PROBABILITÉS ET STATISTIQUE EN S5 IFIPS

J.-P. Lenoir

11 septembre 2007

Chapitre 1

Probabilités

1.1 Probabilité sur un ensemble fini

1.1.1 Evènement aléatoire

Historiquement, la notion de probabilité s'est dégagée à partir d'exemples simples empruntés aux jeux de hasard (le mot hasard vient de l'arabe *az-zahr* : le dé).

Nous allons introduire cette notion en l'associant à un exemple : le jeu de dé.

DÉFINITIONS	EXEMPLE
Une expérience aléatoire est une expérience dont on ne peut prévoir le résultat.	L'expérience est le jet d'un dé cubique ordinaire. Le résultat de l'expérience est le nombre indiqué sur la face supérieure du dé.
On peut alors lui associer alors un univers appelé aussi ensemble fondamental de l'expérience qui est l'ensemble de tous les résultats possibles de l'expérience aléatoire. On le note Ω.	$\Omega = \{1, 2, 3, 4, 5, 6\}$.
Un événement aléatoire est un sous-ensemble de Ω.	L'événement *obtenir un nombre pair* est le sous-ensemble $A = \{2, 4, 5\}$ de Ω.
On dit que l'événement A est réalisé si le résultat de l'expérience appartient à A.	Si la face supérieure du dé indique 5, A n'est pas réalisé. Si elle indique 4, A est réalisé.
Si un événement ne contient qu'un seul élément, on dit que c'est un événement élémentaire.	$B = \{1\}$ est un des 5 événements élémentaires de Ω .

1.1.2 De la fréquence à la probabilité de réalisation d'un évènement aléatoire

La fréquence théorique d'un événement est la limite de la fréquence de réalisation de cet événement lorsque le nombre de répétitions d'une même expérience tend vers l'infini (c'est ce qu'exprime une des lois de la théorie des probabilités appelée la loi faible des grands nombres).

Cela signifie que si l'on veut connaître la fréquence théorique d'apparition du nombre 6 dans notre jet de dé, il suffit de le lancer un grand nombre de fois, 10000 par exemple. La fréquence théorique cherchée, que l'on appellera probabilité de réalisation de l'événement élémentaire $\{6\}$, sera très voisine de la fréquence expérimentale d'apparition du nombre 6 au cours de nos 10000 lancers. Elle sera encore plus voisine de la fréquence expérimentale obtenue lors de 100000 lancers. Le grand nombre de répétitions de l'expérience aléatoire efface la notion de "chance".

La notion de fréquence théorique ou de probabilité va permettre d'indiquer si, lors d'une expérience aléatoire, un événement donné est plus ou moins susceptible d'être réalisé.

Les probabilités peuvent être classées suivant trois critères :
- Une probabilité à priori est une probabilité déterminée à l'avance, sans effectuer aucune expérience.
  Exemple. On peut à priori accorder une probabilité de 0.5 à événement qui consiste à obtenir le côté face d'une pièce de monnaie non truquée.
- La probabilité empirique d'un événement est déterminée à l'aide de l'observation et de l'expérimentation. C'est la valeur limite de la fréquence de réalisation de événement lorsque l'expérience est réalisée un très grand nombre de fois.
  Exemple. Si lorsqu'on a lancé la pièce de monnaie 10000 fois on constate que la fréquence du côté face se stabilise autour de 0.65, il faut envisager de réviser notre probabilité à priori et conclure que la pièce est truquée. Ce type de probabilité joue un rôle important pour les prévisions d'articles en stock chez un détaillant, pour le calcul des primes des compagnies d'assurances, etc...
- La probabilité subjective est le dernier type de probabilité. Elle intervient lorsqu'il est impossible d'établir une probabilité à priori ou une probabilité empirique.
  Exemple. Le directeur d'une entreprise peut en se fiant à son expérience affirmer qu'il y a une probabilité de 0.6 que ses employés déclenchent une grève.

Nous nous intéresserons principalement aux deux premiers types de probabilité.

Vu qu'une probabilité peut être considérée comme une fréquence idéale, on lui connaît d'avance certaines propriétés.
- Une probabilité est une quantité sans dimension.
- Elle est toujours comprise entre 0 et 1.
- L'univers Ω a la probabilité maximum d'être réalisé, car c'est l'événement certain. Sa probabilité de réalisation est donc égale à 1.
- Si A et B sont *incompatibles* (ensembles disjoints), la fréquence de réalisation de l'événement A ou B est la somme de la fréquence de réalisation de A et de la fréquence de réalisation de B.

1.1.3 Propriétés des probabilités d'un évènement aléatoire

Définition 1 *Si l'univers Ω est constitué de n événements élémentaires $\{e_i\}$, une* mesure de probabilité *sur Ω consiste à se donner n nombres $p_i \in [0,1]$, les probabilités des événements élémentaires, tels que*

$$\sum_{i=1}^{n} p_i = 1.$$

Si l'événement A est la réunion disjointe de k événements élémentaires $\{e_i\}$, avec $0 < k < n$, la probabilité de A vaut, par définition,

$$p(A) = p(\bigcup_{i=1}^{k}\{e_i\}) = \sum_{i=1}^{k} p(e_i) = \sum_{i=1}^{k} p_i.$$

Par suite, $0 \leq p(A) \leq 1$.

La signification concrète de la probabilité d'un événement A est la suivante. Dans une expérience aléatoire, plus $p(A)$ est proche de 1, plus A a de chances d'être réalisé ; plus $p(A)$ est proche de 0, moins il a de chances d'être réalisé.

Exemple 2 *Probabilité* uniforme *ou* équiprobabilité *: tous les P_i valent $1/n$. La probabilité d'un sous-ensemble à k éléments vaut alors $p(A) = \dfrac{k}{n} = \dfrac{cardA}{card\Omega}$.*

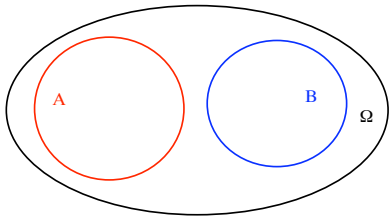
On exprime aussi cette propriété par la formule

$$p(A) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}.$$

Les propriétés suivantes découlent de la définition.

Propriété 3 *Si A et B sont incompatibles, i.e., si leur intersection $A \cap B$ est vide, alors*

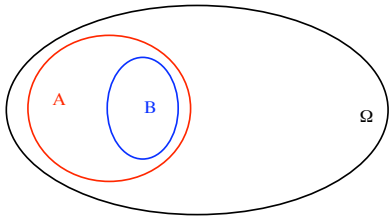
$$p(A \cup B) = p(A) + p(B).$$



Propriété 4 *Si B est un sous-ensemble de A,*

$$B \subseteq A \Rightarrow p(B) \leq p(A).$$

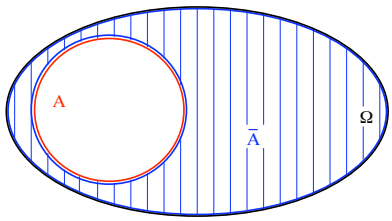
En effet, $B = A \cup (B \setminus A)$. Or A et $B \setminus A$ sont incompatibles ($B \setminus A$ est l'ensemble des éléments de B qui ne sont pas éléments de A). Donc $p(B) = p(A) + p(B \setminus A)$. Comme $p(B \setminus A)$ est positive, on obtient le résultat annoncé.



Propriété 5 *On appelle $\emptyset$ l'évènement impossible, puisqu'il n'est jamais réalisé. Sa probabilité vaut $p(\emptyset) = 0$.*

Propriété 6 *On note $\bar{A}$ l'évènement contraire de A. C'est le complémentaire de A dans Ω. Sa probabilité vaut*

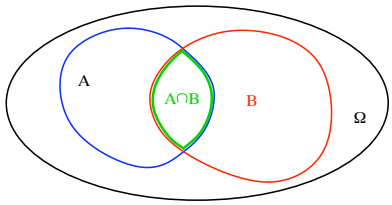
$$p(\bar{A}) = 1 - p(A).$$



Propriété 7 (Théorème des probabilités totales). *Si A et B sont deux sous-ensembles de Ω,*

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$

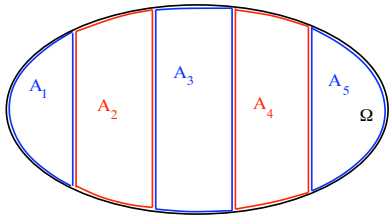
Preuve. $p(A) = p(A \setminus B) + p(A \cap B)$ car $A \setminus B$ et $A \cap B$ sont incompatibles. De même $p(B) = p(B \setminus A) + p(A \cap B)$ car $B \setminus A$ et $A \cap B$ sont incompatibles. De plus, $p(A \cup B) = p(A \setminus B) + p(B \setminus A) + p(A \cap B)$, car $A \setminus B$, $B \setminus A$ et $A \cap B$ sont incompatibles. En additionnant, il vient $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.■



Propriété 8 (Généralisation du théorème des probabilités totales ou règle de l'addition). *Si* $A_1, \ldots, A_k$ *forment une* partition *de* Ω*, i.e. ils sont deux à deux disjoints* ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$)*, et* $\Omega = \bigcup_{j=1}^{k} A_j$*, alors*

$$\sum_{j=1}^{k} p(A_j) = 1.$$

Dans cette situation, on dit parfois que les A_j forment un *système complet d'évènements.*



1.2 Probabilité conjointe

La probabilité que deux événements A et B se réalisent est appelée probabilité conjointe de A et B, notée $p(A \cap B)$ et s'énonçant probabilité de A et B. Le calcul de cette probabilité s'effectue de manière différente selon que A et B sont dépendants ou indépendants, c'est-à-dire selon que la réalisation de l'un influence ou non celle de l'autre.

1.2.1 Evénements indépendants

Exemple 9 *Je lance un dé rouge et un dé vert et je cherche la probabilité d'obtenir un total de 2. Je dois donc obtenir 1 avec chacun des deux dés. La probabilité d'obtenir 1 avec le dé rouge est* $1/6$ *et demeurera* $1/6$ *quelque soit le résultat du dé vert. Les deux événements "obtenir 1 avec le dé rouge" et "obtenir 1 avec le dé vert" sont indépendants.*

Propriété 10 *Si deux événements sont indépendants, la probabilité qu'ils se réalisent tous les deux est égale au produit de leurs probabilités respectives. On peut donc écrire :*

$$p(A \cap B) = p(A) \times p(B).$$

Dans notre exemple : $p(\text{total} = 2) = p(\text{dé vert} = 1) \times p(\text{dé rouge} = 1) = 1/36$.

Remarque 11 *Les tirages avec remise constituent une bonne illustration d'événements indépendants.*

1.2.2 Evénements dépendants - probabilité conditionnelle

Si deux événements sont dépendants plutôt qu'indépendants, comment calculer la probabilité que les deux se réalisent, puisque la probabilité de réalisation de l'un dépend de la réalisation de l'autre ? Il nous faut connaître pour cela le degré de dépendance des deux événements qui est indiqué par la notion de probabilité conditionnelle.

Définition 12 *Soient A et B deux événements, A étant supposé de probabilité non nulle. On appelle* probabilité conditionnelle *de B par rapport à A, la probabilité de réalisation de l'événement B sachant que A est réalisé. On la note*

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

$p(B|A)$ se lit *p de B si A* ou *p de B sachant A.*

Remarque 13 *L'application : $p_B : A \mapsto p_B(A) = p(A|B)$, $\Omega \to [0,1]$, est une probabilité sur Ω et vérifie toutes les propriétés d'une probabilité.*

Théorème 1 (Théorème des probabilités composées ou règle de la multiplication).

$$p(A \cap B) = p(B|A)p(A) = p(A|B)p(B).$$

En voici une généralisation. Soit $A_1, \ldots, A_k$ un système complet d'évènements. Alors

$$p(B) = \sum_{j=1}^{k} p(B \cap A_j) = \sum_{j=1}^{k} p(A_j)p(B|A_j).$$

Théorème 2 (Formule de Bayes). *Soit $A_1, \ldots, A_k$ un système complet d'évènements. Soit E un évènement de probabilité non nulle. Alors*

$$p(A_j|E) = \frac{p(A_j \cap E)}{p(E)} = \frac{p(A_j)p(E|A_j)}{\sum_{i=1}^{k} p(A_i)p(E|A_i)}.$$

Remarque 14 *Les tirages sans remise constituent une bonne illustration d'événements dépendants.*

Exercice 15 *Une urne contient 5 boules noires et 3 boules blanches. Quelle est la probabilité d'extraire 2 boules blanches en 2 tirages ?*

Solution de l'exercice 15. *Tirage sans remise.*

Appelons B_1, l'événement : obtenir une boule blanche au premier tirage.

Appelons B_2, l'événement : obtenir une boule blanche au deuxième tirage.

La probabilité cherchée $p(B_1 \cap B_2)$ est égale à $p(B_1) \times p(B_2|B_1)$. Or $p(B_1)$ vaut 3/8 et $p(B_2|B_1)$ est égale à 2/7 puisque lorsqu'une boule blanche est sortie au premier tirage, il ne reste plus que 7 boules au total, dont 2 seulement sont blanches. On conclut que $p(B_1 \cap B_2) = \frac{3}{8} \times \frac{2}{7} = \frac{3}{28}$.

1.3 Comment aborder un exercice de probabilités ?

Dans de nombreux problèmes, la recherche des solutions peut être facilitée par la démarche suivante.

