noeud au hasard (ici, nous avons choisi le noeud «dyspnée»). Ce noeud demande à ses voisins de lui transmettre les messages qu'ils ont reçus. Les voisins, avant de répondre, demandent eux-mêmes à leurs voisins (sauf le noeud qui demande les messages) de leur fournir les messages qu'ils ont reçu, et ainsi de suite, jusqu'à ce que l'on atteigne un noeud (un noeud à une extrémité du réseau) dont le seul voisin est celui qui demande à ce qu'on lui transmette un message. Dans ce cas, le noeud à l'extrémité du réseau répond. Lorsqu'un noeud a reçu tous les messages qu'il attendait, il renvoit le message qu'il a fabriqué au noeud qui le lui demandait, et ainsi de suite. C'est ce qui correspond aux étapes 1 et 2 de la figure 21 : «dyspnée» demande à «tuberculose», «cancer» et «bronchite» de lui transmettre des messages. «bronchite» peut répondre tout de suite ($\pi_1(B)$) car il est à une extrémité du réseau. «cancer» n'est pas à une extrémité et demande donc à «fumer» de lui transmettre un message mais «fumer» n'a, pour l'instant, aucune information à propager. «tuberculose» n'est pas non plus à une extrémité et demande donc à «séjour en Asie» de lui transmettre un message. «séjour en Asie» qui est à une extrémité du réseau s'exécute ($\pi_1(A)$). Quand «tuberculose» a reçu le message, il en envoie un à son tour à «dyspnée» ($\pi_2(A)$). Lorsque le noeud «dyspnée» a reçu tous les messages, il connaît toutes les informations entrées dans le réseau. Il renvoit alors à ses voisins des messages sur l'ensemble des informations qu'ils n'avaient pas encore à leur disposition (c'est l'étape 3). À leur tour, ces noeuds renvoient à leur voisins des messages sur l'ensemble des informations qu'ils n'avaient pas encore eues, et ainsi de suite jusqu'à ce que tous les noeuds aient reçu des messages sur l'ensemble des informations entrées dans le réseau (sur la figure 21, cela correspond à l'étape 4). On obtient donc l'algorithme suivant :

Algorithme 2 (Calcul des probabilités *a posteriori*)

1. *On choisit un noeud X au hasard.*
2. *Ce noeud demande à ses voisins de lui envoyer un message. À leur tour, ceux-ci demandent à leurs voisins (sauf X) de leur envoyer des messages, et ainsi de suite : chaque noeud demande à tous ses voisins, excepté le noeud qui lui a demandé d'envoyer un message, de lui envoyer un message. Les noeuds aux extrémités du réseau ne peuvent plus demander de message. Ils émettent alors leur message. Lorsqu'un noeud a reçu tous les messages qu'il attendait, il envoie alors le message qu'on lui avait demandé, et ainsi de suite jusqu'à X.*
3. *Lorsque X a reçu tous les messages qu'il avait demandé, il distribue des messages à tous ses voisins. À leur tour, ceux-ci distribuent des messages vers leurs voisins, excepté X, et ainsi de suite, jusqu'à ce qu'on ne puisse plus envoyer de message.*

La phase 2 s'appelle une collecte d'informations et la phase 3 s'appelle une distribution d'informations.

Il reste encore à déterminer ce que contiennent les différents messages transitant sur le réseau. Pour répondre à cette question, il convient de comprendre quelle est la signification de ces messages. À partir de maintenant, pour simplifier les notations, nous noterons π_{XY} (resp. λ_{XY}) le message envoyé du noeud X vers son enfant (resp. parent) Y. La figure 22 va nous aider à comprendre quel est le contenu des messages : l'arc (X, Y) sépare le réseau

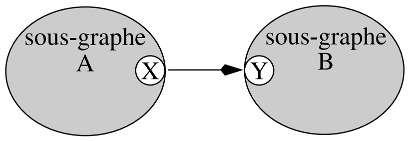


FIG. 22 – De la signification du π et du λ.

en deux sous-réseaux, désignés par les lettres A et B. L'ensemble des informations/observations du sous-réseau A

va avoir un certain impact sur X. Eh bien c'est cet impact que le vecteur π_{XY} propage vers Y. De même, λ_{YX} propage vers X l'impact des observations du sous-réseau B. Par conséquent, π_{XY} est totalement indépendant de λ_{YX}, et ces deux vecteurs peuvent même être calculés sur des machines fonctionnant en parallèle. Considérons maintenant la figure 23 : un noeud X a pour parents $Y_1, \ldots, Y_n$ et pour enfants $Z_1, \ldots, Z_p$. Il désire envoyer

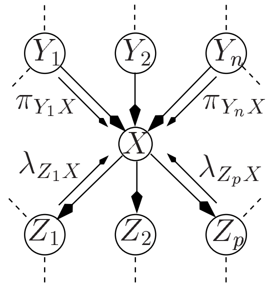


FIG. 23 – Les messages à envoyer.

un message λ_{XY_1} vers Y_1. Il doit donc considérer comme sous-réseau A la partie du réseau de la figure 23 en haut à gauche de Y_1, et comme sous-réseau B le reste du réseau de la figure 23. Donc, le message λ_{XY_1} doit impérativement renfermer les informations contenues dans tous les π_{Y_jX}, pour $j \geq 2$, car les Y_j ou leurs parents peuvent avoir reçu des informations. Pour les mêmes raisons, λ_{XY_1} doit aussi renfermer les informations contenues dans tous les λ_{Z_iX}. Les équations aux dimensions vont nous montrer comment réaliser cela : Lorsque les Z_i ont des informations à transmettre à X, ils envoient des messages λ_{Z_iX} vers X. Ce dernier fait alors des opérations matricielles entre λ_{Z_iX} et son hypermatrice de probabilité conditionnelle, à savoir une hypermatrice de taille $|X| \times |Y_1| \times \cdots \times |Y_n|$. Pour que cela ait un sens, il faut donc que λ_{Z_iX} ait pour dimension une de celles de l'hypermatrice. Il ne serait pas logique que cette dimension soit un des $|Y_j|$ puisque les enfants de X n'ont *a priori* rien à voir avec les parents de X. Donc la «logique» impose que tout vecteur λ_{Z_iX} ait pour dimension $|X|$. Passons maintenant aux parents de X. Y_j envoie un message π_{Y_jX} à X pour le mettre au courant des nouvelles informations rentrées dans le réseau. Pour les mêmes raisons que précédemment, le vecteur π_{Y_jX} ne peut avoir pour dimension que $|Y_j|$ ou $|X|$. S'il avait pour dimension $|X|$, on serait tout de même embêtés car, dans ce cas, on ne voit pas très bien comment obtenir des probabilités marginales *a posteriori* pour X : en effet, pour cela, il faudrait éliminer de l'hypermatrice de probabilité conditionnelle stockée dans X toutes les dimensions sauf celle de X. Or si tous les messages λ_{Z_iX}, π_{Y_jX} ont pour dimension $|X|$, aucune opération matricielle entre ces vecteurs et l'hypermatrice de probabilité conditionnelle ne permet d'éliminer les dimensions $|Y_1|$, $|Y_2|$,..., $|Y_n|$. Donc, pour pouvoir calculer les probabilités *a posteriori*, il faut **obligatoirement** que les vecteurs π_{Y_jX} aient pour dimension $|Y_j|$.

Ces préliminaires étant établis, passons maintenant aux calculs proprement dits. Les π_{Y_jX} ont donc pour taille $|Y_j|$. Or, X ne connaît a priori cette dimension que par l'intermédiaire de sa matrice de probabilité conditionnelle. Étant donné que λ_{XY_1} a pour dimension $|Y_1|$, il va falloir faire disparaître de la matrice de probabilité conditionnelle les dimensions en Y_j, $j \geq 2$. Le seul opérateur que l'on ait défini qui nous permette cela est le produit hypermatrice-vecteur. On devra donc calculer

$$\Big(\Big(\big(\llbracket \text{Prob}(x|y_1, y_2, \ldots, y_n) \rrbracket_{XY_1\ldots Y_n} \otimes \pi_{Y_2X}\big) \otimes \pi_{Y_3X}\Big) \otimes \cdots\Big) \otimes \pi_{Y_nX}.$$

Après calcul, on obtient une matrice de taille $|X||Y_1|$. Il faut donc encore éliminer la dimension en X (donc faire un produit matrice-vecteur) pour obtenir un message ayant la bonne taille.

Notons E_X la nouvelle information rentrant en X. Celle-ci sera un vecteur de 0 et/ou de 1 de taille $|X|$. Les 1 représentent les valeurs observées ou possibles et les 0 les valeurs de X qui sont impossibles. Par exemple, si une variable X peut prendre trois valeurs (x_1, x_2, x_3) et si l'on découvre que, d'après nos observations, seules les valeurs x_1 et x_3 sont possibles, alors $E_X = (1, 0, 1)$.

Revenons à notre matrice de taille $|X| \times |Y_1|$. On veut toujours éliminer la dimension en X. Tous les λ_{Z_iX} ainsi que E_X ont pour taille $|X|$. On va donc constituer un vecteur de taille $|X|$ intégrant toutes leurs informations, et puis on va multiplier ce vecteur par la matrice de taille $|X||Y_1|$. On obtiendra alors un vecteur de taille Y_1, qui a en outre le bon goût de renfermer exactement les informations que l'on veut faire passer à Y_1. Donc :

$$\lambda_{XY_1} = \Big(\Big(\Big(\llbracket \mathrm{Prob}(x|y_1,\ldots,y_n)\rrbracket_{XY_1\ldots Y_n} \otimes \pi_{Y_2X}\Big) \otimes \cdots\Big) \otimes \pi_{Y_nX}\Big) \otimes \big(\lambda_{Z_1X} \odot \lambda_{Z_2X} \odot \cdots \odot \lambda_{Z_pX} \odot E_X\big)\,.$$

De la même manière, les équations aux dimensions nous montrent que

$$\pi_{XZ_1} = \Big(\Big(\Big(\llbracket \mathrm{Prob}(x|y_1,\ldots,y_n)\rrbracket_{XY_1\ldots Y_n} \otimes \pi_{Y_1X}\Big) \otimes \cdots\Big) \otimes \pi_{Y_nX}\Big) \odot \lambda_{Z_2X} \odot \cdots \odot \lambda_{Z_pX} \odot E_X\,.$$

Voilà, maintenant vous savez exactement comment réaliser le mécanisme de propagation d'information dans les réseaux bayésiens : ce sont les algorithmes 1 et 2 et les messages décrits ci-dessus.

5 Échantillonnage et variables aléatoires

Dans tout ce que nous avons vu jusqu'à maintenant, nous avons toujours considéré que toutes les données concernant une variable X ou Y étaient disponibles sur l'ensemble de la population étudiée. Or, bien souvent ce n'est pas le cas : la plupart du temps, les populations sont de tailles trop importantes pour que l'on puisse connaître la valeur prise par la variable pour chacun des individus. Par exemple, pour estimer quel homme (ou femme) politique va gagner aux prochaines élections présidentielles, on ne peut demander à chaque individu de la population française ses intentions de votes (cela demanderait 60 millions de questionnaires !). C'est pourquoi, en statistiques, on se borne à étudier un échantillon de la population :

Définition 25 : *Un échantillon d'une population est un sous-ensemble représentatif de la population.*

L'introduction des échantillons pose trois problèmes majeurs :
- Comment déterminer un échantillon ?
- À partir d'une population, que peut-on dire sur un échantillon ?
- À partir d'un échantillon, que peut-on dire de la population ?

La réponse à la première question sera traitée dans la section 5.1. La deuxième question sera examinée dans les sections 5.2 et 6. Elle peut être particulièrement intéressante dans le contrôle qualité. La troisième fera l'objet de la section 7, et s'applique particulièrement bien aux problèmes de sondage.

Exemple 21 : Supposons qu'une entreprise fabrique des roulements à billes d'un diamètre de 10cm et les distribue par paquets de 10000. Les machines ne produisent jamais exactement les mêmes pièces, aussi peut-il arriver qu'un roulement soit défectueux. On accepte qu'un paquet puisse avoir jusqu'à 3 roulements défectueux, mais au delà, les clients deviennent mécontents. D'après les caractéristiques des machines, on sait que le pourcentage de paquets «défectueux» est en moyenne de 0,2% avec un écart-type de 0,003%. En théorie, l'entreprise devrait donc tester tous les paquets qu'elle envoie à ses clients, mais cela s'avérerait trop coûteux. Elle va donc procéder à des contrôles statistiques : l'ensemble des paquets produits par l'entreprise forme la population statistique. À partir de cette population, on peut déduire théoriquement le nombre de roulements défectueux que l'on devrait observer si l'on testait quelques paquets (échantillon) choisis au hasard dans la population. On en teste 100 et on compare par rapport au résultat théorique. On peut alors en déduire le nombre de paquets contenant plus de 3 roulements défectueux. ♦

5.1 Prélèvement d'un échantillon

La qualité essentielle d'un échantillon est que sa composition soit due au hasard. Idéalement, si on assimilait la population à une énorme urne contenant N boules, on devrait construire les échantillons en choisissant au hasard n boules. Un échantillon créé de cette manière (qui assure à chaque individu de la population la même chance d'appartenir à l'échantillon) s'appelle un échantillon aléatoire.

Définition 26 : *Il y a deux façons de prélever un échantillon aléatoire de n individus parmi une population de N individus :*

1. *On peut commencer par choisir un individu au hasard, noter la valeur que prend la variable pour celui-ci, remettre l'individu dans la population et réitérer le processus. L'échantillon aléatoire est alors dit avec remise. Il est alors possible que le même individu apparaisse plusieurs fois dans l'échantillon.*
2. *On peut aussi utiliser le même procédé mais ne pas remettre dans la population les individus après qu'ils aient été selectionnés. On dit alors que l'échantillon aléatoire est sans remise.*

Exemple 22 : Soit S une population de 5 individus, A, B, C, D, E, telle que A et D sont fumeurs alors que B, C, E sont non-fumeurs. Considérons que les individus sont des boules dans une urne et tirons au hasard un échantillon aléatoire avec remise de taille 3 : nous tirons donc une première boule, A, nous la reposons dans l'urne ; nous retirons une autre boule, E, que nous reposons à nouveau dans l'urne ; enfin nous tirons une troisième boule, E. Nous avons donc prélevé l'échantillon AEE. À chaque tirage, on tire une boule parmi 5. Il y a donc $5 \times 5 \times 5 = 125$ échantillons possibles :

AAA	BAA	CAA	DAA	EAA
AAB	BAB	CAB	DAB	EAB
AAC	BAC	CAC	DAC	EAC
AAD	BAD	CAD	DAD	EAD
AAE	BAE	CAE	DAE	EAE
ABA	BBA	CBA	DBA	EBA
ABB	BBB	CBB	DBB	EBB
ABC	BBC	CBC	DBC	EBC
ABD	BBD	CBD	DBD	EBD
ABE	BBE	CBE	DBE	EBE
ACA	BCA	CCA	DCA	ECA
ACB	BCB	CCB	DCB	ECB
ACC	BCC	CCC	DCC	ECC
ACD	BCD	CCD	DCD	ECD
ACE	BCE	CCE	DCE	ECE
ADA	BDA	CDA	DDA	EDA
ADB	BDB	CDB	DDB	EDB
ADC	BDC	CDC	DDC	EDC
ADD	BDD	CDD	DDD	EDD
ADE	BDE	CDE	DDE	EDE
AEA	BEA	CEA	DEA	EEA
AEB	BEB	CEB	DEB	EEB
AEC	BEC	CEC	DEC	EEC
AED	BED	CED	DED	EED
AEE	BEE	CEE	DEE	EEE

TAB. 12: Échantillons avec remise

Tirons au hasard un échantillon aléatoire sans remise de taille 3 : on tire une boule parmi les 5 que contient l'urne, disons la boule C. On ne la replace pas dans l'urne. On tire à nouveau une boule de l'urne, qui n'en contient déjà plus 5 mais 4. Après le tirage, il en reste 3. On tire à nouveau une boule de l'urne. Il y a donc $5 \times 4 \times 3 = 60$ échantillons sans remise possibles :

ABC	BAC	CAB	DAB	EAB
ABD	BAD	CAD	DAC	EAC
ABE	BAE	CAE	DAE	EAD
ACB	BCA	CBA	DBA	EBA
ACD	BCD	CBD	DBC	EBC
ACE	BCE	CBE	DBE	EBD
ADB	BDA	CDA	DCA	ECA
ADC	BDC	CDB	DCB	ECB
ADE	BDE	CDE	DCE	ECD
AEB	BEA	CEA	DEA	EDA
AEC	BEC	CEB	DEB	EDB
AED	BED	CED	DEC	EDC

TAB. 13: Échantillons sans remise

♦

Une fois l'échantillon constitué, on va calculer ses tendances centrales et de dispersion, et on va essayer d'en déduire celles de la population entière (ce doit être possible puisque l'échantillon est supposé représenter la population entière). Autrement dit, on va suivre la chronologie suivante :

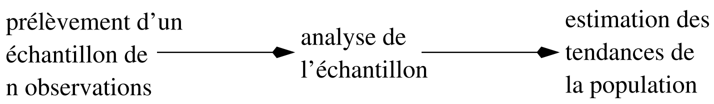


FIG. 24: démarche du statisticien.

5.2 Population et échantillon

Reprenons l'exemple 22. Nous avons tiré 125 échantillons aléatoires avec remise et nous avons noté le nombre de fois où chaque échantillon a été observé :

AAA	0	BAA	0	CAA	0	DAA	1	EAA	1
AAB	1	BAB	2	CAB	2	DAB	1	EAB	2
AAC	0	BAC	1	CAC	2	DAC	2	EAC	0
AAD	2	BAD	1	CAD	0	DAD	1	EAD	1
AAE	1	BAE	0	CAE	1	DAE	1	EAE	0
ABA	2	BBA	0	CBA	3	DBA	1	EBA	0
ABB	1	BBB	1	CBB	0	DBB	0	EBB	0
ABC	0	BBC	0	CBC	2	DBC	2	EBC	1
ABD	2	BBD	3	CBD	1	DBD	0	EBD	0
ABE	1	BBE	0	CBE	2	DBE	1	EBE	1
ACA	0	BCA	2	CCA	0	DCA	3	ECA	0
ACB	3	BCB	2	CCB	1	DCB	1	ECB	2
ACC	0	BCC	1	CCC	0	DCC	0	ECC	0
ACD	3	BCD	0	CCD	1	DCD	0	ECD	0
ACE	2	BCE	3	CCE	0	DCE	0	ECE	1
ADA	1	BDA	3	CDA	2	DDA	3	EDA	2
ADB	0	BDB	1	CDB	1	DDB	0	EDB	1
ADC	1	BDC	2	CDC	3	DDC	0	EDC	1
ADD	0	BDD	0	CDD	1	DDD	0	EDD	0
ADE	0	BDE	2	CDE	0	DDE	3	EDE	1
AEA	0	BEA	1	CEA	0	DEA	2	EEA	0
AEB	2	BEB	0	CEB	3	DEB	1	EEB	0
AEC	1	BEC	1	CEC	2	DEC	1	EEC	2
AED	2	BED	1	CED	1	DED	0	EED	1
AEE	1	BEE	0	CEE	0	DEE	1	EEE	2

Le tableau ci-dessus représente donc un échantillon de la population. La méthode de construction du tableau suit les préceptes de la définition 26, à savoir que l'échantillon est tiré avec remise. Conclusion : il représente la population S. Le problème consiste maintenant à définir un outil nous permettant d'exploiter ce tableau.

Ici, la taille de l'échantillon n'est pas très élevée (taille = effectif = 125). Mais lorsqu'elle l'est, on peut dire, grâce à la loi des grands nombres en probabilité (cf. section 3.4), que la fréquence d'observation de n'importe quel triplet, prenons par exemple AAD, correspond à la probabilité que l'événement «je tire AAD dans la population» soit réalisé (ici, cette probabilité est égale à 2/125).

Considérons maintenant que l'on s'intéresse au nombre de fumeurs dans l'échantillon. On peut utiliser une variable aléatoire X qui à chaque triplet associe le nombre de fumeurs que compte le triplet. Par exemple, X associerait à ABD le nombre 2 car seuls A et D fument. On peut alors calculer la probabilité que la variable aléatoire X soit égale à une valeur valeur quelconque x :

$$\text{Prob}(X = x) = \text{Prob}(X^{-1}(x)) = \frac{\text{nombre de triplets qui ont exactement } x \text{ fumeurs}}{\text{nombre de triplets total}}.$$

On compte donc le nombre d'échantillons qui contiennent 0, 1, 2 ou 3 fumeurs. On obtient alors le résultat suivant :

27 échantillons contiennent 0 fumeur,
54 échantillons en contiennent 1,
36 échantillons contiennent 2 fumeurs et
8 échantillons en contiennent 3.

En utilisant la notation de la section 3.7, ce tableau se résume alors à :

$$\begin{aligned} \mathrm{Prob}(X = 0) &= \tfrac{27}{125}, \\ \mathrm{Prob}(X = 1) &= \tfrac{54}{125}, \\ \mathrm{Prob}(X = 2) &= \tfrac{36}{125}, \\ \mathrm{Prob}(X = 3) &= \tfrac{8}{125}. \end{aligned}$$

Définition 27 : *En statistiques, une* <u>*variable aléatoire*</u> *est une variable pour laquelle on peut déduire (par calcul) une distribution de probabilité, tandis qu'une* <u>*variable statistique*</u> *est une variable pour laquelle on établit* <u>*par observation*</u> *une distribution de probabilité.*

Dans l'exemple ci-dessus, la variable statistique correspond à l'ensemble des 125 triplets de personnes possibles : c'est ce que l'on observe lors de la création de l'échantillon. X est une variable aléatoire puisqu'on calcule sa loi de probabilité à partir de celle de la variable statistique.

Pour arriver à cette loi de probabilité, nous avons supposé que l'on pouvait confondre fréquences et probabilités parce que la taille de l'échantillon était suffisamment grande. On peut maintenant faire la manipulation inverse, à savoir considérer que la loi de probabilité ci-dessus correspond en fait aux fréquences du caractère *«nombre de fumeurs dans un triplet»*. On pourrait donc construire un échantillon de taille 125 contenant 27 «0», 54 «1», 36 «2» et 8 «3». Dans ce cas, on peut utiliser la section 2, et en déduire que le nombre moyen de fumeurs dans les triplets est égal à :

$$\frac{27}{125} \times 0 + \frac{54}{125} \times 1 + \frac{36}{125} \times 2 + \frac{8}{125} \times 3 = 1,2.$$

De même l'écart-type du *«nombre de fumeurs dans un triplet»* serait égal à :

$$\sqrt{\frac{27}{125} \times (0-1,2)^2 + \frac{54}{125} \times (1-1,2)^2 + \frac{36}{125} \times (2-1,2)^2 + \frac{8}{125} \times (3-1,2)^2 = \frac{90}{125}} = \sqrt{0,72} \approx 0,8485.$$

Conclusion : il y a en moyenne 1,2 fumeur par triplet dans l'échantillon, avec un écart-type de 0,8485. Remarquons qu'il y avait 2 fumeurs parmi les 5 personnes de la population, ce qui nous donnait, pour l'ensemble de la population, une moyenne de 1,2 fumeurs par triplet. On voit donc que l'échantillon observé représente «bien» la population totale.

Les petites manipulations auxquelles nous venons de nous adonner reviennent à dire que nous avons calculé la moyenne et l'écart-type de la *variable aléatoire* X comme si celle-ci avait été une *variable statistique*. Dans ce cas, pour bien faire ressortir que X est une variable aléatoire et non une variable statistique, on ne parlera pas de moyenne, mais d' $\boxed{\text{espérance}}$.

Définition 28 : *Soit X une variable aléatoire réelle, c'est-à-dire prenant des valeurs dans $\mathbb{R}$. Soient $\{x_1, \ldots, x_I\}$ ses valeurs. Notons p_i la probabilité que X prenne la valeur x_i, autrement dit $p_i = \mathrm{Prob}(x_i)$. Alors, l'espérance de X, que l'on note $E(X)$, est égale à :*

$$E(X) = \sum_{i=1}^{I} p_i x_i.$$

La variance de X, que l'on note $V(X)$, est égale à :

$$V(X) = \sum_{i=1}^{I} p_i (x_i - E(X))^2 = \sum_{i=1}^{I} p_i x_i^2 - (E(X))^2.$$

On manipule donc de la même manière les variables aléatoires et les variables statistiques. Seules les notations changent :

aléatoire	statistique
E(X)	μ_X
V(X)	σ_X^2

5.3 Techniques pratiques d'échantillonnage

En théorie, sélectionner un échantillon est relativement simple : il faut faire en sorte que tous les individus dans la population aient une chance identique d'appartenir à l'échantillon. Pour cela, il suffit de choisir *aléatoirement* des individus dans l'ensemble de la population. Cependant en pratique, c'est une opération assez difficile à réaliser car faire en sorte que tous les individus aient la même chance d'appartenir à l'échantillon relève de la quadrature du cercle.

Exemple 23 : Un magicien vous présente un jeu de cartes, l'étale devant vous et vous demande d'en choisir une. En théorie chaque carte devrait avoir la même chance d'être sélectionnée. Cependant, si le magicien réalise cette expérience un grand nombre de fois, il s'apercevra que les cartes se trouvant aux extrémités sont plus rarement sélectionnées que celles qui se trouvent vers le milieu du jeu.

Demandez autour de vous à des personnes de choisir un nombre entre 0 et 10. Vous vous apercevrez que le 0 et le 10 sont beaucoup moins choisis que les autres chiffres, alors même qu'en théorie ils ont autant de chances d'être sélectionnés que les autres. ♦

Peut-être l'exemple ci-dessous sera-t-il encore plus convaincant :

Exemple 24 : On mène une étude pour savoir si les français pensent que les personnes gagnant plus de 400000F par an devraient payer plus d'impôts. Allons donc dans un quartier ouvrier. Sélectionnons des gens au hasard et posons-leur cette question. Il y a fort à parier que la réponse sera «oui». Reposons maintenant la même question à des gens choisis aléatoirement dans Neuilly. La réponse tendra plutôt vers le «non». Pourtant nous avons sélectionné les gens au hasard dans les deux cas.

On mène maintenant une enquête pour déterminer le nombre moyen de kilomètres que font les gens en taxi par semaine dans Paris. Le problème, dans ce cas, va être de faire en sorte que les personnes qui ne prennent jamais le taxi aient ni plus ni moins de chances de faire partie de l'échantillon que ceux qui le prennent régulièrement. ♦

Les exemples ci-dessus montrent qu'il faut faire attention lorsque l'on définit la procédure d'échantillonnage : ce n'est pas parce que la théorie des probabilités nous affirme que tous les individus ont autant de chance d'appartenir à l'échantillon qu'en pratique, si l'on créé l'échantillon n'importe comment, l'équiprobabilité sera vérifiée.

5.3.1 Caractériser la population totale

Avant de pouvoir choisir de manière équiprobable les individus qui vont faire partie de l'échantillon, il convient de savoir sur quelle population on travaille. En principe, la population est connue au travers de fichiers de données. Par exemple, si l'on veut estimer les intentions de vote des français lors des prochaines élections présidentielles, on peut connaître l'ensemble de la population grâce aux divers fichiers administratifs (impôts, recensement, numéros INSEE...).

Cependant, avant de sélectionner des individus à partir de ces fichiers, il convient d'examiner ces derniers afin de déterminer s'ils n'auraient pas quelques déficiences. Par exemple, les prochaines élections présidentielles auront lieu en 2002. À cette époque, certaines personnes seront mortes et, logiquement, ne voteront plus ; d'autres, trop jeunes actuellement, auront atteint l'âge légal pour voter. Ainsi, le fichier actuellement disponible ne reflètera pas exactement la population lors du vote de 2002. Il ne représente donc qu'une approximation de la population sur laquelle on devrait faire l'étude.

Exemple 25 : En Angleterre, les sondages d'opinion pour déterminer les résultats d'élections sont effectués sur un nombre relativement restreint de votants. La population est connue grâce à une étude menée chaque année fin novembre, qui détermine la liste des électeurs dans chaque région. Cette liste est publiée au mois de mars suivant, soit 4 mois après. Quand elle est publiée, elle n'est donc pas exactement à jour. Juste avant que la prochaine liste soit publiée, elle est donc périmée de 16 mois. Sachant qu'environ 0,5% de la population déménage et change de région chaque mois (d'après un rapport du Royal Survey), on peut déduire que 8% de la liste des électeurs est périmée. Ajoutons à cela que 4% des gens qui devraient figurer sur la liste n'y figurent pas parce qu'ils n'ont pas rempli certains documents administratifs, et on peut déduire que seulement 88% de la liste est valide. ♦

Donc la première étape pour déterminer un échantillon consiste à caractériser une approximation de la population totale (à travers des fichiers de données), et à estimer la déviation entre cette approximation et la vraie population.

5.3.2 échantillonnage systématique

En pratique, choisir un échantillon aléatoirement n'est pas possible, à moins que l'on ait un «bon» fichier caractérisant la population totale, et que celle-ci soit relativement petite. On se reporte donc le plus souvent à des méthodes qui ne sont pas vraiment aléatoires mais plutôt quasi-aléatoires. Pour cela, une des idées les plus répandues consiste à intégrer dans l'échantillon tous les $n^{\text{ème}}$ éléments du fichier. Par exemple, si on a une population de 100 personnes et que l'on désire récupérer un échantillon de 10 personnes, on peut très bien sélectionner la personne numéro 7, la $17^{\text{ème}}$, la $27^{\text{ème}}$... Cette méthode s'appelle $\boxed{\text{l'échantillonnage systématique}}$.

Cette technique est très populaire en statistiques mais, pour qu'elle donne des résultats corrects, il faut tout de même s'assurer que le fichier caractérisant la population ne contient pas des configurations cycliques : supposons par exemple que l'on fasse une étude immobilière et que le fichier des maisons à notre disposition est tel que toutes les 10 entrées correspondent à des maisons d'angle. Par malchance, on désire utiliser l'échantillonnage systématique afin de créer un échantillon contenant 10% de la population. Dans ce cas, soit toutes les maisons de l'échantillon seront des maisons d'angle, soit il n'y aura aucune maison d'angle. Dans les deux cas, l'échantillon ne représentera pas l'ensemble de la population. Il faut donc prêter une attention toute particulière afin d'éviter les configurations cycliques.

L'échantillonnage systématique est très utilisé dans les cas où la population est homogène. Il peut être utilisé par exemple dans une banque pour essayer de voir si des erreurs ont été commises quant à l'état des comptes des clients.

5.3.3 Échantillonnage stratifié

Il arrive souvent en statistiques que l'on ait à faire des études sur des populations hétérogènes. Par exemple, lorsque l'on demande aux gens s'ils pensent que les personnes gagnant 400000F devraient payer plus d'impôts, la partie aisée de la population répondra «non» tandis que les couches pauvres de la population répondront «oui». Dans un tel cas, $\boxed{\text{l'échantillonnage stratifié}}$ donnera de meilleurs résultats que l'échantillonnage systématique. L'idée consiste à séparer la population hétérogène en un ensemble de sous-populations homogènes et de créer séparément des échantillons dans chacune des sous-populations, puis de rassembler ces échantillons pour former l'échantillon représentant la population totale.

Exemple 26 : On essaye d'estimer l'opinion des étudiants sur l'éducation. Dans notre étude, nous décidons que ces opinions divergent suivant que l'on est lycéen, étudiant à l'université ou dans une école privée. Nous savons que 20% des étudiants sont à l'université, 10% dans des écoles privées et 70% sont lycéens. On veut sélectionner un échantillon de 2000 étudiants. La première étape consiste à dire que 20% de l'échantillon, soit 400 étudiants, doit être constitué d'étudiants de l'université, 200 (10%) doivent provenir d'écoles privées, et 1400 (70%) doivent être lycéens. Dans une deuxième étape, on peut constituer, par la méthode de notre choix, un échantillon pour chaque $\boxed{\text{strate}}$ (pour chaque sous-population). ♦

L'une des applications principales de cette méthode se situe dans l'évaluation de stocks de marchandises. On sépare alors les marchandises en catégories de prix, et on crée des échantillons pour chaque catégorie. Elle sert aussi énormément lorsqu'on essaye d'évaluer l'opinion des gens sur des sujets pour lesquels les caractéristiques sociales ont beaucoup d'importance. Dans tous les cas, ce qui importe pour utiliser cette méthode, c'est d'avoir suffisamment d'informations pour savoir quelles strates l'on doit créer.

5.3.4 Échantillonnage multi-étapes

L'un des problèmes qui se pose souvent en échantillonnage réside dans le fait que la population sur laquelle on travaille est disséminée dans tout un pays ou, pire encore, sur tout un continent. Aussi, la personne qui est chargée de réaliser l'échantillon passe-t-elle plus de temps à voyager qu'à poser aux gens qu'elle a sélectionné les questions relevant de l'enquête qu'elle réalise. Le coût pour créer l'échantillon devient alors tout à fait prohibitif (coût des voyages + paye de la personne qui réalise l'échantillonnage pendant plusieurs semaines parce que les voyages prennent du temps).

L'échantillonnage multi-étapes a été développé pour remédier à cela. Plutôt que de passer du temps à voyager pour effectuer un sondage sur des gens disséminés à travers toute la malaisie, ne serait-il pas plus commode d'interviewer 100 personnes vivant dans 20 régions différentes ? À condition que l'on puisse montrer que l'échantillon ainsi créé est représentatif de la population, on se sera épargné beaucoup d'efforts et on aura économisé pas mal d'argent.

En fait, l'idée de l'échantillonnage multi-étapes est similaire à la stratification, mais on divise la population suivant des critères géographiques plutôt que sur des caractéristiques sociales. Ainsi, la première étape de sélection de l'échantillon consiste à séparer la zone dans laquelle réside la population que l'on étudie en un ensemble de régions. Ensuite, dans chaque région, on sélectionne un échantillon représentatif (de la région). On peut pour cela appliquer une stratification spécifique à la région.

6 Loi binômiale, loi de Poisson et loi normale

Jusqu'à maintenant, lorsque nous avons étudié des variables statistiques, nous connaissions l'ensemble de la population et, par là même, les lois de probabilité auxquelles elles obéissaient. Dans la section précédente, nous avons vu qu'en réalité, les tailles des populations étant souvent très élevées, nous ne pouvions travailler que sur des échantillons. Dans ce cas, les distributions de probabilité de la population ne sont plus connues avec exactitude, et l'un des problèmes auxquels est confronté le statisticien consiste à déterminer ces distributions.

La démarche du statisticien consiste alors à travailler par analogie. Prenons deux exemples : 1) vous voulez déterminer les probabilités que certains numéros sortent à la roulette à une table donnée du casino de Deauville ; 2) vous jouez avec des dés (vous ne savez pas s'ils sont pipés) et vous voulez déterminer la probabilité que les dés tombent sur certaines faces. Ces deux exemples ne sont pas fondamentalement différents : dans les deux cas, pour déterminer les probabilités qui vous intéressent, vous allez répéter des expériences (faire tourner la roulette, jeter les dés), en noter les résultats, et recommencer, le résultat d'une expérience étant indépendant de la précédente. On peut donc se dire qu'il doit exister une classe de distributions de probabilités dépendant d'un certain nombre de paramètres, et que les lois pour la roulette et pour les dés doivent appartenir toutes les deux à cette classe, leur différence résidant uniquement dans la valeur des paramètres.

Dans cette section, nous allons voir les trois classes de distributions les plus utilisées en statistiques : les lois binômiale et de Poisson en ce qui concerne les variables (statistiques ou aléatoires) discrètes, et la loi normale pour les variables continues.

6.1 La loi binômiale

Définition 29 : *On appelle* <u>*épreuve de Bernoulli*</u> *une expérience aléatoire qui ne peut prendre que deux résultats. Ceux-ci ont alors pour nom* succès *et* échec*. Nous noterons p la probabilité de succès, et $q = 1 - p$ la probabilité d'échec.*

Exemple 27 : On sait qu'aux dernières élections, 49% des électeurs ont voté pour le parti A, 38% pour le parti B, 10% pour le parti C, et 3% pour la parti D. L'expérience dans laquelle on choisit au hasard un électeur et on lui demande s'il a voté pour le parti A est une épreuve de Bernoulli. En effet, elle ne peut prendre que 2 valeurs : succès si l'électeur a bien voté pour le parti A, et échec sinon. Ainsi, $p = 49\%$ et $q = 51\%$. De la même manière, l'expérience qui consiste à lui demander s'il a voté pour le parti A ou le parti C est aussi une épreuve de Bernoulli (dont les probabilités p et q sont respectivement égales à $49\% + 10\% = 59\%$ et $100\% - 59\% = 41\%$). Par contre, l'expérience qui consiste à choisir un électeur au hasard et à lui demander pour quel parti il a voté n'est pas une épreuve de Bernoulli car elle peut avoir 4 résultats : parti A, parti B, parti C et parti D. ♦

Exemple 28 : Reprenons notre exemple de la roulette du casino de Deauville. 37 numéros peuvent sortir, mais nous voulons parier sur le 11. Dans ce cas, on peut considérer l'expérience de Bernoulli dans laquelle succès représente le fait que la roulette s'arrête sur le 11, et échec le fait qu'un autre numéro sorte. Alors $p = \frac{1}{37}$ et $q = \frac{36}{37}$. ♦

Définition 30 : *Une* <u>*épreuve binômiale*</u> *est une expérience dans laquelle :*

1. *on répète n fois la même épreuve de Bernoulli,*
2. *les probabilités p et q restent inchangées pour chaque épreuve de Bernoulli,*
3. *les épreuves de Bernoulli sont toutes réalisées indépendamment les unes des autres.*

Exemple 27 (suite) : Reprenons l'exemple 27. L'expérience qui consiste à choisir au hasard dans la rue 40 personnes et à leur demander si elles ont voté pour le parti A est une épreuve binômiale. En effet, 1) on répète 40 fois l'épreuve de Bernoulli de l'exemple 27 ; 2) pour chaque épreuve de Bernoulli, il y a toujours 49% de chances que la personne réponde qu'elle a voté pour le parti A et 51% de chances qu'elle réponde qu'elle a voté pour un autre parti ; 3) les réponses des personnes sont indépendantes les unes des autres. ♦

Exemple 29 : Une entreprise fabrique des machines à laver, et l'on sait que 5% de l'ensemble des machines sont défectueuses. 10 machines sont choisies au hasard sur la chaîne de production et sont examinées afin de déterminer si elles sont défectueuses ou non. Est-ce une épreuve binômiale ? 1) On répète 10 fois la même expérience. De plus, chaque expérience ne peut avoir que deux résultats possibles : soit la machine est défectueuse (succès), soit elle ne l'est pas (échec) ; 2) chaque machine a $p = 5\%$ de chances d'être défectueuse, et $q = 95\%$ de chances d'être bonne ; 3) le fait qu'une machine soit défectueuse ou non n'a aucune incidence sur l'état des autres machines, les épreuves de Bernoulli sont donc toutes indépendantes. Conclusion : on a bien affaire à une épreuve binômiale. ♦

Exemple 22 (suite) : Reprenons l'exemple des fumeurs de la page 55 : dans une population de 5 individus, A, B, C, D, E, seuls A et D fument. On réalise un prélèvement sans remise (cf. tableau 13) de 3 individus parmi A, B, C, D et E, et l'on s'intéresse au nombre de fumeurs que l'échantillon contient. Est-ce une épreuve binômiale ? 1) On peut considérer que le tirage de chacun des 3 individus de l'échantillon est une épreuve de Bernoulli : en effet, c'est le nombre de fumeurs de l'échantillon qui nous intéresse. Donc si l'on pose que succès = fumeur et échec = non fumeur, le tirage ne peut avoir que deux résultats. La constitution de l'échantillon s'apparente donc à une expérience dans laquelle on effectuerait 3 épreuves de Bernoulli. 2) Lors du tirage du premier individu, on a $p = \frac{2}{5}$ de sélectionner un fumeur et $q = \frac{3}{5}$ de sélectionner un non-fumeur. Si l'on a choisi un fumeur, étant donné que l'échantillon est créé sans remise, il n'y a plus qu'une chance sur cinq de sélectionner un fumeur comme deuxième individu de l'échantillon. Sinon, il n'y a plus que deux chances sur cinq pour sélectionner un non-fumeur. Par conséquent, les probabilités p et q des épreuves de Bernoulli ne sont pas toutes identiques. Donc la constitution de l'échantillon n'est pas une épreuve binômiale.

Par contre, si l'on avait constitué l'échantillon avec remise, alors on aurait bien affaire à une épreuve binômiale car les probabilités p et q seraient respectivement égales à $\frac{2}{5}$ et $\frac{3}{5}$, quelles que soient les épreuves de Bernoulli. ♦

> **Définition 31 :** *Soit X la variable aléatoire définie comme le nombre de succès d'une épreuve binômiale constituée de n épreuves de Bernoulli. Une telle variable suit une <u>loi binômiale</u> de paramètres n et p. On le note de la manière suivante : $X \sim \mathrm{Bin}(n; p)$.*

Ainsi, dans l'exemple de l'échantillon de fumeurs avec remise, si X représente le nombre de fumeurs de l'échantillon, $Y \sim \mathrm{Bin}(3; \frac{2}{5})$. Dans l'exemple de l'élection, si X représente le nombre d'individus, parmi les 100 interrogés, qui ont déclaré avoir voté pour le parti A, alors $X \sim \mathrm{Bin}(100; 0,49)$.

Passons maintenant à l'expression mathématique de la loi binômiale. X représente le nombre de succès obtenus après la réalisation de n épreuves de Bernoulli. Donc la variable aléatoire X peut prendre les valeurs 0 (aucune épreuve n'a été couronnée de succès), 1 (une seule épreuve a eu pour résultat succès, les autres ayant eu pour résultat échec), 2,..., n. Établir l'expression de la loi binômiale consiste à déterminer les probabilités $\mathrm{Prob}(X = 0), \dots, \mathrm{Prob}(X = n)$.

Lorsque X vaut 0, toutes les épreuves ont eu pour résultat un échec. Or, la probabilité qu'une épreuve soit un échec est $q = 1 - p$. De plus, toutes les épreuves sont indépendantes les unes des autres, donc, d'après la définition 21,

$$\mathrm{Prob}(X = 0) = \underbrace{q \times q \times \cdots \times q}_{n \text{ fois}} = q^n.$$

De la même manière, il est évident que :

$$\mathrm{Prob}(X = n) = \underbrace{p \times p \times \cdots \times p}_{n \text{ fois}} = p^n.$$

Pour les autres valeurs de X, la formule est plus complexe. $\mathrm{Prob}(X = k)$, $k \neq 0, n$, est égale à la probabilité que k épreuves de Bernoulli aient été des succès et que $n - k$ aient été des échecs. Supposons que $k = 2$ et $n = 4$. Notons S_i le fait que la $i^{\text{ème}}$ épreuve soit un succès et E_i le cas contraire. Le résultat de l'épreuve binômiale peut alors s'exprimer sous la forme d'un quadruplet recensant les succès et échecs de chacune des épreuves de Bernoulli. Les seules possibilités pour avoir $k = 2$ sont les suivantes : (S_1, S_2, E_3, E_4), (S_1, E_2, S_3, E_4),

(S_1, E_2, E_3, S_4), (E_1, S_2, S_3, E_4), (E_1, S_2, E_3, S_4) et (E_1, E_2, S_3, S_4). La probabilité d'obtenir chacune de ces 6 situations est bien évidemment $p^2 \times q^2$. Donc $\text{Prob}(X = 2) = 6 \times p^2 \times q^2$. Cet exemple montre que le calcul de $\text{Prob}(X = k)$ se ramène au produit du nombre de situations pour lesquelles on a bien k succès par $p^k \times q^{n-k}$.

Considérons que l'épreuve binômiale est un vecteur de n cases dans lesquelles on va placer successivement les résultats des différentes épreuves de Bernoulli. On va d'abord placer les k épreuves couronnées de succès. On peut placer la première dans n cases du vecteurs. Une fois qu'on a trouvé un emplacement pour la première épreuve de Bernoulli, il reste $n - 1$ cases. Donc on a $n - 1$ possibilités pour placer la deuxième. De manière analogue, il ne nous reste plus que $n-2$ possibilités pour la troisième, et ainsi de suite. Le nombre de possibilités pour caser les k épreuves est donc :

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

La formule ci-dessus pose néanmoins un problème : de la manière avec laquelle on a choisi les cases du vecteur, on compte plusieurs fois les mêmes vecteurs : supposons par exemple que $k = 2$ et que ce sont les épreuves 3 et 7 qui ont été couronnées de succès ; dans la formule ci-dessus, on compte une fois le choix case $= 3$ puis case $= 7$, et une fois case $= 7$ puis case $= 3$. La formule ci-dessus compte donc le nombre de possibilités de choisir <u>de manière ordonnée</u> k cases parmi n. Ce qui nous intéresse serait plutôt de choisir k cases parmi n sans tenir compte de l'ordre dans lequel on fait ce choix. Il nous faut donc déterminer le lien entre le nombre d'ensembles de k éléments et le nombre de suites ordonnées de k éléments. On a k possibilités de placer le premier élément ordonné. Une fois qu'il est placé, on a $k - 1$ possibilités pour placer le deuxième, et ainsi de suite. À chaque ensemble de k éléments, on peut donc associer donc $k!$ ensembles ordonnés différents. Par conséquent, le nombre de possibilités d'avoir k épreuves de Bernoulli avec succès parmi les n épreuves est :

$$\frac{n!}{(n-k)!\ k!}$$

La probabilité que nous cherchions est donc la suivante :

Théorème 4 : *Soit X une variable aléatoire suivant une loi binômiale* $\text{Bin}(n; p)$*. Alors*

$$\text{Prob}(X = k) = \begin{cases} \dfrac{n!}{(n-k)!\ k!} \times p^k \times q^{n-k} & \text{si } k \in \{0, \ldots, n\}, \\ 0 & \text{sinon.} \end{cases}$$

Exemple 22 (suite) : Considérons l'épreuve binômiale qui consiste à sélectionner un échantillon avec remise de trois personnes parmi la population A, B, C, D, E. Soit X le nombre de fumeurs de l'échantillon, c'est-à-dire le nombre total de succès (succès = choisir un fumeur) de l'épreuve. On sait que $p = \frac{2}{5}$ car seuls A et D fument. De plus, $n = 3$ puisqu'on ne prélève un échantillon que de 3 personnes. Par conséquent, d'après le théorème 4,

$$\begin{aligned} \text{Prob}(X = 0) &= \frac{3!}{3!\ 0!} \times \left(\frac{2}{5}\right)^0 \times \left(\frac{3}{5}\right)^3 = \frac{27}{125} \\ \text{Prob}(X = 1) &= \frac{3!}{2!\ 1!} \times \left(\frac{2}{5}\right)^1 \times \left(\frac{3}{5}\right)^2 = \frac{54}{125} \\ \text{Prob}(X = 2) &= \frac{3!}{1!\ 2!} \times \left(\frac{2}{5}\right)^2 \times \left(\frac{3}{5}\right)^1 = \frac{36}{125} \\ \text{Prob}(X = 3) &= \frac{3!}{0!\ 3!} \times \left(\frac{2}{5}\right)^3 \times \left(\frac{3}{5}\right)^0 = \frac{8}{125} \end{aligned}$$

On remarquera que ce résultat théorique concorde exactement avec ce que nous avions trouvé en réalisant l'expérience (cf. page 58). ♦

Exemple 30 : Une pizzeria propose à ses clients de leur livrer des pizzas en moins de trente minutes. Lorsque le temps imparti est écoulé, l'entreprise ne fait pas payer le client. Malgré tous ses efforts, en moyenne 2% des pizzas arrivent après la demi-heure fatidique. 10 commandes viennent d'être passées à la pizzeria, qui vont être transportées par 10 livreurs différents. Quelle est la probabilité qu'exactement une pizza arrive en retard ?

On peut considérer que l'on a affaire à une épreuve binômiale. En effet, chaque pizza livrée est une épreuve de Bernoulli (succès = arrivée avant une demi-heure). Chaque livreur a une vitesse indépendante de celle des autres, et a une probabilité de 98% de livrer à temps sa pizza. Ici $p = 98\%$, $n = 10$ et $k = 9$. Donc, la probabilité qu'exactement une pizza arrive en retard est :

$$\text{Prob}(X = 9) = \frac{10!}{1!\ 9!} \times 0,98^9 \times 0,02^1 = 0,1667.$$

Il y a donc 16,67 chances sur 100 qu'une seule pizza soit livrée en retard.

Quelle est la probabilité qu'au plus une pizza arrive après une demi-heure ? Dans ce cas là on cherche $\text{Prob}(X \geq 9) = \text{Prob}(X = 9) + \text{Prob}(X = 10)$. Or,

$$\text{Prob}(X = 10) = \frac{10!}{0!\ 10!} \times 0,98^{10} \times 0,02^0 = 0,8171.$$

Donc la probabilité d'avoir au plus une pizza livrée en retard est de $16,67\% + 81,71\% = 98,38\%$. ♦

Après avoir donné l'expression mathématique de la loi de probabilité d'une variable aléatoire suivant une loi binômiale, un réflexe de statisticien nous pousse furieusement à faire de même avec l'espérance et la variance. Bien entendu, les définitions de ces deux notions n'ont pas changé : soit X une variable aléatoire obéissant à une loi binômiale $\text{Bin}(n; p)$. Alors :

$$E(X) = \sum_{k=0}^{n} k \times \text{Prob}(X = k) = \sum_{k=0}^{n} k \times \frac{n!}{(n-k)!\ k!} \times p^k \times q^{n-k},$$
$$V(X) = \sum_{k=0}^{n} k^2 \times \text{Prob}(X = k) - (E(X))^2 = \sum_{k=0}^{n} k^2 \times \frac{n!}{(n-k)!\ k!} \times p^k \times q^{n-k} - (E(X))^2.$$

Après un calcul un peu complexe et franchement sans intérêt, on montre que ces formules se simplifient et l'on obtient :

Théorème 5 (espérance et variance d'une distribution binômiale) : *Soit X une variable aléatoire suivant une loi binômiale* $\text{Bin}(n; p)$. *Alors $E(X) = np$ et $V(X) = npq$.*

Exemple 31 : Selon une enquête publiée dans le *Time* du 8 novembre 1993, 58% des américains pensent que le clônage humain est une mauvaise chose d'un point de vue moral. On choisit un échantillon de 25 américains. Notons par X la variable aléatoire représentant le nombre de personnes dans cet échantillon qui partagent ce point de vue. On se demande quelle est la moyenne et la variance de X.

Puisque la population américaine est assez grande, on peut assimiler l'échantillon de 25 personnes (qui est manifestement sans remise) à un échantillon avec remise. Dans ce cas, la création de l'échantillon peut se décrire comme une épreuve binômiale. Donc la variable X suit une loi binômiale et on peut appliquer le théorème 5. On a alors :

$$n = 25, \qquad p = 58\%, \qquad q = 42\%, \qquad E(X) = np = 14,5 \qquad V(X) = npq = 6,09.$$

En moyenne, il y a donc 14,5 personnes dans l'échantillon qui pensent que le clônage, c'est mal ! ♦

La loi binômiale, outre le fait qu'elle s'applique dans beaucoup de situations, est une loi sympathique car elle possède de bonnes propriétés. Par exemple,

Théorème 6 : *Soient X et Y deux variables aléatoires suivant respectivement les lois binômiales* $\text{Bin}(n; p)$ *et* $\text{Bin}(m; p)$. *Si X et Y sont indépendantes, c'est-à-dire que la valeur que prend l'une des variables ne dépend absolument pas de la valeur que prend l'autre, alors la variable $Z = X + Y$, qui recense le nombre total de succès des deux épreuves relatives à X et Y, obéit aussi à une loi binômiale : $Z \sim$* $\text{Bin}(n + m; p)$.

Exemple 22 (suite) : Reprenons notre exemple des fumeurs, nous avions vu que la sélection d'un échantillon de taille 3 avec remise était une épreuve binômiale et que si X était la variable correspondant au nombre de fumeurs dans l'échantillon, alors $X \sim \text{Bin}(3; \frac{2}{5})$. Considérons maintenant que l'on réalise de manière tout à fait indépendante 3 épreuves binômiales consistant à sélectionner un échantillon de taille 1. Notons Y le nombre de fumeurs à l'issue de la première épreuve, Z pour la deuxième et T pour la troisième. On a visiblement $Y \sim \text{Bin}(1; \frac{2}{5})$, $Z \sim \text{Bin}(1; \frac{2}{5})$ et $T \sim \text{Bin}(1; \frac{2}{5})$. Les épreuves étant réalisées indépendamment les unes des autres et les probabilités de succès de toutes ces épreuves étant identiques, on peut appliquer le théorème ci-dessus. Par conséquent $Y + Z + T \sim \text{Bin}(3; \frac{2}{5})$, ce qui correspond bien au nombre de fumeurs que nous avions trouvé pour un échantillon de taille 3. ♦

6.2 La loi de Poisson

La loi de Poisson est une autre distribution de probabilité concernant les variables aléatoires discrètes. Elle est aussi importante car elle a de nombreuses applications. Par exemple, supposons que, dans une laverie automatique, une machine à laver tombe en panne en moyenne trois fois par mois. Quelle est la probabilité qu'elle tombe en panne exactement deux fois au cours du prochain mois. Ce problème est un exemple typique de distribution de Poisson. Dans le jargon Poisonnien, les pannes sont appelées $\boxed{\text{occurrences}}$.

La <u>loi de Poisson</u> s'applique dans le cas d'expériences dans lesquelles les occurrences sont totalement aléatoires (sans régularité) et indépendantes les unes des autres.

Par totalement aléatoire, on entend qu'elles ne suivent pas un modèle donné, et qu'elles sont par conséquent totalement imprévisibles. À l'instar de la loi binômiale, une occurrence n'a aucune influence sur les autres occurrences, c'est pour cela qu'elles sont dites indépendantes. On considère toujours les occurrences sur un certain $\boxed{\text{intervalle}}$. Dans l'exemple des machines à laver, cet intervalle est d'un mois. D'une manière générale, l'intervalle considéré peut être temporel, spatial, ou encore volumique. Dans la loi de Poisson, le nombre exact d'occurrences dans un intervalle donné est inconnu. Mais si l'on connaît le nombre moyen d'occurrences pour un intervalle donné, alors on peut calculer la probabilité qu'il y ait x occurrences dans cet intervalle.

Exemple 32 : Considérons le nombre de patients arrivant au service des urgences d'un hôpital pendant un intervalle d'une heure. Dans cet exemple, une occurrence représente l'arrivée d'un patient, l'intervalle est une heure (intervalle temporel). Il est bien évident que les occurrences sont aléatoires : on ne peut déterminer à l'avance quand les patients vont arriver. Elles sont aussi indépendantes : l'arrivée d'un patient n'a pas de relation avec les arrivées des autres patients : les patients arrivent individuellement à l'hôpital. ♦

Exemple 33 : La loi de Poisson s'applique aussi pour quantifier le nombre d'accidents qui vont se produire sur une autoroute sur une période donnée, le nombre de clients qui vont venir dans une épicerie pendant une période d'une semaine, ou bien encore le nombre de télévisions vendues par un magasin sur une semaine.

Par contre, la loi de Poisson ne s'applique pas au nombre de patients venant consulter un médecin généraliste. En effet, les arrivées des patients ne sont pas aléatoires puisqu'elles sont planifiées avec le carnet de rendez-vous du docteur. De même, l'arrivée d'avions sur un aéroport ne peut être modélisée par une loi de Poisson car tous les avions sont planifiés pour arriver à certaines heures : la tour de contrôle connaît le nombre d'arrivées pour chaque période. ♦

Définition 32 : *D'après la loi de Poisson, la probabilité d'avoir x occurrences dans l'intervalle est :*

$$\text{Prob}(x) = \frac{\lambda^x \; e^{-\lambda}}{x!},$$

où λ est le nombre moyen d'occurrences dans l'intervalle.

Exemple 34 : Retournons dans la laverie automatique. La machine à laver tombe en panne en moyenne trois fois par mois. Utilisons la loi de Poisson pour déterminer la probabilité qu'elle va être en panne exactement deux fois le mois prochain. λ = le nombre moyen de pannes par mois = 3. Par conséquent,

$$\text{Prob}(x = 2) = \frac{\lambda^x \; e^{-\lambda}}{x!} = \frac{3^2 \; e^{-3}}{2!} \approx 0,224.$$

Il y a donc à peu près 22 chances sur 100 pour que la machine tombe en panne exactement deux fois le mois prochain. Calculons maintenant la probabilité qu'elle tombe en panne au plus une fois.

$$\text{Prob}(x = 0 \text{ ou } 1) = \text{Prob}(x = 0) + \text{Prob}(x = 1) = \frac{3^0\ e^{-3}}{0!} + \frac{3^1\ e^{-3}}{1!} \approx 0,1992.$$

♦

Exemple 35 : Dans un hôpital, environ 30 personnes se présentent en moyenne au service des urgences le vendredi soir entre 19h et 20h. Bien évidemment, la loi suivie par la variable X =«nombre de personnes se présentant entre 19h et 20h» est une loi de Poisson. Par conséquent, si l'on veut connaître la probabilité qu'un vendredi donné, exactement 15 personnes se présentent au service des urgences, il suffit de calculer la valeur de l'expression suivante :

$$\text{Prob}(X = 15) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{30^{15} e^{-30}}{15!}.$$

De même, la probabilité que le nombre de personnes se présentant aux urgences soit compris entre 10 et 20 est égal à :

$$\text{Prob}(10 \leq X \leq 20) = \sum_{i=10}^{20} \frac{30^i e^{-30}}{i!}.$$

♦

Théorème 7 : *Soit X une variable aléatoire suivant une loi de Poisson. Alors, la moyenne de X est $\mu = \lambda$, et sa variance est $\sigma^2 = \lambda$.*

6.3 La loi normale

Bien entendu, les variables aléatoires discrètes ne sont pas les seules à avoir des classes de distributions de probabilités importantes, les variables continues en ont aussi. Certainement celle qui est la plus connue et la plus utilisée est la loi normale.

Définition 33 : *Une variable aléatoire continue X obéit à une loi normale de paramètres μ et σ^2, ce que l'on note $X \sim \text{N}(\mu; \sigma^2)$, si sa distribution de probabilité est donnée par la courbe délimitée par la fonction de densité :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \textit{pour tout } x \in \mathbb{R}.$$

En fait, cette définition n'est pratiquement jamais utilisée car nous allons voir que la loi normale a des propriétés tout à fait intéressantes qui nous permettront de la calculer à partir de tables statistiques. Par contre, il est intéressant de s'attarder un moment sur la représentation graphique de la fonction de densité :

La distribution normale est en forme de cloche, quels que soient μ et σ. Le centre de symétrie est μ et c'est en ce point que la courbe atteint son maximum. Par conséquent, la moyenne et la médiane d'une variable suivant une loi normale sont égales à μ. La fonction $f(x)$ n'est jamais nulle, mais elle devient extrêmement petite lorsque l'on s'écarte de plus de 3σ de la moyenne. En effet, un calcul intégral montre que :

intervalle	aire sous la courbe dans l'intervalle
$[\mu - \sigma, \mu + \sigma]$	0,683
$[\mu - 2\sigma, \mu + 2\sigma]$	0,954
$[\mu - 3\sigma, \mu + 3\sigma]$	0,997

Nous avons déjà vu que l'aire sous la courbe pour un intervalle donné $[a, b]$ nous donne la probabilité que X prenne sa valeur dans l'intervalle $[a, b]$. Par conséquent, il y a 99,7% de chances que X prenne sa valeur entre $\mu - 3\sigma$ et $\mu + 3\sigma$. Remarquons que c'est bien plus que ce que Bienaymé-Tchébicheff pouvaient prédire : eux pouvaient seulement dire que la probabilité de se situer entre $\mu - 3\sigma$ et $\mu + 3\sigma$ devait au moins être égale à 88,9%.

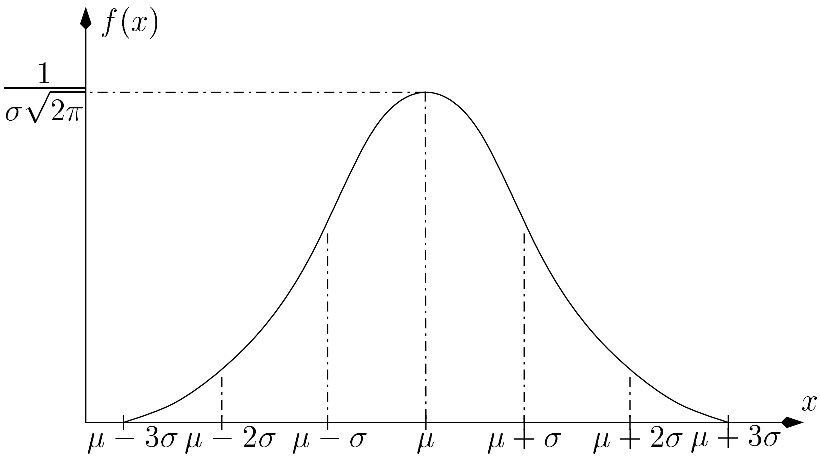


FIG. 25 – Fonction de densité de la loi normale.

6.3.1 Loi normale centrée réduite

Comme nous l'avions fait pour la moyenne et la variance, voyons ce qui se passe lorsqu'une variable aléatoire est une transformée affine d'une autre :

Théorème 8 : *Soit X une variable aléatoire obéissant à une loi* $\mathrm{N}(\mu;\sigma^2)$. *Alors la variable $Y = aX + b$ obéit à une loi* $\mathrm{N}(a\mu + b; a^2\sigma^2)$.

Ce théorème signifie que toute variable résultant d'une transformation affine d'une variable aléatoire obéissant à une loi normale est aussi une variable aléatoire obéissant à une loi normale. Les paramètres de celle-ci se déduisent des résultats que nous avions déjà établis dans les sections précédentes :

$$E(aX + b) = aE(X) + b \qquad \text{et} \qquad V(aX + b) = a^2V(X).$$

L'importance de ce théorème nous est donnée par le corollaire suivant, dans lequel la variable Z et égale à $aX + b$, où $a = 1/\sigma$ et $b = -\mu/\sigma$.

Corollaire 2 : *Soit X une variable aléatoire obéissant à une loi* $\mathrm{N}(\mu;\sigma^2)$. *Alors* $Z = \dfrac{X-\mu}{\sigma}$ *suit la loi* $\mathrm{N}(0;1)$. *On dit alors que Z suit une loi normale centrée (à cause de la moyenne en 0) réduite (à cause du σ^2 égal à 1).*

Ce corollaire nous permet de calculer la probabilité pour qu'une variable $X \sim \mathrm{N}(\mu;\sigma^2)$ prenne sa valeur dans un intervalle donné en ramenant cette probabilité à une probabilité formulée sur une variable $Z \sim \mathrm{N}(0;1)$. D'une manière générale, pour $X \sim \mathrm{N}(\mu;\sigma^2)$,

$$\begin{aligned}
\mathrm{Prob}(a \le X \le b) &= \mathrm{Prob}(a - \mu \le X - \mu \le b - \mu) \\
&= \mathrm{Prob}\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\
&= \mathrm{Prob}\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right).
\end{aligned}$$

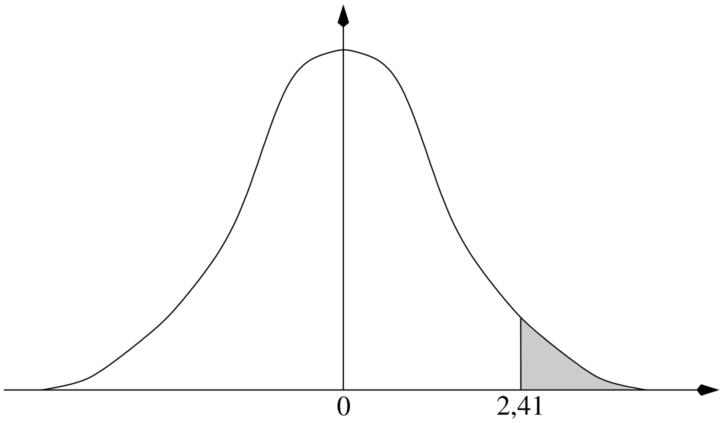
On peut alors utiliser la table de la loi normale centrée réduite située à la fin de cette section pour calculer cette probabilité.