1. Déterminer la liste des événements élémentaires ou décrire le contenu de l'univers Ω.
2. Rechercher la mesure de probabilité associée à cet univers.
   - Soit la probabilité est uniforme et dans ce cas, la probabilité d'un événement A est donnée par $p(A) = \frac{\text{card}A}{\text{card}\Omega}$.
   - Soit on détermine la probabilité de chaque événement élémentaire en n'oubliant pas que la somme de toutes les probabilités de ces événements élémentaires est égale à 1.
3. Identifier correctement le ou les événements dont on cherche à évaluer la probabilité.
4. Utiliser la formule appropriée permettant de calculer la probabilité demandée. On pourra se poser la question suivante : Doit-on calculer la probabilité ?
   - D'un événement élémentaire ?
   - D'un événement contraire ?

– Événements compatibles ou incompatibles (probabilités totales) ?
– Événements dépendants ou indépendants (probabilités composées) ?

Exercice 16 *1. On jette deux dés non pipés. Quelle est la probabilité d'obtenir un total de 7 points ?*

2. Cette fois-ci les dés sont pipés : les numéros pairs sont deux fois plus probables que les numéros impairs. Quelle est la probabilité d'obtenir un total différent de 8 ?

Solution de l'exercice 16. *Dés pipés.*

1. – L'univers est l'ensemble de tous les résultats possibles lorsqu'on jette deux dés. Imaginons que les deux dés sont reconnaissables et les résultats sont donc tous les couples (a, b) où a et b sont des nombres compris entre 1 et 6. Il contient donc 36 éléments. On peut écrire $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$ et $\text{card}(\Omega) = 36$.
   – Tous les résultats possibles sont équiprobables. La mesure de probabilité est donc uniforme sur Ω.
   – L'événement dont on cherche la probabilité est (somme = 7). Il est composé des événements élémentaires $(1, 6)$, $(6, 1)$, $(2, 5)$, $(5, 2)$, $(3, 4)$, $(4, 3)$. Ils sont au nombre de 6. On peut écrire : $\text{card}(\text{somme} = 7) = 6$.
   – Finalement, étant donné que $p(A) = \frac{\text{card}A}{\text{card}\Omega}$, on obtient $p(\text{somme} = 7) = \frac{6}{36} = \frac{1}{6}$.
2. – L'univers est toujours le même.
   – On cherche à déterminer la mesure de probabilité sur Ω dans le cas où les dés sont truqués : elle n'est plus uniforme.
   Il faut répondre à la question : lorsqu'on lance un seul dé, quelle est la probabilité de chaque numéro ?
   – Tous les numéros pairs ont la même probabilité que l'on note p^p ; tous les numéros impairs ont la même probabilité que l'on note p^i. L'énoncé nous permet d'écrire que $p^p = 2p^i$.
   – D'autre part, étant donné que les numéros 1,2,3,4,5,6 constituent l'ensemble des résultats d'un jet de dé, la somme des probabilités de ces 6 résultats vaut 1. D'où $3p^p + 3p^i = 1$, soit encore $9p^i = 1$. D'où $p^i = \frac{1}{9}$ et $p^p = \frac{2}{9}$.
   – L'événement dont on cherche la probabilité est (somme $\neq$ 8). Chercher directement la probabilité de cet événement nous obligerait à considérer beaucoup de cas. Il sera donc plus rapide de déterminer d'abord la probabilité de l'événement contraire (somme = 8). Ce dernier est constitué des événements élémentaires $(2, 6)$, $(6, 2)$, $(3, 5)$, $(5, 3)$, $(4, 4)$.
   – Les résultats des deux dés sont indépendants. Nous pouvons donc affirmer que

$$p(\{(2, 6)\}) = p(\{2\}) \times p(\{6\}) = \frac{2}{9} \times \frac{2}{9} = \frac{4}{81}.$$

   De même, $p(\{(6, 2)\}) = p(\{(4, 4)\}) = \frac{4}{81}$, alors que $p(\{(5, 3)\}) = p(\{(3, 5)\}) = \frac{1}{81}$
   – Finalement $p(\text{somme} = 8) = \frac{14}{81}$ et $p(\text{somme} \neq 8) = \frac{67}{81}$.

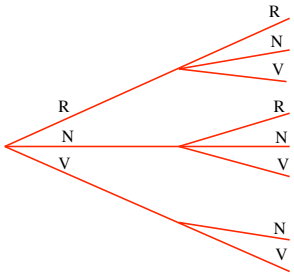
1.4 Techniques de dénombrement

1.4.1 Diagrammes arborescents ou arbres

Exemple 17 *On considère une urne qui contient deux boules rouges, deux noires et une verte. On tire deux boules sans remise. Il s'agit d'une expérience à deux étapes où les différentes possibilités qui peuvent survenir sont représentées par un arbre horizontal.*

On obtient trois branches principales et trois branches secondaires pour chaque étape sauf pour le cas où une verte a été tirée en premier.

Le nombre de branches terminales de cet arbre donne le nombre d'éléments de l'univers.



Lorsqu'on rencontre beaucoup d'étapes dans une expérience et de nombreuses possibilités à chaque étape, l'arbre associé à l'expérience devient trop complexe pour être analysé. Ces problèmes se simplifient à l'aide de formules algébriques, comme on va le voir.

La démonstration de ces formules repose sur le fait que dans le cas d'une expérience à deux étapes, par exemple, un arbre qui aurait r branches principales et s branches secondaires commençant à partir des r branches principales aura rs branches terminales.

1.4.2 Arrangements et permutations

Envisageons un ensemble de n objets différents. Choisissons maintenant r de ces n objets et ordonnons les.

Définition 18 *Une disposition ordonnée de r objets distincts pris parmi n est appelée* arrangement *de r objets pris parmi n (on a obligatoirement $r \leq n$).*

Combien y en a-t-il ?

Pour compter le nombre total d'arrangements de r objets pris parmi n, il suffit de considérer les r positions comme fixées et de compter le nombre de façons dont on peut choisir les objets pour les placer dans ces r positions. C'est une expérience à r étapes où l'on applique la technique du paragraphe précédent. Pour la première position, on a n choix possibles. Pour la deuxième position, on a $n-1$ choix possibles... Pour la r-ième position, on a $n-r+1$ choix possibles. Si on désigne par A_n^r le nombre total d'arrangements cherchés, l'arbre aura A_n^r branches terminales. On conclut

Proposition 19

$$A_n^r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

Rappel 20 *$n!$ (lire "factorielle n") est le produit de tous les entiers jusqu'à n, $n! = n(n-1)(n-2)\cdots 3.2.1$. Par convention, $0! = 1$.*

Exemple 21 *Les arrangements de deux lettres prises parmi 4 lettres $\{a,b,c,d\}$ sont au nombre de $A_4^2 = \frac{4!}{2!} = 12$. Ce sont : $(a,b),(a,c),(a,d),(b,a),(b,c),(b,d),\ (c,a),(c,b),(c,d),(d,a),(d,b),(d,c)$.*

Cas particulier : $r = n$ Il s'agit d'ordonner n objets entre eux, c'est-à-dire d'effectuer une permutation de ces n objets.

Définition 22 *Une permutation de n éléments est une disposition ordonnée de ces n éléments.*

Proposition 23 *Les permutations de n éléments sont au nombre de $A_n^n = n!$.*

1.4.3 Combinaisons

Définition 24 *Un choix de r objets distincts pris parmi n sans tenir compte de leur ordre est appelé combinaison de r objets pris parmi n.*

Dans l'exemple précédent correspondant à l'ensemble des quatre lettres $\{a, b, c, d\}$, la combinaison $\{a, b\}$ est la même que la combinaison $\{b, a\}$ alors que l'arrangement (a, b) est différent de l'arrangement (b, a).

Combien y en a-t-il ? Le nombre total de combinaisons de r objets pris parmi n est noté C_n^r ou $\binom{r}{n}$. Pour trouver l'expression de $\binom{r}{n}$, comparons le nombre d'arrangements et de combinaisons possibles de r objets pris parmi n.

- Dans un arrangement on choisit r objets, puis on tient compte de leur ordre.
- Dans une combinaison seul le choix des r objets compte. Comme le nombre de façons d'ordonner les r objets choisis est $r!$, on conclut qu'à chaque combinaison de r objets pris parmi n, on peut associer $r!$ arrangements et donc qu'il y a $r!$ fois plus d'arrangements que de combinaisons.

On conclut

Proposition 25

$$\binom{r}{n} = \frac{A_n^r}{r!} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}.$$

Exemple 26 *Le nombre de combinaisons de deux lettres prises parmi quatre* $\{a, b, c, d\}$ *est* $\binom{2}{4} = \frac{4!}{2!2!} = 6$. *Ce sont :* $\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}$.

1.4.4 Permutations lorsque certains éléments sont semblables

Dans les paragraphes précédents, on a supposé que les n objets étaient tous différents. Il arrive parfois que les n objets en contiennent un certain nombre qui sont indiscernables.

Supposons qu'il n'y ait que k sortes d'objets distincts sur les n objets. Il y a

- n_1 objets de la 1-ère sorte,
- n_2 objets de la 2-ème sorte....
- n_k objets de la k-ème sorte.

On a bien sûr $n_1 + n_2 + \cdots + n_k = n$.

Pour déterminer le nombre total de permutations distinctes, comparons ce nombre cherché $\mathcal{P}$ avec le nombre obtenu si on supposait les objets différenciés. Plaçons nous dans le cas de l'exemple suivant : On cherche le nombre d'anagrammes du mot $PROBABILITE$.

Choisissons un de ces anagrammes : le plus simple est $PROBABILITE$.

- Si on différencie les lettres B, cette disposition peut provenir des deux permutations $PROB_1AB_2ILITE$ ou $PROB_2AB_1ILITE$, soit 2! possibilités.
- Si on différencie les lettres I, cette disposition peut provenir des deux permutations $PROBABI_1LI_2TE$ ou $PROBABI_2LI_1TE$, soit encore 2! possibilités

A un anagramme correspond donc $2! \times 2! = 4$ permutations, ce qui signifie qu'il y a 4 fois plus de permutations que d'anagrammes. Le mot $PROBABILITE$ comprend 11 lettres. Il y a 11! permutations possibles. On a donc $\frac{11!}{2!2!} = 9979200$ anagrammes possibles.

Cas général. La différenciation des n_1 premiers objets donnera $n_1!$ fois plus d'éléments que ce qu'on cherche, la différenciation des n_2 premiers objets donnera $n_2!$ fois plus d'éléments que ce qu'on cherche, et finalement on trouve que $n!$ est $n_1!n_2!\cdots n_k!$ fois plus grand que le nombre cherché $\mathcal{P}$. On conclut

Proposition 27 *Le nombre d'anagrammes d'un mot de n lettres, comportant seulement $k < n$ lettres distinctes, en nombres $n_1, \ldots, n_k$ est*

$$\mathcal{P} = \frac{n!}{n_1!n_2!\cdots n_k!}.$$

1.4.5 Cas ou les éléments ne sont pas obligatoirement distincts

Combien y a-t-il de manières de choisir r éléments parmi n de façon ordonnée en n'imposant pas qu'ils soient tous distincts les uns des autres ?

En 1ère position, il y a n choix possibles. En 2ème position, il y a encore n choix possibles... En $r$ème position, il y a toujours n choix possibles.

Conclusion : Il y a donc n^r choix pour les r éléments (r peut être supérieur à n dans ce cas).

1.4.6 Récapitulation

	Conditions	Le nombre de tirages possibles est le nombre de :	Un exemple usuel
$p \geq n$	les p éléments ne sont pas nécessairement tous distincts mais sont ordonnés	p-listes d'éléments de E, soit : n^p	tirages successifs avec remise de p objets parmi n
$p < n$	les p éléments sont tous distincts et ordonnés	arrangements de p éléments de E, soit : A_n^p	tirages successifs sans remise de p objets parmi n.
$p = n$	les n éléments sont tous distincts et ordonnés	permutations des n éléments de E, soit : $n!$	anagrammes d'un mot formé de lettres toutes distinctes
$p < n$	les p éléments sont tous distincts et non ordonnés	combinaisons de p éléments de E, soit $\binom{p}{n}$	tirages simultanés de p objets parmi n.

Chapitre 2

Variables aléatoires

2.1 Définition

Exemple 28 *On jette deux fois une pièce de monnaie non truquée, et on s'intéresse au nombre de fois que le côté "face" a été obtenu. Pour calculer les probabilités des divers résultats, on introduira une variable X qui désignera le nombre de "face" obtenu. X peut prendre les valeurs 0,1,2.*

Exemple 29 *On lance une fléchette vers une cible circulaire de rayon égal à 50 cm et on s'intéresse à la distance entre la fléchette et le centre de la cible. On introduira ici une variable X, distance entre l'impact et le centre de la cible, qui peut prendre n'importe quelle valeur entre 0 et 50.*

Dans ces deux cas, X prend des valeurs réelles qui dépendent du résultat de l'expérience aléatoire. Les valeurs prises par X sont donc aléatoires. X est appelée variable aléatoire.

Définition 30 *Soit un univers Ω associé à une expérience aléatoire, sur lequel on a défini une mesure de probabilité. Une* variable aléatoire *X est une application de l'ensemble des événements élémentaires de l'univers Ω vers $\mathbf{R}$ (vérifiant quelques conditions mathématiques non explicitées ici).*

Une variable aléatoire est une variable (en fait une fonction !) qui associe des valeurs numériques à des événements aléatoires.