6.3.2 Utilisation de la table de la loi normale centrée réduite

Cette table nous donne l'aire de la courbe normale centrée réduite à droite d'un point Z_α (c'est-à-dire la quantité $\mathrm{Prob}(Z \ge Z_\alpha)$), où Z_α est un nombre positif ou nul. À partir de là on peut virtuellement calculer n'importe quelle probabilité, comme le montrent les exemples suivants.

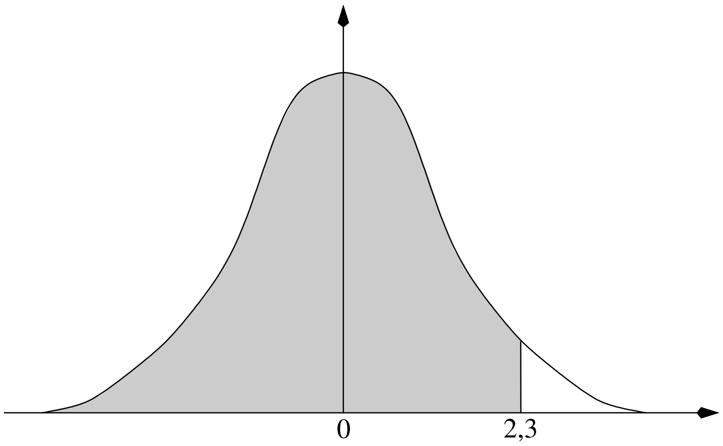
Exemple 36 : Soit $Z \sim \mathrm{N}(0;1)$.

Calculons $\mathrm{Prob}(Z \geq 2,41)$. Cela correspond à l'aire grisée sur la figure ci-dessous :

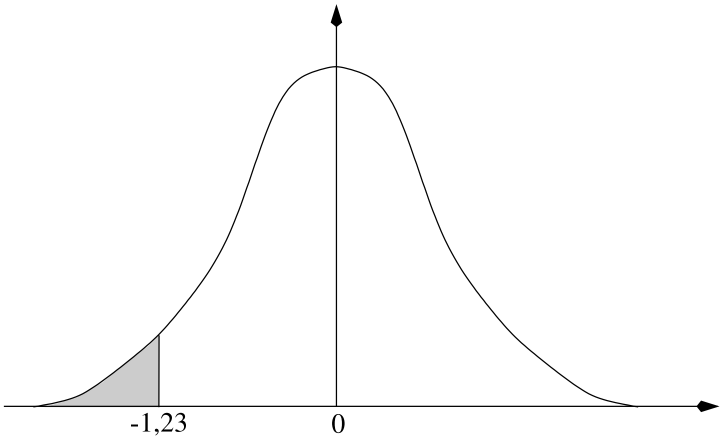


D'après la table page 80, à l'intersection de la ligne 2.4 et de la colonne 0,01, on obtient $\mathrm{Prob}(Z \geq 2,41) = 0,0080$.

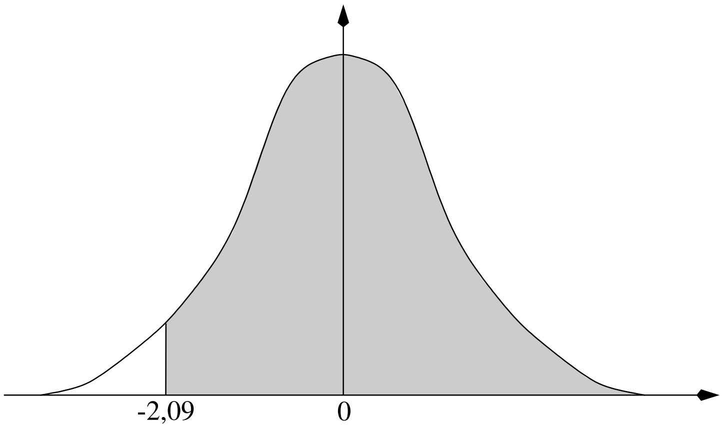
Calculons $\mathrm{Prob}(Z \leq 2,30)$. On remarque que $\mathrm{Prob}(Z \leq 2,30) = 1 - \mathrm{Prob}(Z > 2,30)$. Or, on peut lire $\mathrm{Prob}(Z > 2,30)$ dans la table. Cette probabilité vaut $0,0107$. Donc $\mathrm{Prob}(Z \leq 2,30) = 1 - 0,0107 = 0,9893$. Cela correspond à l'aire grisée sur la figure ci-dessous :



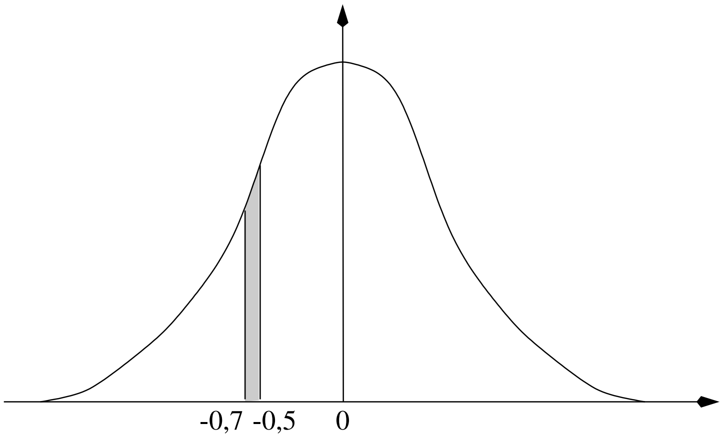
Pour calculer $\mathrm{Prob}(Z \leq -1,23)$, il suffit de constater que $\mathrm{Prob}(Z \leq -1,23) = \mathrm{Prob}(Z \geq 1,23) = 0,1093$. Cela correspond à l'aire grisée ci-dessous :



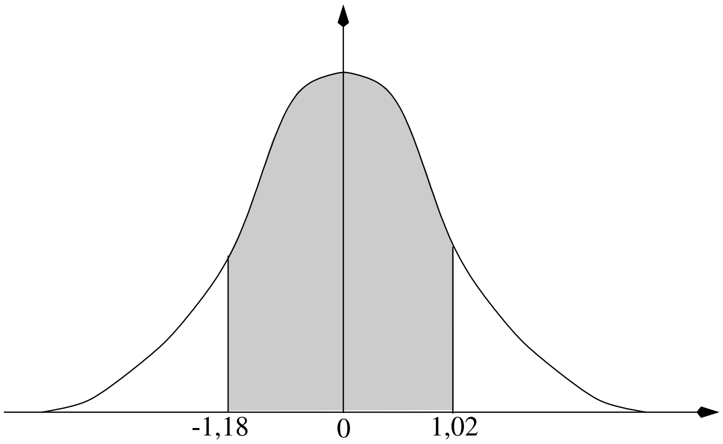
Pour calculer Prob$(Z \geq -2,09)$, on constate que Prob$(Z \geq -2,09) = $ Prob$(Z \leq 2,09) = 1 - $ Prob$(Z > 2,09) = 1 - 0,0183 = 0,9817$.



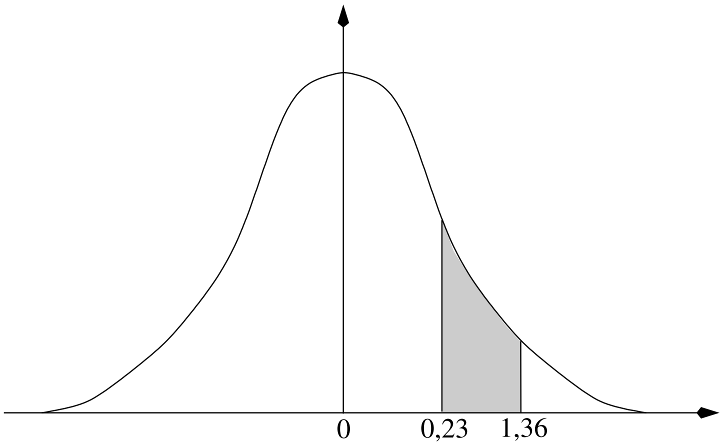
Pour calculer Prob$(-0,70 \leq Z \leq -0,50)$, on constate que Prob$(-0,70 \leq Z \leq -0,50) = $ Prob$(Z \leq -0,50) - $ Prob$(Z < -0,70) = $ Prob$(Z \geq 0,50) - $ Prob$(Z > 0,70) = 0,3085 - 0,2420 = 0,0665$.



Pour calculer Prob$(-1,18 \leq Z \leq 1,02)$, on remarque que Prob$(-1,18 \leq Z \leq 1,02) = 1 - $ Prob$(Z < -1,18) - $ Prob$(Z > 1,02) = 1 - $ Prob$(Z > 1,18) - $ Prob$(Z > 1,02) = 1 - 0,1190 - 0,1539 = 0,7271$.



Pour calculer $\text{Prob}(0,23 \leq Z \leq 1,36)$, on remarque que $\text{Prob}(0,23 \leq Z \leq 1,36) = \text{Prob}(Z \geq 0,23) - \text{Prob}(Z > 1,36) = 0,4090 - 0,0859 = 0,3231$.



♦

À partir des calculs de probabilité sur les variables obéissant aux lois normales centrées réduites, on peut calculer n'importe quelle probabilité de n'importe quelle variable obéissant à une loi normale.

Exemple 37 : Soit $X \sim \text{N}(-4,4)$. Calculons $\text{Prob}(-4,2 \leq X \leq 1,6)$. Pour cela, utilisons le fait que $Z = \frac{X-(-4)}{2} = \frac{X+4}{2} \sim \text{N}(0;1)$. Alors

$$\begin{aligned}
\text{Prob}(-4,2 \leq X \leq 1,6) &= \text{Prob}\left(\frac{-4,2+4}{2} \leq \frac{X+4}{2} \leq \frac{1,6+4}{2}\right) \\
&= \text{Prob}(-0,1 \leq Z \leq 2,8) \\
&= 1 - \text{Prob}(Z < -0,1) - \text{Prob}(Z > 2,8) \\
&= 1 - \text{Prob}(Z > 0,1) - \text{Prob}(Z > 2,8) \\
&= 1 - 0,4602 - 0,0026 = 0,5372.
\end{aligned}$$

♦

La table de la loi normale centrée réduite peut aussi nous permettre de calculer les quantiles d'ordre α de variables aléatoires. Soit $X \sim \text{N}(\mu;\sigma^2)$ une variable aléatoire. Soit x_α le quantile d'ordre α, c'est-à-dire le nombre tel que $\text{Prob}(X < x_\alpha) = 1-\alpha$. Pour des raisons de commodité, étant donné le sens de la table de la loi normale centrée réduite, nous nous référerons au quantile d'ordre $1-\alpha$ de X comme étant le nombre x_α tel que $\text{Prob}(X \geq x_\alpha) = \alpha$. Ainsi, le 94$^{\text{ème}}$ centile de X sera le nombre $x_{0,06}$ tel que $\text{Prob}(X \geq 0,06) = 0,06$. Pour déterminer un tel nombre, il s'agit de voir le lien entre ce nombre et le quantile du même ordre de la variable centrée réduite $Z = \frac{X-\mu}{\sigma} \sim \text{N}(0;1)$. La variable X peut s'exprimer comme $X = \mu + \sigma Z$. Un quantile étant une mesure de position, nous en déduisons la formule suivante :

$$\text{quantile de } X = \mu + \sigma(\text{quantile de } Z).$$

Autrement dit, $x_\alpha = \mu + \sigma z_\alpha$. Donc, pour calculer le quantile d'ordre α d'une variable $X \sim \text{N}(\mu;\sigma^2)$, il suffit de calculer la valeur du quantile pour la variable centrée réduite Z.

Exemple 38 : Soit $X \sim \text{N}(10;25)$. Déterminons la valeur du deuxième quintile de cette variable, c'est-à-dire un nombre tel que la probabilité pour que la variable X soit inférieure à ce nombre est égale à 40%, et la probabilité de lui être supérieure est de 60%.

Nous cherchons donc $x_{0,60}$. On sait que $x_{0,60} = 10 + 5z_{0,60}$. Reste maintenant à calculer $z_{0,60}$. Par symétrie de la loi $\text{N}(0;1)$, $z_{0,60} = -z_{0,40}$. Or nous observons dans la table de la loi normale centrée réduite que $\text{Prob}(Z \geq 0,25) = 0,4013$ et $\text{Prob}(Z \geq 0,26) = 0,3974$. Nous en déduisons que $z_{0,60}$ est compris entre $0,25$ et $0,26$. Si on utilise une règle de 3 (interpolation linéaire), on obtient $z_{0,60} = 0,253$. Par conséquent, $x_{0,60} = 10 - (5 \times 0,253) = 8,735$.

Il est assez légitime de faire une interpolation linéaire car la fonction de densité étant assez régulière, on peut l'approximer avec des segments de droite. C'est d'ailleurs ce principe qu'utilisent les ordinateurs pour tracer des courbes à l'écran. ♦

6.4 Distribution d'une somme de variables aléatoires suivant une loi normale

Nous avions déjà vu que lorsque l'on réalise une enquête statistique sur une population très disséminée géographiquement, il est pratique de scinder celle-ci en petites sous-populations, de les étudier (calculer leurs moyennes, leurs écarts-type...) séparément, et de synthétiser ces études afin de conclure l'enquête sur la population globale. Concrètement, sur l'exemple du nombre de médecins pour 100000 habitants en France (exemple 6 page 17), on effectue l'étude région par région. On peut ainsi déterminer le nombre moyen de médecins pour chaque région et, ensuite, on regroupe toutes ces données pour obtenir le nombre moyen de médecins sur l'ensemble de la France. Cela revient à dire que sur chaque région on étudie une variable aléatoire X_i, et que la variable aléatoire correspondant à l'ensemble de la France est $Y = \sum_i X_i$ (ou $\sum_i a_i X_i + b_i$). Le problème que nous allons nous poser dans cette sous-section est donc le suivant : «si j'ai I variables aléatoires $X_1, \ldots, X_I$ suivant des lois normales, que puis-je dire sur la variable $Y = X_1 + X_2 + \cdots + X_I$?»

Tout d'abord, voyons ce que l'on peut dire de l'espérance de Y :

Soient $X_1, \ldots, X_I$ des variables aléatoires, et soit Y la variable aléatoire somme de ces I variables, c'est-à-dire $Y = X_1 + X_2 + \cdots X_I$. Alors

$$E(Y) = E(X_1 + X_2 + \cdots + X_I) = E(X_1) + E(X_2) + \cdots + E(X_I).$$

En particulier, cette propriété est vraie lorsque les variables $X_1, \ldots, X_I$ suivent une loi normale.

Cette propriété s'étend aussi à la variance :

Soient $X_1, \ldots, X_I$ des variables aléatoires indépendantes, et soit Y la variable aléatoire somme de ces I variables. Alors

$$V(Y) = V(X_1 + X_2 + \cdots + X_I) = V(X_1) + V(X_2) + \cdots + V(X_I).$$

En particulier, cette propriété est vraie lorsque les variables $X_1, \ldots, X_I$ suivent une loi normale.

Certes, les deux propriétés ci-dessus fournissent des informations sur Y, mais pour les établir, on n'a pas besoin que les variables $X_1, \ldots, X_I$ suivent des lois normales. Voyons maintenant ce que l'on peut précisément déduire de l'information «les X_i suivent une loi normale» :

Théorème 9 : *Soient $X_1, \ldots, X_I$ des variables aléatoires indépendantes suivant des lois normales, et soit Y la variable aléatoire somme de ces I variables. Alors Y suit aussi une loi normale.*

Exemple 39 : Soient $X_1 \sim \mathrm{N}(0; 1)$, $X_2 \sim \mathrm{N}(10; 7)$ et $X_3 \sim \mathrm{N}(5; 20)$ trois variables indépendantes suivant des lois normales, et soit $Y = X_1 + X_2 + X_3$. Le théorème 9 nous dit que, dans ce cas, Y suit aussi une loi normale, autrement dit, il existe μ et σ tels que $Y \sim \mathrm{N}(\mu; \sigma^2)$. Or, d'après les deux propriétés au dessus du théorème 9, nous savons que $\mu = E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n) = 0 + 10 + 5 = 15$, et que $\sigma^2 = V(Y) = 1 + 7 + 20 = 28$. Par conséquent, $Y \sim \mathrm{N}(15, 28)$. ♦

Exemple 40 : Vous vous connectez par modem à Jussieu à partir de votre *home sweet home*, et vous exécutez à distance un logiciel sur une des machines du bâtiment 41. Vous savez que le temps d'exécution du logiciel suit une loi normale $\mathrm{N}(2\mathrm{mn}; 0,25\mathrm{mn}^2)$ (on considérera que le temps d'exécution n'est pas toujours le même car vous utilisez une machine multi-tâches). Le temps de transfert du résultat par modem suit, lui aussi, une loi normale $\mathrm{N}(1\mathrm{mn}; 0,01\mathrm{mn}^2)$.

Vous vous posez alors la question métaphysique suivante : si Y désigne la variable représentant le temps au bout duquel vous obtenez le résultat de l'exécution sur votre ordinateur (chez vous), quelle pourrait bien être la loi suivie par Y ? Si l'on note X_1 la variable représentant le temps d'exécution (en local) du logiciel, et si X_2 est la variable aléatoire représentant le temps de transfert, alors on doit avoir $Y = X_1 + X_2$. En supposant que les temps de transfert ne sont pas liés à la machine qui exécute votre logiciel, on obtient $Y \sim \mathrm{N}((2+1)\mathrm{mn}; (0,25 + 0,01)\mathrm{mn}^2) = \mathrm{N}(3\mathrm{mn}; 0,26\mathrm{mn}^2)$. ♦

Exemple 41 : Un pilote de ligne assure régulièrement le trajet Paris-Montpellier. Il s'est amusé à calculer le temps qu'il passe entre le moment où il part de chez lui (à Paris, huit heures du matin) et le moment où il arrive à l'aéroport de Montpellier. Voici le résultat de ses observations : le temps passé dans le RER pour aller jusqu'à Orly suit une loi normale N(35 min, 9 min^2) ; le temps pour préparer le vol/inspecter l'appareil suit une loi N(1 heure, 16 min^2) ; enfin, le temps de vol suit une loi N(1 heure 10 min, 25 min^2).

Le pilote a décidé de donner rendez-vous à l'aéroport de Montpellier à un de ses collègues. Il ne voudrait pas le fixer trop tôt pour ne pas être en retard, ni le fixer trop tard car cela l'obligerait à attendre. Pour l'aider à choisir l'heure du rendez-vous, calculons la probabilité que le pilote arrive :
1/ entre 10h31 et 10h52.
2/ après 11h.
3/ avant 10h40.
Soient $T1$, T_2, T_3 les variables «temps passé dans le RER», «temps pour préparer le vol» et «temps de vol». On sait que :

$$T_1 \sim N(35 \text{ min}, 9 \text{ min}^2) \qquad T_2 \sim N(60 \text{ min}, 16 \text{ min}^2) \qquad T_3 \sim N(70 \text{ min}, 25 \text{ min}^2).$$

Par conséquent, la variable X = «temps total pour arriver à Montpellier» vérifie les relations suivantes :

$$X = T_1 + T_2 + T_3 \qquad X \sim N((35+60+70) \text{ min}, (9+16+25) \text{ min}^2) = N(165 \text{ min}, 50 \text{ min}^2).$$

Par la suite, tous les calculs seront exprimés en minutes et minutes^2.

$$\begin{aligned}
\text{1/ Prob(10h30} \leq \text{arrivée à Montpellier} \leq \text{10h52)} &= \text{Prob}(151 \leq X \leq 172)\\
&= \text{Prob}\left(\frac{151-165}{\sqrt{50}} \leq \frac{X-165}{\sqrt{50}} \leq \frac{172-165}{\sqrt{50}}\right)\\
&= \text{Prob}\left(-1,98 \leq \frac{X-165}{\sqrt{50}} \leq 0,99\right)\\
&= 1 - 0,0239 - 0,1611 = 0,8150.
\end{aligned}$$

$$\begin{aligned}
\text{2/ Prob(arrivée à Montpellier} \geq \text{11h)} &= \text{Prob}(X \geq 180)\\
&= \text{Prob}\left(\frac{X-165}{\sqrt{50}} \geq \frac{180-165}{\sqrt{50}}\right)\\
&= \text{Prob}\left(\frac{X-165}{\sqrt{50}} \geq 2,12\right)\\
&= 0,017.
\end{aligned}$$

$$\begin{aligned}
\text{3/ Prob(arrivée à Montpellier} \leq \text{10h40)} &= \text{Prob}(X \leq 160)\\
&= \text{Prob}\left(\frac{X-165}{\sqrt{50}} \leq \frac{160-165}{\sqrt{50}}\right)\\
&= \text{Prob}\left(\frac{X-165}{\sqrt{50}} \leq -0,71\right)\\
&= \text{Prob}\left(\frac{X-165}{\sqrt{50}} \geq 0,71\right) = 0,2389.
\end{aligned}$$

♦

Remarquons que le théorème 9 possède un homologue pour la loi binômiale : le théorème 6. Ces deux théorèmes sont relativement intéressants puisqu'ils permettent d'étudier de grosses populations à partir de petites sous-populations. Ils posent toutefois un léger problème : la population globale peut être répartie géographiquement de manière très hétérogène (il y a par exemple fort à parier que la région Ile de France compte une sous-population plus importante que la Corse). Il n'est donc pas forcément très judicieux de représenter la population globale avec la variable $Y = \sum_{i \in I} X_i$: il paraît en effet plus logique de pondérer les X_i en fonction de la taille des sous-populations. Dans ce cas, $Y = \sum_{i \in I} a_i X_i$. On montre que le théorème 9 peut alors s'énoncer sous la forme :

Théorème 10 : *Soient $X_1, \ldots, X_I$ des variables aléatoires indépendantes suivant des lois normales, et soit Y la variable aléatoire définie par $Y = \sum_{i \in I} a_i X_i + b_i$, où $a_1, \ldots, a_I$ sont des nombres réels. Alors Y suit aussi une loi normale.*

Ce théorème découle trivialement des théorèmes 9 et 8. En effet, lorsque l'on multiplie une variable aléatoire $X_i \sim \mathrm{N}(\mu;\sigma^2)$ par a_i et qu'on lui rajoute b_i, d'après le théorème 8, la variable aléatoire $Y_i = a_iX_i + b_i$ résultante suit la loi $\mathrm{N}(a_i\mu + b_i; a_i^2\sigma^2)$. On peut donc appliquer le théorème 9 avec $Y = \sum_{i\in I} Y_i$.

Graphiquement, on voit bien sur la figure ci-dessous, que si l'on diminue le paramètre σ, tout en conservant μ constant, cela revient juste à étirer vers le haut le centre de la courbe de densité, le centre de symétrie reste toutefois au même endroit sur l'axe des x. Si, au contraire, on modifie μ, tout en conservant σ constant, on ne fait que déplacer horizontalement la courbe de densité.

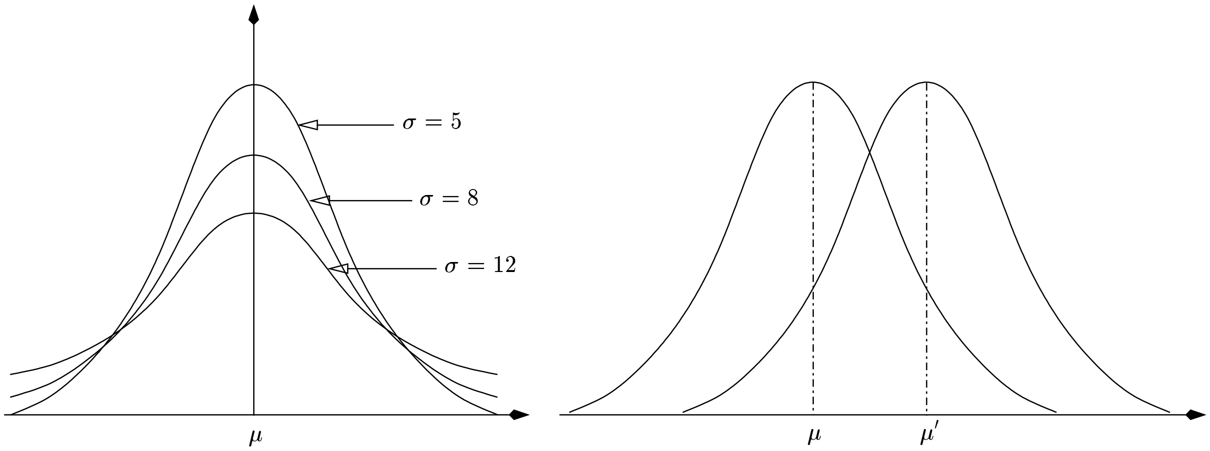


Fig. 26 – Modification des paramètres μ et σ.

6.5 La loi normale comme limite d'autres lois

La loi normale est une loi très importante en statistiques, non seulement parce qu'elle s'applique dans beaucoup de situations où les variables aléatoires sont continues, mais aussi (et peut-être surtout) parce qu'elle est une limite vers laquelle tendent un certain nombre d'autres lois, en particulier des lois concernant des variables discrètes. Ainsi, le théorème ci-dessous, l'un des théorèmes fondamentaux de la statistique, montre que, si l'on a I variables $X_1, \ldots, X_I$, indépendantes et identiquement distribuées, et ce quelle que soit la loi de probabilité qu'elles suivent, alors si I est suffisamment grand, la somme de ces variables suit une loi normale.

Théorème 11 (Théorème central limite) : *Soient $X_1, \ldots, X_I$ des variables aléatoires indépendantes et identiquement distribuées (même loi de probabilité, même espérance, même variance). Soit $Y = X_1 + \cdots + X_I$. Alors, si I est suffisamment grand (en principe $I \geq 30$) :*

$$\frac{Y - E(Y)}{\sqrt{V(Y)}} \sim \mathrm{N}(0;1).$$

Exemple 42 : On réalise l'expérience qui consiste à lancer 4 dés à 6 faces totalement indépendants. On note X_i le résultat du $i^{\text{ème}}$ dé. Les dés ne sont pas pipés, donc chaque X_i suit la distribution :

x_i	1	2	3	4	5	6
$\mathrm{Prob}(X_i = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

On s'intéresse aux distributions de probabilité suivies par les variables aléatoires $Y_i = \sum_{j\leq i} X_i$. On a bien entendu

$$\mathrm{Prob}(Y_i = y_i) = \sum_{x_1+\cdots+x_i=y_i} \mathrm{Prob}(X_1 = x_1, \ldots, X_i = x_i).$$

Et puisque les variables X_i sont indépendantes,

$$\mathrm{Prob}(Y_i = y_i) = \sum_{x_1+\cdots+x_i=y_i} \mathrm{Prob}(X_1 = x_1) \times \cdots \times \mathrm{Prob}(X_i = x_i).$$

Un simple calcul combinatoire nous donne alors les lois de probabilités suivantes :

y_1	1	2	3	4	5	6
$\mathrm{Prob}(Y_1 = y_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

y_2	2	3	4	5	6	7	8	9	10	11	12
$\mathrm{Prob}(Y_2 = y_2)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

y_3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\mathrm{Prob}(Y_3 = y_3)$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

y_4	4	5	6	7	8	9	10	11	12	13	14
$\mathrm{Prob}(Y_4 = y_4)$	$\frac{1}{1296}$	$\frac{5}{1296}$	$\frac{10}{1296}$	$\frac{20}{1296}$	$\frac{35}{1296}$	$\frac{56}{1296}$	$\frac{80}{1296}$	$\frac{104}{1296}$	$\frac{125}{1296}$	$\frac{140}{1296}$	$\frac{146}{1296}$

y_4	15	16	17	18	19	20	21	22	23	24
$\mathrm{Prob}(Y_4 = y_4)$	$\frac{140}{1296}$	$\frac{125}{1296}$	$\frac{104}{1296}$	$\frac{80}{1296}$	$\frac{56}{1296}$	$\frac{35}{1296}$	$\frac{20}{1296}$	$\frac{10}{1296}$	$\frac{4}{1296}$	$\frac{1}{1296}$

Si l'on représente graphiquement ces distributions de probabilité, on obtient alors les 4 figures ci-dessous :

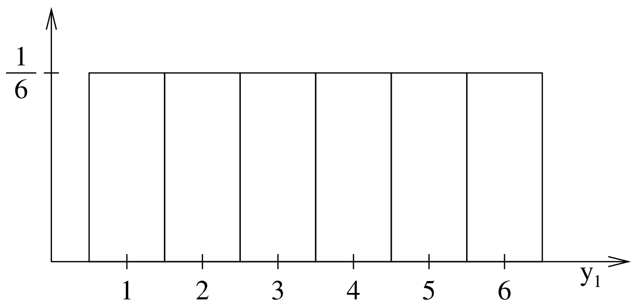


FIG. 27: Distribution de probabilité de Y_1.

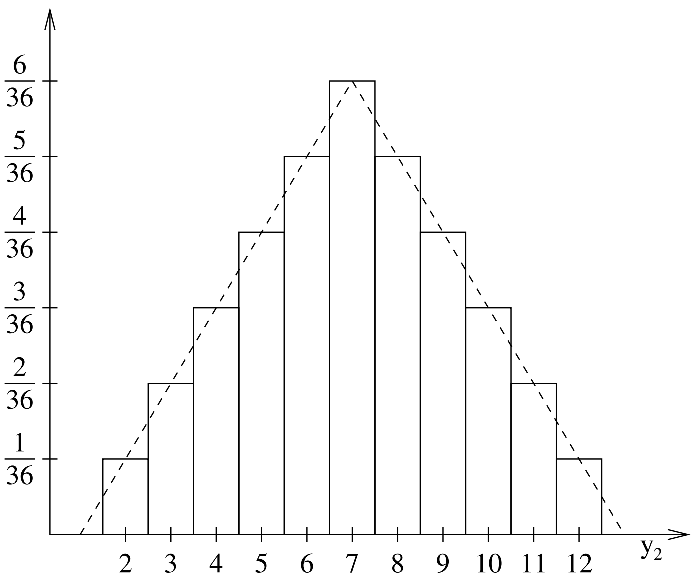


FIG. 28: Distribution de probabilité de Y_2.

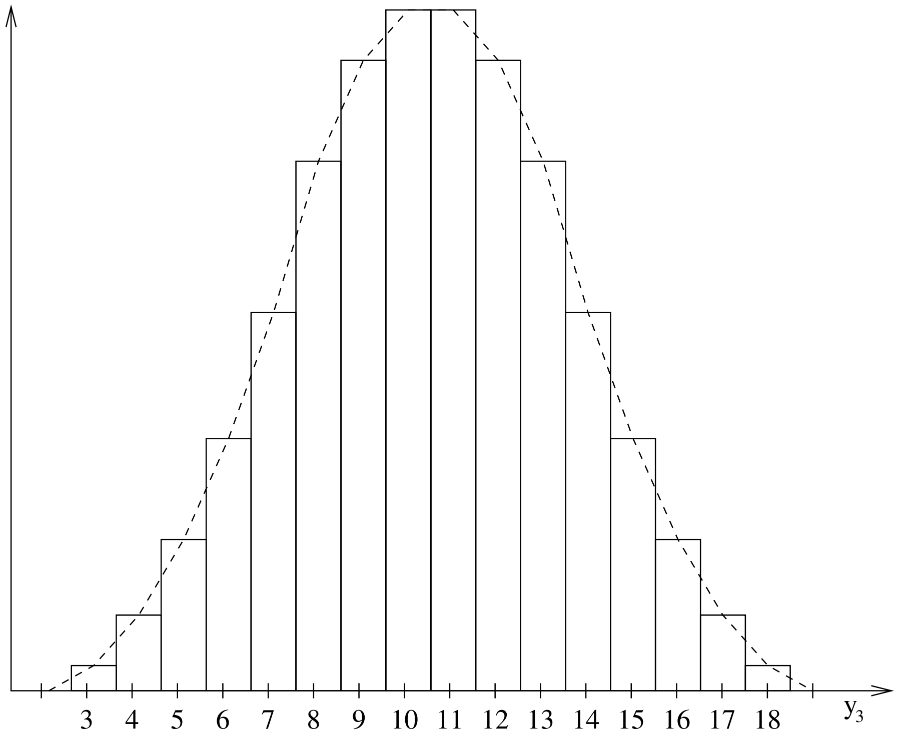


FIG. 29: Distribution de probabilité de Y_3.

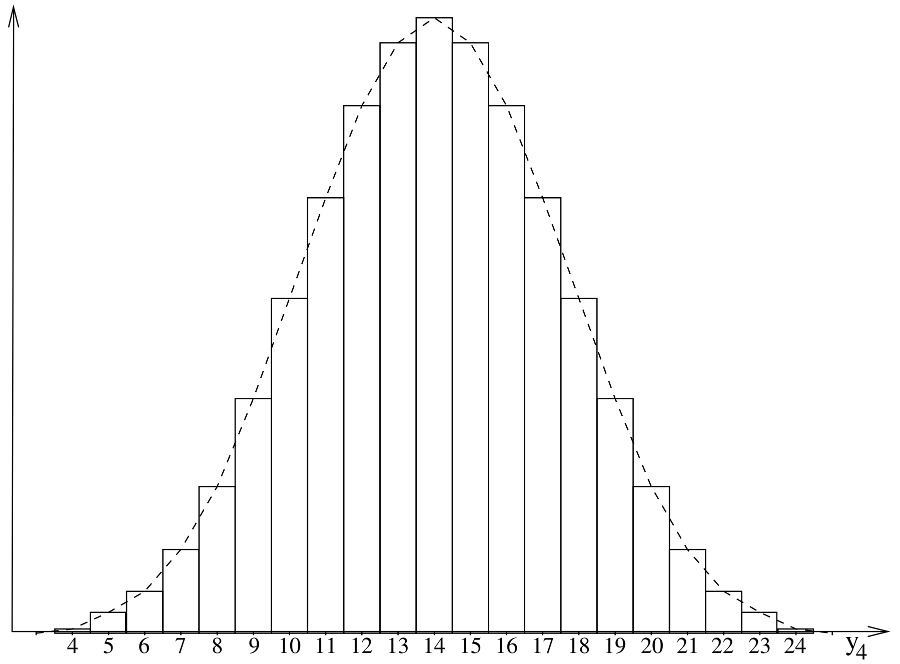
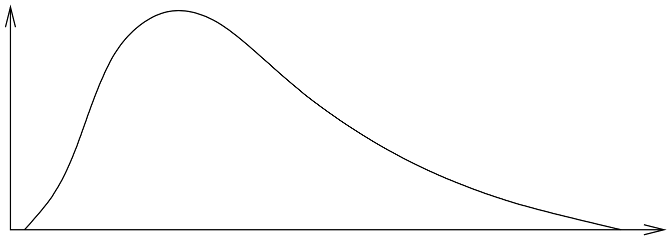


FIG. 30: Distribution de probabilité de Y_4.

On peut remarquer sur ces figures qu'effectivement les courbes ont tendance à se rapprocher de plus en plus de la courbe en forme de cloche de la loi normale. ♦

On peut interpréter le théorème 11 de la manière suivante : quelle que soit la distribution de probabilité d'une population, si l'on extrait de cette population un échantillon suffisamment grand, alors la somme Y (mais aussi la moyenne Y/I) de cet échantillon suit approximativement une loi normale.

Exemple 43 : Le loyer moyen payé par l'ensemble des locataires à Paris est de 4500F/mois, l'écart-type sur l'ensemble des loyés payés est de 700F. Toutefois, il y a beaucoup plus de loyers faibles (studios, 2/3 pièces) que de loyers élevés, si bien que la courbe de densité des loyers ressemble à :



Une société désire louer 35 appartements pour ses employés dans Paris. Elle aimerait connaître la distribution du coût entraîné par ces locations.

On sélectionne donc un échantillon de 35 appartements. D'après le théorème 11, si Y représente le coût recherché, Y suit une loi normale. L'espérance de Y correspond à ce que vaut Y en moyenne, donc, on peut l'approximer par 35×4500F/mois (car les loyers sont tous supposés indépendants). De même, grâce à l'indépendance entre les loyers, $V(Y)$ est égal à la somme des variances de chacun des loyers, ce que l'on peut approximer par $35 \times 700^2(\text{F/mois})^2$. On peut donc en déduire que $Y \sim \text{N}(157500\text{F/mois}; 171540000(\text{F/mois})^2)$. La société peut alors calculer la probabilité que son budget de 130000F/mois ne soit pas dépassé. ♦

Le théorème centrale limite a pour conséquence que, lorsque n est grand, la distribution d'une variable suivant une loi binômiale $\text{Bin}(n;p)$ tend vers une loi normale. On peut comprendre cette propriété en se rappelant qu'une variable suivant la loi $\text{Bin}(n;p)$ est la somme des succès de n épreuves de Bernoulli, chacune de ces épreuves ayant p chances de succès. On peut donc assimiler $X \sim \text{Bin}(n;p)$ à une somme de variables aléatoires $X_i \sim \text{Bin}(1;p)$. Dans ce cas, on peut appliquer le théorème central limite, qui nous indique que X doit tendre vers une loi normale.

Corollaire 3 : *Soit X une variable aléatoire suivant une loi $\text{Bin}(n;p)$. Alors, lorsque n est suffisamment grand,*

$$\frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{npq}} \sim \text{N}(0;1).$$

Par «n suffisamment grand», on entend $n \geq 30$, $np \geq 5$ et $nq \geq 5$. Lorsque les deux dernières contraintes ne sont pas respectées, il faut que n soit bien supérieur à 30.

6.6 Exercices

Exercice 1 Identifiez lesquelles parmi les variables suivantes obéissent à une loi binômiale. Le cas échéant, identifiez les paramètres de la loi, la variance et l'espérance de ces variables.

1/ Selon Statistique Canada, en 1976, la distribution de la langue maternelle à Montréal s'établissait ainsi : Française : 65,3%, Anglaise : 21,7%, autres : 13%. On définit les variables aléatoires suivantes pour un échantillon avec remise de 100 Montréalais :

- X : nombre de personnes dans l'échantillon dont la langue maternelle est le français.
- Y : nombre de personnes dans l'échantillon dont la langue maternelle est l'anglais.
- Z : nombre de personnes dans l'échantillon dont la langue maternelle est le français ou l'anglais.

2/ On choisit au hasard et avec remise trois cartes dans un jeu de 52 cartes.

- X : nombre d'as dans l'échantillon.
- Y : nombre de rois dans l'échantillon.
- W : nombre de paires dans l'échantillon.
- Z : la différence entre le nombre d'as et le nombre de rois dans l'échantillon.

Exercice 2 Pour la livraison qu'il assure à une usine, un fabricant affirme que seulement 2% des pièces livrées sont défectueuses. Pour une livraison de 500 pièces, on décide de vérifier si les dires du fabricant sont exacts en

prélevant un échantillon avec remise de 25 pièces. Soit X le nombre de pièces défectueuses dans l'échantillon.
1/ identifiez la distribution de X.
2/ quelle est la probabilité pour qu'il y ait au moins une pièce défectueuse ?
3/ quelles sont l'espérance et la variance de X.

Exercice 3 En moyenne, deux comptes sont ouverts chaque jour dans une succursale de la BNP. Quelle est la probabilité pour qu'il y ait, un jour donné, exactement 6 ouvertures de comptes ; au plus 3 ouvertures de comptes ; au moins 7 ouvertures de comptes ?

Exercice 4 Une personne travaillant dans une compagnie d'assurances vend en moyenne 1,2 polices d'assurances par jour.
1/ En utilisant la loi de Poisson, trouver la probabilité que cette personne ne vende aucune assurance un jour donné.
2/ Quelle est la moyenne et la variance de la variable «nombre de polices vendues par jour» ?

Exercice 5 Soit X une variable aléatoire suivant une loi normale de paramètres $\mu = 8,3$ et $\sigma^2 = 0,16$. Calculez :
1/ $\text{Prob}(7,5 < X \leq 9,1)$.
2/ $\text{Prob}(8,4 < X < 8,8)$.
3/ $\text{Prob}(X < 7,2)$.
4/ $\text{Prob}(X \geq 8,6)$.

Exercice 6 On constate que la loi N(80; 49) approche très bien la distribution du poids en kg des individus d'une population donnée. Calculez
1/ la proportion d'individus dont le poids observé est de 82 kg.
2/ la proportion d'individus dont le poids est supérieur à 90 kg.
3/ la proportion d'individus dont le poids est supérieur à 70 kg, tout en étant inférieur à 90 kg.

Exercice 7 Après avoir pris connaissance des notes (sur 100) obtenues par ses étudiants à un examen, un prof décide d'ajouter à la note de chaque étudiant 10% de sa note actuelle et 5 points. Sachant qu'initialement la distribution des notes était approchée par une loi N(55; 100),
1/ quelle est la distribution des notes après modification ?
2/ quel est le pourcentage d'échecs initial, si la note de passage est fixée à 60 ?
3/ quel est le pourcentage d'échecs après modification ?

Exercice 8 Un étudiant doit répondre à un QCM comportant 100 questions. Il y a trois réponses possibles par question, mais une seule est correcte. L'étudiant n'ayant pas préparé l'examen (honte sur lui !) décide de répondre à l'aide d'un dé à six faces. Si les faces 1 ou 2 tombent, il coche la case 1, si ce sont les faces 3 ou 4, il coche la case 2, et si ce sont les faces 5 ou 6, il coche la case 3.
1/ Quelle est la probabilité qu'une question donnée soit correctement remplie ?
2/ Quelle est la loi de probabilité suivie par la variable «nombre de réponses correctes» ?
3/ Une réponse correcte vaut 1 point et une réponse incorrecte en vaut 0. Quelle est la probabilité que l'étudiant ait la moyenne à son examen ?

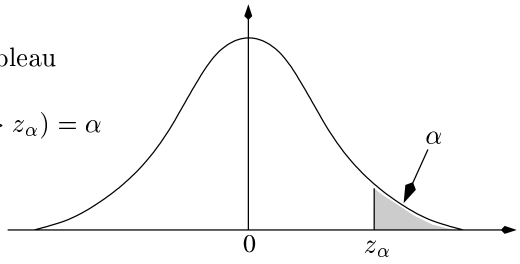
Exercice 9 Le bureau de la statistique du Canada a recensé le nombre de voyages (déplacements de plus de 75KM) effectués par les québécois durant l'été 1978. Approximativement, les résultats obtenus sont les suivants :

Voyages	Effectif
0	2700000
1	1600000
2	700000
3	400000
4	200000
5 à 9	500000
plus de 10	200000

On prélève un échantillon de la population de taille 1000, et on considère la variable X représentant le nombre de personnes dans l'échantillon n'ayant effectué aucun voyage.
1/ Quelle est la probabilité que X soit situé entre 250 et 500 ?
2/ Quelle est la probabilité que X soit inférieur à 400 ?

6.7 Table de la loi normale centrée réduite

valeurs dans le tableau
ci-dessous : les α
tels que $\mathrm{Prob}(Z > z_\alpha) = \alpha$



z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0859	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0466	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0063
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0047
2,6	0,0047	0,0045	0,0044	0,0043	0,0042	0,0041	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0014	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
3,6	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,7	0,0001	0,0001	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

7 Estimation d'une population à partir d'échantillons

Jusqu'à maintenant, on est toujours parti d'une population dont on connaissait les caractéristiques, et on déterminait ce que l'on pouvait dire d'un échantillon donné. Ces techniques sont intéressantes dans le contrôle de qualité, où l'on estime connaître le nombre moyen de pièces défectueuses fabriquées par les machines d'une usine. En sélectionnant de petits échantillons au hasard, on peut contrôler que le nombre de pièces défectueuses correspond bien à ce que le calcul théorique (à partir de la population globale) prévoyait. S'il n'y a pas correspondance, cela signifie probablement qu'une des machines a besoin d'une révision.

Cependant, dans beaucoup d'études statistiques, on ne connaît pas la population et ce sont précisément ses caractéristiques que l'on veut déterminer en prélevant des échantillons. C'est le cas par exemple lorsqu'on réalise des sondages : on cherche ce que les français vont voter aux prochaines élections présidentielles. Il est bien évident que l'on ne connaît pas les intentions de vote de l'ensemble de la population. Des sondages d'opinion vont permettre de constituer des échantillons représentatifs et, à partir de ceux-ci, on va essayer de reconstituer ce que pense l'ensemble de la population. Le but de la section 7 est de fournir les outils statistiques adéquats pour réaliser cela.

Il convient maintenant de s'interroger sur ce que désignent les «caractéristiques de la population». En fait, il n'y a pas beaucoup d'innovations à ce niveau : ce que l'on essaye d'estimer habituellement sont la moyenne μ d'une variable aléatoire dans la population, sa variance, et, éventuellement, sa proportion de succès p.

7.1 Moyenne estimée à partir d'échantillons

Les résultats que nous allons voir ci-après ne sont valables que pour des échantillons prélevés **avec remise**. En effet, le fait de remettre après tirage les individus dans l'urne qui nous sert à constituer l'échantillon permet d'utiliser des propriétés mathématiques intéressantes (en particulier, la probabilité de tirer un individu au $i^{\text{ème}}$ tirage ne dépend pas des tirages précédents), qui vont nous permettent d'établir le lien entre échantillon et population globale. Ce lien n'est plus vérifié lorsque les échantillons sont constitués sans remise. Toutefois, nous avons vu que quand la taille de l'échantillon est très petite par rapport à celle de la population[6], il y a très peu de différence entre la distribution de probabilité obtenue à partir d'un échantillon avec remise et celle obtenue à partir d'un échantillon sans remise. En pratique, on réalise à peu près toujours des échantillons sans remise, mais avec des tailles très inférieures à la taille de la population.

Dans la suite, nous appellerons μ et σ^2 la moyenne et la variance d'une variable aléatoire X sur l'ensemble de la population. Lorsque l'on tire les n individus de l'échantillon, on observe n valeurs de X : $x_1, \ldots, x_n$. Comme l'échantillon est sélectionné avec remise, on peut considérer que chaque x_i est la valeur prise par une variable X_i de même distribution que X, et que tous les X_i sont indépendants. Si l'on note $\overline{X}$ la variable aléatoire correspondant à la moyenne de l'échantillon, on obtient :

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Dans ce cas, il est facile de déterminer l'espérance de $\overline{X}$:

$$E(\overline{X}) = \frac{1}{n} E(X_1 + X_2 + \cdots + X_n).$$

Or, toutes les variables X_i sont indépendantes. Donc :

$$\begin{aligned} E(\overline{X}) &= \tfrac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right), \\ &= \tfrac{1}{n}\left(\mu + \mu + \cdots + \mu\right) = \mu. \end{aligned}$$

[6] On considère généralement que si un échantillon sans remise a une taille inférieure à $1/20^{\text{ème}}$de la taille de la population, on peut l'assimiler à un échantillon avec remise.

De manière analogue, la variance de $\overline{X}$ est déterminée par :

$$\begin{aligned} V(\overline{X}) &= V\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right), \\ &= \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n), \\ &= \frac{1}{n^2}\left(V(X_1) + V(X_2) + \cdots + V(X_n)\right), \\ &= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}. \end{aligned}$$

D'où le théorème suivant :

Théorème 12 : *Soit X une variable d'espérance μ et de variance σ^2. Soit $\overline{X}$ la variable aléatoire «moyenne d'un échantillon de taille n prélevé avec remise sur X». Alors :*

$$E(\overline{X}) = \mu \qquad\qquad V(\overline{X}) = \frac{\sigma^2}{n}.$$

Le théorème ci-dessus est valide quelle que soit la distribution de X. Nous avons vu dans la section précédente que la loi normale était une loi très importante en statistiques. Voyons donc ce que l'on peut déduire sur $\overline{X}$ lorsque X suit cette loi :

Théorème 13 : *Soit X une variable aléatoire suivant une loi normale $\mathrm{N}(\mu;\sigma^2)$. Soit $\overline{X}$ la variable aléatoire «moyenne d'un échantillon de taille n prélevé avec remise sur X». Alors $\overline{X} \sim \mathrm{N}(\mu;\frac{\sigma^2}{n})$.*

Démonstration : On peut considérer que

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Or, les variables X_i sont indépendantes (car l'échantillonnage est réalisé avec remise) et elles suivent une loi normale. Donc, d'après le théorème 9, page 72, $\overline{X}$ suit une loi normale. De plus, d'après le théorème 12, ses paramètres ne peuvent être que μ et $\frac{\sigma^2}{n}$. ♦

Exemple 44 : À l'issue d'un concours pour rentrer aux ENSI (un groupement d'écoles d'ingénieurs), les notes des copies suivent une loi normale $\mathrm{N}(11,4;36)$. Les correcteurs ont eu chacun la lourde charge de corriger 100 copies. On se demande quelle est la probabilité pour qu'un correcteur donné ait eu une moyenne sur l'ensemble de ses copies inférieure à 10, ou comprise entre 11,5 et 12,5.

On peut considérer les 100 copies comme un échantillon de taille 100 prélevé sur l'ensemble des copies du concours. D'après le théorème précédent, la moyenne $\overline{X}$ sur ces 100 copies suit une loi normale $\mathrm{N}(11,4;\frac{36}{100})$. Par conséquent,

$$\mathrm{Prob}(\overline{X} \le 10) = \mathrm{Prob}\left(\frac{\overline{X} - 11,4}{\sqrt{\frac{36}{100}}} \le \frac{10 - 11,4}{\sqrt{\frac{36}{100}}}\right) = 0,0099.$$

La probabilité est très faible car avec 100 copies, on peut être à peu près certain d'obtenir un $\overline{X}$ très proche de μ.

$$\mathrm{Prob}(11,5 \le \overline{X} \le 12,5) = \mathrm{Prob}\left(\frac{11,5 - 11,4}{\sqrt{\frac{36}{100}}} \le \frac{\overline{X} - 11,4}{\sqrt{\frac{36}{100}}} \le \frac{12,5 - 11,4}{\sqrt{\frac{36}{100}}}\right) \approx 0,4325 - 0,0336 = 0,3989.$$

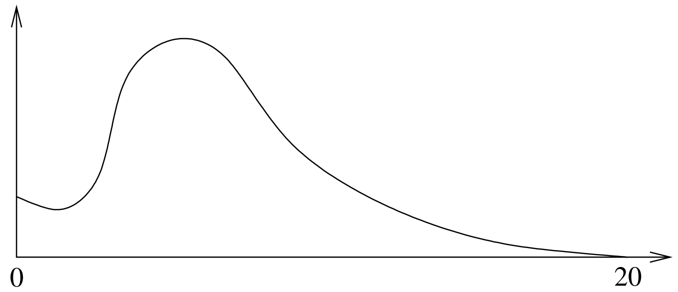
♦

Lorsque la variable X ne suit pas une loi normale, on peut utiliser le théorème central limite pour déduire la loi suivie par $\overline{X}$:

Théorème 14 : *Soit X une variable aléatoire dont l'espérance et la variance sont respectivement μ et σ^2. Soit $\overline{X}$ la variable aléatoire «moyenne d'un échantillon de taille n, où n est suffisamment grand ($n \geq 30$ si la distribution de X n'est pas trop dissymétrique, $n \geq 50$ sinon), sur X». Alors,*

$$\frac{\overline{X} - E(\overline{X})}{\sqrt{V(\overline{X})}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0;1).$$

Exemple 45 : Reprenons l'exemple des copies des ENSI. Les correcteurs sont assez sévères, ils mettent donc plus de notes basses que de notes élevées. De plus, les très mauvaises copies les énervent et ils ont tendance à les sous-noter. Globalement, la courbe de densité des notes a l'aspect suivant :



On ne peut pas vraiment dire que cette courbe ressemble à la courbe d'une loi normale. On ne peut donc appliquer le théorème 13. Par contre, on peut appliquer le théorème ci-dessus car la taille de l'échantillon est suffisamment grande. On peut donc tout de même dire que

$$\frac{\overline{X} - 11,4}{\frac{6}{10}} \sim \mathrm{N}(0;1).$$

Réciproquement, si l'on observe que l'échantillon des 100 copies suit la loi ci-dessus, on peut affirmer que la moyenne de l'ensemble des copies de la population est 11,4. ♦

7.2 Proportion de succès

Considérons une étude statistique dans laquelle on désire identifier la distribution de la variable $\overline{P}$ «proportion de succès dans un échantillon (avec remise) de taille n». On note p la proportion de succès dans la population. Comme dans la sous-section précédente, on peut condisérer que $\overline{P}$ est la moyenne des variables aléatoires P_i reflétant les succès pour chaque individu. Dans ce cas, puisque l'échantillon est sélectionné avec remise, les P_i sont indépendants et suivent tous une loi binômiale $\mathrm{Bin}(1,p)$. Lorsque n est suffisamment grand, et p n'est ni trop petit, ni trop grand ($np \geq 5$ et $nq \geq 5$), alors on peut appliquer le théorème central limite, qui nous dit que :

$$\frac{\overline{P} - E(\overline{P})}{\sqrt{V(\overline{P})}} \sim \mathrm{N}(0;1).$$

Or, si les P_i suivent une loi $\mathrm{Bin}(1,p)$, $\overline{P}$ doit suivre la loi $\frac{1}{n}\mathrm{Bin}(n;p)$. Par conséquent, $E(\overline{P}) = \frac{np}{n}$ et $V(\overline{P}) = \frac{npq}{n^2} = \frac{pq}{n}$. Donc le théorème suivant est valide :

Théorème 15 : *Soit $\overline{P}$ la proportion de succès dans un échantillon (avec remise) de taille n suffisamment grande. Soit p la proportion de succès de la population totale. Alors :*

$$\frac{\overline{P} - E(\overline{P})}{\sqrt{V(\overline{P})}} = \frac{\overline{P} - p}{\sqrt{\frac{pq}{n}}} \sim \mathrm{N}(0;1).$$

Exemple 46 : D'après un sondage, 18% des personnes actives pensent que leur carrière professionnelle est enrichissante à la fois d'un point de vue personnel et d'un point de vue financier. Supposons que ce résultat soit vérifié pour l'ensemble de la population. On sélectionne un échantillon de 100 personnes. Est-on en droit d'affirmer que dans cet échantillon, entre 20 et 40 personnes partagent ce point de vue ?

Soit $\overline{P}$ la proportion dans l'échantillon des personnes qui partagent ce point de vue. $p = 0,18$, $q = 1 - 0,18 = 0,82$ et $n = 100$. On peut appliquer le théorème ci-dessus car on a bien $n \geq 30$, $np = 5,4 \geq 5$ et $nq = 24,6 \geq 5$. Donc

$$\frac{\overline{P} - 0,18}{0,03842} \sim \mathrm{N}(0;1).$$

$$\begin{aligned}\mathrm{Prob}(0,2 \leq \overline{P} \leq 0,4) &= \mathrm{Prob}\left(\tfrac{0,2-0,18}{0,03842} \leq \tfrac{\overline{P}-0,18}{0,03842} \leq \tfrac{0,4-0,18}{0,03842}\right)\\ &= \mathrm{Prob}\left(0,52 \leq \tfrac{\overline{P}-0,18}{0,03842} \leq 5,723\right) \approx 0,3015.\end{aligned}$$

L'affirmation paraît donc pour le moins douteuse. ♦

7.3 Estimation de μ et de p à partir d'un échantillon

D'après ce qui précède, on se doute bien que si l'on sélectionne un échantillon de taille n suffisamment grand, la moyenne $\overline{X}$ obtenue sur l'échantillon va se rapprocher de celle de la population. L'intuition nous suggère donc d'estimer μ par la moyenne de $\overline{X}$. Cette intuition peut heureusement se justifier théoriquement. En fait, cette propriété est vraie parce que $\overline{X}$ est un estimateur non biaisé, c'est-à-dire que l'espérance de $\overline{X}$ est réellement égale à μ. C'est ce critère qui est utilisé pour déterminer un «bon» estimateur :

Un estimateur non biaisé est un estimateur dont l'espérance tend vers le paramètre que l'on essaye d'estimer.

Il serait fort gênant d'utiliser des estimateurs biaisés. En effet, cela signifierait qu'en moyenne les estimations obtenues à partir des échantillons sur-évalueraient ou sous-évalueraient le paramètre (μ ou p) à estimer.

Nous avions vu que l'espérance et la variance de $\overline{X}$ sont respectivement égales à μ et $\frac{\sigma^2}{n}$. Cela signifie que lorsque n augmente, la variance de l'échantillon diminue. Or, celle-ci représente l'étalement des valeurs de $\overline{X}$ autour de sa moyenne. On voit donc que plus n est grand, plus les valeurs de $\overline{X}$ se concentrent autour de la moyenne, qui, par bonheur, est le paramètre à estimer. On dit alors que $\overline{X}$ est un estimateur convergent.

De manière analogue, $\overline{P}$ est un bon estimateur de la proportion p de succès de la population. La variable $\overline{P}$ est, elle aussi, sans biais car $E(\overline{P}) = p$. De plus, $\overline{P}$ a le bon goût d'être un estimateur convergent puisque $V(\overline{P}) = \frac{pq}{n}$ diminue lorsque n grandit.

Exemple 47 : Dans une ville de 50000 habitants, un médecin a signalé 3 cas d'une maladie qui peut s'avérer contagieuse dans certaines conditions. Le maire se demande s'il doit ou non lancer une grande campagne de dépistage, un test de dépistage coûtant relativement cher. Pour cela, il demande à son statisticien de service de l'aider à prendre sa décision. Celui-ci demande à 100 personnes de passer le test. Sur ces 100 personnes, 2 se révèlent être positifs au test. Donc on peut estimer que p est à peu près égal à $\frac{2}{100}$, et que seulement 2% des 50000 habitants, soit 1000 personnes, seront positives au test. ♦

D'une manière générale, en statistiques, on essaye toujours de se ramener à des estimateurs non biaisés. Si l'estimateur trouvé est biaisé, il faut alors **absolument** quantifier le biais, et proposer un nouvel estimateur qui corrige ce biais. C'est le cas précisément lorsqu'on essaye d'estimer σ^2. Prenons un exemple simple :

Exemple 48 : On considère une population de 4 nombres 1,2,3 et 4. La moyenne de cette population est donc

$$\mu = \frac{1+2+3+4}{4} = \frac{5}{2}.$$

De même, la variance σ^2 est égale à

$$\sigma^2 = \frac{\left(1-\frac{5}{2}\right)^2 + \left(2-\frac{5}{2}\right)^2 + \left(3-\frac{5}{2}\right)^2 + \left(4-\frac{5}{2}\right)^2}{4} = \frac{5}{4}.$$

On sélectionne avec remise des échantillons de taille 2 dans cette population. On a noté dans le tableau ci-dessous les échantillons ainsi que leur espérance et leur variance :

échantillon	espérance	variance
1 1	$E(\overline{X})=1$	$V(\overline{X})=0$
1 2	$E(\overline{X})=\frac{3}{2}$	$V(\overline{X})=\frac{1}{4}$
1 3	$E(\overline{X})=2$	$V(\overline{X})=1$
1 4	$E(\overline{X})=\frac{5}{2}$	$V(\overline{X})=\frac{9}{4}$
2 1	$E(\overline{X})=\frac{3}{2}$	$V(\overline{X})=\frac{1}{4}$
2 2	$E(\overline{X})=2$	$V(\overline{X})=0$
2 3	$E(\overline{X})=\frac{5}{2}$	$V(\overline{X})=\frac{1}{4}$
2 4	$E(\overline{X})=3$	$V(\overline{X})=1$
3 1	$E(\overline{X})=2$	$V(\overline{X})=1$
3 2	$E(\overline{X})=\frac{5}{2}$	$V(\overline{X})=\frac{1}{4}$
3 3	$E(\overline{X})=3$	$V(\overline{X})=0$
3 4	$E(\overline{X})=\frac{7}{2}$	$V(\overline{X})=\frac{1}{4}$
4 1	$E(\overline{X})=\frac{5}{2}$	$V(\overline{X})=\frac{9}{4}$
4 2	$E(\overline{X})=3$	$V(\overline{X})=1$
4 3	$E(\overline{X})=\frac{7}{2}$	$V(\overline{X})=\frac{1}{4}$
4 4	$E(\overline{X})=4$	$V(\overline{X})=0$

Il est facile de vérifier que la moyenne sur l'ensemble des échantillons est bien égale à μ. Calculons maintenant la moyenne des variances :

$$\frac{1}{16}\left[0+\frac{1}{4}+1+\frac{9}{4}+\frac{1}{4}+0+\frac{1}{4}+1+1+\frac{1}{4}+0+\frac{1}{4}+\frac{9}{4}+1+\frac{1}{4}+0\right]=\frac{10}{16}=\frac{5}{8}.$$

Conclusion : l'espérance des variances des échantillons ne correspond pas à la variance de la population. La variance d'un échantillon n'est donc pas un bon estimateur de la variance de la population. ♦

L'exemple précédent peut se comprendre de la manière suivante : on avait vu dans les sections précédentes la formule suivante :

$$V(X_1+\cdots+X_n)=V(X_1)+\cdots+V(X_n).$$

Et c'est en fait cette formule que l'on aurait aimé appliquer pour dire que la variance de l'échantillon permettait d'estimer la variance de la population. Malheureusement, cette formule ne s'applique que lorsque les X_i sont indépendants. Or, ils ne le sont plus lorsque l'on examine tous les échantillons possibles : le dernier échantillon est déterminé lorsqu'on a examiné tous les autres. Cela suggère de n'examiner que le nombre d'échantillons possibles moins 1. Mathématiquement, cela revient à formuler le théorème suivant :

Théorème 16 : *Soit X une variable aléatoire définie sur une certaine population. Appelons σ^2 la variance de X sur cette population. On sélectionne (avec remise) un échantillon de taille n. Soient $X_1,\dots,X_n$ les valeurs de X correspondant à chacun des individus de l'échantillon, et soit $\overline{X}$ la moyenne de ces valeurs. Alors :*

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

Autrement dit, $\frac{n}{n-1}$ fois la variance de X sur l'échantillon, que l'on appelle la variance corrigée, est un «bon» estimateur de la variance de X sur l'ensemble de la population.

7.4 Exercices

Exercice 1 On estime que, dans une certaine population, 20% des individus de 19 ans et plus sont célibataires. On prélève dans cette population un échantillon avec remise de 250 personnes.
1/ Donnez la loi approximative de $\overline{P}$, la proportion de célibataires dans l'échantillon.
2/ Calculez Prob($\overline{P}\geq 0,21$).
3/ Calculez Prob($0,18\leq\overline{P}\leq 0,22$).