Par convention, une variable aléatoire sera représentée par une lettre majuscule X alors que les valeurs particulières qu'elle peut prendre seront désignées par des lettres minuscules $x_1, x_2, \ldots, x_i, \ldots, x_n$.

Les deux variables aléatoires définies dans les exemples 28 et 29 sont de natures différentes. La première est discrète, la seconde continue.

2.2 Variables aléatoires discrètes

Définition 31 *Une* variable aléatoire discrète *est une variable aléatoire qui ne prend que des valeurs entières, en nombre fini ou dénombrable.*

Pour apprécier pleinement une variable aléatoire, il est important de connaître quelles valeurs reviennent le plus fréquemment et quelles sont celles qui apparaissent plus rarement. Plus précisément, on cherche les probabilités associées aux différentes valeurs de la variable

Définition 32 *Associer à chacune des valeurs possibles de la variable aléatoire la probabilité qui lui correspond, c'est définir la* loi de probabilité *ou la* distribution de probabilité *de la variable aléatoire.*

Pour calculer la probabilité que la variable X soit égale à x, valeur possible pour X, on cherche tous les événements élémentaires e_i pour lesquels $X(e_i) = x$, et on a

$$p(X = x) = \sum_{i=1}^{k} p(\{e_i\}),$$

si $X = x$ sur les événements élémentaires $e_1, e_2, \ldots, e_k$.

La *fonction de densité discrète* f est la fonction de $\mathbf{R}$ dans $[0, 1]$, qui à tout nombre réel x_i associe $f(x_i) = p(X = x_i)$. On a bien sûr $\sum_i f(x_i) = 1$.

Exemple 33 Cas de l'exemple 28.
La variable X = nombre de côtés "face" peut prendre les valeurs 0,1,2.

$$\begin{aligned} f(0) &= p(X = 0) = p((pile, pile)) = \frac{1}{4}; \\ f(1) &= p(X = 1) = p((pile, face)) + p((face, pile)) = \frac{1}{2}; \\ f(2) &= p(X = 2) = p((face, face)) = \frac{1}{4}; \\ f(x) &= 0 \text{ si } x \notin \{0, 1, 2\}. \end{aligned}$$

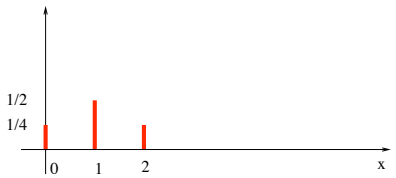
On présente sa distribution de probabilité dans un tableau.

x	0	1	2	total
$f(x) = p(X = x)$	1/4	1/2	1/4	1

2.2.1 Représentation graphique de la distribution de probabilité

Elle s'effectue à l'aide d'un diagramme en bâtons où l'on porte en abscisses les valeurs prises par la variable aléatoire et en ordonnées les valeurs des probabilités correspondantes.

Dans l'exemple du jet de pièces :



2.2.2 Fonction de répartition

En statistique descriptive, on a introduit la notion de fréquences cumulées croissantes. Son équivalent dans la théorie des probabilités est la fonction de répartition.

Définition 34 *La* fonction de répartition *d'une variable aléatoire X indique pour chaque valeur réelle x la probabilité que X prenne une valeur au plus égale à x. C'est la somme des probabilités des valeurs de X jusqu'à x. On la note F.*

$$\forall x \in \mathbf{R}, \quad F(x) = p(X \leq x) = \sum_{x_i \leq x} p(X = x_i).$$

La fonction de répartition est toujours croissante, comprise entre 0 et 1 et se révélera un instrument très utile dans les travaux théoriques.

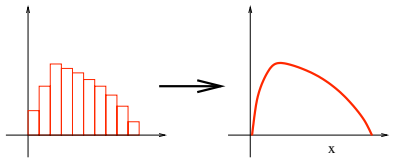
2.3 Variables aléatoires continues

Définition 35 *Une variable aléatoire est dite* continue *si elle peut prendre toutes les valeurs d'un intervalle fini ou infini.*

2.3.1 Fonction de densité de probabilité

Dans le cours de statistique descriptive, nous avons appris à représenter la distribution d'une variable statistique continue (ou à caractère continu) à l'aide d'un histogramme de fréquences, qui est une série de rectangles. L'aire de chaque rectangle est proportionnelle à la fréquence de la classe qui sert de base au rectangle.

Si l'on augmentait indéfiniment le nombre d'observations en réduisant graduellement l'intervalle de classe jusqu'à ce qu'il soit très petit, les rectangles correspondant aux résultats vont se multiplier tout en devenant plus étroits et à la limite vont tendre à se fondre en une surface unique limitée d'une part par l'axe des abscisses et d'autre part par une courbe continue.



On abandonne alors la notion de valeur individuelle et l'on dit que la distribution de probabilité est continue. La courbe des fréquences relatives idéalisée est alors la courbe représentative d'une fonction de densité de probabilité f.

Pour une variable statistique continue, l'aire des rectangles était un témoin fidèle de la fréquence de chaque classe. Il en va en de même pour une variable aléatoire continue mais il faudra raisonner à présent sur des classes infiniment petites d'amplitude dx. L'élément infinitésimal d'aire $f(x)\,dx$ représente la probabilité que X appartienne à un intervalle d'amplitude dx,

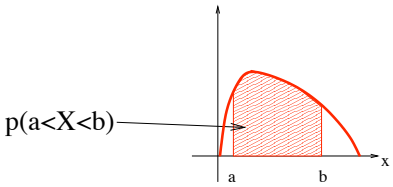
$$p(x < X \leq x + dx) = f(x)\,dx.$$

f a donc les propriétés suivantes :

a) La courbe d'une fonction de densité de probabilité est toujours située au dessus de l'axe des abscisses donc f est une fonction toujours positive.

b) La probabilité que la variable aléatoire X soit comprise entre les limites a et b c'est-à-dire $p(a \leq X \leq b)$, est égale à l'aire entre l'axe des abscisses, la courbe représentative de la fonction de densité de probabilité et les droites d'équations $x = a$ et $x = b$,

$$p(a < X \leq b) = \int_a^b f(x)\,dx$$



c) L'aire totale comprise entre la courbe et l'axe des abscisses est égale à 1 :

$$\int_{\mathbf{R}} f(x)\,dx = 1.$$

d) Alors qu'une probabilité est sans dimension, une densité de probabilité a pour dimension celle de l'inverse de X : $[X^{-1}]$.

e) Il résulte de a et c qu'une densité de probabilité est une fonction *intégrable* au sens de Lebesgue sur $\mathbf{R}$.

Remarque 36 *Le cas où on a une courbe continue est un cas théorique qui supposerait :*
- *un nombre infini de mesures de la variable statistique*
- *une sensibilité très grande de l'appareil de mesure.*

Nous supposerons toutefois que nous sommes dans ce cas lorsque nous serons en présence d'un grand nombre de mesures.

2.3.2 Fonction de répartition

De même que pour les variables aléatoires discrètes, on peut définir la fonction de répartition F de la variable continue X qui permet de connaître la probabilité que X soit inférieure à une valeur donnée :

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

Propriété 37 *1. F est continue et croissante sur $\mathbf{R}$.*

2. $\forall x \in \mathbf{R}, \quad F'(x) = f(x)$.

3. $\lim\limits_{x\to-\infty} F(x) = 0, \quad \lim\limits_{x\to+\infty} F(x) = 0$.

4. $p(a \leq X \leq b) = F(b) - F(a)$.

Exercice 38

Soit f la fonction définie sur $\mathbf{R}$ par $f(x) = ke^{-x}$ si $x \geq 0$, $f(x) = 0$ sinon.

1. Déterminer k pour que f soit la fonction de densité de probabilité d'une variable aléatoire X.
2. Déterminer la fonction de répartition de la variable X.
3. Calculer $p(1 < X < 2)$.

Solution de l'exercice 38. *Densité de probabilité.*

1. f doit être une fonction positive, donc il nous faut impérativement trouver pour k une valeur positive. Une fonction de densité de probabilité doit vérifier $\int_{\mathbf{R}} f(x)\,dx = 1$, donc $\int_0^{+\infty} ke^{-x}\,dx = 1$. Il en résulte que $k = 1$.
2. Par définition la fonction de répartition de X est la fonction F définie par

$$\begin{aligned} F(x) &= \int_0^x e^{-t}\,dt = 1 - e^{-x} \text{ si } x > 0, \\ &= 0 \text{ sinon.} \end{aligned}$$

3.

$$p(1 < X < 2) = \int_1^2 e^{-x}\,dx = e^{-1} - e^{-2} \sim 0.23.$$

2.4 Couples de variables aléatoires

Il existe beaucoup de situations où l'intérêt se porte sur la réalisation conjointe d'événements associés à deux (ou plusieurs) variables aléatoires. Nous allons envisager deux cas : celui où les variables sont discrètes et celui où elles sont continues.

2.4.1 Couples de variables aléatoires discrètes

Loi de probabilité conjointe

Considérons deux variables aléatoires discrètes X et Y. Il nous faut pour modéliser le problème une fonction qui nous donne la probabilité que $(X = x_i)$ en même temps que $(Y = y_j)$. C'est la loi de probabilité conjointe.

Définition 39 *Soient X et Y deux variables aléatoires discrètes dont l'ensemble des valeurs possibles sont respectivement $\{x_1, x_2, \ldots, x_n\}$ et $\{y_1, y_2, \ldots, y_m\}$. Associer à chacune des valeurs possibles (x_i, y_j) du couple (X, Y), la probabilité $f(x_i, y_j)$, c'est définir la* loi de probabilité conjointe *ou* fonction de densité conjointe *des variables aléatoires X et Y,*

$$f(x_i, y_j) = p((X = x_i) \text{ et } (Y = y_j)).$$

Le couple (X, Y) s'appelle variable aléatoire à deux dimensions et peut prendre $m \times n$ valeurs.

Proposition 40 *1. Pour tout $i = 1, 2, \ldots, n$ et $j = 1, 2, \ldots, m$, $0 \leq f(x_i, y_j) \leq 1$.*

2. $\sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_j) = 1.$

On peut représenter graphiquement f sous forme d'un diagramme en bâtons en trois dimensions.

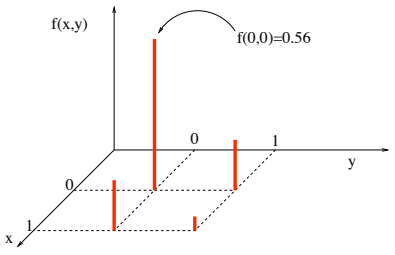
Exemple 41 *Soit X le nombre de piques obtenus lors du tirage d'une carte dans un jeu ordinaire de 52 cartes et Y le nombre de piques obtenus dans un deuxième tirage, la première carte n'étant pas remise. X et Y ne prennent que les valeurs 0 (pas de pique) ou 1 (un pique).*

Détermination de la loi du couple (X, Y).

$$\begin{aligned} f(0,0) &= \frac{39}{52} \times \frac{38}{51} = 0.56, \quad f(1,0) = \frac{13}{52} \times \frac{39}{51} = 0.19, \\ f(0,1) &= \frac{39}{52} \times \frac{13}{51} = 0.19, \quad f(1,1) = \frac{39}{52} \times \frac{12}{51} = 0.06. \end{aligned}$$

On vérifie que la somme de ces valeurs est égale à 1.

On représente f sous forme d'un diagramme en bâtons en trois dimensions.



Loi de probabilité marginale

Lorsqu'on connaît la loi conjointe des variables aléatoires X et Y, on peut aussi s'intéresser à la loi de probabilité de X seule et de Y seule. Ce sont les lois de probabilité marginales.

Définition 42 *Soit (X, Y) une variable aléatoire à deux dimensions admettant comme loi de probabilité conjointe $f(x, y)$. Alors, les lois de probabilité marginales de X et Y sont définies respectivement par*

$$\begin{aligned} f_X(x_i) &= p(X = x_i) = \sum_{j=1}^{m} f(x_i, y_j) \text{ pour } i = 1, 2, \ldots, n, \\ f_Y(y_j) &= p(Y = y_i) = \sum_{i=1}^{n} f(x_i, y_j) \text{ pour } j = 1, 2, \ldots, m. \end{aligned}$$

Si la loi de probabilité conjointe du couple (X, Y) est présentée dans un tableau à double entrée, nous obtiendrons la loi de probabilité marginale f_X de X en sommant les $f(x_i, y_j)$, suivant l'indice j (par colonnes) et celle de Y, f_Y, en sommant les $f(x_i, y_j)$ suivant l'indice i (par lignes).

Exemple 43 *Lois marginales de l'exemple 41.*

	$Y = y_1 = 0$	$Y = y_2 = 1$	$f_X(x_i)$
$X = x_1 = 0$	0.56	0.19	0.75
$X = x_2 = 1$	0.19	0.06	0.25
$f_Y(y_j)$	0.75	0.25	1.00

Remarque 44 *Le couple (X, Y) et les deux variables X et Y constituent trois variables aléatoires distinctes. La première est à deux dimensions, les deux autres à une dimension.*

Loi de probabilité conditionnelle

Nous avons vu dans le paragraphe précédent comment déterminer la probabilité de réalisation de deux événements lorsqu'ils sont dépendants l'un de l'autre. Pour cela, nous avons introduit la notion de probabilité conditionnelle en posant

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

La notion équivalente dans le cas d'un couple de variables aléatoires est celle de loi de probabilité conditionnelle permettant de mesurer la probabilité que X soit égale à une valeur donnée lorsqu'on connaît déjà la valeur que prend Y.