Exercice 2 200 personnes ont participé à un tournoi d'échecs. On a demandé leur âge à 15 d'entre elles. Voici ce que l'on a obtenu :

22	35	20	18	17
19	43	37	21	23
32	27	29	25	19

Estimer la moyenne d'âge de l'ensemble du tournoi et l'écart-type de l'âge des participants.

Exercice 3 Dans une usine qui fabrique des pièces automobiles, on sait par expérience qu'environ 3% des pièces sont défectueuses. Dans un lot de 50000 pièces, on sélectionne un échantillon de 100 pièces et on en trouve 10 défectueuses. Est-ce alarmant ?

Exercice 4 Dans la région de Carbost Village (Écosse), la distillerie Talisker s'est appropriée dans le passé 30% du marché des whiskies. Elle veut savoir quelle est sa part de marché à la date d'aujourd'hui. Pour cela, elle fait réaliser une enquête statistique dans laquelle on a demandé à 1000 consommateurs de whisky s'ils ont acheté du Talisker ou un autre whisky. Il apparaît que 290 consommateurs ont acheté du Talisker. Qu'en concluez-vous ?

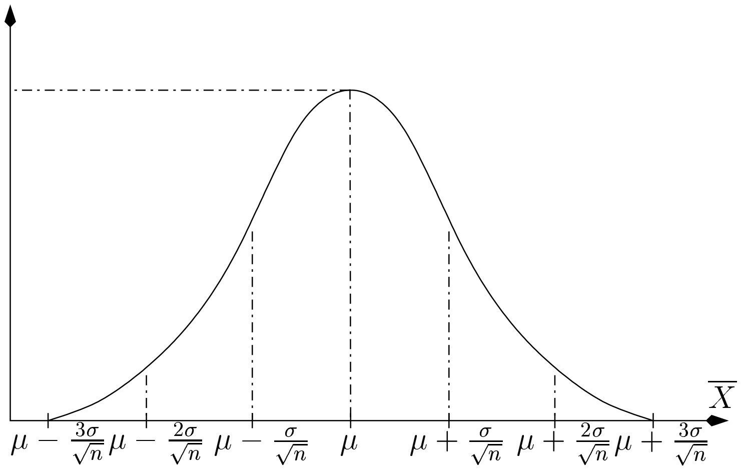
8 Les intervalles de confiance et la loi de Student

Nous avons vu dans la section précédente comment l'on pouvait estimer les caractéristiques d'une population (μ, σ et p) à partir d'échantillons. Il est bien évident que plus la taille de l'échantillon est grande, plus cette estimation est bonne. Cela suggère que l'on devrait pouvoir quantifier la «qualité» de nos estimations. C'est le but de cette section. Pour cela, contrairement à la section précédente dans laquelle les paramètres étaient estimés grâce à une valeur unique : $\mu = 3$ (on parle alors d'$\boxed{\text{estimation ponctuelle}}$), nous allons ici estimer les paramètres grâce à des intervalles : «95% des échantillons prédisent que $\mu \in [2,4;3,6]$». C'est ce que l'on appelle des $\boxed{\text{intervalles de confiance}}$.

Commençons par ne considérer que des échantillons de taille n prélevés avec remise sur une variable $X \sim \mathrm{N}(\mu;\sigma^2)$.

8.1 Lorsque X suit une loi normale

D'après la section précédente, nous savons que $\overline{X} \sim \mathrm{N}(\mu;\frac{\sigma^2}{n})$. Donc, si l'on examinait tous les échantillons possibles de taille n, on obtiendrait la distribution suivante :



Bien que μ soit l'espérance de $\overline{X}$, en pratique, il y a très peu de chances pour que la moyenne d'un échantillon tiré au hasard soit exactement μ. Voilà pourquoi, on préfère en statistiques estimer μ grâce à un intervalle. Construisons un tel intervalle :
En centrant et en réduisant $\overline{X}$, nous avons vu que $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0;1)$. Grâce à la table de la loi normale, on en déduit que

$$\mathrm{Prob}\left(-1,96 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1,96\right) = 95\%.$$

Autrement dit,

$$\mathrm{Prob}\left(\overline{X} - 1,96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1,96\frac{\sigma}{\sqrt{n}}\right) = 95\%.$$

Donc pour 95% des échantillons prélevés, μ est situé entre $\overline{X} - 1,96\frac{\sigma}{\sqrt{n}}$ et $\overline{X} + 1,96\frac{\sigma}{\sqrt{n}}$. On peut donc choisir un échantillon au hasard et déterminer où se situe μ avec une très bonne précision. En suivant ce même principe, on peut déterminer la probabilité que μ appartienne à n'importe quel intervalle.

Exemple 49 : Considérons un échantillon de taille 25 prélevé avec remise sur une variable $X \sim \mathrm{N}(\mu;100)$.
Pour estimer μ par un intervalle de confiance de 90%, on peut procéder de la manière suivante :

$$X \sim \mathrm{N}(\mu;100) \text{ et } n = 25 \Rightarrow \frac{\overline{X}-\mu}{10/\sqrt{25}} = \frac{\overline{X}-\mu}{2} \sim \mathrm{N}(0;1).$$

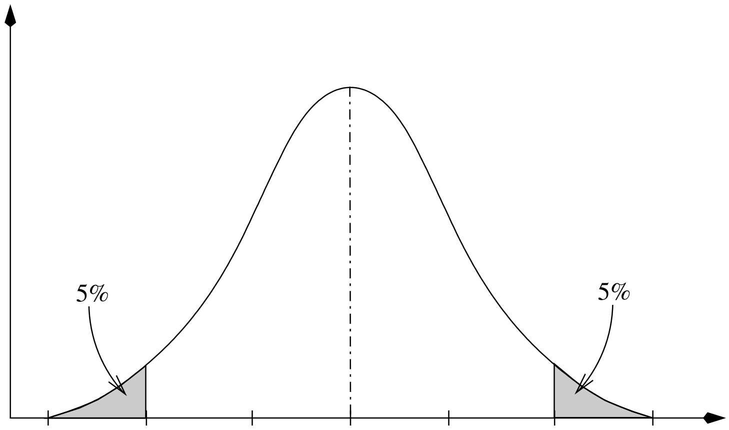
D'après la table de la loi normale,

$$0,90 \quad = \quad \mathrm{Prob}(-1,645 \leq Z \leq 1,645),$$

$$\begin{aligned}
&= \text{Prob}\left(-1,645 \leq \frac{\overline{X}-\mu}{2} \leq 1,645\right), \\
&= \text{Prob}(\overline{X} - (1,645 \times 2) \leq \mu \leq \overline{X} + (1,645 \times 2)), \\
&= \text{Prob}(\overline{X} - 3,29 \leq \mu \leq \overline{X} + 3,29).
\end{aligned}$$

Ainsi, pour 90% des échantillons, si $\overline{x}$ représente la moyenne de l'échantillon, $\mu \in [\overline{x} - 3,29; \overline{x} + 3,29]$. ♦

Généralisons un peu la méthode ci-dessus : comment avons-nous déterminé que l'on avait $\text{Prob}(-1,645 \leq Z \leq 1,645) = 0,90$? Tout simplement en se référant à la table de la loi normale et à la figure ci-dessous :



On voit donc que pour trouver les bornes supérieures et inférieures de $\frac{\overline{X}-\mu}{2}$, c'est-à-dire 1,645, il suffit de rechercher dans la table de la loi normale le quantile $z_{5\%}$, c'est-à-dire le nombre tel que $\text{Prob}(Z > z_{5\%}) = 5\%$. En généralisant, si l'on recherche l'intervalle de confiance de niveau de confiance $1 - \alpha$, il faut rechercher dans la table de la loi normale le quantile $z_{\alpha/2}$. Autrement dit,

> Une estimation de μ par intervalle de confiance, de niveau de confiance $1 - \alpha$, fondée sur un échantillon de taille n prélevé avec remise sur une variable $X \sim \text{N}(\mu; \sigma^2)$ est donné par :
>
> $$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Il est important de rappeler ici la signification du niveau de confiance $1-\alpha$. En effet, très souvent, on entend dire que «le niveau de confiance représente la probabilité que μ appartienne à l'intervalle de confiance». Cette interprétation de $1 - \alpha$ n'a aucun sens car μ n'est pas une variable aléatoire et donc on ne peut calculer de probabilité sur μ. C'est comme si l'on parlait de la probabilité que le nombre 3 appartienne à l'ensemble $[1, 5]$: cette affirmation est soit vraie, soit fausse, mais il n'y a pas de probabilité là-dessus. Non, la vraie signification du niveau de confiance est la suivante :

> Le niveau de confiance $1 - \alpha$ représente la proportion des échantillons possibles pour lesquels l'estimation de μ par l'intervalle de confiance est correcte.

Dans l'exemple ci-dessus, on peut donc prendre des échantillons de taille 25, en estimer la moyenne $\overline{x}$ et affirmer que $\mu \in [\overline{x} - 3,29; \overline{x} + 3,29]$. Dans 90% des cas, cette affirmation sera correcte, mais dans 10% des cas, nous nous serons trompés.

Exemple 50 : Chaque année, plusieurs milliers de candidats se présentent au concours d'entrée aux ENSI. Par expérience, on sait que chaque année la variance sur les notes obtenues par les candidats est à peu près de 16. Un correcteur a corrigé l'ensemble de ses 100 copies et a obtenu une moyenne de 8,75. Il se demande quelle pourrait bien être la moyenne sur l'ensemble des copies du concours. On peut supposer que les copies suivent une loi normale $\text{N}(\mu; 16)$.

D'après ce qui précède, il y a une proportion $1 - \alpha$ d'échantillons pour lesquels

$$\mu \in \left[\overline{x} - z_{\alpha/2}\frac{4}{10}; \overline{x} + z_{\alpha/2}\frac{4}{10}\right].$$

Par conséquent, on peut dresser le tableau suivant, qui recense le pourcentage d'échantillons pour lesquels le calcul est correct :

% d'échantillons	intervalle de confiance		
50%	$[8,75 - 0,674 \times 0,4; 8,75 + 0,674 \times 0,4]$	$=$	$[8,48; 9,02]$
75%	$[8,75 - 1,15 \times 0,4; 8,75 + 1,15 \times 0,4]$	$=$	$[8,29; 9,21]$
80%	$[8,75 - 1,28 \times 0,4; 8,75 + 1,28 \times 0,4]$	$=$	$[8,24; 9,26]$
90%	$[8,75 - 1,645 \times 0,4; 8,75 + 1,645 \times 0,4]$	$=$	$[8,09; 9,41]$
95%	$[8,75 - 1,96 \times 0,4; 8,75 + 1,96 \times 0,4]$	$=$	$[7,96; 9,53]$
99%	$[8,75 - 2,575 \times 0,4; 8,75 + 2,575 \times 0,4]$	$=$	$[7,72; 9,78]$

Le correcteur peut donc en déduire qu'il y a peu de chances que la moyenne des copies dépasse les 10/20.

♦

8.2 Lorsque l'on ne connaît pas la loi suivie par X

Bien évidemment, dans la pratique, il est rare de pouvoir affirmer que la variable X suit, sur l'ensemble de la population, une loi normale. Par contre, nous avons vu d'après le théorème central limite de la section 6 que la distribution de $\overline{X}$ sur un échantillon de taille n est :

$$\overline{X} \sim \mathrm{N}\left(\mu; \frac{\sigma^2}{n}\right),$$

où μ et σ^2 sont l'espérance et la variance de la variable X. Par conséquent, par un raisonnement analogue à celui de la sous-section précédente, on peut en déduire que :

> Si X est une variable d'espérance μ et de variance σ^2, et si $\overline{X}$ représente la moyenne d'un échantillon avec remise de taille n sur X. Alors, une estimation de μ par intervalle de confiance, de niveau $1 - \alpha$, est donnée par :
>
> $$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Exemple 51 : Des études biométriques ont montré qu'habituellement la taille d'un adulte de sexe masculin est une variable d'écart-type 5cm dans une population occidentale. On a prélevé un échantillon de 144 étudiants à Jussieu et on a obtenu une moyenne de 178,4cm. Estimons la moyenne de la population estudiantine de Jussieu avec un intervalle de confiance de 95%. D'après le théorème central limite, il suffit de calculer $z_{0,025}$. D'après la table de la loi normale, on obtient $z_{0,025} = 1,96$. Donc l'intervalle recherché est :

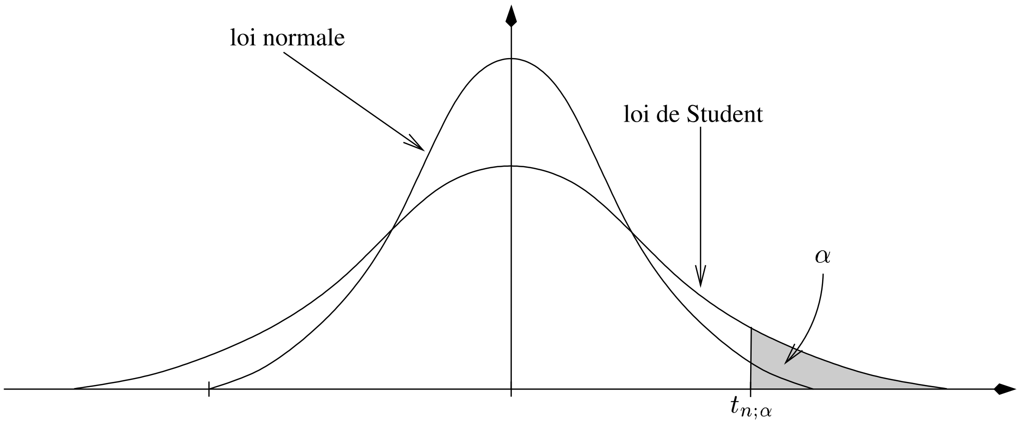
$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \left[178,4 - 1,96 \times \frac{5}{12}; 178,4 + 1,96 \times \frac{5}{12}\right] = [177,58; 179,22].$$

♦

Toutes les estimations ci-dessus posent tout de même un problème : elles ne sont valides que lorsque l'on connaît la variance de X sur l'ensemble de la population. Or, bien souvent, on ne connaît pas sa valeur. Comment faire dans ce cas ? Eh bien tout simplement, nous allons reprendre l'estimation de la variance que nous avions vu dans la section précédente : la variance corrigée. $\boxed{\text{Dans la suite, nous noterons } S^2 \text{ la variance corrigée}}$. Puisque $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0; 1)$, nous aimerions obtenir un résultat tel que $\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathrm{N}(0; 1)$. Malheureusement, ce résultat n'est pas correct, sauf lorsque la taille de l'échantillon est grande (en principe, lorsqu'elle est supérieure à 75). C'est pourquoi les statisticiens ont étudié avec attention la distribution de $\frac{\overline{X} - \mu}{S/\sqrt{n}}$. Celle-ci a été trouvée et publiée en 1908 par W. S. Gossett, sous le pseudonyme de *Student*. La loi en question est donc connue sous le nom de *loi de Student*. Elle ressemble beaucoup à la loi normale (forme de cloche, symétrie par rapport à $\mu = 0$), mais elle est plus évasée et plus applatie que la loi normale $\mathrm{N}(0; 1)$.

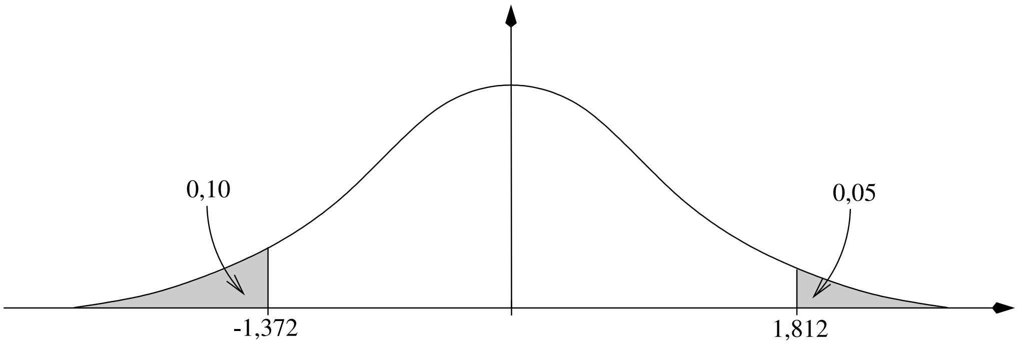
Définition 34 (Loi de Student) : *La loi de Student ne possède qu'un seul paramètre n : que l'on appelle le nombre degrés de liberté. La loi à n degrés est notée T_n. L'espérance d'une variable obéissant à une loi T_n est 0, et sa variance est $\frac{n}{n-2}$ pour $n > 2$. La table de la loi de Student est fournie à la fin de cette section. Plus n est grand, plus T_n se rapproche de* $\mathrm{N}(0;1)$.

En principe, pour une loi de Student T_n, on note le quantile d'ordre $1-\alpha$ de la manière suivante : $t_{n;\alpha}$. À l'instar de la loi normale, ce quantile est tel que $\mathrm{Prob}(Y > t_{n;\alpha}) = \alpha$. On peut le représenter sur la figure ci-dessous :



Exemple 52 : Soit $Y \sim T_{10}$. Déterminons grâce à la table de la loi de Student les valeurs de $t_{10;0,05}$ et de $t_{10;0,90}$, c'est-à-dire des 95$^{\text{ème}}$et 10$^{\text{ème}}$centiles de Y.

La valeur de $t_{10;0,05}$ se lit directement sur la table en regardant l'intersection de la ligne 10 et de la colonne 0,05. On trouve ainsi que $t_{10;0,05} = 1,812$. D'après la figure ci-dessous, on voit bien que $t_{10;0,90} = -t_{10;0,10}$. Or, d'après la table, $t_{10;0,10} = -1,372$. Donc $t_{10;0,90} = 1,372$.



♦

La dernière ligne de la table de la loi de Student (celle correspondant à $n = +\infty$) donne en fait les valeurs des quantiles de la loi normale $\mathrm{N}(0;1)$ puisque c'est vers cette loi que tend la distribution de Student.

Théorème 17 : *Soit $\overline{X}$ la variable «moyenne d'un échantillon avec remise de taille n sur une variable X». Soit S^2 la variance corrigée de l'échantillon.*

$$\text{Si } X \sim \mathrm{N}(\mu;\sigma^2) \text{ alors } \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}.$$

Attention : *il faut utiliser la loi de Student avec $n-1$ degrés de liberté et non pas avec n degrés.*

Autrement dit, si l'on sait que X suit une loi normale dont on ne connaît ni la moyenne, ni la variance, on peut tout de même donner un intervalle de confiance de niveau $1-\alpha$ pour l'estimation de la moyenne μ :

$$\begin{aligned} 1-\alpha &= \mathrm{Prob}\left(-t_{n-1;\alpha/2} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{n-1;\alpha/2}\right) \\ &= \mathrm{Prob}\left(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right). \end{aligned}$$

Ainsi, on obtient :

> **Théorème 18 :** *Lorsque l'on sait que X suit une loi normale, dont on ne connaît ni la moyenne, ni la variance, l'intervalle de confiance pour μ, de niveau de confiance $1-\alpha$, est donné par :*
>
> $$\left[\overline{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right].$$

Exemple 53 : Le laboratoire d'informatique de Paris 6 possède un parc non négligeable de stations de travail Sun. Un des ingénieurs système a sous sa responsabilité 49 stations. Il a déterminé que la fréquence des pannes importantes de ses Suns est en moyenne de 54 mois, avec un écart-type de 8 mois. Le directeur du laboratoire désire acheter du matériel informatique et lui demande quelle est la fréquence des pannes des Suns.

L'ingénieur système considère donc ses machines comme un échantillon de 49 individus. Il suppose que X, la fréquence des pannes de la population totale (l'ensemble des stations de travail dans le monde entier), suit une loi normale. Grâce à son échantillon, il peut déterminer que l'intervalle de confiance de niveau $1-\alpha$ est :

$$\left[54 - t_{48;\alpha/2} \times \frac{8}{7}; 54 + t_{48;\alpha/2} \times \frac{8}{7}\right].$$

Par conséquent, il peut répondre avec un risque d'erreur de 5% que

$$\mu \in \left[54 - 2,011 \times \frac{8}{7}; 54 + 2,011 \times \frac{8}{7}\right] = [51,7; 56,3].$$

♦

Exemple 54 : Un distributeur de boissons est réglé de telle sorte que la quantité X de liquide qu'il verse dans un gobelet est distribuée selon une loi normale. Afin d'affiner les réglages, un technicien prélève un échantillon de 10 boissons. Il obtient sur cet échantillon une moyenne de 20 cl avec un écart-type de 1,65 cl. Il veut estimer avec un risque d'erreur de 5% la moyenne μ de la quantité de boisson versée par le distributeur.

D'après le théorème 18, le technicien peut déduire que l'intervalle de confiance à 95% est :

$$\left[20 - t_{9;0,025} \times \frac{1,65}{\sqrt{10}}; 20 + t_{9;0,025} \times \frac{1,65}{\sqrt{10}}\right] = \left[20 - 2,262 \times \frac{1,65}{\sqrt{10}}; 20 + 2,262 \times \frac{1,65}{\sqrt{10}}\right] = [18,82; 21,18].$$

♦

Exemple 55 : Afin de mettre au point le réglage d'une machine devant usiner des pièces de 6 cm de diamètre, un technicien fait fonctionner celle-ci sur un échantillon de 9 pièces. Le diamètre des pièces de l'échantillon est en moyenne de 6,2 cm, avec une variance corrigée de 16 mm^2. On suppose que la variable $X =$ «diamètre moyen d'une pièce usinée par la machine» suit une loi normale. Le technicien veut calculer un intervalle de confiance de niveau de confiance 95% de la variable X.

Puisqu'on ne connaît que la variance corrigée, on peut dire que $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$, avec $n = 9$ et $S^2 = 16$ mm^2. Donc, si l'on exprime tout en mm :

$$\begin{aligned}
95\% &= \text{Prob}\left(T_{8;0,975} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq T_{8;0,025}\right) \\
&= \text{Prob}\left(-2,306 \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq 2,306\right) \\
&= \text{Prob}\left(\overline{X} - 2,306\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + 2,306\frac{S}{\sqrt{n}}\right) \\
&= \text{Prob}\left(62 - 2,306\frac{4}{\sqrt{9}} \leq \mu \leq 62 + 2,306\frac{4}{\sqrt{9}}\right) = \text{Prob}(58,93 \leq \mu \leq 65,07).
\end{aligned}$$

Par conséquent, l'intervalle de confiance est : [58,93 mm, 65,07 mm]. ♦

Il reste encore un cas à étudier : celui où la variable X n'est pas distribuée selon une loi normale et sa variance n'est pas connue. Dans un tel cas, il est possible de montrer mathématiquement que lorsque n est suffisamment grand ($n > 75$), alors

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim \mathrm{N}(0;1).$$

Ce résultat n'est pas correct lorsque n est petit. On a donc :

> **Théorème 19 :** *Lorsque X ne suit pas une loi normale et que σ^2 n'est pas connu, si $n > 75$, l'intervalle de confiance pour μ, de niveau de confiance $1-\alpha$, est donné par :*
>
> $$\left[\overline{x} - z_{\alpha/2} \times \frac{s}{\sqrt{n}}; \overline{x} + z_{\alpha/2} \times \frac{s}{\sqrt{n}}\right].$$

On peut résumer l'ensemble des résultats que nous avons obtenus dans cette section avec le tableau ci-dessous.

Intervalles de confiance pour μ de niveau de confiance $1-\alpha$		
Situation	Loi utilisée	Bornes de l'intervalle
σ^2 connue $X \sim$ loi normale ou n grand (> 75)	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0;1)$	$\overline{x} \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$
σ^2 inconnue $X \sim$ loi normale	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$	$\overline{x} \pm t_{n-1;\alpha/2} \times \frac{s}{\sqrt{n}}$
σ^2 inconnue n très grand (> 75)	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim \mathrm{N}(0;1)$	$\overline{x} \pm z_{\alpha/2} \times \frac{s}{\sqrt{n}}$

8.3 Intervalle de confiance pour *p*

Pour déterminer les intervalles de confiance pour μ, nous avons abondamment utilisé les résultats de la section 7. Pour déterminer ceux pour la proportion p de succès, nous allons procéder de la même manière. Nous avions vu page 83 que lorsque n est suffisamment grand ($n > 30$) et lorsque p n'est ni trop grand ($nq \geq 5$) ni trop petit ($np \geq 5$), alors

$$\frac{\overline{P}-p}{\sqrt{\frac{pq}{n}}} \sim \mathrm{N}(0;1),$$

où $\overline{P}$ est la variable «proportion de succès dans un échantillon de taille n prélevé avec remise». En procédant comme dans les sous-sections précédentes, on obtient :

$$1-\alpha = \mathrm{Prob}\left(-z_{\alpha/2} \leq \frac{\overline{P}-p}{\sqrt{\frac{pq}{n}}} \leq z_{\alpha/2}\right),$$

où $z_{\alpha/2}$ est le quantile d'ordre $1-\alpha/2$. En isolant p, on obtient :

$$1-\alpha = \mathrm{Prob}\left(\overline{P} - z_{\alpha/2}\sqrt{\frac{pq}{n}} \leq p \leq \overline{P} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right).$$

Autrement dit,

$$p \in \left[\overline{P} - z_{\alpha/2}\sqrt{\frac{pq}{n}}; \overline{P} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right]$$

pour une proportion $1-\alpha$ de tous les échantillons possibles de taille n.

Cette équation pose toutefois un problème : on essaye d'estimer p et l'on retrouve justement p dans les bornes de l'intervalle : pour estimer l'intervalle de confiance pour p de niveau $1-\alpha$, il faut déjà connaître p. Pour s'en «sortir» mathématiquement, il suffit de trouver l'ensemble des p qui vérifient l'équation ci-dessus, ce qui revient à résoudre le système d'inéquations d'inconnue p suivant :

$$\begin{cases} p \geq \overline{P} - z_{\alpha/2}\sqrt{\frac{pq}{n}}, \\ p \leq \overline{P} + z_{\alpha/2}\sqrt{\frac{pq}{n}}. \end{cases}$$

En pratique, toutefois, on ne procède pas de cette manière car le calcul est trop fastidieux. Étant donné que la plupart du temps nous avons à notre disposition des échantillons de tailles relativement importantes, il est usuel d'approximer $\sqrt{\frac{pq}{n}}$ par $\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$, où $\overline{p}$ est la proportion de succès observée dans l'échantillon. On a alors la propriété suivante.

Lorsque n est grand ($n > 30$) et p ni trop grand, ni trop petit ($np \geq 5$, $nq \geq 5$), alors un intervalle de confiance pour p de niveau de confiance d'environ $1-\alpha$ («environ» à cause de l'approximation de p par $\overline{p}$) est donné par :

$$\left[\overline{p} - z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}; \overline{p} + z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$

Exemple 56 : Un sondage réalisé sur un échantillon de 200 électeurs d'un département a montré que 38% des gens comptent voter pour le parti A aux prochaines élections. À partir de cette information, il est possible d'estimer, avec un niveau de confiance de 95%, la proportion p d'électeurs dans le département qui comptent voter pour le parti A. En effet, cet intervalle est :

$$\left[0,38 - 1,96 \times \sqrt{\frac{0,38 \times 0,62}{200}}; 0,38 + 1,96 \times \sqrt{\frac{0,38 \times 0,62}{200}}\right] = [0,313; 0,447].$$

♦

Insistons sur le fait que cette technique de calcul ne fonctionne que si n est suffisamment grand pour que l'on puisse sans risque approximer $\sqrt{\frac{pq}{n}}$ par $\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$.

8.4 Précision d'une estimation par intervalle de confiance

Un examen rapide des divers intervalles présentés dans cette section montre que, pour un niveau de confiance donné, la précision de l'estimation obtenue augmente lorsque la taille de l'échantillon augmente. En effet, tous les intervalles présentés sont de la forme $\left[a - \frac{b}{\sqrt{n}}; a + \frac{b}{\sqrt{n}}\right]$, où a correspond à la valeur dans l'échantillon du paramètre que l'on veut estimer sur l'ensemble de la population, et où b est un nombre positif.

Exemple 57 : Considérons la forme des intervalles de confiance pour μ, de niveau de confiance 95%, lorsqu'un échantillon de taille n est prélevé avec remise sur une variable $X \sim \mathrm{N}(\mu; 100)$. L'intervalle de confiance pour μ admet alors pour bornes $\overline{x} \pm \frac{1,96 \times 10}{\sqrt{n}}$. Par conséquent, si $n = 100$, les bornes de l'intervalle sont $\overline{x} \pm 1,96$, mais si $n = 400$, elle sont $\overline{x} \pm 0,98$. ♦

On appelle erreur d'estimation la différence entre les bornes de l'intervalle et son centre (soit la moitié de l'écart entre les bornes de l'intervalle).

Dans l'exemple ci-dessus, lorsque n vaut 100, l'erreur d'estimation est de 1,96, alors qu'elle est de 0,98 lorsque n vaut 400. À partir de la notion d'erreur d'estimation, on peut aisément calculer à partir de quelle taille d'échantillon une précision d'estimation voulue peut être atteinte pour un niveau de confiance donné.

Exemple 58 : Déterminons à partir de quelle taille d'échantillon une estimation de μ, par intervalle de confiance de niveau 95%, aura une erreur de 1,0 lorsque la variable sur laquelle l'échantillon est prélevé a une variance $\sigma^2 = 100$. En supposant que n est assez grand pour que l'on puisse utiliser l'intervalle borné par $\overline{x} \pm 1,96 \times \frac{\sigma}{\sqrt{n}}$, on constate que l'erreur d'estimation est $1,96 \times \frac{\sigma}{\sqrt{n}} = \frac{10 \times 1,96}{\sqrt{n}}$. Une erreur de 1,0 est observée lorsque $\frac{19,6}{\sqrt{n}} = 1$, ou encore lorsque $n = 384,16$. La taille d'un échantillon étant entière, on peut donc en déduire que si l'on prélève un échantillon d'au moins 385 individus, alors, pour 95% des échantillons, $\mu \in [\overline{x} - 1; \overline{x} + 1]$. ♦

D'une manière générale,

> Lorsque μ peut être estimée par l'intervalle $\left[\overline{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right]$,
>
> alors toute taille d'échantillon n telle que $n \geq \left(z_{\alpha/2} \times \frac{\sigma}{e}\right)^2$
>
> permet une erreur d'estimation inférieure ou égale à e.

Exemple 59 : Combien de dossiers doit-on échantillonner afin d'avoir une erreur d'au plus 0,5, dans une estimation par intervalle de confiance de niveau 99%, du résultat moyen obtenu par un candidat s'étant présenté aux concours des ENSI, si l'on sait que les résultats suivent approximativement une loi normale et qu'ils s'étalent de 3 à 18 ?

Dans cet exemple, la valeur de σ^2 n'est pas connue mais nous pouvons en déduire une valeur approximative. En effet, nous savons que 97% de la population d'une loi normale est située entre $\mu - 3\sigma$ et $\mu + 3\sigma$. Autrement dit, la quasi-totalité de la population est située sur une plage de 6σ. Par conséquent, on peut poser que $18 - 3 = 15 \approx 6\sigma$, d'où une valeur approximative pour σ de $2,5$. La variable étudiée obéissant à une loi normale, le niveau de confiance 99% est atteint pour $z_{0,005} = 2,576$. Par conséquent, il faut prélever un échantillon de taille n telle que :

$$n \geq \left(2,576 \times \frac{2,5}{0,5}\right)^2 \approx 165,89.$$

Donc tout échantillon de taille supérieure ou égale à 166 permettra d'obtenir une erreur inférieure à 0,5. ♦

Il est également possible de déduire une procédure à suivre pour déterminer une taille d'échantillon permettant une précision donnée d'estimation pour une proportion de succès p. Cette procédure s'appuie sur la propriété algébrique suivante : $x(1-x) \leq 1/4$ pour tout $x \in \mathbb{R}$. De cette propriété, nous pouvons déduire que l'erreur d'estimation

$$z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

donnée par un intervalle de confiance de la forme

$$\left[\overline{p} - z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}; \overline{p} + z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right]$$

est inférieure ou égale à

$$z_{\alpha/2}\sqrt{\frac{1}{4n}} = \frac{z_{\alpha/2}}{2\sqrt{n}}.$$

D'où une taille d'échantillon n telle que $\frac{z_{\alpha/2}}{2\sqrt{n}} \leq e$ permet d'obtenir une proportion dont l'erreur d'estimation est inférieure à e.