Définition 45 *Soit la variable aléatoire (X, Y) à deux dimensions admettant comme loi conjointe $f(x, y)$ et comme lois marginales $f_X(x)$ et $f_Y(y)$. Si l'on suppose que la probabilité que X prenne la valeur x_i n'est pas nulle, alors la probabilité conditionnelle de $(Y = y_j)$ sachant que $(X = x_i)$ s'est réalisé est définie par*

$$f(y_j|x_i) = \frac{f(x_i, y_j)}{f_X(x_i)}.$$

Les probabilités $f(y_j|x_i)$ associées aux différentes valeurs possibles y_j de Y constituent la loi de probabilité conditionnelle de Y.

De même, si l'on suppose que la probabilité que Y prenne la valeur y_j n'est pas nulle, alors la probabilité conditionnelle de $(X = x_i)$ sachant que $(Y = y_j)$ s'est réalisé est définie par

$$f(x_i|y_j) = \frac{f(x_i, y_j)}{f_Y(y_j)}.$$

Les probabilités $f(x_i|y_j)$ associées aux différentes valeurs possibles x_i de X constituent la loi de probabilité conditionnelle de X.

Cas de variables aléatoires indépendantes

Lorsque deux variables aléatoires X et Y sont indépendantes, la loi conditionnelle de X, pour toute valeur de Y, est identique à la loi marginale de X et lorsque la loi conditionnelle de Y, pour toute valeur de X, est identique à la loi marginale de Y. Autrement dit,

Proposition 46 *Soit (X, Y) une variable aléatoire à deux dimensions admettant comme loi de probabilité conjointe la fonction $f(x, y)$ et comme lois de probabilité marginales $f_X(x)$ et $f_Y(y)$. Les variables aléatoires X et Y sont indépendantes si et seulement si les probabilités conjointes sont égales au produit des probabilités marginales, $f(x_i, y_j) = f_X(x_i) \times f_Y(y_j)$ pour toutes les valeurs (x_i, y_j).*

Pour conclure que deux variables ne sont pas indépendantes, il suffit de trouver une valeur du couple (X, Y) pour laquelle la relation précédente n'est pas satisfaite.

Exemple 47 *Dans l'exemple du tirage de cartes, nous savons, par exemple que* $p((X = 0)$ *et* $(Y = 1)) = f(0, 1) = 0.19$*, alors que* $p(X = 0) \times p(Y = 1) = f_X(0) \times f_Y(1) = 0.75 \times 0.25 = 0.188$.

Conclusion : les variables X et Y sont dépendantes, ce qui paraît cohérent étant donné que le tirage était effectué sans remise.

2.4.2 Couple de variables aléatoires continues

Dans le cas de deux variables continues X et Y, le couple (X, Y) est dit continu.

Fonction de densité de probabilité conjointe

La distribution de probabilité conjointe de X et de Y est décrite par une *fonction de densité de probabilité conjointe* $f(x, y)$ définie pour chaque valeur (x, y) du couple (X, Y). La fonction f détermine une surface au-dessus de l'ensemble des valeurs (x, y).

On a $p((X, Y) \in D)$ = volume sous la surface représentative de f et au-dessus du domaine D du plan (xOy).

Dans le cas où D est un rectangle $[c, d] \times [u, v]$,

$$p((c < X \leq d) \text{ et } (u < Y \leq v)) = \int_{[c,d]\times[u,v]} f(x, y)\, dx\, dy.$$

Propriété 48 *1. Pour tout couple* $(x, y) \in \mathbf{R}^2$, $f(x, y) \geq 0$.

2. $\int_{\mathbf{R}^2} f(x, y)\, dx\, dy = 1$.

3. Il en résulte qu'une densité de probabilité conjointe est une fonction intégrable *au sens de Lebesgue sur* $\mathbf{R}^2$.

Densité de probabilité marginale

De même que pour les couples de variables aléatoires discrètes, on définit les fonctions densités de probabilité marginales et conditionnelles. Pour les définir dans le cas continu, il suffit de remplacer les sommes du cas discret par des intégrales.

Définition 49 *Soit* (X, Y) *une variable aléatoire continue à deux dimensions admettant comme densité de probabilité conjointe la fonction* $f(x, y)$*. Alors, les* densités de probabilité marginales *de* X *et* Y *sont définies respectivement par*

$$\begin{aligned} f_X(x) &= \int_{\mathbf{R}} f(x, y)\, dy \text{ pour } x \in \mathbf{R}, \\ f_Y(y) &= \int_{\mathbf{R}} f(x, y)\, dx \text{ pour } y \in \mathbf{R}. \end{aligned}$$

Variables indépendantes

Proposition 50 *Deux variables continues* X *et* Y *sont indépendantes si et seulement si la fonction de densité de probabilité conjointe est égale au produit des fonctions de densité marginales.*

Autrement dit, pour tout couple $(x, y) \in \mathbf{R}^2$,

$$f(x, y) = f_X(x) f_Y(y).$$

Dans ce cas, le théorème de Fubini-Tonelli donne

$$\begin{aligned} p((c < X \leq d) \text{ et } (u < Y \leq v)) &= \int_{[c,d]\times[u,v]} f(x,y)\,dx\,dy \\ &= \left(\int_c^d f_X(x)\,dx\right)\left(\int_u^v f_Y(y)\,dy\right). \end{aligned}$$

2.5 Paramètres descriptifs d'une distribution

En statistique descriptive, nous avons caractérisé les distributions statistiques des valeurs observées par certains nombres représentatifs qui résumaient de façon commode et assez complète la distribution. Pour apprécier la tendance centrale d'une série d'observations, nous avons employé, entre autres, la moyenne arithmétique et pour caractériser la dispersion de la série autour de la moyenne, nous avons introduit la variance ou l'écart quadratique moyen.

2.5.1 Espérance mathématique d'une distribution de probabilité

Si l'on s'imagine que le nombre d'observations croît indéfiniment (on passe d'un échantillon de taille n à la population toute entière), les fréquences observées vont tendre vers les probabilités théoriques et on admet que la moyenne calculée sur l'échantillon de taille n va tendre vers une valeur limite qui sera la moyenne de l'ensemble des valeurs de la population entière. On l'appelle espérance mathématique de la variable aléatoire X, car c'est la valeur moyenne que l'on s'attend à avoir dans un échantillon de grande taille.

Définition 51 *1. Cas d'une variable discrète*

- *Soit X une variable aléatoire discrète qui prend un nombre fini de valeurs $x_1, x_2, \ldots, x_n$ et dont la loi de probabilité est f : $f(x_i) = p(X = x_i)$. L'*espérance mathématique *de X, notée $E(X)$, est définie par*

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

- *Si la variable aléatoire X prend un nombre dénombrable de valeurs $x_1, x_2, \ldots, x_n, \ldots$, son* espérance mathématique *est alors définie par $E(X) = \sum_{i=1}^{\infty} x_i f(x_i)$, à condition que la série converge absolument.*

2. *Cas d'une variable continue. Si la variable aléatoire X est continue et a pour fonction de densité de probabilité f,* son espérance mathématique *est*

$$E(X) = \int_{\mathbf{R}} x f(x)\,dx,$$

pourvu que la fonction $x \mapsto xf(x)$ soit intégrable sur $\mathbf{R}$.

2.5.2 Variance d'une distribution de probabilités

En effectuant le même raisonnement que précédemment pour passer d'un échantillon de taille n à la population totale, on suppose que la variance calculée sur l'échantillon tend vers une limite lorsque le nombre d'observations tend vers l'infini. Cette limite est appelée variance de la variable aléatoire X.

Définition 52 – *On appelle* variance *de la variable aléatoire X la valeur moyenne des carrés des écarts à la moyenne,*

$$Var(X) = E\left((X - E(X))^2\right).$$

Le calcul de la variance se simplifie en utilisant l'expression :

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

– *On appelle* écart-type *de la variable aléatoire X la racine carrée de sa variance.*

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

Dans le cas d'une variable aléatoire continue,

Dans le cas d'une variable aléatoire discrète finie,

$$Var(X) = \sum_{i=1}^{n}(x_i - E(X))^2 f(x_i) = \left(\sum_{i=1}^{n} x_i^2 f(x_i)\right) - E(X)^2.$$

Dans le cas d'une variable aléatoire continue,

$$Var(X) = \int_{\mathbf{R}}(x - E(X))^2 f(x)\,dx = \left(\int_{\mathbf{R}} x^2 f(x)\,dx\right) - E(X)^2.$$

2.5.3 Fonction caractéristique d'une distribution de probabilité

Définition 53 *On appelle* fonction caractéristique *de la variable aléatoire X la fonction ξ_X définie sur $\mathbf{R}$ par*

$$\xi_X(u) = E(e^{-2i\pi uX}).$$

Dans le cas d'une variable aléatoire discrète finie,

$$\xi_X(u) = \sum_{i=1}^{n} f(x_i)e^{-2i\pi u x_i}.$$

Dans le cas d'une variable aléatoire continue,

$$\xi_X(u) = \int_{\mathbf{R}} e^{-2i\pi ux} f(x)\,dx.$$

Dans le cas continu, on constate que $\xi_X = Ff$ est la transformée de Fourier de la densité de probabilité. Elle existe donc toujours puisque la densité de probabilité est intégrable au sens de Lebesgue.

Proposition 54 *Soit X une variable aléatoire qui possède une espérance et une variance. Alors*

$$\begin{aligned} E(X) &= -\frac{1}{2i\pi}\frac{d}{du}\xi_X(u)_{|u=0}, \\ Var(X) &= -\frac{1}{4\pi^2}\left(\frac{d^2}{du^2}\xi_X(u)_{|u=0} - (\frac{d}{du}\xi_X(u)_{|u=0})^2\right). \end{aligned}$$

Preuve. Dans le cas discret,

$$\begin{aligned} \frac{d}{du}\xi_X(u) &= \sum_{i=1}^{n} -2i\pi x_i f(x_i)e^{-2i\pi u x_i} = -2i\pi E(Xe^{-2i\pi uX}), \\ \frac{d^2}{du^2}\xi_X(u) &= \sum_{i=1}^{n} -4\pi^2 x_i^2 f(x_i)e^{-2i\pi u x_i} = -4\pi^2 E(X^2 e^{-2i\pi uX}), \end{aligned}$$

d'où

$$\begin{aligned}\xi''_X(0) - (\xi'_X(0))^2 &= -4\pi^2 E(X^2) + 4\pi^2 E(X)^2 \\ &= -4\pi^2 (E(X^2) - E(X)^2) \\ &= -4\pi^2 Var(X).\end{aligned}$$

Dans le cas continu, le calcul est le même, au moyen de la formule pour la dérivée de la transformée de Fourier.■

Remarque 55 *Une loi de probabilité régit le comportement d'une variable aléatoire. Cette notion abstraite est associée à la population, c'est-à-dire à l'ensemble de tous les résultats possibles d'un phénomène particulier. C'est pour cette raison que l'espérance et la variance de la loi de probabilité, qui n'ont aucun caractère aléatoire, sont appelés paramètres de la distribution de probabilité.*

2.5.4 Propriétés de l'espérance mathématique et de la variance

Résumons les principales propriétés de ces deux paramètres dans un tableau.

Changement d'origine	Changement d'échelle	Transformation affine
$E(X+c) = E(X) + c$	$E(aX) = aE(X)$	$E(aX+c) = aE(X) + c$
$Var(X+c) = Var(X)$	$Var(aX) = a^2 Var(X)$	$Var(aX+c) = a^2 Var(X)$
$\sigma(X+c) = \sigma(X)$	$\sigma(aX) = \lvert a\rvert \sigma(X)$	$\sigma(aX+c) = \lvert a\rvert \sigma(X)$
$\xi_{X+c}(u) = e^{-2i\pi cu}\xi_X(u)$	$\xi_{aX}(u) = \xi_X(au)$	$\xi_{aX+c}(u) = e^{-2i\pi cu}\xi_X(au)$

Définition 56
- *Une variable aléatoire X est dite* centrée *si son espérance mathématique est nulle.*
- *Une variable aléatoire X est dite* réduite *si son écart-type est égal à 1.*
- *Une variable aléatoire centrée réduite est dite* standardisée.

A n'importe quelle variable aléatoire X, on peut associer la variable standardisée

$$Z = \frac{X - E(X)}{\sigma(X)}.$$

En divisant la variable centrée par son écart-type, une valeur située à un écart-type de la moyenne sera ramenée à 1, une autre située à deux écarts-types sera ramenée à 2 : l'échelle de référence, ou unité de mesure, d'une variable centrée-réduite est l'écart-type.

Les valeurs des variables centrées-réduites sont complètement indépendantes des unités de départ. Une mesure exprimée en mètres ou en centimètres donne exactement la même variable centrée-réduite. On peut ainsi faire des comparaisons entre variables de natures différentes. Si un enfant est à +3 écarts-types de la moyenne pour sa taille et +1 écart-type pour son poids, on sait qu'il est plus remarquable par sa taille que par son poids.

L'examen des variables centrées-réduites est très pratique pour déceler les valeurs "anormalement" grandes ou "anormalement" petites.

Le passage d'une variable aléatoire X à une variable standardisée est requis pour l'utilisation de certaines tables de probabilité. C'est le cas pour l'utilisation de la table de la loi normale que nous traiterons dans le prochain chapitre.

Combinaisons de plusieurs variables aléatoires

1. Somme et différence.
   Dans tous les cas,

$$\begin{aligned}E(X+Y) &= E(X) + E(Y), \\ E(X-Y) &= E(X) - E(Y).\end{aligned}$$

Dans le cas de variables *indépendantes* :

$$\begin{aligned} Var(X+Y) &= Var(X)+Var(Y), \\ Var(X-Y) &= Var(X)+Var(Y). \end{aligned}$$

2. Produit. Dans le cas de variables *indépendantes*,

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

3. Conséquence. Dans le cas de variables *indépendantes*,

$$\xi_{X+Y} = \xi_X \xi_Y.$$

Ceci montre que, dans le cas de variables continues indépendantes, la densité de probabilité de $X+Y$ est le *produit de convolution* des densités de probabilité de X et de Y.

Covariance de deux variables aléatoires

Lorsque deux variables aléatoires ne sont pas indépendantes, il existe une caractéristique qui permet de déterminer l'intensité de leur dépendance. C'est la covariance.

Définition 57 *La covariance de deux variables aléatoires X et Y est définie par*

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

Proposition 58 *1. Si deux variables aléatoires sont indépendantes, leur covariance est nulle.*

2. Attention : La réciproque n'est pas vraie. Deux variables de covariance nulle ne sont pas obligatoirement indépendantes.

3. Si deux variables aléatoires sont dépendantes,

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

Chapitre 3

Principales distributions de probabilités

Introduction

De nombreuses situations pratiques peuvent être modélisées à l'aide de variables aléatoires qui sont régies par des lois spécifiques. Il importe donc d'étudier ces modèles probabilistes qui pourront nous permettre par la suite d'analyser les fluctuations de certains phénomènes en évaluant, par exemple, les probabilités que tel événement ou tel résultat soit observé.

La connaissance de ces lois théoriques possède plusieurs avantages sur le plan pratique :

- Les observations d'un phénomène particulier peuvent être remplacées par l'expression analytique de la loi où figure un nombre restreint de paramètres (1 ou 2, rarement plus).
- La loi théorique agit comme modèle (idéalisation) et permet ainsi de réduire les irrégularités de la distribution empirique. Ces irrégularités sont souvent inexplicables et proviennent de fluctuations d'échantillonnage, d'imprécision d'appareils de mesure ou de tout autre facteur incontrôlé ou incontrôlable.
- Des tables de probabilités ont été élaborées pour les lois les plus importantes. Elles simplifient considérablement les calculs.

Ce cours présente trois distributions discrètes : la distribution binomiale, la distribution géométrique et la distribution de Poisson. Puis il aborde deux distributions continues : la distribution exponentielle et la distribution normale. Il importe de bien comprendre quelles sont les situations concrètes que l'on peut modéliser à l'aide de ces distributions. Viennent enfin trois distributions théoriques dont la fonction n'est pas de modéliser mais de servir d'outils dans les problèmes d'estimation et de test.

3.1 Distribution binomiale

3.1.1 Variable de Bernoulli ou variable indicatrice

Définition

Définition 59 *Une variable aléatoire discrète qui ne prend que les valeurs 1 et 0 avec les probabilités respectives p et $q = 1 - p$ est appelée variable de Bernoulli.*

Exemple 60 *Une urne contient deux boules rouges et trois boules vertes. On tire une boule de l'urne. La variable aléatoire X = nombre de boules rouges tirées est une variable de Bernoulli. On a : $p(X = 1) = 2/5 = p$, $p(X = 0) = 3/5 = q$.*

Plus généralement, on utilisera une variable de Bernoulli lorsqu'on effectue une épreuve qui n'a que deux issues : le succès ou l'échec. Une telle expérience est alors appelée épreuve de Bernoulli. On affecte alors 1 à la variable en cas de succès et 0 en cas d'échec.

Distribution de probabilités

x	0	1
$f(x) = p(X = x)$	q	p

Paramètres de la distribution

On calcule

$$\begin{aligned} E(X) &= 0.q + 1.p = p, \\ V(X) &= E(X^2) - E(X)^2 = (0^2 q + 1^2 p) - p^2 = p - p^2 = pq, \\ \xi_X(u) &= E(e^{-2i\pi u X}) = 1.q + e^{-2i\pi u} p = q + p\cos(2\pi u) + ip\sin(2\pi u). \end{aligned}$$

$E(X) = p$	$V(X) = pq$	$\sigma(X) = \sqrt{pq}$	$\xi_X(u) = q + pe^{-2i\pi u}$

3.1.2 Distribution binomiale

Situation concrète

a) On effectue une épreuve de Bernoulli. Elle n'a donc que deux issues : le succès avec une probabilité p ou l'échec avec une probabilité q.

b) On répète n fois cette épreuve.

c) Les n épreuves sont indépendantes entre elles, ce qui signifie que la probabilité de réalisation de l'événement "succès" est la même à chaque épreuve et est toujours égale à p.

Dans cette situation, on s'intéresse à la variable X = "nombre de succès au cours des n épreuves".

Distribution de probabilités

Appelons X_i les variables de Bernoulli associées à chaque épreuve. Si la i-ème épreuve donne un succès, X_i vaut 1. Dans le cas contraire X_i vaut 0. La somme de ces variables comptabilise donc le nombre de succès au cours des n épreuves. On a donc $X = X_1 + X_2 + \cdots + X_n$. X peut prendre $n + 1$ valeurs : $0, 1, \ldots, n$.

Cherchons la probabilité d'obtenir k succès, c'est-à-dire $p(X = k)$.

La probabilité d'avoir k succès suivis de $n - k$ échecs est $p^k q^{n-k}$ car ces résultats sont indépendants les uns des autres.

La probabilité d'avoir k succès et $n - k$ échecs dans un autre ordre de réalisation est toujours $p^k q^{n-k}$. Donc tous les événements élémentaires qui composent l'événement $(X = k)$ ont même probabilité.

Combien y en a-t-il ' Autant que de façons d'ordonner les k succès par rapport aux $n-k$ échecs ' Il suffit de choisir les k places des succès parmi les n possibles et les $n - k$ échecs prendront les places restantes. Or il y a $\binom{k}{n}$ manières de choisir k places parmi n.

Finalement, on obtient

$$p(X = k) = \binom{k}{n} p^k q^{n-k}.$$

On dit que la variable aléatoire X suit une *loi binomiale de paramètres n et p*. On note $X \hookrightarrow B(n, p)$.

Remarque : L'adjectif binomial vient du fait que lorsqu'on somme toutes ces probabilités, on retrouve le développement du binôme de Newton,

$$\sum_{k=0}^{n} \binom{k}{n} p^k q^{n-k} = (p + q)^n = 1.$$

NB : La loi binomiale est tabulée en fonction des 2 paramètres n et p.

Paramètres descriptifs de la distribution

Nous savons que $X = X_1 + \cdots + X_n$ avec $E(X_i) = p$ pour $i = 1, 2, \ldots, n$, donc $E(X) = E(X_1) + \cdots + E(X_n) = np$.

Les variables X_i sont indépendantes et $Var(X_i) = pq$ pour $i = 1, 2, \ldots, n$, donc $Var(X) = Var(X_1) + \cdots + Var(X_n) = npq$. D'autre part, les fonctions caractéristiques se multiplient, donc $\xi_X(u) = (q + pe^{-2i\pi u})^n$.

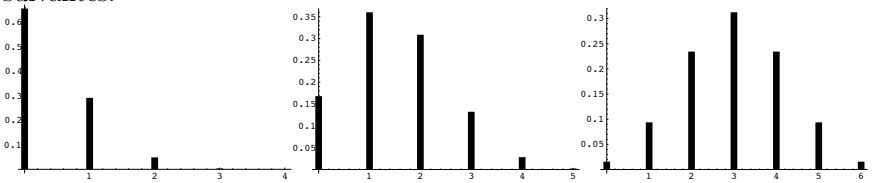
$E(X) = np$	$V(X) = npq$	$\sigma(X) = \sqrt{npq}$	$\xi_X(u) = (q + pe^{-2i\pi u})^n$

Remarque 61 *La formule donnant l'espérance semble assez naturelle. En effet, le nombre moyen de succès (qui correspond à la signification de l'espérance) est intuitivement égal au produit du nombre d'essais par la probabilité de réalisation d'un succès.*

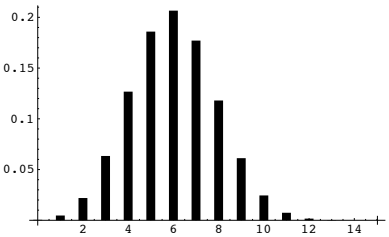
Propriétés de la distribution binomiale

Forme de la distribution binomiale

La représentation graphique de la distribution de la loi binomiale est habituellement présentée sous la forme d'un diagramme en bâtons. Puisque la loi dépend de n et p, nous aurons diverses représentations graphiques si nous faisons varier n et/ou p comme c'est le cas pour les figures suivantes.



$n = 4$, $p = 0.1$ $n = 5$, $p = 0.3$ $n = 6$, $p = 0.5$



$n = 15$, $p = 0.4$

On peut effectuer plusieurs remarques à propos de ces diagrammes.

a) La forme de la distribution est symétrique si $p = 1/2$, quelque soit n.

b) Elle est dissymétrique dans le cas où $p \neq 1/2$. Si p est inférieur à 0.50, les probabilités sont plus élevées du côté gauche de la distribution que du côté droit (asymétrie positive). Si p est supérieur à 1/2, c'est l'inverse (asymétrie négative).

c) La distribution tend à devenir symétrique lorsque n est grand. De plus, si p n'est pas trop voisin de 0 ou 1, elle s'approchera de la distribution de la loi normale que l'on verra plus loin dans ce chapitre.

Somme de deux variables binomiales

Si X_1 et X_2 sont des variables *indépendantes* qui suivent des lois binomiales $B(n_1, p)$ et $B(n_2, p)$ respectivement, alors $X_1 + X_2$ suit une loi binomiale $B(n_1 + n_2, p)$.

Cette propriété s'interprète facilement : si X_1 représente le nombre de succès en n_1 épreuves identiques indépendantes et X_2 en n_2 épreuves indépendantes entre elles et indépendantes des premières avec la même probabilité de succès que les premières, alors $X_1 + X_2$ représente le nombre de succès en $n_1 + n_2$ épreuves identiques et indépendantes.

3.2 Distribution géométrique

3.2.1 Situation concrète

a) On effectue une épreuve de Bernoulli. Elle n'a donc que deux issues : le succès avec une probabilité p ou l'échec avec une probabilité $q = 1 - p$.

b) On répète l'épreuve jusqu'à l'apparition du premier succès.

c) Toutes les épreuves sont indépendantes entre elles.

Dans cette situation, on s'intéresse à la variable $X =$ "nombre de fois qu'il faut répéter l'épreuve pour obtenir le premier succès".

Remarque 62 *On est donc dans les mêmes hypothèses que pour la loi binomiale, mais le nombre d'épreuves n'est pas fixé à l'avance. On s'arrête au premier succès.*

3.2.2 Distribution de probabilités

L'ensemble des valeurs prises par X est $1, 2, 3, \ldots$. On cherche la probabilité d'avoir recours à n épreuves pour obtenir le premier succès.

Ce succès a une probabilité de réalisation de p. Puisque c'est le premier, il a été précédé de $n - 1$ échecs qui ont chacun eu la probabilité q de se produire. Étant donné l'indépendance des épreuves, on peut dire que la probabilité de réalisation de $n - 1$ échecs suivis d'un succès est le produit des probabilités de réalisation de chacun des résultats,

$$p(X = n) = q^{n-1}p.$$

On dit que la variable aléatoire X suit une *loi géométrique de paramètre p*. On note $X \hookrightarrow G(p)$.

Remarque 63 *L'appellation* géométrique *vient du fait qu'en sommant toutes les probabilités, on obtient une série géométrique. En effet,*

$$\sum_{n=1}^{+\infty} q^{n-1}p = \frac{p}{1-q} = 1.$$

3.2.3 Paramètres descriptifs de la distribution

On calcule

$$\begin{aligned} \xi_X(u) &= \sum_{n=1}^{\infty} q^{n-1}pe^{-2i\pi un} \\ &= pe^{-2i\pi u}\sum_{k=0}^{\infty} q^k e^{-2i\pi uk} \\ &= \frac{pe^{-2i\pi u}}{1-qe^{-2i\pi u}}, \end{aligned}$$

et on en tire, en dérivant par rapport à u en $u = 0$, l'espérance et la variance.

$E(X) = 1/p$	$Var(X) = q/p^2$	$\sigma(X) = \sqrt{q}/p$	$\xi_X(u) = \frac{pe^{-2i\pi u}}{1-qe^{-2i\pi u}}$

Remarque 64 *On peut interpréter l'expression de l'espérance de façon intuitive. En effet en n épreuves, on s'attend à obtenir np succès et par conséquent, le nombre moyen d'épreuves entre deux succès devrait être* $\frac{n}{np} = \frac{1}{p}$.