> Toute taille d'échantillon n telle que $n \geq \left(\frac{z_{\alpha/2}}{2e}\right)^2$ permet d'obtenir une proportion de succès par un intervalle de confiance de niveau de confiance $1 - \alpha$ dont l'erreur d'estimation est inférieure ou égale à e.

Exemple 60 : À partir de quelle taille d'échantillon peut-on estimer une proportion de succès p, à l'aide d'un intervalle de confiance de niveau de confiance 95%, et obtenir une erreur d'estimation d'au plus 1% ?

En supposant que n est assez grand et que la proportion étudiée n'est ni trop grande ni trop petite, l'intervalle de confiance est

$$\left[\overline{p} - 1,96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}; \overline{p} + 1,96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$

La taille de l'échantillon nous est alors donnée par la formule suivante :

$$n \geq \left(\frac{1,96}{2 \times 0,01}\right)^2 = 9604.$$

♦

8.5 Exercices

Exercice 1 Un enseignant sait que, pour un certain examen, la distribution des notes est une loi normale d'écart-type 12. Construisez un intervalle de confiance de niveau de confiance 95% pour la moyenne de l'examen en utilisant l'échantillon suivant :

18	17	4	12	10
7	13	13	9	2
5	15	11	13	14
12	8	9	11	16

Si l'on désirait une erreur d'estimation inférieure à 0,5, quelle devrait être la taille minimale de l'échantillon ?

Exercice 2 Dans une entreprise, l'écart-type des salaires est de 7500F. On prélève un échantillon de 250 employés et on obtient un salaire annuel moyen sur l'échantillon de 90000F. Construisez un intervalle de confiance de niveau de confiance 99% pour le salaire moyen de cette entreprise.

Exercice 3 On vérifie si le volume moyen de bière versée dans les bouteilles par une machine respecte bien les standards de la compagnie. L'écart-type du volume de liquide versé par la machine est de 3 ml. On prélève un échantillon de 125 bouteilles provenant de la production de cette machine et on obtient une moyenne de 337 ml. Construisez l'intervalle de confiance de niveau de confiance 95% pour le volume moyen de bière versée dans les bouteilles.

Exercice 4 Afin de connaître la proportion de pièces défectueuses reçues d'un manufacturier lors d'une commande de plusieurs milliers de pièces, un acheteur prélève un échantillon de taille 250 pièces dans la livraison. Il s'avère que 7% des pièces de l'échantillon sont défectueuses alors que le manufacturier avait assuré qu'au plus 6% des pièces seraient défectueuses. L'acheteur est-il en droit de protester ?

Exercice 5 Une compagnie d'assurances désire estimer le pourcentage de conducteurs ayant subi un accident de circulation l'année dernière. La compagnie compte utiliser un intervalle de confiance de 95% et obtenir une erreur d'estimation d'au plus 3%. Calculez la taille minimale de l'échantillon à constituer.

Exercice 6 Un échantillon de 625 électeurs est prélevé afin de déterminer la proportion des électeurs favorables au PACS. Sur les 625 personnes interrogées, 360 se déclarent favorables au PACS.

1/ Estimez à l'aide d'un intervalle de confiance de niveau 95% la proportion véritable des électeurs favorables au PACS.

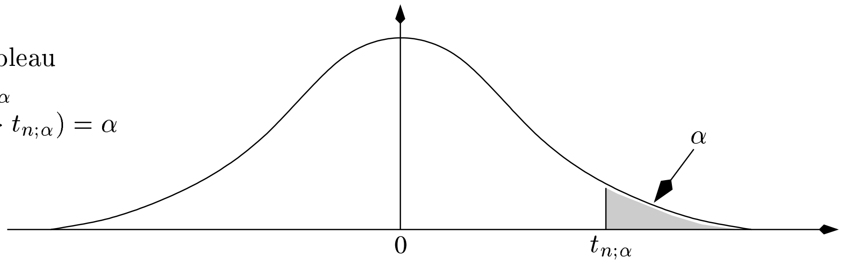
2/ Quelle taille minimale d'échantillon supplémentaire devez-vous prélever pour que l'erreur d'estimation de la proportion soit inférieure à 2% ?

Exercice 7 Les organisateurs d'un salon désirent étudier certaines caractéristiques de la population de leurs visiteurs. Il veulent en particulier estimer le salaire annuel moyen μ de cette population, ainsi que la proportion p

de propriétaires d'automobiles dans cette population. On peut considérer que l'écart-type sur les salaires est d'environ 40000F. Les organisateurs exigent à la fois une erreur d'estimation inférieure à 5000F pour le salaire et inférieure à 5% pour la proportion p. Pour un niveau de confiance de 95%, quelle taille d'échantillon faut-il prélever pour satisfaire à ces deux exigences ?

8.6 Table de la loi de Student

valeurs dans le tableau
ci-dessous : les $t_{n;\alpha}$
tels que $\text{Prob}(Z > t_{n;\alpha}) = \alpha$



$n \setminus \alpha$	0,10	0,05	0,025	0,01	0,005	0,001
1	3,078	6,314	12,706	31,821	63,657	318,309
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,106	4,025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385
31	1,309	1,696	2,040	2,453	2,744	3,375
32	1,309	1,694	2,037	2,449	2,738	3,365
33	1,308	1,692	2,035	2,445	2,733	3,356
34	1,307	1,691	2,032	2,441	2,728	3,348
35	1,306	1,690	2,030	2,438	2,724	3,340
36	1,306	1,688	2,028	2,434	2,719	3,333
37	1,305	1,687	2,026	2,431	2,715	3,326

$n \setminus \alpha$	0,10	0,05	0,025	0,01	0,005	0,001
38	1,304	1,686	2,024	2,429	2,712	3,319
39	1,304	1,685	2,023	2,426	2,708	3,313
40	1,303	1,684	2,021	2,423	2,704	3,307
41	1,303	1,683	2,020	2,421	2,701	3,301
42	1,302	1,682	2,018	2,418	2,698	3,296
43	1,302	1,681	2,017	2,416	2,695	3,291
44	1,301	1,680	2,015	2,414	2,692	3,286
45	1,301	1,679	2,014	2,412	2,690	3,281
46	1,300	1,679	2,013	2,410	2,687	3,277
47	1,300	1,678	2,012	2,408	2,685	3,273
48	1,299	1,677	2,011	2,407	2,682	3,269
49	1,299	1,677	2,010	2,405	2,680	3,265
50	1,299	1,676	2,009	2,403	2,678	3,261
51	1,298	1,675	2,008	2,402	2,676	3,258
52	1,298	1,675	2,007	2,400	2,674	3,255
53	1,298	1,674	2,006	2,399	2,672	3,251
54	1,297	1,674	2,005	2,397	2,670	3,248
55	1,297	1,673	2,004	2,396	2,668	3,245
56	1,297	1,673	2,003	2,395	2,667	3,242
57	1,297	1,672	2,002	2,394	2,665	3,239
58	1,296	1,672	2,002	2,392	2,663	3,237
59	1,296	1,671	2,001	2,391	2,662	3,234
60	1,296	1,671	2,000	2,390	2,660	3,232
61	1,296	1,670	2,000	2,389	2,659	3,229
62	1,295	1,670	1,999	2,388	2,657	3,227
63	1,295	1,669	1,998	2,387	2,656	3,225
64	1,295	1,669	1,998	2,386	2,655	3,223
65	1,295	1,669	1,997	2,385	2,654	3,220
66	1,295	1,668	1,997	2,384	2,652	3,218
67	1,294	1,668	1,996	2,383	2,651	3,216
68	1,294	1,668	1,995	2,382	2,650	3,214
69	1,294	1,667	1,995	2,382	2,649	3,213
70	1,294	1,667	1,994	2,381	2,648	3,211
71	1,294	1,667	1,994	2,380	2,647	3,209
72	1,293	1,666	1,993	2,379	2,646	3,207
73	1,293	1,666	1,993	2,379	2,645	3,206
74	1,293	1,666	1,993	2,378	2,644	3,204
75	1,293	1,665	1,992	2,377	2,643	3,202
∞	1,282	1,645	1,960	2,326	2,576	3,090

9 Les tests d'hypothèse

9.1 Qu'est-ce qu'un test d'hypothèse ?

Comme on l'a vu jusqu'à maintenant, en statistique, on se pose souvent des questions sur la valeur des paramètres p, μ, σ^2... et il n'est pas rare que l'on ait des décisions à prendre concernant ces valeurs. Par exemple, dans l'exercice 4 de la section précédente, un manufacturier nous assurait que p était égal à 6% et l'on avait obtenu sur un échantillon de taille 250 une valeur de p égale à 7%. On devait alors décider si le manufacturier nous mentait ou non, autrement dit s'il était plus probable que p vale 6% ou bien que p soit supérieur à 6%. Nous allons développer dans cette section des outils pour répondre à ce type de question.

Définition 35 : *Un test d'hypothèse est une règle de décision permettant de déterminer laquelle parmi deux hypothèses concernant la valeur d'un paramètre (p, μ, σ^2) est la plus plausible.*

La première étape dans la construction d'un test d'hypothèse, et peut-être la plus compliquée, consiste à identifier les deux hypothèses et à les formuler dans le langage statistique.

Les deux hypothèses à confronter seront toujours notées H_0 et H_1. H_0 est appelée l'hypothèse nulle et H_1 la contre-hypothèse. Ces deux hypothèses doivent impérativement être mutuellement exclusives.

En principe, H_0 est l'hypothèse que l'on essaye de vérifier. Elle est souvent fondée sur des études statistiques antérieures alors que H_1 reflète souvent l'impression du statisticien (l'expérimentateur) sur une modification de la valeur d'un paramètre. Dans l'exercice 4 de la section précédente, ce sont les expériences précédentes qui permettent au manufacturier de nous assurer que p vaut 6% ($H_0 : p = 6\%$). Mais il se peut très bien qu'une des machines de l'usine défaille, et dans ce cas, le nombre de pièces défectueuses peut augmenter ($H_1 : p > 6\%$).

Lorsqu'une hypothèse ne spécifie qu'une seule valeur du paramètre ($H_0 : \mu = 5$), on dit que $\boxed{\text{l'hypothèse est simple}}$. Dans le cas contraire, elle est $\boxed{\text{composée}}$. Lorsque les valeurs du paramètre sous H_1 sont toutes soit plus grandes, soit plus petites, que celles sous H_0, on dit que le $\boxed{\text{test est unilatéral}}$. Lorsque certaines valeurs du paramètre sous H_1 sont plus grandes et d'autres plus petites que la valeur spécifiée sous H_0, on dit que le test est $\boxed{\text{bilatéral}}$.

Exemple 61 :

$H_0 : \mu = 4$ $H_1 : \mu = 6$	hypothèse simple hypothèse simple	test unilatéral
$H_0 : \mu = 4$ $H_1 : \mu > 4$	hypothèse simple hypothèse composée	test unilatéral
$H_0 : \mu = 4$ $H_1 : \mu \neq 4$	hypothèse simple hypothèse composée	test bilatéral
$H_0 : \mu = 4$ $H_1 : \mu > 3$	hypothèse simple hypothèse composée	formulation incorrecte : les hypothèses ne sont pas mutuellement exclusives

♦

La formulation des hypothèses est l'une des étapes les plus importantes dans la construction d'un test d'hypothèse. Aussi n'est-il pas inutile de s'y attarder un peu. Formulons donc dans chacun des cas suivants les hypothèses H_0 et H_1 :

1. Une association de consommateurs examine un échantillon de 100 bouteilles d'un certain Bordeaux afin de déterminer si la quantité de vin contenu dans les bouteilles est bien égale à 75cl.
2. Un économiste enquête, auprès d'un échantillon de 400 individus de la population active, pour savoir si le taux de chômage, qui était de 10% le mois dernier, s'est modifié.
3. Un producteur de lait examine un échantillon de 30 bouteilles remplies par une machine afin de vérifier si celle-ci est bien réglée pour verser en moyenne 1/2 litre par bouteille.

Dans le cas 1, le paramètre étudié est clairement $\mu = E(X)$, où X est la quantité de vin dans les bouteilles de Bordeaux. Le rôle de l'association de consommateurs étant de s'assurer que ces derniers ne soient pas lésés, les hypothèses que l'on veut confronter ici sont donc $H_0 : \mu = 75$cl versus $H_1 : \mu < 75$cl. Le test est unilatéral.
Dans le cas 2, le paramètre étudié est p, la proportion de chômeurs dans la population active. L'économiste veut vérifier si le taux de chômage a été modifié (que ce soit une hausse ou une baisse de ce taux). Donc $H_0 : p = 10\%$ versus $H_1 : p \neq 10\%$. Le test est ici bilatéral.
Dans le cas 3, le paramètre étudié est $\mu = E(X)$, où X est le volume de lait versé par la machine dans les bouteilles. Pour le producteur, la machine est déréglée à partir du moment où le volume versé n'est pas en moyenne d'un demi litre. Donc, ici, $H_0 : \mu = 0,5$l versus $H_1 : \mu \neq 0,5$l.

On voit donc que la formulation des hypothèses dépend de l'enquête que l'on réalise : dans le cas 1, si l'on est un consommateur, on voudra tester $H_0 : \mu = 75$cl versus $H_1 : \mu < 75$cl ; par contre, si l'on est producteur, on testera $H_0 : \mu = 75$cl versus $H_1 : \mu \neq 75$cl.

9.2 La règle de décision du test

La règle de décision du test est fondée sur les résultats de l'échantillonnage. Ainsi, pour un test concernant la valeur de μ, c'est la valeur $\overline{x}$ prise par la moyenne de l'échantillon qui permet de prendre la décision, et dans un test sur p, c'est la proportion de succès $\overline{p}$ dans l'échantillon qui est examinée. Il est important de noter que les résultats de l'échantillonnage sont examinés après la formulation des hypothèses, et non avant. Les valeurs du paramètre sous celles-ci ne doivent pas être fixées à partir du résultat observé à partir de l'échantillon.

Pour construire la règle de décision, il faut en quelque sorte déterminer quelles sont les valeurs qu'il est peu probable que le paramètre étudié (par exemple $\overline{x}$) prenne dans l'échantillon si l'hypothèse H_0 est vraie. Pour cela, il faut examiner quelle est la distribution de l'estimateur du paramètre dans l'échantillon ($\overline{X}$) lorsque H_0 est vraie et déterminer une $\boxed{\text{région critique}}$, ou $\boxed{\text{région de rejet}}$ de H_0, telle que si la valeur prise par $\overline{X}$ est dans cette région, il est peu probable que H_0 soit vraie. La région critique doit tenir compte de la forme de la contre-hypothèse pour que le rejet de H_0 signifie que H_1 est un choix plausible.

Exemple 62 : Supposons qu'un échantillon de taille 25 soit examiné afin d'effectuer un test sur la moyenne μ d'une variable $X \sim \mathrm{N}(\mu; 100)$. Les hypothèses du test sont $H_0 : \mu = 10$ versus $H_1 : \mu > 10$. Sous H_0,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 10}{10/5} = \frac{\overline{X} - 10}{2} \sim \mathrm{N}(0; 1)$$

car, sous H_0, $X \sim \mathrm{N}(10; 100)$. Il est peu probable que $\overline{X}$ prenne une valeur éloignée de plus de 2 écart-types de sa moyenne car $\overline{X}$ suit une loi normale (la probabilité pour que cela arrive est de 4,56%). Il est donc peu probable que $\overline{X}$ prenne une valeur inférieure à 6 ou supérieure à 14.

Étant donné que la contre-hypothèse est $\mu > 10$, la loi de $\overline{X}$ sous H_1 a la même forme que sous H_0, mais elle est distribuée plus à droite. La région critique pourrait être «rejeter H_0 si $\overline{x} > 14$». Il est peu probable que la valeur $\overline{x}$ prise par $\overline{X}$ soit dans cette région si H_0 est vraie, et si tel était le cas, H_1 serait un choix plus plausible.

Une région critique de la forme «rejeter H_0 si $\overline{x} < 6$» n'aurait, quant à elle, aucun sens car H_1 ne serait pas un choix plausible. ♦

D'une manière générale, les règles de décision sont les suivantes :

Hypothèses	Règle de décision
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	«rejeter H_0 si $\overline{x} > c$», où c est un nombre plus grand que μ_0.
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	«rejeter H_0 si $\overline{x} < c$», où c est un nombre plus petit que μ_0.
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	«rejeter H_0 si $\overline{x} < c_1$ ou $c_2 < \overline{x}$», où c_1 et c_2 sont des nombres respectivement plus petit et plus grand que μ_0, et également éloignés de celui-ci.

9.3 Les erreurs pouvant résulter de la prise de décision

La région critique (ou région de rejet) d'un test est construite de manière à ce qu'il soit peu probable de rejeter H_0 alors que celle-ci est vraie. Pour autant, il reste un risque d'erreur. Le tableau ci-dessous résume ce qui peut résulter d'une prise de décision :

Décision prise \ Réalité	H_0 est vraie	H_1 est vraie
H_0 est rejetée	mauvaise décision : erreur de type I	bonne décision
H_0 n'est pas rejetée	bonne décision	mauvaise décision : erreur de type II

Comme l'indique ce tableau, deux types d'erreurs sont possibles : l'erreur de type I, qui consiste à rejeter H_0 alors que cette hypothèse est vraie, et l'erreur de type II, qui consiste à ne pas rejeter H_0 alors que c'est H_1 qui est vraie. Bien évidemment, même si le test a été construit de sorte que de telles erreurs soient peu probables, il existe toujours un risque qu'elles soient réalisées. Voilà pourquoi le statisticien définit les probabilités de réalisation de ces erreurs :

$$\begin{aligned}\alpha &= \text{probabilité de réaliser une erreur de type I}\\ &= \text{probabilité de rejeter } H_0 \text{ sachant que } H_0 \text{ est vraie}\\ &= \text{Prob(rejeter } H_0|H_0 \text{ est vraie)},\end{aligned}$$

$$\begin{aligned}\beta &= \text{probabilité de réaliser une erreur de type II}\\ &= \text{probabilité de rejeter } H_1 \text{ sachant que } H_1 \text{ est vraie}\\ &= \text{Prob(rejeter } H_1|H_1 \text{ est vraie)}.\end{aligned}$$

Un test sera alors d'autant plus une bonne règle de décision que les valeurs de α et de β sont petites. La probabilité α de réaliser une erreur de type I est habituellement assez simple à contrôler étant donné que l'on construit la région critique afin de minimiser les risques d'erreur de type I. Cette probabilité répond aux doux noms de $\boxed{\text{niveau de signification}}$ et de $\boxed{\text{niveau de test}}$.

Exemple 62 (suite) : Calculons le niveau de signification associé au test de l'exemple 62. On sait que la variable étudiée X suit une loi normale $\text{N}(\mu; 100)$. Nous avions déterminé les hypothèses $H_0 : \mu = 10$ versus $H_1 : \mu > 10$. La taille de l'échantillon est $n = 25$ et la région critique est «rejeter H_0 si $\overline{x} > 14$». Pour calculer α, nous allons donc nous placer sous l'hypothèse H_0. Alors $Z = \frac{\overline{X} - 10}{2} \sim \text{N}(0; 1)$. Par conséquent,

$$\begin{aligned}\alpha &= \text{Prob(rejeter } H_0|H_0 \text{ est vraie)}\\ &= \text{Prob}(\overline{X} > 14|\mu = 10)\\ &= \text{Prob}\left(\frac{\overline{X} - 10}{2} > \frac{14 - 10}{2}|\mu = 10\right)\\ &= \text{Prob}(Z > 2) = 0,0228.\end{aligned}$$

Autrement dit, il y a un risque de 2,28% de commettre une erreur de type I. ♦

Le calcul de β est plus compliqué car H_1 est une hypothèse composée. Dans un tel cas, on calcule β pour plusieurs valeurs du paramètre sous la contre-hypothèse. Cela nous permet d'obtenir une courbe des valeurs de β en fonction des valeurs du paramètre. En fait, pour des raisons calculatoires, on s'intéresse plutôt aux valeurs prises par $1-\beta$. La courbe correspondante est appelée $\boxed{\text{la courbe de puissance du test}}$. $1-\beta$ est appelé $\boxed{\text{la puissance du test}}$. Elle correspond à la probabilité de prendre une décision correcte lorsque H_1 est vraie, autrement dit,

$$\begin{aligned}1-\beta &= 1 - \text{Prob(rejeter } H_1|H_1 \text{ est vraie)}\\ &= \text{Prob(rejeter } H_0|H_1 \text{ est vraie)}.\end{aligned}$$

Exemple 62 (suite) : Calculons la courbe de puissance du test de l'exemple 62. Pour cela, considérons une valeur de μ sous H_1. Par exemple, $\mu = 11$. Alors,

$$1 - \beta(11) = \text{Prob}(\text{rejeter } H_0 | H_1 : \mu = 11 \text{ est vraie}).$$

Or, si H_1 est vraie, $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{\overline{X} - 11}{2} \sim \text{N}(0; 1)$. Donc,

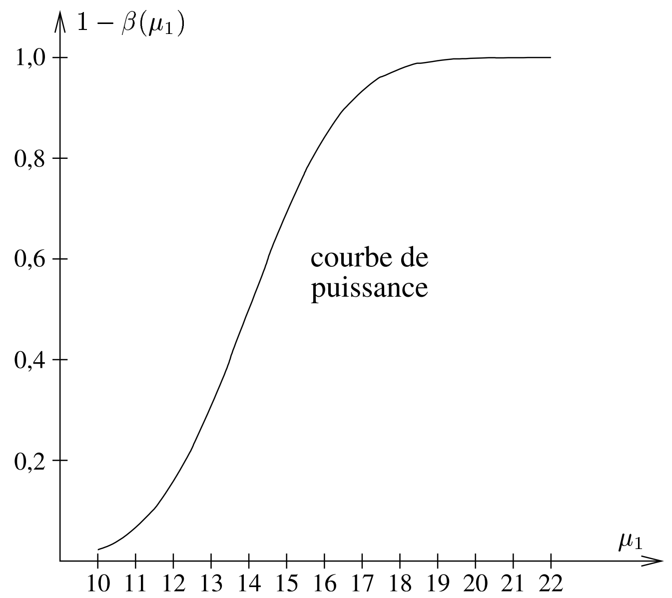
$$\begin{aligned} 1 - \beta(11) &= \text{Prob}(\overline{X} > 14 | \mu = 11) \\ &= \text{Prob}\left(\frac{\overline{X} - 11}{2} > \frac{14 - 11}{2} | \mu = 11\right) \\ &= \text{Prob}(Z > 1,5) = 0,0668. \end{aligned}$$

Plus généralement, si μ_1 est une valeur de μ sous H_1, nous obtenons :

$$1 - \beta(\mu_1) = \text{Prob}\left(\frac{\overline{X} - \mu_1}{2} > \frac{14 - \mu_1}{2}\right).$$

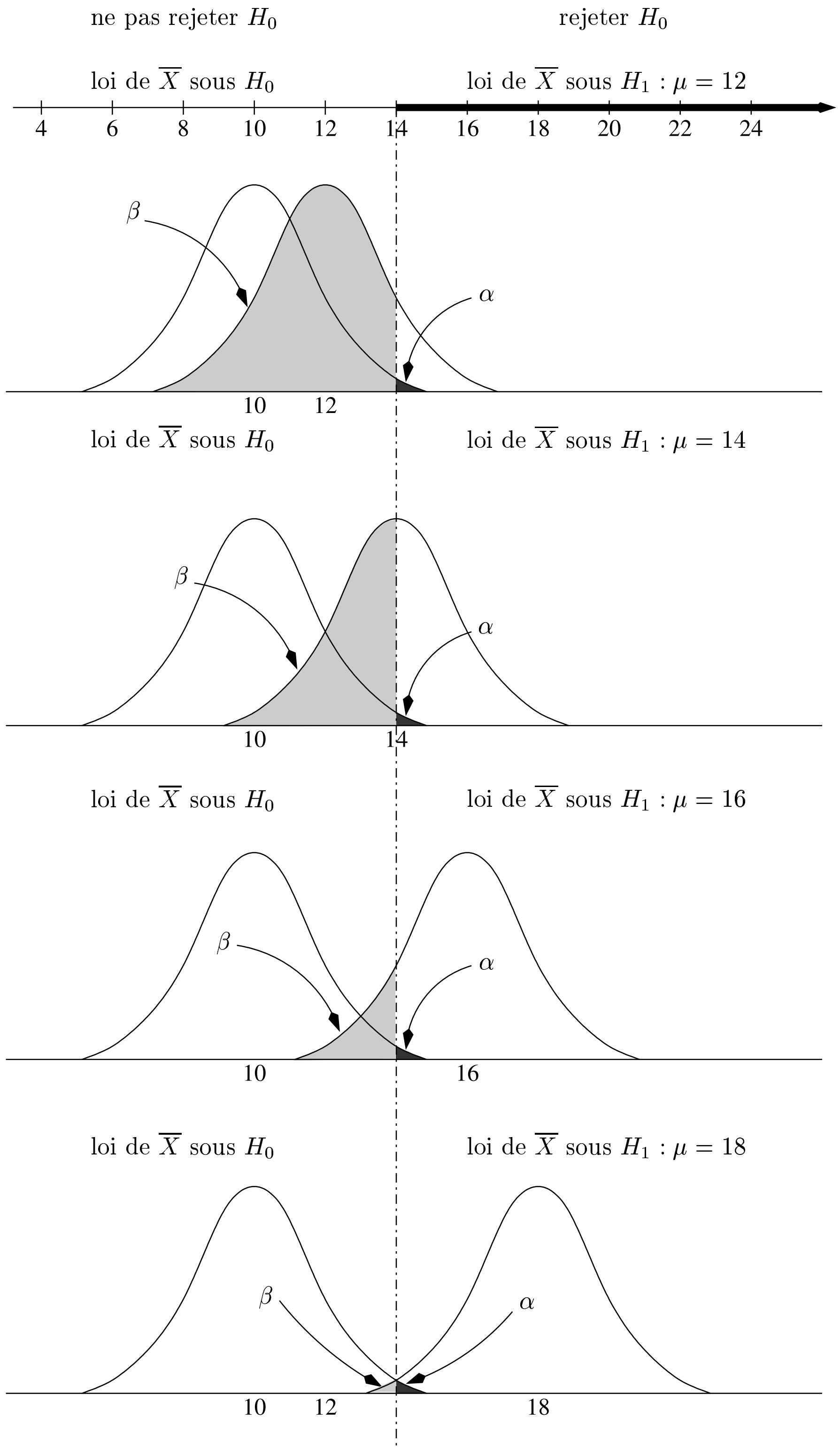
L'ensemble des valeurs que peut prendre μ_1 nous est fourni par l'hypothèse $H_1 : \mu_1 > 10$. Cela correspond à l'intervalle $]10, +\infty[$. On peut donc dresser le tableau suivant, et la courbe de puissance du test :

μ_1	$z_1 = \frac{14 - \mu_1}{2}$	$1 - \beta(\mu_1) = \text{Prob}(Z > z_1)$	$\beta(\mu_1)$
10	2,0	0,0228	0,9772
11	1,5	0,0668	0,9332
12	1,0	0,1587	0,8413
13	0,5	0,3085	0,6915
14	0,0	0,5000	0,5000
15	-0,5	0,6915	0,3085
16	-1,0	0,8413	0,1587
17	-1,5	0,9332	0,0668
18	-2,0	0,9772	0,0228
19	-2,5	0,9938	0,0062
20	-3,0	0,9986	0,0014
21	-3,5	0,9998	0,0002
22	-4,0	1,0000	0,0000



Notons que la valeur minimale que peut prendre $1 - \beta$ correspond exactement à la valeur de α. ♦

Remarquons que plus la valeur de μ sous H_1 s'éloigne de la valeur de μ sous H_0, plus la puissance du test est élevée. Autrement dit, la probabilité de commettre une erreur de type II est très forte lorsque la contre-hypothèse et l'hypothèse nulle sont très proches l'une de l'autre. Cela n'est pas surprenant puisque, dans un tel cas, les distributions de $\overline{X}$ sous l'une ou l'autre des hypothèses se confondent presque. Il est alors assez difficile de rejeter H_0 lorsque H_1 est vraie. On peut illustrer cela sur les figures suivantes :



9.4 Exemple de calcul de la région critique et de la prise de décision

En pratique, la région critique d'un test se construit après que le niveau de test ait été fixé et non l'inverse comme cela a été fait jusqu'à maintenant. Généralement, on fixe la valeur de α à 0,10, 0,05, 0,02 ou 0,01, en fonction de la gravité des conséquences d'une erreur de type I. Une fois la région critique construite, la puissance du test est examinée pour les valeurs de la contre-hypothèse pour lesquelles la réalisation d'une erreur de type II a des conséquences fâcheuses.

Exemple 63 : La loi oblige tout automobiliste à contracter une assurance. La prime exigée annuellement d'un assuré dépend de plusieurs facteurs : la zone habitée, le type de véhicule, l'utilisation à des fins commerciales ou non, la distance estimée que parcourra l'assuré... Il est presque impossible d'estimer la distance parcourue par un automobiliste pour une année donnée. Voilà pourquoi tous les assurés d'un véhicule non utilisé à des fins commerciales se voient imposer le même montant sur ce point. Celui-ci est fonction de la distance moyenne parcourue annuellement par les automobilistes de cette catégorie. Des études ont montré que celle-ci était de 18000 km avec un écart-type de 5000 km. Le montant que l'on prévoit d'exiger est de 13 centimes du km, autrement dit $18000 \times 0,13 = 2340$F. Le montant de cette prime a continuellement augmenté ces dernières années, de telle sorte que l'opinion publique commence à être très mécontente et à exercer de fortes pressions sur les compagnies d'assurances pour qu'elles baissent leurs tarifs.

C'est ainsi que la MAIF est priée de réévaluer tous les facteurs considérés dans le calcul de la prime. Le plus vulnérable de ces facteurs est précisément la distance parcourue annuellement. Un statisticien est donc chargé de réexaminer le bien-fondé de l'estimation à 18000 km de la moyenne contestée. La démarche qu'il compte suivre est de prélever rapidement un échantillon de 400 individus afin de tester si la moyenne a effectivement diminué. Si tel est le cas, une étude plus exhaustive, menée sur un grand nombre d'assurés, sera entreprise afin d'estimer très précisément la valeur de la moyenne. Sinon, ce facteur ne sera pas révisé.