3.2.4 Propriété remarquable de la distribution géométrique

La propriété la plus importante de la loi géométrique est sans doute d'être *sans mémoire*. En effet, la loi de probabilité du nombre d'épreuves à répéter jusqu'à l'obtention d'un premier succès dans une suite d'épreuves de Bernoulli identiques indépendantes est la même quel que soit le nombre d'échecs accumulés auparavant. Mathématiquement, cela se traduit par

$$p(X > n + m|X > n) = p(X > m).$$

On comprend intuitivement que cela découle de l'indépendance des épreuves qui sont toutes identiques. La loi géométrique est la seule distribution de probabilité discrète qui possède cette propriété. En effet, si une variable aléatoire Y à valeurs dans $\mathbf{N}$ satisfait, pour tous n, $m \in \mathbf{N}$, $p(Y > n+m|Y > n) = p(Y > m)$, alors $p(Y > m+n) = p(Y > n)p(Y > m)$. Posant $q = p(Y > 1)$, il vient $p(Y > n) = q^n$, d'où $p(Y = n) = p(Y > n-1) - p(Y > n) = q^{n-1} - q^n = q^{n-1}(1-q)$.

3.3 Distribution de Poisson

La loi de Poisson est attribuée à Simeon D. Poisson, mathématicien français (1781-1840). Cette loi fut proposée par Poisson dans un ouvrage qu'il publia en 1837 sous le titre "Recherche sur la probabilité de jugements en matière criminelle et en matière civile".

3.3.1 Situation concrète

Beaucoup de situations sont liées à l'étude de la réalisation d'un événement dans un intervalle de temps donné (arrivée de clients qui se présentent à un guichet d'une banque en une heure, apparitions de pannes d'un réseau informatique en une année, arrivée de malades aux urgences d'un hôpital en une nuit,....). Les phénomènes ainsi étudiés sont des phénomènes d'attente.

Pour décrire les réalisations dans le temps d'un événement donné, on peut

- soit chercher le nombre de réalisations de l'événement dans un intervalle de temps donné qui est distribué suivant une loi de Poisson.
- soit chercher le temps entre deux réalisations successives de l'événement qui est distribué suivant une loi exponentielle (voir section suivante).

On va voir que la première loi (loi de Poisson) peut être interprétée comme un cas limite d'une loi binomiale et la seconde comme un cas limite d'une loi géométrique.

3.3.2 Processus de Poisson

Précisons les hypothèses faites relativement à la réalisation de l'événement qui nous intéresse.

1. Les nombres de réalisations de l'événement au cours d'intervalles de temps disjoints sont des variables aléatoires indépendantes, c'est-à-dire que le nombre de réalisations au cours d'un intervalle de temps est indépendant du nombre de réalisations au cours d'intervalles de temps antérieurs.
2. La probabilité pour que l'événement se réalise une fois, au cours d'un petit intervalle de temps Δt, est proportionnelle à l'amplitude de l'intervalle et vaut $\alpha\Delta t$, où α est une valeur positive que l'on suppose constante tout au long de la période d'observation.
3. Il est très rare d'observer plus d'une fois l'événement au cours d'un petit intervalle de temps Δt, c'est-à-dire que la probabilité pour que l'événement se réalise plus d'une fois au cours de l'intervalle de temps Δt est négligeable.

Les hypothèses 1., 2., 3. caractérisent ce qu'on appelle un *processus de Poisson*. α est une constante du processus qui représente le nombre moyen de réalisations par unité de temps et que l'on appelle l'*intensité* du processus.

Sous ces hypothèses, la variable aléatoire X = "nombre de fois où l'événement considéré se réalise au cours d'un intervalle de temps de durée t" est distribuée suivant une loi de Poisson de paramètre $\lambda = \alpha t$.

3.3.3 Distribution de probabilités

Nous cherchons à déterminer la loi de probabilité de la variable X = "nombre de réalisations d'un événement donné pendant un intervalle de temps t", sachant que le nombre moyen de réalisations de cet événement par unité de temps est α. Or, nous connaissons déjà la loi de probabilités de la variable Y = "nombre de réalisations d'un événement donné, de probabilité p, au cours de n essais". Il s'agit d'une loi binomiale $B(n,p)$.

Pour comprendre la relation entre ces deux lois, divisons l'intervalle de temps de longueur t, en n petits intervalles de temps disjoints de longueur $\Delta t = t/n$ pour n assez grand.

L'hypothèse 3. permet d'affirmer que dans chacun de ces n petits intervalles il n'y a principalement que deux possibilités : l'événement se réalise une fois ou ne se réalise pas (cela sera d'autant plus vrai que n est grand). Dans chaque intervalle, la variable "nombre de réalisations de l'événement" est une variable de Bernoulli.

L'hypothèse 2. permet d'affirmer que dans chacun de ces n petits intervalles, la probabilité de réalisation de l'événement est constante et égale à $\alpha\Delta t = \alpha t/n$. Les variables de Bernoulli ont donc toutes le même paramètre $p = \alpha t/n$.

L'hypothèse 1. permet d'affirmer que les n variables de Bernoulli sont indépendantes.

La somme de ces n variables de Bernoulli indépendantes de même paramètre $p = \alpha t/n$ est une variable qui suit la loi binomiale $B(n, \alpha t/n)$ et qui représente approximativement le nombre de réalisations de l'événement dans l'intervalle de temps t. Si on choisit n de plus en plus grand, on a de plus en plus d'intervalles, la probabilité de réalisations de l'événement dans chaque intervalle est de plus en plus petite et la distribution $B(n, \alpha t/n)$ se rapproche de plus en plus de la distribution que l'on cherche à déterminer, c'est-à-dire de la distribution de Poisson de paramètre αt. On conclut

Définition 65 *On peut considérer la loi de Poisson de paramètre λ comme la loi limite d'une loi binomiale $B(n, \lambda/n)$ lorsque n tend vers l'infini, le produit des paramètres $n.\lambda/n$ restant toujours constant égal à λ.*

On écrit $X \hookrightarrow P(\lambda)$.

Proposition 66 *La loi de Poisson de paramètre λ est donnée par*

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Preuve. Si Y suit une loi $B(n, \lambda/n)$, on sait que

$$\begin{aligned} P(Y = k) &= \binom{k}{n} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} \\ &= (1 - \frac{\lambda}{n})^{n-k}\frac{\lambda^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k} \\ &= (1 - \frac{\lambda}{n})^{n}\frac{\lambda^k}{k!}(1 - \frac{\lambda}{n})^{-k}[\frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-k+1}{n}]. \end{aligned}$$

Chaque terme du produit entre crochets tend vers 1 lorsque n tend vers l'infini. Il y a k termes, c'est-à-dire un nombre fini. Donc le crochet tend vers 1. De même, $(1 - \frac{\lambda}{n})^{-k}$ tend vers 1. De plus,

$$\ell n((1 - \frac{\lambda}{n})^n) = n\ell n(1 - \frac{\lambda}{n}) \sim n \times (-\frac{\lambda}{n})$$

tend vers $-\lambda$ lorsque n tend vers l'infini, donc $(1 - \frac{\lambda}{n})^n$ tend vers $e^{-\lambda}$. On conclut que $P(Y = k)$ tend vers $e^{-\lambda}\lambda^k/k!$.

Remarque 67 *Il existe des tables donnant la fonction de densité et la fonction de répartition de la loi de Poisson en fonction des différentes valeurs de λ (pour $\lambda \leq 15$).*

3.3.4 Paramètres descriptifs de la distribution

On calcule, lorsque $X \hookrightarrow P(\lambda)$,

$$\xi_X(u) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{-2i\pi uk} = e^{-\lambda} e^{\lambda e^{-2i\pi u}}.$$

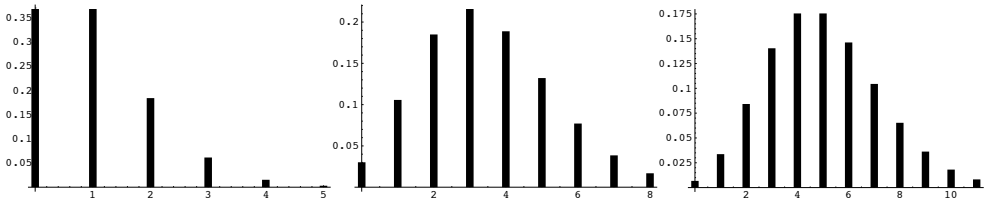
On en déduit le tableau

$E(X) = \lambda$	$Var(X) = \lambda$	$\sigma(X) = \sqrt{\lambda}$	$\xi_X(u) = e^{\lambda(e^{-2i\pi u}-1)}$

La loi $P(\lambda)$ est la loi limite de la loi $B(n, \lambda/n)$ lorsque n tend vers l'infini. On constate que l'espérance mathématique et la variance de la loi $B(n, \lambda/n)$ convergent vers celles de la loi $P(\lambda)$ lorsque n tend vers l'infini. Cela peut se vérifier directement, en appliquant le théorème de convergence dominée pour la mesure Peigne de Dirac (interversion d'une sommation et d'une limite).

3.3.5 Propriétés de la distribution de Poisson

Allure de la distribution



$\lambda = 1$ $\lambda = 3.5$ $\lambda = 5$

Commentaires.
- En général, le diagramme est dissymétrique par rapport à λ avec étalement vers la droite. Les valeurs élevées d'une variable de Poisson sont peu rencontrées.
- A mesure que λ augmente, la forme de la distribution tend à devenir symétrique et s'approche de celle de la loi normale que nous traiterons plus loin dans ce chapitre. Cela est vérifié pour $\lambda \geq 10$ et même acceptable pour $\lambda \geq 5$.

Approximation de la loi binomiale par la loi de Poisson

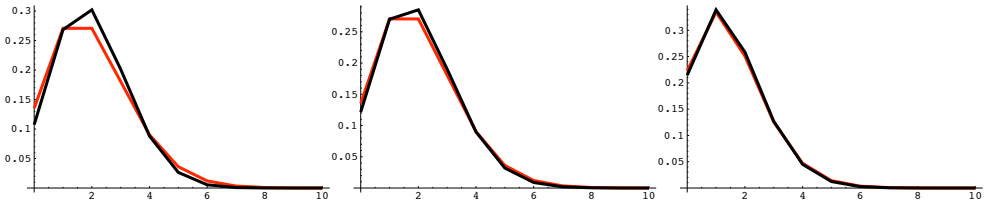
La loi binomiale dépend de deux paramètres n et p. Bien qu'il existe quelques tables, elle est n'est pas simple à utiliser. La loi de Poisson ne dépend que d'un paramètre ce qui la rend plus pratique. Il faut donc avoir toujours présent à l'esprit que, lorsque les conditions le permettent, on peut avoir intérêt à remplacer une loi binomiale par une loi de Poisson.

Lorsque n est grand et p petit, de telle façon que le produit $np = \lambda$ reste petit par rapport à n, la loi binomiale $B(n, p)$ peut être approchée par la loi de Poisson $P(\lambda)$ (revoir ce qui a été dit sur ce sujet dans le paragraphe "Distribution de probabilités"). Cette approximation s'appliquant lorsque p est petit, la loi de Poisson est appelée la loi des événements rares. En pratique, l'approximation est valable si $n > 20$, $p \leq 0.1$ et $np \leq 5$.

On approche la loi $B(n,p)$ par la loi $P(np)$ dès que $n > 20$, $p \leq 0.1$ et $np \leq 5$.

RÈGLE IMPORTANTE. Lorsqu'on approche une loi par une autre, on choisit le ou les paramètres de la loi approchante de manière que l'espérance (et la variance lorsqu'on a suffisamment de paramètres) de la loi approchante soit égale à l'espérance (et la variance) de la loi approchée.

Comparaison des distributions.



En rouge	$B(10, 0.2)$	$B(20, 0.1)$	$B(30, 0.05)$
En noir	$P(2)$	$P(2)$	$P(1.5)$
	mauvaise approximation	approximation acceptable	bonne approximation

Somme de deux lois de Poisson

Si X_1 et X_2 sont des variables aléatoires *indépendantes* qui suivent des lois de Poisson de paramètres respectifs λ_1 et λ_2, alors $X_1 + X_2$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

3.3.6 Importance pratique de la loi de Poisson

Le physicien et le technologue rencontrent la loi de Poisson dans de nombreuses circonstances.

Le comptage en radioactivité, microanalyse X, analyse SIMS...

Un détecteur compte le nombre de particules qu'il reçoit pendant un temps de comptage t_0. Ce nombre X obéit à une loi de Poisson dont le paramètre λ est le produit du flux moyen de particules α pendant le temps de comptage t_0.

Conséquence. L'espérance mathématique de X est αt_0 et l'écart-type sur X est $\sqrt{\alpha t_0}$. Il en résulte que la variabilité de X (quotient de l'espérance sur l'écart-type) devient très grande quand le nombre moyen de particules observé est petit. Si les valeurs prises par X sont proches de 10 correspondant à une espérance αt_0 proche de 10, l'écart-type est proche de 3, ce qui donne une variabilité de 30%. Ceci est parfaitement visible sur le "profil SIMS[1]" présenté ci-dessous. Le bruit devient important lorsque X est entre 10 et 15.

PROFIL S.I.M.S.

Dans la mesure du possible, l'expérimentateur sera alors amené à augmenter le temps de comptage pour augmenter les valeurs de X.

Contrôle de qualité

On effectue un contrôle de qualité sur des composants fabriqués par une unité de production. Un composant est réputé bon (ou mauvais) lorsqu'il satisfait (ou non) à des spécifications. Le critère de choix est donc qualitatif. Le taux moyen de rebuts (pièces mauvaises) est appelé p. On effectue un tirage de n pièces. La probabilité de trouver k pièces mauvaises dans cet échantillon de taille n est donnée par une loi de Poisson $P(np)$, car on peut approcher la loi binomiale par la loi de Poisson (*loi des événements rares*).

Lorsque np est petit, le rapport devient grand et la variabilité sur k rend le contrôle imprécis. Ceci explique que, dans un contrôle de qualité, la taille des échantillons tirés de la population des pièces fabriquées doit être au moins de l'ordre de 100.

Dénombrement d'événements survenant dans un espace délimité

La loi de Poisson permet de modéliser aussi bien le nombre d'événements survenant pendant un temps donné que le nombre d'événements survenant dans un espace délimité.

[1] S.I.M.S. (Secondary Ions Mass Spectroscopy) : méthode permettant une analyse quantitative des impuretés dans une masse solide. Un faisceau d'ions primaires abrase le matériau. Parmi les atomes arrachés à la cible, certains sont ionisés (ions secondaires). Ils sont triés en masse au moyen d'une déflexion magnétique et comptés pendant des intervalles de temps t_0 au cours du processus d'abrasion. On obtient ainsi le profil de concentration de l'impureté dans le matériau.

Par exemple, si on appelle X le nombre de particules bombardant une cible de surface S soumise à une irradiation de fluence F (mesurée en m^{-2}), alors X suit une loi $P(FS)$. FS est donc analogue au αt_0 du comptage en temps.

La loi de Poisson sert donc aussi à décrire des phénomènes de localisation spatiale et non plus seulement temporelle, c'est-à-dire qu'elle modélisera aussi bien le nombre d'accidents qui peuvent survenir en une matinée que le nombre d'accidents qui peuvent survenir sur une section donnée d'autoroute.

3.4 Distribution exponentielle

3.4.1 Situation concrète

On se place dans le cas d'un phénomène d'attente décrit au paragraphe 3 et on s'intéresse à la variable aléatoire qui représente le temps d'attente pour la réalisation d'un événement ou le temps d'attente entre la réalisation de deux événements successifs. Si on se place dans le cas où l'intensité α du processus de Poisson est constante, ce temps d'attente suit une loi exponentielle de paramètre α.

Exemple. Lorsque l'événement attendu est la mort d'un individu (ou la panne d'un équipement), α s'appelle le taux de mortalité (ou le taux de panne). Dire qu'il a une valeur constante, c'est supposer qu'il n'y a pas de vieillissement (ou pas d'usure s'il s'agit d'un équipement), la mort ou la panne intervenant de façon purement accidentelle.

3.4.2 Distribution de probabilité

On veut déterminer la loi de la variable T = "temps d'attente entre la réalisation de deux événements successifs" où le nombre moyen de réalisations de l'événement par unité de temps est α. Pour cela, nous allons procéder comme dans le paragraphe 3 : Si t est la longueur de l'intervalle de temps sur lequel dure notre étude, nous le divisons en n petits intervalles de longueur t/n. Appelons X la variable aléatoire représentant le nombre d'intervalles de temps que l'on doit laisser s'écouler pour obtenir la réalisation de l'événement suivant. Chaque intervalle possédant la même probabilité $\alpha t/n$ de voir l'événement se produire, X suit, par définition, la loi géométrique de paramètre $\alpha t/n$. Le temps d'attente T est alors le produit de ce nombre d'intervalles par le temps de chaque intervalle, $T = \alpha t/n \times X$.

On cherche $p(T > t_0) = p(X > nt_0/t)$, et lorsque n tend vers l'infini on obtient

$$p(T > t_0) = e^{-\alpha t_0} = \int_{t_0}^{+\infty} \alpha e^{-\alpha u}\, du.$$

Ceci nous permet d'affirmer que la fonction de densité de la variable T est $f(x) = \alpha e^{-\alpha x}$ si $x > 0$, 0 sinon.

Définition 68 *La loi exponentielle de paramètre α décrit la distribution d'une variable continue X qui ne prend que des valeurs positives selon la fonction de densité $f(x) = \alpha e^{-\alpha x}$. On la note $Exp(\alpha)$.*

3.4.3 Paramètre descriptifs de la distribution

On vient de voir que la loi exponentielle est la loi limite d'une loi géométrique. On a $T \sim \frac{t}{n}X$ où X suit la loi géométrique de paramètre $\frac{\alpha t}{n}$. Or $E(T) \sim \frac{t}{n}E(X) = \frac{t}{n}(\frac{\alpha t}{n})^{-1} = \frac{1}{\alpha}$ et $Var(T) \sim \frac{t^2}{n^2}Var(X) = \frac{t^2}{n^2}\frac{1-\frac{\alpha t}{n}}{(\frac{\alpha t}{n})^2} \sim \frac{1}{\alpha^2}$.

Proposition 69

$$E(T) = \frac{1}{\alpha}, \quad Var(T) = \frac{1}{\alpha^2}.$$

Remarque 70 *On peut très facilement retrouver ces résultats en effectuant un calcul direct à partir de la fonction de densité en utilisant les formules de définition de l'espérance et la variance (peut-être pourriez-vous le faire à titre d'exercice')*

3.4.4 Propriétés de la distribution exponentielle

1. Comme la loi géométrique, la loi exponentielle est sans mémoire. C'est la seule loi continue qui possède cette propriété. Elle provient bien entendu du fait que le paramètre α est constant.
2. Une somme de n variables indépendantes de même loi exponentielle de paramètre α n'est pas une variable de loi exponentielle mais une variable qui suit une loi gamma de paramètres n et α. Une telle loi est aussi appelée loi d'ERLANG d'ordre n. Elle représente le temps d'attente requis avant qu'un événement se réalise n fois .
3. La loi exponentielle est aussi couramment utilisée dans les problèmes de datation en géochronologie.

3.5 Distribution normale

Du point de vue historique, la nature et l'importance exceptionnelle de cette loi furent pressenties en 1773 par Abraham de Moivre lorsqu'il considéra la forme limite de la loi binomiale.

En 1772, Simon Laplace l'étudia dans sa théorie des erreurs. Mais c'est seulement en 1809 pour Carl Friedrich Gauss et en 1812 pour Simon Laplace qu'elle prit sa forme définitive. C'est ainsi qu'on l'appelle tantôt loi de Laplace, tantôt loi de Gauss, tantôt loi de Laplace-Gauss. On trouve aussi l'expression, consacrée par une longue tradition, de loi normale (ce qui ne signifie pas pour autant que les autres lois soient « anormales »). Elle jouit d'une importance fondamentale car un grand nombre de méthodes statistiques reposent sur elle. Ceci est lié au fait qu'elle intervient comme loi limite dans des conditions très générales.

Pour faire ressortir toute son importance et sa forme, W.J. Youden, du National Bureau of Standards, a eu l'ingénieuse idée de la présenter telle qu'elle apparaît ci-dessous.

La
loi normale
des erreurs
constitue l'une
des généralisations
les plus étendues de
la philosophie naturelle
dans l'histoire de l'humanité.
Elle est un outil précieux pour la
recherche en sciences physiques et
sociales ainsi qu'en médecine, en agriculture
et en génie. Elle est indispensable à l'analyse et à
l'interprétation des données obtenues par l'observation ou
l'expérience.

3.5.1 Situation concrète

On rencontre souvent des phénomènes complexes qui sont le résultat de causes nombreuses, d'effet faible, et plus ou moins indépendantes. Un exemple typique est celui de l'erreur commise sur la mesure d'une grandeur physique. Cette erreur résulte d'un grand nombre de facteurs tels que : variations incontrôlables de la température ou de la pression, turbulence atmosphérique, vibrations de l'appareil de mesure, etc...Chacun des facteurs a un effet faible, mais l'erreur résultante peut ne pas être négligeable. Deux mesures faites dans des conditions que l'expérimentateur considère comme identiques pourront alors donner des résultats différents.

Donc dès que nous seront dans une situation où la distribution dépend de causes

- en grand nombre et indépendantes,
- dont les effets s'additionnent,
- dont aucune n'est prépondérante,

alors nous serons en présence de la distribution normale. C'est le cas, par exemple :

- En métrologie, pour la distribution des erreurs d'observation.
- En météorologie, pour la distribution de phénomènes aléatoires tels que la température et la pression.
- En biologie, pour la distribution de caractères biométriques comme la taille ou le poids d'individus appartenant à une population homogène. En technologie, pour la distribution des cotes des pièces usinées.
- En économie, pour les fluctuations accidentelles d'une grandeur économique (production, ventes, ...) autour de sa tendance, etc.....

3.5.2 Distribution de probabilité

Définition 71 *Une variable aléatoire continue suit une loi normale si l'expression de sa fonction de densité de probabilités est de la forme :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}, \quad x \in \mathbf{R}.$$

La loi dépend des deux réels m et σ appelés paramètres de la loi normale. On la note $\mathcal{N}(m, \sigma)$.

Remarque 72

1. *Une fonction de densité de probabilité étant toujours positive, le paramètre σ est donc un réel strictement positif.*
2. *On démontre que f est bien une fonction de densité de probabilité car $\int_{\mathbf{R}} f(x)\,dx = 1$. Pour le démontrer on utilise que $\int_{\mathbf{R}} e^{-x^2/2}\,dx = \sqrt{2\pi}$ (c'est l'intégrale de Gauss).*
3. *La loi normale étant tabulée, cette expression nous sera de peu d'utilité. Il est important néanmoins de préciser à quoi correspondent m et σ.*

3.5.3 Paramètre descriptifs de la distribution

La fonction caractéristique d'une variable normale standard X vaut

$$\xi_X(u) = e^{-2i\pi mu - 2\pi^2\sigma^2u^2}.$$

On en déduit, à l'aide de la formule qui exprime espérance et variance à partir des dérivées de la fonction caractéristique, que

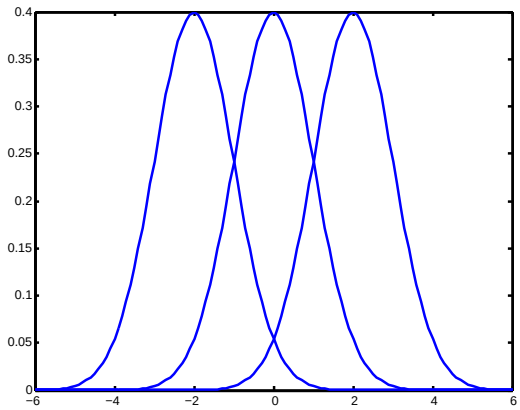
$$E(X) = m, \quad Var(X) = \sigma^2, \quad \sigma(X) = \sigma.$$

On peut aussi faire le calcul directement, à partir de l'intégrale de Gauss.

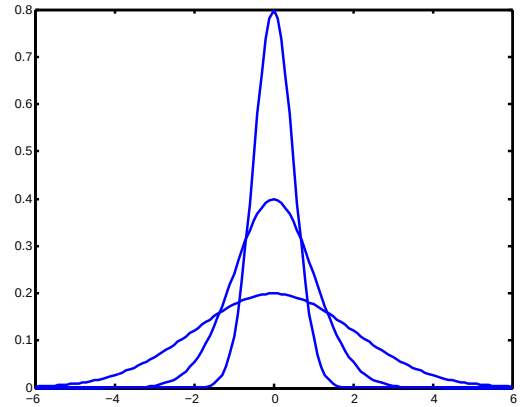
3.5.4 Propriétés de la distribution normale

Forme de la distribution normale

La fonction de densité de probabilités de la loi normale a la forme d'une « courbe en cloche ». En fait il ne s'agit pas d'une courbe unique mais plutôt d'une famille de courbes dépendant de m et σ.



Ecarts types identiques, espérances -2, 0, 2 différentes



Espérances identiques, écarts types 0.5, 1, 2 différents

On peut effectuer quelques remarques à propos de ces courbes.

a) La distribution est symétrique par rapport à la droite d'équation $x = m$. Donc l'aire sous la courbe de part et d'autre de cette droite est égale à 0.5.

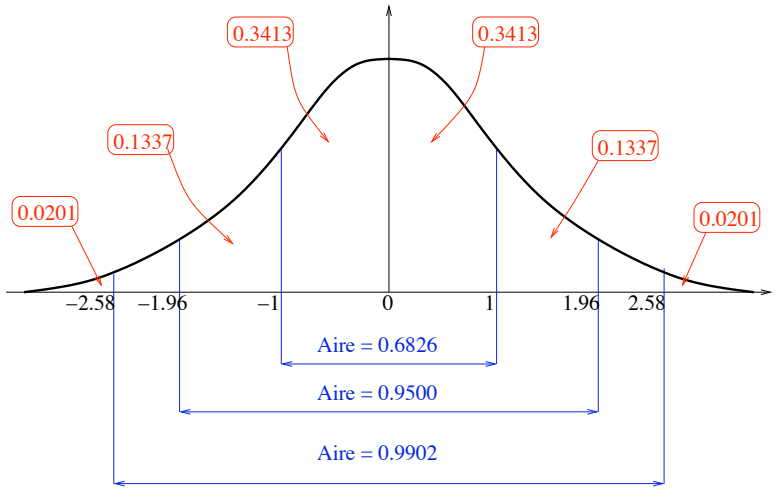
b) La distribution est d'autant plus étalée que σ est grand.

c) L'axe des abscisses est une asymptote et l'aire sous la courbe à l'extérieur de l'intervalle$[m - 3\sigma, m + 3\sigma]$ est négligeable.

Pour fixer les idées, on peut indiquer que

$$\begin{aligned} p(m - \sigma < X < m + \sigma) &= 0.6826 \\ p(m - 2\sigma < X < m + 2\sigma) &= 0.9544 \\ p(m - 3\sigma < X < m + 3\sigma) &= 0.9974. \end{aligned}$$

Cela peut être visualisé sur le graphique ci-après.