La variable qu'étudie le statisticien est X : la distance parcourue en 2000 (dernière année complète sur laquelle on peut fonder l'étude), par un véhicule utilisé à des fins non commerciales. Il décide pour l'instant de ne pas remettre en cause l'estimation de l'écart-type σ de X (5000 km). Les hypothèses qu'il formule sont les suivantes :

$$H_0 : \mu = 18000 \quad \text{versus} \quad H_1 : \mu < 18000, \text{ où } \mu = E(X).$$

Il s'interroge sur les conséquences d'une erreur de type I afin de fixer α. Une erreur de type I (H_0 serait rejetée alors qu'elle est vraie) entraînerait la réalisation de l'étude exhaustive pour rien, donc une dépense inutile, et porterait atteinte à sa réputation. Il préfère donc fixer α à 0,01. Il peut alors fixer la région critique du test : celle-ci est de la forme «rejeter H_0 si $\overline{x} < c$» étant donné que les valeurs de μ sous H_1 sont plus petites que celles sous H_0. De plus, $n = 400$ est assez grand pour que le théorème central limite s'applique. Ainsi, sous H_0,

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 18000}{5000/\sqrt{400}} = \frac{\overline{X} - 18000}{250} \sim \mathrm{N}(0;1).$$

D'où

$$\begin{aligned}
\alpha = 0,01 &= \mathrm{Prob}(\overline{X} < c | \mu = 18000) \\
&= \mathrm{Prob}\left(\frac{\overline{X} - 18000}{250} < \frac{c - 18000}{250} | \mu = 18000\right) \\
&= \mathrm{Prob}\left(Z < \frac{c - 18000}{250}\right) \\
&= \mathrm{Prob}(Z < -2,326).
\end{aligned}$$

Par conséquent, $\frac{c - 18000}{250} = -2,326$, ou encore $c = 17418,5$. La règle de décision suggérée par la région critique est donc :

«rejeter H_0 si $\overline{x} < 17418,5$».

Notre statisticien doit maintenant examiner la puissance de son test afin de voir si sa règle de décision est «solide». Il réfléchit alors sur les conséquences d'une erreur de type II (rejeter H_1 alors que H_1 est vraie). Si une telle erreur se produisait, les automobilistes n'obtiendraient pas une réduction du prix de la prime alors qu'il y auraient droit. Si la diminution à laquelle ils avaient droit se chiffrait à 125F ou moins, on ne pourrait pas parler de conséquences sérieuses. Le tarif étant de 13 centimes au km, une diminution de 125F correspondrait à $125/0,13 = 961,54$ km. Le statisticien vérifie donc la puissance du test pour une valeur de $\mu = 18000 - 961,54 = 17038,46$ km.

$$\begin{aligned}
1-\beta(17038,46) &= \text{Prob}(\text{rejeter } H_0|H_1 : \mu = 17038,46 \text{ est vraie}) \\
&= \text{Prob}(\overline{X} < 17418,5|\mu = 17038,46) \\
&= \text{Prob}\left(\frac{\overline{X}-17038,46}{250} < \frac{17418,5-17038,46}{250}|\mu = 17038,46\right) \\
&= \text{Prob}(Z < 1,52) \\
&= 0,9357.
\end{aligned}$$

Étant donné qu'il sait que la puissance d'un test augmente lorsque la valeur de μ sous H_1 s'éloigne de celle sous H_0, il est rassuré car pour toute diminution de la moyenne parcourue permettant de diminuer les tarifs de 125F ou plus, c'est-à-dire pour toute valeur de $\mu \leq 17038,46$, la probabilité de commettre une erreur de type II est inférieure à 1 - 0,9357 = 6,43%. Il peut alors procéder à l'échantillonnage puisque le test construit est «bon». ♦

Lorsque la règle de décision est construite, la dernière étape d'un test consiste à prélever l'échantillon et à prendre une décision. Si la valeur observée $\overline{x}$ de $\overline{X}$, ou $\overline{p}$ pour $\overline{P}$, permet de rejeter H_0, le test est dit **significatif**.

Exemple 63 (suite) : Après avoir prélevé son échantillon de 400 automobilistes, le statisticien a observé une moyenne $\overline{x} = 17230$ km. Sa règle de décision était de rejeter H_0 si $\overline{x}$ était inférieur à $17418,5$ km. Il conclut donc qu'il faut rejeter l'hypothèse H_0. Le test est significatif. Autrement dit, la différence entre la valeur observée dans l'échantillon et la valeur 18000 est statistiquement significative. ♦

Remarquons que la taille de l'échantillon et la valeur du niveau α influent sur la qualité du test, c'est-à-dire sur la puissance du test. En effet, pour un niveau α fixé, lorsque la taille de l'échantillon augmente, la variance de $\overline{X}$ (σ^2/n) diminue, le profil de la distribution de $\overline{X}$ est plus resserré et les distributions de $\overline{X}$, centrées à μ_0 sous H_0 vraie et à μ_1 sous H_1 vraie, se distinguent plus facilement. Dans ce cas, la puissance du test augmente.

Pour une taille d'échantillon fixe, la grandeur de la région critique (rejeter H_0) augmente ou diminue selon que la valeur du niveau α augmente ou diminue. La puissance $1-\beta$ étant associée à cette même région critique, elle varie dans le même sens que le niveau α.

La puissance d'un test augmente dans le même sens que le niveau α et que la taille de l'échantillon.

9.5 Les tests usuels sur μ

Dans cette sous-section, nous allons identifier les formes exactes des régions critiques des tests usuels sur μ. Commençons par le cas où la variance σ^2 de la variable étudiée est connue. Si celle-ci obéit à une loi normale, ou bien si la taille n de l'échantillon est suffisamment grande pour que le théorème central limite s'applique, on a :

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \text{N}(0;1).$$

Supposons que le niveau de test est α et que l'hypothèse nulle est $H_0 : \mu = \mu_0$. Étudions la forme de la région critique du test en considérant trois cas pour la contre-hypothèse.

premier cas : supposons que la contre-hypothèse est $H_1 : \mu > \mu_0$. Alors la région critique est de la forme «rejeter H_0 si $\overline{x} > c$». La valeur de c est déterminée en explicitant la signification de α :

$$\begin{aligned}
\alpha &= \text{Prob(rejeter } H_0 | H_0 \text{ est vraie)} \\
&= \text{Prob}(\overline{X} > c | \mu = \mu_0) \\
&= \text{Prob}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_0\right) \\
&= \text{Prob}\left(Z > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= \text{Prob}(Z > z_\alpha).
\end{aligned}$$

On en déduit que $\dfrac{c - \mu_0}{\sigma/\sqrt{n}} = z_\alpha$, ou encore que $c = \mu_0 + z_\alpha \sigma/\sqrt{n}$.

deuxième cas : supposons que H_1 est de la forme $\mu < \mu_0$. Alors, d'une manière similaire au cas précédent, on montre que la région critique est de la forme «rejeter H_0 si $\overline{x} < c$» où $c = \mu_0 - z_\alpha \sigma/\sqrt{n}$.

troisième cas : supposons que H_1 est de la forme $\mu \neq \mu_0$. Alors la région critique est de la forme «rejeter H_0 si $\overline{x} < c_1$ ou si $\overline{x} > c_2$». Les valeurs de c_1 et c_2 sont obtenues en explicitant la valeur de α et en attribuant une probabilité $\alpha/2$ au rejet par chacun des côtés :

$$\begin{aligned}
\alpha &= \text{Prob(rejeter } H_0 | H_0 \text{ est vraie)} \\
&= \text{Prob}(\overline{X} < c_1 | \mu = \mu_0) + \text{Prob}(\overline{X} > c_2 | \mu = \mu_0) \\
&= \text{Prob}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{c_1 - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_0\right) + \text{Prob}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c_2 - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_0\right) \\
&= \text{Prob}\left(Z < \frac{c_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \text{Prob}\left(Z > \frac{c_2 - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= \text{Prob}(Z < -z_{\alpha/2}) + \text{Prob}(Z > z_{\alpha/2}).
\end{aligned}$$

On en déduit :

$$c_1 = \mu_0 - z_{\alpha/2}\sigma/\sqrt{n} \quad \text{et} \quad c_2 = \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}.$$

Les formes de tests à utiliser lorsque σ^2 est inconnue se déduisent de manière analogue, mais en utilisant les lois de probabilité que nous avions vu dans la section précédente :

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim T_{n-1} \qquad \text{(valide si } X \text{ suit une loi normale)}$$

ou bien

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathrm{N}(0; 1) \qquad \text{(valide si } n \text{ est très grand).}$$

On peut donc résumer les formes de toutes les régions critiques dans le tableau suivant.

Régions critiques des test usuels sur μ lorsque la taille de l'échantillon $= n$, le niveau $= \alpha$, et $H_0 : \mu = \mu_0$

Situation	Loi utilisée	Forme de la région critique		
		$H_1 : \mu < \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu \neq \mu_0$
σ^2 connue et $X \sim$ normale ou n grand	$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0;1)$	$\overline{x} < \mu_0 - z_\alpha \sigma/\sqrt{n}$	$\overline{x} > \mu_0 + z_\alpha \sigma/\sqrt{n}$	$\overline{x} < \mu_0 - z_{\alpha/2} \sigma/\sqrt{n}$ ou $\overline{x} > \mu_0 + z_{\alpha/2} \sigma/\sqrt{n}$
σ^2 inconnue et $X \sim$ normale	$\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim T_{n-1}$	$\overline{x} < \mu_0 - t_{n-1;\alpha} s/\sqrt{n}$	$\overline{x} > \mu_0 + t_{n-1;\alpha} s/\sqrt{n}$	$\overline{x} < \mu_0 - t_{n-1;\alpha/2} s/\sqrt{n}$ ou $\overline{x} > \mu_0 + t_{n-1;\alpha/2} s/\sqrt{n}$
σ^2 inconnue et n très grand	$\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathrm{N}(0;1)$	$\overline{x} < \mu_0 - z_\alpha s/\sqrt{n}$	$\overline{x} > \mu_0 + z_\alpha s/\sqrt{n}$	$\overline{x} < \mu_0 - z_{\alpha/2} s/\sqrt{n}$ ou $\overline{x} > \mu_0 + z_{\alpha/2} s/\sqrt{n}$

Exemple 64 : Vous voulez investir à la bourse. Afin d'optimiser vos profits, vous relevez pendant deux semaines le cours du CAC40. Au début des deux semaines, celui-ci vaut 6155 points. Dans l'échantillon de 15 jours, le CAC40 vaut en moyenne 6170 points, avec un écart-type de 6 points. Vous ne voulez investir que si le CAC40 est à la hausse. Sachant que la variable X = «valeur du CAC40» suit une loi normale, vous effectuez donc un test d'hypothèse de niveau de confiance 99% pour savoir si le CAC40 a augmenté :
X = «valeur du CAC40» suit une loi normale.

$$H_0 : \mu = 6155 \qquad \text{versus} \qquad H_1 : \mu > 6155$$

$$\begin{aligned}
0,01 &= \text{Prob(rejeter } H_0 | H_0 \text{ est vraie)} \\
&= \text{Prob}\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > c \,\middle|\, \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0;1) \right) \\
&= \text{Prob}\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > 2,325 \right) \\
&= \text{Prob}\left(\overline{X} > \mu + 2,325\sigma/\sqrt{n} \right) \\
&= \text{Prob}\left(\overline{X} > 6155 + 2,325 \times 6/\sqrt{15} \right) = \text{Prob}\left(\overline{X} > 6158,6 \right).
\end{aligned}$$

Donc si $\overline{X} \leq 6158,6$, on doit rejeter la perspective d'une hausse. Or ce n'est pas le cas ici, donc il est plus probable qu'il y ait effectivement eu une hausse.

Maintenant, pour vous assurer que le test était bien significatif, vous voulez calculer la puissance du test pour $\mu = 6170$:

$$\begin{aligned}
1-\beta(6170) &= \text{Prob(rejeter } H_0 | H_1 \text{ est vraie)} \\
&= \text{Prob}(\overline{X} > 6158,6 | \mu = 6170) \\
&= \text{Prob}\left(\frac{\overline{X}-6170}{\sigma/\sqrt{n}} > \frac{6158,6-6170}{\sigma/\sqrt{n}} \,\middle|\, \frac{\overline{X}-6170}{\sigma/\sqrt{n}} \sim N(0;1) \right) \\
&= \text{Prob}\left(\frac{\overline{X}-6170}{6/\sqrt{15}} > \frac{6158,6-6170}{6/\sqrt{15}} \,\middle|\, \frac{\overline{X}-6170}{6/\sqrt{15}} \sim N(0;1) \right) \\
&= \text{Prob}\left(\frac{\overline{X}-6170}{6/\sqrt{15}} > -7,36 \,\middle|\, \frac{\overline{X}-6170}{6/\sqrt{15}} \sim N(0;1) \right) = 1.
\end{aligned}$$

On peut donc en conclure que le test était bien significatif. ♦

9.6 Les tests sur p

Lorsque la taille de l'échantillon est grande et que la proportion de succès p sur laquelle on veut tester une hypothèse n'est ni trop petite ni trop grande, on peut utiliser la formule que nous avions vue dans la section précédente, à savoir :

$$Z = \frac{\overline{P}-p}{\sqrt{\frac{pq}{n}}} \sim \mathrm{N}(0;1).$$

En procédant comme dans la sous-section précédente, nous pouvons déduire que la région critique d'un test de niveau α, fondée sur un échantillon de taille n, pour $H_0 : p = p_0$, consiste à rejeter H_0 dans les cas suivants :

région critique	contre-hypothèse
$\overline{p} < p_0 - z_\alpha\sqrt{\frac{p_0q_0}{n}}$	$H_1 : p < p_0$
$\overline{p} > p_0 + z_\alpha\sqrt{\frac{p_0q_0}{n}}$	$H_1 : p > p_0$
$\overline{p} < p_0 - z_{\alpha/2}\sqrt{\frac{p_0q_0}{n}}$ ou $\overline{p} > p_0 + z_{\alpha/2}\sqrt{\frac{p_0q_0}{n}}$	$H_1 : p \neq p_0$

Exemple 65 : Actuellement, la proportion de gens atteints d'une certaine maladie chez qui le traitement usuel permet d'enrayer le mal en moins de deux mois est de 36%. Afin de juger l'efficacité d'un nouveau médicament, on traite avec celui-ci un échantillon de 100 patients nouvellement atteints par cette maladie. Dans quelle proportion de ceux-ci doit-on observer une guérison en moins de deux mois pour qu'à un niveau de 5% on puisse affirmer que ce nouveau traitement permet une guérison rapide plus souvent que ne le fait le traitement usuel ?

Soit p la proportion des patients traités grâce au nouveau médicament pour lesquels la guérison se produit en moins de deux mois. Les hypothèses à tester sont : $H_0 : p = 0,36$ versus $H_1 : p > 0,36$. La taille n de l'échantillon est grande, p n'est ni trop petit, ni trop grand sous H_0. On peut donc considérer que sous H_0,

$$Z = \frac{\overline{P} - p}{\sqrt{\frac{pq}{n}}} \sim \mathrm{N}(0;1).$$

La région critique du test est décrite par «rejeter H_0 si $\overline{p} > c$» où

$$c = p_0 + z_{0,05}\sqrt{\frac{p_0 q_0}{n}} = 0,36 + \left(1,645 \times \sqrt{\frac{0,36 \times 0,64}{100}}\right) = 0,439.$$

Il faudra donc observer une proportion de guérisons rapides supérieure à 0,439 pour pouvoir conclure par ce test que le nouveau médicament est meilleur que l'ancien. ♦

9.7 Exercices

Exercice 1 On croit savoir que la population des gens mariés dans une population adulte est de 70%. Un sociologue désire confronter les hypothèses suivantes pour la proportion p de gens mariés :

$$H_0 : p = 70\% \quad \text{versus} \quad H_1 : p < 70\%.$$

Pour ce faire, le sociologue prélève un échantillon de 500 individus et utilise la règle de décision suivante : «rejeter H_0 si $\overline{p} < 67\%$».
1/ Calculer la valeur de α.
2/ Calculer la puissance du test pour $p = 65\%$.

Exercice 2 On prélève un échantillon de 16 étudiants dans un groupe ayant subi un examen. Les notes obtenues (sur 100) sont :

97	68	60	75	82	81	80	69
64	80	68	70	72	76	74	63

La distribution des notes est une loi normale de variance 95.
1/ Peut-on affirmer que la moyenne du groupe est supérieure à 70 si on utilise un test de niveau de signification de 5% ?
2/ Calculez la puissance du test lorsque la contre-hypothèse spécifie une moyenne de 74.

Exercice 3 Après avoir effectué des modifications mécaniques sur un modèle de voiture, un constructeur affirme que la consommation d'essence en l/100 km a diminué. Une association de consommateurs examine la consommation d'essence des voitures d'un échantillon de ce modèle. Elle obtient le résultat suivant :

10,35	10,29	10,36	10,35	10,35
10,26	10,33	10,35	10,29	10,38
10,37	10,39	10,30	10,34	10,32
10,39	10,38	10,32	10,35	10,34
10,35	10,33	10,39	10,34	10,36

On sait qu'habituellement la distribution de la consommation d'essence suit une loi normale. Sachant qu'avant les modifications, la moyenne était de 10,4 l/100 km, et en utilisant un test de niveau de signification de 1%, peut-on conclure que le constructeur a raison ?

Exercice 4 Lors des dernières élections, 45% des électeurs ont voté pour le parti A. Ce parti commande un sondage afin de savoir si sa popularité a augmenté. Dans un échantillon de 1254 électeurs, 588 affirment qu'ils voteraient pour ce parti s'il y avait à nouveau des élections. À un niveau de 5%, peut-on conclure que la popularité du parti est à la hausse depuis les dernières élections ?

10 Loi du χ^2, intervalle de confiance sur σ^2, test d'ajustement et d'indépendance

Nous avons vu dans les sections 8 et 9 comment réaliser des intervalles de confiance et des tests d'hypothèse sur les paramètres μ et p. Il manquait à notre éventail d'outils statistiques celui nous permettant de déterminer ces intervalles et ces tests pour le paramètre σ^2. L'un des objectifs de cette section est de combler cette lacune. Pour cela, nous allons introduire une nouvelle loi, à savoir la loi du χ^2. Cette loi a des propriétés tout à fait intéressantes. En particulier, nous verrons qu'elle permet d'établir si une variable suit une certaine distribution, ou bien encore de tester si deux caractères d'une même population sont indépendants. Mais commençons par décrire ce qu'est la loi du χ^2.

10.1 La loi du χ^2

Outre l'omniprésente loi normale, et les deux autres lois que nous avons rencontrées, i.e., la loi binômiale et la loi de Student, il existe une quatrième loi très utilisée en statistiques : la loi du χ^2. À l'instar de la loi de Student, elle n'a qu'un seul paramètre, $n \in \mathbb{N}^*$, que l'on nomme $\boxed{\text{nombre de degrés de liberté}}$. Pour un n donné, la distribution du χ^2 est notée : χ^2_n.

Définition 36 : *La loi χ^2_n représente la distribution d'une variable aléatoire continue pouvant prendre comme valeur tout nombre positif. L'espérance d'une telle variable est n, et sa variance $2n$.*

La distribution du χ^2 n'est pas symétrique : si l'on tire 1000 nombres au hasard suivant cette loi, on doit obtenir une moyenne de n. Donc, on ne peut tirer que des nombres compris entre 0 et $1000 \times n$. Il est évident que l'on ne pourra pas tirer beaucoup de nombres élevés (au plus 1 nombre supérieur à $500 \times n$ par exemple). Par contre, cette restriction ne s'applique pas aux nombres petits. La distribution du χ^2 doit donc être plus «élevée» vers la gauche que vers la droite. Cela se vérifie effectivement sur la figure 31.

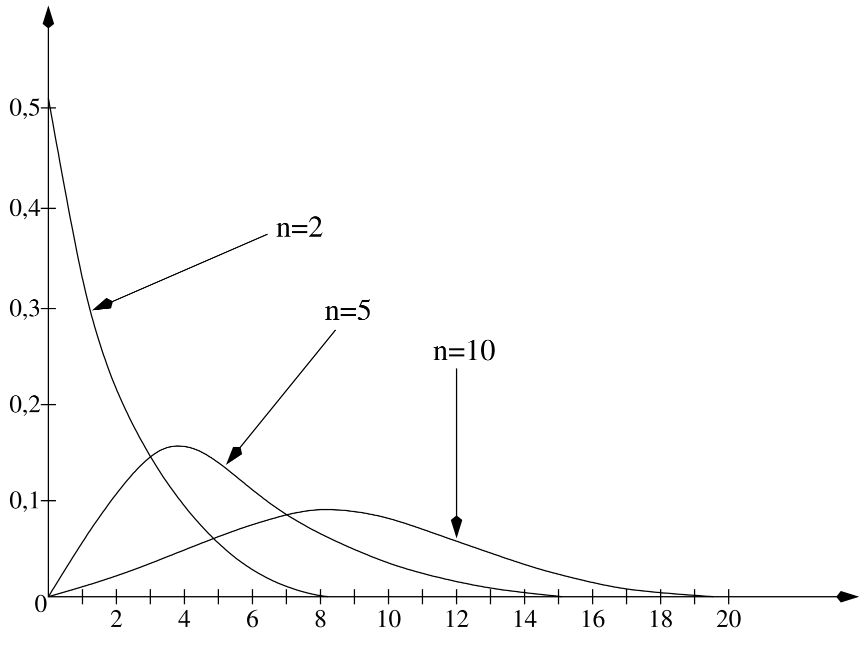


FIG. 31 – Distribution de $X \sim \chi^2_n$ pour quelques valeurs de n.

On peut voir sur cette figure que lorsque la valeur de n augmente, la distribution prend de plus en plus la forme de cloche qu'a la loi normale de moyenne n et de variance $2n$.

Comme pour les autres lois, lorsque nous utiliserons celle-ci, nous nous servirons de sa table de distribution. Cette dernière est formée sur le même principe que les tables précédentes : le quantile d'ordre $1-\alpha$ d'une variable aléatoire X suivant une loi χ^2_n est noté $c_{n;\alpha}$ et correspond au nombre réel positif tel que $\text{Prob}(X > c_{n;\alpha}) = \alpha$ (cf. la figure ci-après).

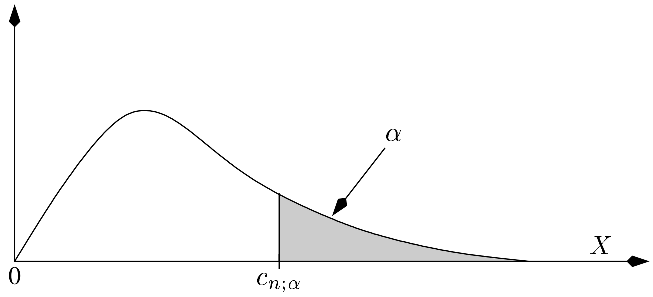


FIG. 32 – Les quantiles de la loi du χ^2.

Toutefois, comme la loi du χ^2 ne nous servira que pour construire des intervalles de confiance et des régions critiques de tests d'hypothèses, nous n'aurons besoin que des quantiles situés aux extrémités de la distribution. C'est pourquoi la table fournie à la fin de cette section ne contient pas tous les quantiles. De plus, elle est limitée à $n = 30$ car lorsque n est supérieur à 30, on peut approximer les quantiles de la manière suivante :

Lorsque n est supérieur à 30, on a :

$$c_{n;\alpha} \approx \frac{1}{2}\left[z_\alpha + \sqrt{2n-1}\right]^2,$$

où z_α est le quantile d'ordre $1-\alpha$ d'une variable aléatoire Z suivant une loi $\text{N}(0;1)$.

Voyons tout de suite des exemples de quantiles de variables suivant la loi du χ^2.

Exemple 66 : Soit X une variable aléatoire suivant une loi du χ^2 de 10 degrés de liberté. Déterminons les $10^{\text{ème}}$ et $95^{\text{ème}}$ centiles de X. Ces quantiles correspondent respectivement à $c_{10;0,9}$ et $c_{10;0,05}$. Recherchons-les dans la table située à la fin de la section. Le premier se trouve à l'intersection de la ligne $n = 10$ et de la colonne $\alpha = 0,90$, et le deuxième à l'intersection de la ligne $n = 10$ et de la colonne $\alpha = 0,05$. On obtient donc $c_{10;0,9} = 4,87$ et $c_{10;0,05} = 18,3$.

Soient $Y \sim \chi^2_{70}$ et $T \sim \chi^2_{100}$. Déterminons, là encore, les $10^{\text{ème}}$ et $95^{\text{ème}}$ centiles de ces variables. Les nombres de degrés de liberté des lois suivies par Y et T sont supérieurs à 30. On va donc utiliser l'approximation $c_{n;\alpha} \approx \frac{1}{2}\left[z_\alpha + \sqrt{2n-1}\right]^2$. Pour déterminer $c_{70;0,90}$, on cherche donc $z_{0,90}$. On sait que, dans la loi normale, $z_{0,90} = -z_{0,10}$ et, d'après la table de la loi normale, $z_{0,10} = 1,282$. Donc

$$c_{70;0,90} \approx \frac{1}{2}\left[-1,282 + \sqrt{2\times 70-1}\right]^2 \approx 55,21.$$

On calcule les autres quantiles de la même manière :

$$\begin{aligned}
c_{70;0,05} &\approx \tfrac{1}{2}\left[1,645 + \sqrt{2\times 70-1}\right]^2 \approx 90,25,\\
c_{100;0,90} &\approx \tfrac{1}{2}\left[-1,282 + \sqrt{2\times 100-1}\right]^2 \approx 82,24,\\
c_{100;0,05} &\approx \tfrac{1}{2}\left[1,645 + \sqrt{2\times 100-1}\right]^2 \approx 124,06.
\end{aligned}$$

♦

Connaissant la loi du χ^2, nous pouvons maintenant définir les intervalles de confiance et les tests d'hypothèses pour la variance.

10.2 Intervalles de confiance et tests d'hypothèses sur σ^2

Reprenons la méthode que nous avions utilisée dans les deux sections précédentes pour définir les intervalles de confiance et tests d'hypothèses sur la moyenne : nous étions partis du constat que

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \text{N}(0;1).$$

De là, nous pouvions calculer grâce à la table de la loi normale

$$\text{Prob}\left(-z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha,$$

ou encore

$$\text{Prob}\left(\overline{X} - \frac{\sigma}{\sqrt{n}}\ z_{\alpha/2} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}}\ z_{\alpha/2}\right) = 1-\alpha.$$

Cela nous permettait de déterminer l'intervalle de confiance de niveau de confiance $1-\alpha$:

$$\left[-\frac{\sigma}{\sqrt{n}}\ z_{\alpha/2}; \overline{X} + \frac{\sigma}{\sqrt{n}}\ z_{\alpha/2}\right].$$

De même, pour les tests d'hypothèse, nous avions utilisé le fait que $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \text{N}(0;1)$, pour en déduire les probabilités (d'après la table de la loi normale) que l'hypothèse H_0 soit plus ou moins plausible que H_1.

On voit donc que ce qui importe pour définir les intervalles de confiance et les tests d'hypothèse sur un paramètre, disons y, c'est de connaître une distribution de probabilité, appelons-la $\mathcal{L}$, qui est suivie par une variable aléatoire $f(y,\overline{Y})$ dépendant de y et de l'estimation de y obtenue sur un échantillon (que nous noterons $\overline{Y}$). Grâce à la table de la loi $\mathcal{L}$, on peut donc calculer, pour tout γ, les valeurs de α et de β pour lesquelles on a :

$$\text{Prob}(\alpha \leq f(y,\overline{Y}) \leq \beta) = \gamma.$$

Pour une valeur de $\overline{Y}$ fixée, $f(y,\overline{Y})$ ne dépend que d'un seul paramètre : y. Si l'on note $f_{\overline{Y}}(y) = f(y,\overline{Y})$ et si l'on suppose que $f_{\overline{Y}}$ est inversible, alors :

$$\text{Prob}\left(f_{\overline{Y}}^{-1}(\alpha) \leq y \leq f_{\overline{Y}}^{-1}(\beta)\right) = \gamma.$$

On peut alors aisément déduire de cette équation les intervalles de confiance et les régions critiques des tests d'hypothèses. La fonction $f(\cdot,\cdot)$ et la loi $\mathcal{L}$ pour le paramètre σ^2 nous sont données par le théorème suivant :

Théorème 20 : *Soit $\overline{X}$ la variable «moyenne d'un échantillon aléatoire de taille n prélevé avec remise sur une variable X». Soit S^2 la variance corrigée d'un tel échantillon :*

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

Si $X \sim \text{N}(\mu;\sigma^2)$, alors :

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Grâce à ce théorème, on peut déduire que :

$$\begin{aligned} 1-\alpha &= \text{Prob}\left(c_{n-1;1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq c_{n-1;\alpha/2}\right), \\ &= \text{Prob}\left(\frac{(n-1)S^2}{c_{n-1;\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{c_{n-1;1-\alpha/2}}\right). \end{aligned}$$

D'où

Soit X une variable aléatoire suivant une loi $\text{N}(\mu;\sigma^2)$. Soit s^2 la variance corrigée observée sur un échantillon de taille n. Alors un intervalle de confiance pour σ^2 de niveau de confiance $1-\alpha$ est donné par :

$$\left[\frac{(n-1)s^2}{c_{n-1;\alpha/2}}; \frac{(n-1)s^2}{c_{n-1;1-\alpha/2}}\right].$$

Exemple 53 (suite) : Reprenons l'exemple de la page 91 : un ingénieur système a sous sa responsabilité 49 stations SUN. Il a déterminé que la fréquence des pannes importantes de ses machines est en moyenne de 54 mois, avec un écart-type de 8 mois. Le directeur lui demande quelle est la fréquence des pannes des Suns (dans l'absolu). L'ingénieur système considère ses machines comme un échantillon de 49 individus. Il suppose que X, la fréquence des pannes de la population totale (l'ensemble des stations de travail dans le monde entier), suit une loi normale $\mathrm{N}(\mu, \sigma^2)$. Grâce à son échantillon, il peut déterminer que l'intervalle de confiance pour σ^2 de niveau $1-\alpha$ est :

$$\left[\frac{(49-1)\times 8^2}{c_{48;\alpha/2}}; \frac{(49-1)\times 8^2}{c_{48;1-\alpha/2}}\right].$$

Puisque $n > 30$, il peut approximer les valeurs de $c_{n;\alpha/2}$ et de $c_{n;1-\alpha/2}$. L'intervalle de confiance devient alors :

$$\left[\frac{(49-1)\times 8^2}{\frac{1}{2}\left[z_{\alpha/2}+\sqrt{2\times 49-1}\right]^2}; \frac{(49-1)\times 8^2}{\frac{1}{2}\left[z_{1-\alpha/2}+\sqrt{2\times 49-1}\right]^2}\right].$$

Par conséquent, il peut répondre avec un risque d'erreur de 5% que

$$\sigma^2 \in \left[\frac{3072}{\frac{1}{2}\left[1,96+\sqrt{97}\right]^2}; \frac{3072}{\frac{1}{2}\left[-1,96+\sqrt{97}\right]^2}\right] = [44,06; 98,72],$$

ou encore que $\sigma \in [6,64; 9,94]$. ♦

Exemple 54 (suite) : Reprenons l'exemple du distributeur de boissons de la page 91 : celui-ci est réglé de telle sorte que la quantité X de liquide qu'il verse dans un gobelet est distribuée selon une loi normale. Afin d'affiner les réglages, un technicien prélève un échantillon de 10 boissons. Il obtient sur cet échantillon une moyenne de 20 cl avec un écart-type de 1,65 cl. Il veut estimer avec un risque d'erreur de 5% la variance de la quantité de boisson versée par le distributeur. L'intervalle de confiance est :

$$\left[\frac{9\times 1,65^2}{c_{9;0,025}}; \frac{9\times 1,65^2}{c_{9;0,975}}\right].$$

D'après la table de la loi du χ^2, on obtient l'intervalle :

$$\left[\frac{24,5025}{19}; \frac{24,5025}{2,7}\right] = [1,29; 9,07].$$

♦

De manière analogue, on peut construire des tests d'hypothèses lorsque la variable est distribuée selon une loi normale. En effet, pour tester $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$, on peut utiliser une région critique de la forme «rejeter H_0 si $s^2 > k$» (où s^2 représente la variance corrigée observée sur l'échantillon). Pour déterminer la valeur de k, on utilise le théorème 20 :

$$\text{Sous l'hypothèse } H_0,\ \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}.$$

Par conséquent, le niveau de test α est égal à :

$$\begin{aligned}
\alpha &= \text{Prob(rejeter } H_0 | H_0 \text{ est vraie)} \\
&= \text{Prob}(S^2 > k | \sigma^2 = \sigma_0^2) \\
&= \text{Prob}\left(\frac{(n-1)S^2}{\sigma_0^2} > \frac{(n-1)k}{\sigma_0^2}\right) \\
&= \text{Prob}\left(\frac{(n-1)S^2}{\sigma_0^2} > c_{n-1;\alpha}\right).
\end{aligned}$$

On peut généraliser cette méthode de calcul à tous les tests d'hypothèse possibles et imaginables. Cela nous conduit au tableau suivant :

Contre-hypothèse	forme de la région critique
$H_1 : \sigma^2 < \sigma_0^2$	$\dfrac{(n-1)s^2}{\sigma_0^2} < c_{n-1;1-\alpha}$
$H_1 : \sigma^2 > \sigma_0^2$	$\dfrac{(n-1)s^2}{\sigma_0^2} > c_{n-1;\alpha}$
$H_1 : \sigma^2 \neq \sigma_0^2$	$\dfrac{(n-1)s^2}{\sigma_0^2} < c_{n-1;1-\alpha/2}$ ou $\dfrac{(n-1)s^2}{\sigma_0^2} < c_{n-1;\alpha/2}$

TAB. 14 – Régions critiques des tests d'hypothèses de niveau α sur σ^2, lorsqu'un échantillon de taille n est prélevé sur une variable $X \sim \mathrm{N}(\mu;\sigma^2)$, et lorsque l'hypothèse nulle est de la forme $H_0 : \sigma^2 = \sigma_0^2$.

Exemple 63 (suite) : Reprenons l'exemple de la page 104 : Le montant de l'assurance pour une automobile utilisée à des fins non commerciales est fonction de la distance moyenne parcourue annuellement par les automobilistes. Des études datant de quelques années ont montré que celle-ci était de 18000 km avec un écart-type de 5000 km. On suppose que la variable «distance parcourue par les automobilistes» suit une loi normale. Le statisticien de la page 104 se demande si l'écart-type doit être revu à la baisse. Pour cela, il a prélevé un échantillon de 400 individus et il a obtenu une variance corrigée égale à (4700 km)^2. Les hypothèses qu'il formule sont donc les suivantes :

$$H_0 : \sigma^2 = 5000^2 \qquad \text{versus} \qquad H_1 : \sigma^2 < 5000^2,$$

et il doit utiliser la première ligne du tableau ci-dessus. Cela lui fournit la région critique :

$$\frac{(400-1)\times 4700^2}{5000^2} < c_{400-1;1-\alpha}.$$

Puisque $n > 30$, on utilise l'approximation de c :

$$\frac{399\times 4700^2}{5000^2} < \frac{1}{2}\left[z_{1-\alpha} + \sqrt{2\times 400 - 1}\right]^2.$$

Si notre statisticien veut un niveau de test égal à 95%, il obtient :

$$\frac{399\times 4700^2}{5000^2} = 362,5564 \qquad \text{et} \qquad \frac{1}{2}\left[z_{1-\alpha} + \sqrt{2\times 400 - 1}\right]^2 = \frac{1}{2}\left[1,645 + \sqrt{799}\right]^2 = 447,351.$$

Le statisticien peut en conclure (avec un taux d'erreur de 5%) que la variance a effectivement baissé. ♦

Précisons que les résultats ci-dessus **ne sont absolument pas valables** si la variable X n'est pas distribuée selon une loi normale. Dans ce cas, la connaissance de la loi de probabilité suivie par X permet de construire intervalles de confiance et tests d'hypothèses, mais cela s'avère beaucoup plus compliqué que ce que nous avons vu.

Comme annoncé dans l'introduction de cette section, la loi du χ^2 a de nombreuses applications en dehors de la construction d'intervalles de confiance ou de tests d'hypothèses sur σ^2. Nous allons consacrer la fin de cette section à en étudier deux particulièrement intéressantes d'un point de vue statistique.

10.3 Test d'ajustement du χ^2

Très souvent dans les sections précédentes, nous avions supposé que telle ou telle variable aléatoire suivait une loi (très souvent normale, mais aussi binômiale). Nous allons voir maintenant comment utiliser ce que l'on appelle le $\boxed{\text{test d'ajustement du } \chi^2}$ pour vérifier une telle hypothèse. En réalité, ce test permet de confronter les deux hypothèses suivantes :

H_0 : La variable étudiée obéit à la distribution théorique spécifiée

versus

H_1 : La variable étudiée n'obéit pas à la distribution théorique spécifiée.

L'idée de ce test consiste à regrouper les n données de l'échantillon en I classes et à comparer à l'aide d'une statistique d'ajustement les effectifs de classe observés dans l'échantillon avec les effectifs de classe que l'on devrait obtenir si H_0 était vraie (on parle alors d' effectif espéré). Notons n_i, $i = 1,\ldots,I$, les effectifs observés , et ν_i, $i = 1, \ldots, I$, les effectifs espérés.

La statistique d'ajustement utilisée pour vérifier si la variable obéit à la distribution théorique, que l'on notera A, est :

$$A = \sum_{i=1}^{I} \frac{(n_i - \nu_i)^2}{\nu_i}.$$

On peut démontrer que si la taille n de l'échantillon est grande ($n \geq 30$) et si les effectifs espérés ν_i sont tous supérieurs ou égaux à 5, alors la statistique d'ajustement A suit sous H_0 une loi du χ^2_d.
Le nombre de degrés de liberté d dépend du nombre de classes I, et de la façon dont la distribution théorique est spécifiée en H_0 : si la spécification permet de déduire directement la valeur des ν_i, alors $d = I - 1$; si, par contre, il faut procéder à l'estimation de r paramètres pour calculer les ν_i, alors $d = I - 1 - r$.

Lorsque H_0 est vraie, il devrait y avoir peu de différence entre les n_i et les ν_i. Donc la statistique d'ajustement devrait être petite. On rejette donc H_0 lorsque A prend des valeurs trop grandes, c'est-à-dire lorsque A est plus grand qu'une certaine quantité c. Pour connaître la valeur de c, il suffit d'examiner ce que signifie le niveau du test :

$$\alpha = \text{Prob}(\text{rejeter } H_0 | H_0 \text{ est vraie}) = \text{Prob}(A > c | H_0 \text{ est vraie}) = \text{Prob}(A > c | A \sim \chi^2_d).$$

Par conséquent, $c = c_{d;\alpha}$.

Pour vérifier si H_0 est vraie ou si H_1 est plus plausible, il faut donc commencer par calculer la statistique d'ajustement A. Pour cela, on doit calculer les ν_i. À cette fin, on va être amené à estimer r paramètres. Lorsque A est calculé, on calcule $d = I - 1 - r$ et l'on note la valeur de $c = c_{d;\alpha}$. Si $A \leq c$, on déduit que H_0 est vraie, sinon c'est H_1 qui est vraie.

Les deux exemples suivants vont nous permettre d'illustrer cette démarche.

Exemple 67 : Dans un supermarché, on maintient 8 caisses de plus de 10 articles en opération durant les nocturnes du jeudi. Normalement, la clientèle devrait se répartir uniformément entre les caisses. Afin de vérifier cela, on a recensé le nombre de clients passés à chacune des caisses un jeudi soir. Les résultats suivants ont été observés :

Numéro de la caisse	Nombre de clients
1	72
2	70
3	71
4	52
5	45
6	59
7	67
8	48
Total	484

Les hypothèses que l'on veut confronter sont les suivantes :

H_0 : la clientèle se répartit uniformément entre les 8 caisses.

versus

H_1 : la clientèle ne se répartit pas uniformément entre les 8 caisses.

Si H_0 est vraie, les clients se répartissent uniformément entre toutes les caisses. On devrait donc avoir dans chacune des caisses un effectif de $1/8^{\text{ème}}$ de la population, soit $\nu_i = 484/8 = 60,5$. Par conséquent,

$$A = \frac{(72-60,5)^2}{60,5} + \frac{(70-60,5)^2}{60,5} + \frac{(71-60,5)^2}{60,5} + \frac{(52-60,5)^2}{60,5}$$
$$+ \frac{(45-60,5)^2}{60,5} + \frac{(59-60,5)^2}{60,5} + \frac{(67-60,5)^2}{60,5} + \frac{(48-60,5)^2}{60,5} = 15,56$$

Notons que chacun des ν_i est supérieur à 5 et que la taille de l'échantillon (484) est supérieur à 30. Par conséquent, on peut affirmer que A suit approximativement une loi du χ^2_d. Pour calculer les ν_i, nous n'avons pas eu besoin d'estimer de paramètres. Donc le nombre de degrés de liberté de la loi est égal à $I-1 = 8-1 = 7$.

Si l'on fixe le niveau du test à $\alpha = 0,05$, la région critique du test correspond à rejeter H_0 si A prend une valeur supérieure à $c_{7;0,05} = 14,1$. Or $A = 15,56$. On peut donc rejeter H_0, c'est-à-dire conclure que la clientèle n'est pas répartie uniformément entre les 8 caisses. ♦

Exemple 68 : Les tailles de 100 clientes sélectionnées au hasard dans le supermarché ont été regroupées en 12 classes :

classe de taille (en cm)	nombre de clientes
$[145,5;148,5[$	1
$[148,5;151,5[$	3
$[151,5;154,5[$	4
$[154,5;157,5[$	12
$[157,5;160,5[$	16
$[160,5;163,5[$	22
$[163,5;166,5[$	19
$[166,5;169,5[$	11
$[169,5;172,5[$	6
$[172,5;175,5[$	2
$[175,5;178,5[$	2
$[178,5;181,5[$	2
Total	100

Considérant ces 100 observations comme un échantillon représentatif de la clientèle féminine du supermarché, on peut confronter les hypothèses suivantes sur la variable X définie comme la taille X en cm d'une cliente :

$$H_0 : X \text{ suit une loi normale}$$
versus
$$H_1 : X \text{ ne suit pas une loi normale.}$$

Tout comme dans l'exemple précédent, les effectifs observés sont donnés dans l'énoncé du problème. Par contre, le calcul des effectifs espérés ν_i est plus délicat. Pour le réaliser, il nous faut calculer les proportions p_i de clientes qui, théoriquement si H_0 est vraie, devraient correspondre à chacune des classes. ν_i sera alors égal à $p_i\times$ la taille de l'échantillon, soit $100 \times p_i$.

La première difficulté vient du fait que la loi théorique spécifiée sous H_0 correspond à la distribution d'une variable pouvant prendre n'importe quelle valeur réelle. Donc, sous H_0, il pourrait y avoir des clientes plus petites que 145,5 cm, ou bien plus grandes que 181,5 cm. Afin d'être certain que la somme des p_i sur l'intervalle $[145,5;181,5[$ est bien égale à 1, on doit considérer que la première classe représente la classe «moins de 148,5 cm» et la dernière «plus de 178,5 cm».

La deuxième difficulté vient de ce que l'énoncé de H_0 ne permet pas de calculer immédiatement les p_i. En effet, pour les calculer, il faudrait d'abord connaître les paramètres μ et σ^2 de la loi normale spécifiée sous H_0. Il nous faut donc estimer ces deux paramètres. D'après la section 7, on peut estimer μ par la moyenne de l'échantillon observé, et σ^2 par sa variance corrigée. Ici, un simple calcul nous donne $\mu = 162,66$ et $\sigma^2 = 40,71$.

On peut maintenant s'adonner au calcul des p_i : $X \sim \mathrm{N}(162,66; 40,71)$. Donc $\frac{X-162,66}{\sqrt{40,71}} \sim \mathrm{N}(0;1)$. D'où

$$\begin{aligned}
p_1 &= \mathrm{Prob}(X < 148,5) = \mathrm{Prob}\left(\frac{X-162,66}{\sqrt{40,71}} < \frac{148,5-162,66}{\sqrt{40,71}}\right) \\
&= \mathrm{Prob}\left(\frac{X-162,66}{\sqrt{40,71}} < -2,22\right) = 0,0132. \\
p_2 &= \mathrm{Prob}(148,5 \leq X < 151,5) = \mathrm{Prob}\left(\frac{148,5-162,66}{\sqrt{40,71}} \leq \frac{X-162,66}{\sqrt{40,71}} < \frac{151,5-162,66}{\sqrt{40,71}}\right) \\
&= \mathrm{Prob}\left(-2,22 \leq \frac{X-162,66}{\sqrt{40,71}} < 1,75\right) = 0,0269.
\end{aligned}$$

De manière analogue, on calcule les autres p_i. On obtient alors le résultat suivant :

classe de taille (en cm)	effectif observé n_i	proportion théorique p_i	effectif espéré ν_i
moins de $148,5$	1	0,0132	1,32
$[148,5; 151,5[$	3	0,0269	2,69
$[151,5; 154,5[$	4	0,0602	6,02
$[154,5; 157,5[$	12	0,1087	10,87
$[157,5; 160,5[$	16	0,1579	15,79
$[160,5; 163,5[$	22	0,1848	18,48
$[163,5; 166,5[$	19	0,1740	17,40
$[166,5; 169,5[$	11	0,1320	13,20
$[169,5; 172,5[$	6	0,0805	8,05
$[172,5; 175,5[$	2	0,0396	3,96
$[175,5; 178,5[$	2	0,0156	1,56
plus de $178,5$	2	0,0066	0,66
Total	100	1	100

On ne peut pas encore calculer A car certaines classes ont moins de 5 individus. Il va donc falloir effectuer des regroupements de classes. Nous allons donc regrouper les trois premières classes ($\nu_1+\nu_2+\nu_3 = 10,03 > 5$) et les trois dernières ($\nu_{10}+\nu_{11}+\nu_{12} = 6,18 > 5$). On obtient donc le tableau suivant :

indice de classe	classe de taille (en cm)	effectif observé n_i	proportion théorique p_i	effectif espéré ν_i
1	moins de $154,5$	8	0,1003	10,03
2	$[154,5; 157,5[$	12	0,1087	10,87
3	$[157,5; 160,5[$	16	0,1579	15,79
4	$[160,5; 163,5[$	22	0,1848	18,48
5	$[163,5; 166,5[$	19	0,1740	17,40
6	$[166,5; 169,5[$	11	0,1320	13,20
7	$[169,5; 172,5[$	6	0,0805	8,05
8	plus de $172,5$	6	0,0618	6,18
	Total	100	1	100

Il est maintenant possible de calculer $A = \sum_{i=1}^{8} \frac{(n_i-\nu_i)^2}{\nu_i}$. On obtient $A = 2,25$. Avec ces classes, tous les effectifs espérés sont supérieurs à 5. L'échantillon a plus de 30 individus. Donc A suit une loi du χ^2. On a dû estimer deux paramètres μ et σ^2. Donc le nombre de degrés de liberté est $8-1-2=5$. Par conséquent, $A \sim \chi^2_5$. En fixant le niveau du test α à 0,05, la région critique du test correspond à rejeter H_0 si A prend une valeur supérieure à $c_{5;0,05} = 11,1$. Or, ici, $A = 2,25$. Donc on ne peut rejeter H_0. On peut donc en conclure que la variable «taille des clientes» suit une loi normale. ♦

10.4 Test d'indépendance

La dernière application de la loi du χ^2 que nous allons voir permet de tester si deux caractères provenant d'une même population sont indépendants, c'est-à-dire si la valeur que prend l'un influence la valeur que prend l'autre.

Pour procéder à un tel test, on prélève un échantillon de n individus dans la population et on note pour chacun de ces individus la valeur des deux caractères. Les n résultats sont habituellement présentés dans un tableau rectangulaire dont le nombre de lignes correspond au nombre de classes utilisées pour regrouper les valeurs du premier caractère, et dont le nombre de colonnes correspond au nombre de classes utilisées pour regrouper le deuxième caractère. D'une manière générale, si on note X le premier caractère et Y le second, et si l'on note $A_1, A_2, \ldots, A_I$ les classes du premier caractère et $B_1, B_2, \ldots, B_J$ les classes du second, les résultats de l'échantillonnage sont représentés en remplissant le tableau ci-dessous, que l'on appelle tableau de contingence :

$X\backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}

Les n_{ij} représentent le nombre d'individus pour lesquels X appartient à la classe A_i et Y appartient à la classe B_j. L'idée du test d'indépendance est tout simplement de construire un test opposant les hypothèses H_0 : X et Y sont indépendants, et H_1 : X et Y ne sont pas indépendants. Pour cela, on va comparer les effectifs n_{ij} observés avec les effectifs espérés ν_{ij} que l'on obtiendrait si H_0 était vraie. On calcule alors une statistique d'ajustement :

$$A = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-\nu_{ij})^2}{\nu_{ij}} = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{\nu_{ij}}\right) - n.$$

On peut démontrer que si la taille n de l'échantillon est grande et si tous les effectifs espérés sous H_0, ν_{ij}, sont supérieurs ou égaux à 5, alors la statistique d'ajustement suit approximativement une loi χ^2_d, où le nombre de degrés de liberté d est égal à $(I-1)\times(J-1)$. La région critique du test de niveau α est alors : «on rejette H_0 si A prend une valeur supérieure à $c_{d;\alpha}$».

La difficulté dans la construction d'un test d'indépendance réside dans le calcul des effectifs ν_{ij} espérés théoriquement si H_0 est vraie. Celui-ci nécessite la détermination des proportions théoriques p_{ij} d'individus pour lesquels, si H_0 était vraie, X appartiendrait à la classe A_i et Y à la classe B_j. En termes de probabilités,

$$p_{ij} = \text{Prob}(X \in A_i, Y \in B_j).$$

Or, si H_0 est vraie, d'après la page 33, on a :

$$\text{Prob}(X \in A_i, Y \in B_j) = \text{Prob}(X \in A_i) \times \text{Prob}(Y \in B_j).$$

Or il est relativement aisé de déterminer $\text{Prob}(X \in A_i)$ et $\text{Prob}(Y \in B_j)$. En effet, $\text{Prob}(X \in A_i)$ s'estime par la proportion des individus de l'échantillon pour lesquels $X \in A_i$. Autrement dit, cette probabilité est estimée en divisant la somme des effectifs observés sur la $i^{\text{ème}}$ ligne du tableau de contingence, que l'on note $n_{i\cdot}$, par la taille n de l'échantillon. De même, $\text{Prob}(Y \in B_j)$ est estimée par la division de la somme des effectifs observés sur la $j^{\text{ème}}$ colonne du tableau de contingence, que l'on note $n_{\cdot j}$, par la taille n de l'échantillon :

$$\text{Prob}(X \in A_i) = \frac{n_{i\cdot}}{n} = \frac{\sum_{j=1}^{J} n_{ij}}{n} \qquad \text{et} \qquad \text{Prob}(Y \in B_j) = \frac{n_{\cdot j}}{n} = \frac{\sum_{i=1}^{I} n_{ij}}{n}$$

On peut alors compléter le tableau de contingence avec les $n_{i\cdot}$ et les $n_{\cdot j}$, pour obtenir :

$X\backslash Y$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	*total*
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_I	n_{I1}	n_{I2}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\cdot}$
total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot J}$	n

Il est alors assez simple de calculer les ν_{ij} :

$$\nu_{ij} = np_{ij} = n \times \frac{n_{i\cdot}}{n} \times \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} \times n_{\cdot j}}{n}.$$

Exemple 69 : Un échantillon de 200 contribuables est prélevé afin de vérifier si le revenu brut annuel d'un individu est un caractère dépendant du niveau de scolarité de l'individu. Les résultats de l'échantillon sont résumés dans le tableau 15. Celui-ci nous donne les valeurs des effectifs observés n_{ij} ainsi que les

niveau de scolarité (en années) / classe de revenu (en F)	[0,7[	[7,12[	[12,14[	14 et plus	total
[0;75000[	17	14	9	5	45
[75000;12000[	12	37	11	5	65
[12000;20000[	7	20	20	8	55
200000 et plus	4	9	10	12	35
total	40	80	50	30	n=200

TAB. 15 – Tableau recensant les valeurs des n_{ij}.

sommes en ligne $n_{i\cdot}$ et en colonne $n_{\cdot j}$. Les effectifs espérés théoriquement si H_0 est vraie s'obtiennent en multipliant entre eux les $n_{i\cdot}$ et les $n_{\cdot j}$, et en divisant le tout par n. Par exemple,

$$\nu_{12} = \frac{n_{1\cdot} \times n_{\cdot 2}}{n} = \frac{45 \times 80}{200} = 18,$$

$$\nu_{34} = \frac{n_{3\cdot} \times n_{\cdot 4}}{n} = \frac{55 \times 30}{200} = 8,25.$$

Après calcul de tous les ν_{ij}, on obtient le tableau 16.

niveau de scolarité (en années) / classe de revenu (en F)	[0,7[	[7,12[	[12,14[	14 et plus	total
[0;75000[	9,00	18,00	11,25	6,75	45
[75000;12000[	13,00	26,00	16,25	9,75	65
[12000;20000[	11,00	22,00	13,75	8,25	55
200000 et plus	7,00	14,00	8,75	5,25	35
total	40	80	50	30	n=200

TAB. 16: Tableau recensant les valeurs des ν_{ij}.

La taille de l'échantillon est grande (200). De plus, on peut remarquer dans le tableau 16 que tous les ν_{ij} sont supérieurs ou égaux à 5. Par conséquent, sous H_0, la statistique d'ajustement suivante

$$A = \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{n_{ij}^2}{\nu_{ij}} \right) - n$$

suit une loi χ^2_d, où $d = (I-1) \times (J-1) = 3 \times 3 = 9$ (puisqu'il y a $I = 4$ classes de valeurs pour le premier caractère, et $J = 4$ classes pour le deuxième caractère). En utilisant un niveau $\alpha = 0,05$, on rejette H_0 si la valeur de A est supérieure à $c_{9;0,05} = 16,9$. Si l'on calcule A, on trouve la valeur : 34,06. Par conséquent, on doit rejeter H_0 : on conclut qu'il y a une dépendance entre le revenu et le niveau de scolarité dans la population étudiée (quelque part, c'est tout de même rassurant!). ♦

10.5 Exercices

Exercice 1 Des tests ont été effectués sur un échantillon de 25 voitures d'un même modèle afin de déterminer la consommation d'essence de ce type de véhicule. Les résultats, en nombre de litres pour 100 km, sont consignés dans le tableau suivant :

8,1	8,2	7,9	8,4	8,4
8,2	8,3	8,0	8,0	8,2
8,6	8,4	7,9	8,3	8,0
8,3	8,1	8,5	8,2	8,2
8,1	8,2	8,5	8,3	8,4

Donnez des estimations ponctuelles et par intervalle de confiance de la moyenne μ et de la variance σ^2 de la consommation d'essence. On utilisera un niveau $\alpha = 0,05$ et on supposera que la variable consommation d'essence suit une loi normale.

Exercice 2 Il y a cinq ans, les statisticiens du laboratoire d'informatique de Paris 6 (LIP6) ont établi que les scores des chercheurs à Tetris suivaient une loi $N(134500; 125000)$. Récemment, ils ont choisi 30 personnes au hasard et leur ont demandé quels scores ils avaient actuellement au Tetris. Ils ont ainsi obtenu une moyenne de 142300 et une variance corrigée de 112700. Au niveau 5%, testez s'il y a eu une modification des paramètres μ et σ^2.

Exercice 3 Tirer des nombres aléatoires sur ordinateur est, contrairement à ce que l'on pourrait croire, une tâche assez complexe à réaliser. Un nouvel algorithme a été développé à cet effet et nous nous demandons si les nombres qu'il fournit sont bien aléatoires. Il nous faudra une batterie de tests pour pouvoir décider si l'algorithme est bon. Parmi ceux-ci, il en existe un relativement simple : supposons que l'algorithme ne nous fournit que des nombres entre 0 et 9. Si ces nombres sont vraiment tirés aléatoirement, alors en en tirant suffisamment, on devrait se rapprocher d'une loi uniforme : probabilité $1/10^{\text{ème}}$ de choisir n'importe quel nombre entre 0 et 9. On a donc fait tourné l'algorithme 260 fois et on a obtenu les résultats suivants :

Valeur	effectif
0	28
1	18
2	23
3	33
4	25
5	33
6	18
7	19
8	25
9	28

À l'aide d'un test d'ajustement, testez si les nombres 0 à 9 sont distribués uniformément. Utilisez un niveau de test $\alpha = 0,05$.

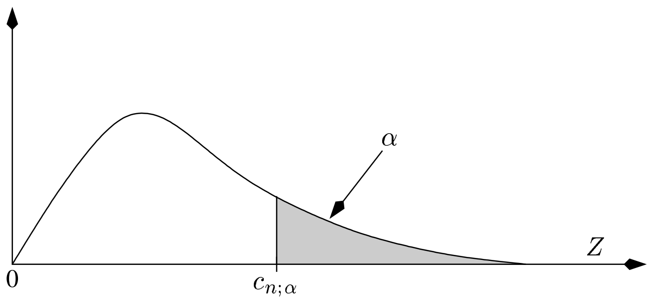
Exercice 4 Dans un échantillon aléatoire de 240 personnes, on a recueilli l'information suivante sur l'âge et sur le type de sport le plus fréquemment pratiqué à Jussieu :

âge / activité sportive	moins de 20 ans	[20; 25[ans	[25; 30[ans	plus de 30 ans
jogging	15	20	15	30
natation	15	10	20	25
ping pong	20	10	30	30

Au niveau $\alpha = 0,05$, peut-on conclure que le type de sport le plus fréquemment pratiqué à Jussieu est dépendant de l'âge du capitaine ?

10.6 Table de la loi du χ^2

valeurs dans le tableau
ci-dessous : les $c_{n;\alpha}$
tels que $\text{Prob}(Z > c_{n;\alpha}) = \alpha$

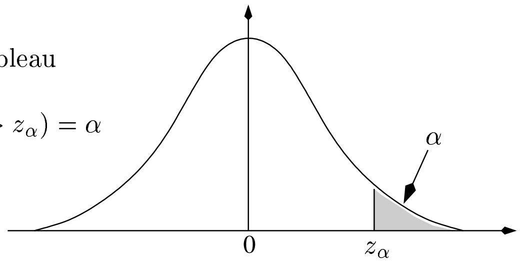


$n \setminus \alpha$	0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,0000393	0,000157	0,000982	0,00393	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	40,3	43,8	47,0	50,9	53,7

11 Liste des tables des lois usuelles

11.1 Table de la loi normale centrée réduite

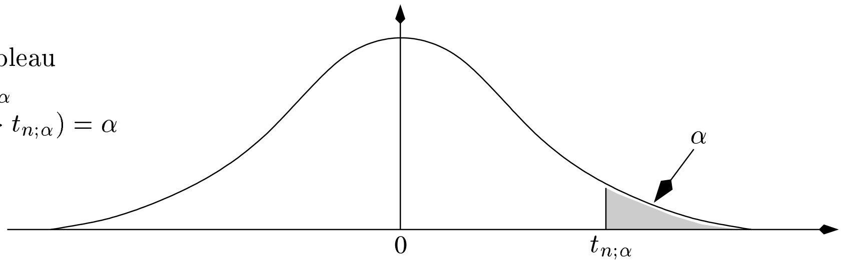
valeurs dans le tableau
ci-dessous : les α
tels que $\mathrm{Prob}(Z > z_\alpha) = \alpha$



z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0859	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0466	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0063
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0047
2,6	0,0047	0,0045	0,0044	0,0043	0,0042	0,0041	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0014	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
3,6	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,7	0,0001	0,0001	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

11.2 Table de la loi de Student

valeurs dans le tableau
ci-dessous : les $t_{n;\alpha}$
tels que $\text{Prob}(Z > t_{n;\alpha}) = \alpha$

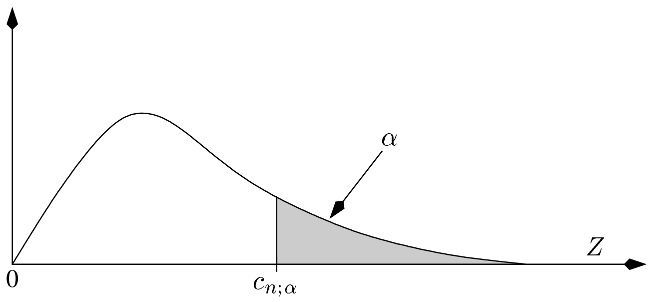


$n \setminus \alpha$	0,10	0,05	0,025	0,01	0,005	0,001
1	3,078	6,314	12,706	31,821	63,657	318,309
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,106	4,025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385
31	1,309	1,696	2,040	2,453	2,744	3,375
32	1,309	1,694	2,037	2,449	2,738	3,365
33	1,308	1,692	2,035	2,445	2,733	3,356
34	1,307	1,691	2,032	2,441	2,728	3,348
35	1,306	1,690	2,030	2,438	2,724	3,340
36	1,306	1,688	2,028	2,434	2,719	3,333
37	1,305	1,687	2,026	2,431	2,715	3,326

$n \setminus \alpha$	0,10	0,05	0,025	0,01	0,005	0,001
38	1,304	1,686	2,024	2,429	2,712	3,319
39	1,304	1,685	2,023	2,426	2,708	3,313
40	1,303	1,684	2,021	2,423	2,704	3,307
41	1,303	1,683	2,020	2,421	2,701	3,301
42	1,302	1,682	2,018	2,418	2,698	3,296
43	1,302	1,681	2,017	2,416	2,695	3,291
44	1,301	1,680	2,015	2,414	2,692	3,286
45	1,301	1,679	2,014	2,412	2,690	3,281
46	1,300	1,679	2,013	2,410	2,687	3,277
47	1,300	1,678	2,012	2,408	2,685	3,273
48	1,299	1,677	2,011	2,407	2,682	3,269
49	1,299	1,677	2,010	2,405	2,680	3,265
50	1,299	1,676	2,009	2,403	2,678	3,261
51	1,298	1,675	2,008	2,402	2,676	3,258
52	1,298	1,675	2,007	2,400	2,674	3,255
53	1,298	1,674	2,006	2,399	2,672	3,251
54	1,297	1,674	2,005	2,397	2,670	3,248
55	1,297	1,673	2,004	2,396	2,668	3,245
56	1,297	1,673	2,003	2,395	2,667	3,242
57	1,297	1,672	2,002	2,394	2,665	3,239
58	1,296	1,672	2,002	2,392	2,663	3,237
59	1,296	1,671	2,001	2,391	2,662	3,234
60	1,296	1,671	2,000	2,390	2,660	3,232
61	1,296	1,670	2,000	2,389	2,659	3,229
62	1,295	1,670	1,999	2,388	2,657	3,227
63	1,295	1,669	1,998	2,387	2,656	3,225
64	1,295	1,669	1,998	2,386	2,655	3,223
65	1,295	1,669	1,997	2,385	2,654	3,220
66	1,295	1,668	1,997	2,384	2,652	3,218
67	1,294	1,668	1,996	2,383	2,651	3,216
68	1,294	1,668	1,995	2,382	2,650	3,214
69	1,294	1,667	1,995	2,382	2,649	3,213
70	1,294	1,667	1,994	2,381	2,648	3,211
71	1,294	1,667	1,994	2,380	2,647	3,209
72	1,293	1,666	1,993	2,379	2,646	3,207
73	1,293	1,666	1,993	2,379	2,645	3,206
74	1,293	1,666	1,993	2,378	2,644	3,204
75	1,293	1,665	1,992	2,377	2,643	3,202
∞	1,282	1,645	1,960	2,326	2,576	3,090

11.3 Table de la loi du χ^2

valeurs dans le tableau
ci-dessous : les $c_{n;\alpha}$
tels que $\text{Prob}(Z > c_{n;\alpha}) = \alpha$



$n \setminus \alpha$	0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,0000393	0,000157	0,000982	0,00393	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	40,3	43,8	47,0	50,9	53,7

Index