Loi $\mathcal{N}(0,1)$.

d) σ représente la différence des abscisses entre le sommet de la courbe et le point d'inflexion.

e) La longueur à mi-hauteur de la courbe (L.M.H. ou en anglais F.W.H.M. Full Width Half Maximum) vaut 2.35σ. Cette distance est souvent employée par le spectroscopiste pour déterminer expérimentalement σ. Cette méthode doit cependant être utilisée avec précaution car il faut s'assurer que les "bruits" permettent d'observer correctement le "pied" de la courbe.

Somme de deux variables normales

Soient X_1 et X_2 deux variables indépendantes. Si X_1 suit $\mathcal{N}(m_1, \sigma_1)$ et X_2 suit $\mathcal{N}(m_2, \sigma_2)$, alors $X_1 + X_2$ suit $\mathcal{N}(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Loi normale centrée réduite ou loi normale standardisée

Nous avons vu dans le chapitre 2 qu'à toute variable aléatoire X ,on pouvait associer une variable dite standardisée $\dfrac{X - E(X)}{\sigma(X)}$ d'espérance nulle et de variance unité (ceci résultait des propriétés de translation et de changement d'échelle).

On montre assez facilement que si on effectue cette transformation sur une variable suivant une loi normale, la variable standardisée suit encore une loi normale mais cette fois-ci de paramètres 0 et 1. La loi standardisée est appelée loi normale centrée réduite, et notée $\mathcal{N}(0,1)$. Donc si X suit $\mathcal{N}(m, \sigma)$, on pose $T = \dfrac{X - m}{\sigma}$ et T suit $\mathcal{N}(0,1)$.

On peut résumer la correspondance de la façon suivante :

$X \rightharpoonup \mathcal{N}(m, \sigma)$ $E(X) = m$ $Var(X) = \sigma^2$	$T = \frac{X-m}{\sigma}$	$T \rightharpoonup \mathcal{N}(0,1)$ $E(T) = 0$ $Var(T) = 1$

Il faut garder à l'esprit que concrètement T est le nombre d'écarts-type entre la valeur de X et la moyenne.

La loi $\mathcal{N}(0,1)$ est tabulée à l'aide la fonction de répartition des valeurs positives. Elle donne les valeurs de $\Phi(t) = p(0 \leq T \leq t) = \int_0^t \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$ pour $t > 0$. Ce nombre représente l'aire sous la courbe représentative de la distribution et au dessus de l'intervalle $[0, t]$. Pour cette raison la table de la loi normale est aussi appelée table d'aires. Cette table ne dépend d'aucun paramètre, mais permet cependant de déterminer les probabilités de n'importe quelle distribution normale !

Comment utiliser la table d'aires ?

La première colonne de la table indique les unités et les dixièmes des valeurs de T alors que les centièmes des valeurs de T se lisent sur la ligne supérieure de la table. La valeur trouvée à l'intersection de la ligne et de la colonne adéquates donne l'aire cherchée.

a) Je cherche la valeur de A l'intersection de la ligne "0.5" et de la colonne "0.00", je lis 0.1915.

b) Je cherche la valeur de $p(-0.5 \leq T \leq 0)$. J'utilise la symétrie de la courbe par rapport à l'axe des ordonnées et j'en conclus que $p(-0.5 \leq T \leq 0) = p(0 \leq T \leq 0.5) = 0.1915$. Et que pensez-vous de la valeur de $p(-0.5 < T < 0)$?

c) Je cherche la valeur de $p(-2.24 \leq T \leq 1.12)$. L'aire cherchée correspond à la somme suivante

$$p(-2.24 \leq T \leq 1.12) = p(-2.24 \leq T \leq 0) + p(0 < T \leq 1.12) = 0.4875 + 0.3686 = 0.8561.$$

d) Je cherche la valeur de $p(1 \leq T \leq 2)$. L'aire cherchée correspond à la différence suivante

$$p(1 \leq T \leq 2) = p(0 \leq T \leq 2) - p(0 \leq T \leq 1) = 0.4772 - 0.3413 = 0.1359.$$

e) Je cherche la valeur t de T telle que $p(0 \leq T \leq t) = 0.4750$. C'est le problème inverse de celui des exemples précédents. Il s'agit de localiser dans la table l'aire donnée et de déterminer la valeur de T correspondante. Je trouve $t = 1.96$.

Remarque 73 *Si la valeur de l'aire ne peut se lire directement dans les valeurs de la table, on pourra toujours effectuer une interpolation linéaire entre deux valeurs adjacentes ou prendre la valeur la plus proche.*

3.6 Approximation par des lois normales

3.6.1 Théorème central limite (ou de tendance normale)

Théorème 3 *Hypothèses : Soit une suite de variables aléatoires $X_1, X_2, \ldots, X_n$ vérifiant les conditions suivantes :*

1. *Les variables sont indépendantes.*
2. *Leurs espérances mathématiques $m_1, m_2, \ldots, m_n$ et leurs variances $Var(X_1)$, $Var(X_2)$,... $Var(X_n)$ existent toutes.*
3. $\lim_{n\to\infty} \dfrac{\sup_k Var(X_k)}{\sum_{j=1}^{n} Var(X_j)} = 0.$

Conclusion : La distribution de la variable somme $X = X_1 + X_2 + \cdots + X_n$ se rapproche de la distribution normale lorsque n tend vers l'infini.

Idée de la preuve, dans le cas particulier où les variables X_k ont toutes la même loi.

Quitte à standardiser, on peut supposer que X_k est d'espérance nulle et de variance 1. Alors sa fonction caractéristique a un développement limité, au voisinage de $u = 0$, de la forme

$$\xi_{X_k}(u) = 1 - 2\pi^2 u^2 + \cdots$$

Posons

$$Y = \frac{1}{\sqrt{n}} X = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

Alors

$$\begin{aligned} \xi_Y(u) &= \xi_X(\frac{u}{\sqrt{n}}) \\ &= \xi_{X_k}(\frac{u}{\sqrt{n}})^n \\ &= (1 - \frac{2\pi^2 u^2}{n} + \cdots)^n \\ &\sim e^{-2\pi^2 u^2}, \end{aligned}$$

qui est la fonction caractéristique de la distribution normale standard $\mathcal{N}(0,1)$.

Remarque 74 *C'est ce théorème très important qui nous permet d'affirmer que la situation concrète énoncée au début de ce paragraphe nous met en présence d'une loi normale. En effet,*
- *$X_1, X_2, \ldots, X_n$ correspondent aux différents facteurs de fluctuations.*
- *Le grand nombre de causes est assuré par le fait que n tend vers l'infini.*
- *L'indépendance parle d'elle-même.*
- *Additionner les effets revient à considérer la variable somme.*
- *Dire qu'aucun facteur n'est prépondérant est traduit par l'hypothèse (3) du théorème.*

En pratique, ceci se vérifie dès que $n \geq 30$.

3.6.2 Approximation de la loi binomiale par la loi normale

Une variable qui suit une loi binomiale $\mathcal{B}(n,p)$ peut toujours être considérée comme une somme de n variables de Bernoulli indépendantes de même paramètre p.

$$X = X_1 + \cdots + X_n,$$

où X_i sont des variables de Bernoulli. Les hypothèses du théorème centrale limite étant vérifiées, on peut affirmer que, lorsque n tend vers l'infini, la loi binomiale $\mathcal{B}(n,p)$ tend vers une loi normale. La loi normale qui l'approche le mieux est celle qui possède la même espérance np et le même écart-type $\sqrt{npq}$, $q = 1-p$.

Or la distribution binomiale est asymétrique sauf lorsque $p = 1/2$. La distribution normale, elle, est symétrique. L'approximation sera valable lorsque p n'est pas trop voisin de 0 ou 1 et sera d'autant meilleure que p est proche de $1/2$ et que n est grand.

En pratique :

On approche la loi $\mathcal{B}(n,p)$ par la loi $\mathcal{N}(np, \sqrt{npq})$ dès que $\begin{cases} n \geq 30 \\ np \geq 15 \\ nq \geq 15 \end{cases}$

3.6.3 La correction de continuïté

Cette approximation pose deux problèmes.

1. *On remplace une distribution concernant un nombre fini de valeurs par une distribution sur* **R** *tout entier.*
   Étant donné qu'à l'extérieur de l'intervalle $[m-3\sigma, m+3\sigma]$ la distribution normale est presque nulle, cela ne pose pas de problèmes.
2. *On remplace une distribution discrète par une distribution continue.*
   Il nous faut donc appliquer ce qu'on appelle une *correction de continuité*. Si on nomme X la variable binomiale et Y la variable normale, on remplacera une valeur k de X par un intervalle de Y centré sur k et d'amplitude 1, ce qui signifie que l'on écrit

   $$p(X = k) \simeq p(k - \frac{1}{2} < Y < k + \frac{1}{2}).$$

   Dans la pratique lorsque n est très grand, cette correction n'est pas nécessaire. On l'effectuera cependant si on souhaite une grande précision.

Remarque 75 *Remplacer une loi binomiale par une loi normale simplifie considérablement les calculs.*

En effet les tables de la loi binomiale dépendent de deux paramètres et les valeurs de n dans ces tables sont limitées supérieurement par 20. La loi normale, elle, après standardisation ne dépend d'aucun paramètre .

3.6.4 Approximation de la loi de Poisson par la loi normale

On démontre qu'on peut aussi approcher la loi de Poisson par la loi normale pour les grandes valeurs du paramètre de la loi de Poisson. La seule qui puisse convenir est celle qui a même espérance et même variance. On approche donc la loi $\mathcal{P}(\lambda)$ par la loi $\mathcal{N}(\lambda, \sqrt{\lambda})$. En pratique, cela s'applique dès que $\lambda \geq 16$.

On approche la loi $\mathcal{P}(\lambda)$ par la loi $\mathcal{N}(\lambda, \sqrt{\lambda})$ dès que $\lambda \geq 16$

Remarque 76 *La loi de Poisson étant elle aussi une loi discrète, on peut avoir à appliquer la correction de continuité.*

3.7 Quelques conseils pour résoudre les problèmes

Voici, lorsqu'elle s'applique, une méthode de travail qui peut guider votre démarche.

1. Suite à l'énoncé du problème, identifier correctement à l'aide de mots la variable aléatoire que vous allez considérer.
2. Préciser les valeurs possibles que peut prendre cette variable.
3. Identifier correctement la loi de probabilité qu'elle suit en essayant de reconnaître dans le problème une situation type.
4. Déterminer les paramètres de la loi.
5. Utiliser les formules théoriques ou les tables pour déterminer les probabilités demandées. Face à de longs calculs et en l'absence de tables correspondant à vos ou votre paramètre, penser à approcher votre loi par une autre.

3.7.1 Quelques exercices types

Exercice 77 *Supposons qu'une tentative pour obtenir une communication téléphonique échoue (par exemple, parce que la ligne est occupée) avec la probabilité 0.25 et réussisse avec la probabilité 0.75. On suppose que les tentatives sont indépendantes les unes des autres. Quelle est la probabilité d'obtenir la communication si l'on peut effectuer trois tentatives au maximum ?*

Solution de l'exercice 77. *3 essais.*

Nous nous intéressons à la variable $X =$ « nombre de tentatives nécessaires pour obtenir la communication », ce que l'on peut considérer comme le nombre d'essais à faire pour obtenir le premier succès. X suit une loi géométrique de paramètre $p = 0.75$.

On cherche à déterminer $p(X \leq 3) = p(X = 1) + p(X = 2) + p(X = 3)$.

- On peut obtenir la communication au 1er essai. On a pour cela une probabilité $p(X = 1) = 0.75$.
- On peut obtenir la communication au 2ème essai. On a pour cela une probabilité $p(X = 2) = 0.25 \times 0.75 = 0.1875$.
- On peut obtenir la communication au 3ème essai. On a pour cela une probabilité $p(X = 3) = 0.252 \times 0.75 = 0.0469$.

Finalement la probabilité d'obtenir la communication en trois essais maximum est $0.75 + 0.1875 + 0.0469 = 0.9844$ soit 98.5 %.

Exercice 78 *Un fabricant de pièces de machine prétend qu'au plus 10% de ses pièces sont défectueuses. Un acheteur a besoin de 120 pièces. Pour disposer d'un nombre suffisant de bonnes pièces, il en commande 140. Si l'affirmation du fabricant est valable, quelle est la probabilité que l'acheteur reçoive au moins 120 bonnes pièces ?*

Solution de l'exercice 78. *Bonnes pièces.*

Appelons X la variable aléatoire correspondant au "nombre de bonnes pièces dans le lot de 140 pièces".

X prend ses valeurs entre 0 et 140. De plus pour chaque pièce, on n'a que deux éventualités : elle est bonne ou elle est défectueuse. La probabilité qu'une pièce soit défectueuse est 0.1. Par conséquent elle est bonne avec la probabilité 0.9. On est donc dans une situation type : X suit la loi binomiale $\mathcal{B}(140, 0.9)$ de paramètres $n = 140$ et $p = 0.9$.

On veut déterminer la probabilité que l'acheteur reçoive au moins 120 bonnes pièces sur les 140, soit $X \geq 120$. A priori, il nous faudrait calculer la somme des probabilités $p(X = 120) + p(X = 121) + \cdots + p(X = 140)$, ce qui serait épouvantablement long. On approxime donc la loi binomiale par une loi tabulée.

Comme $n \geq 30$, $np = 126 \geq 15$ et $nq = 14$, on pourra approcher la loi binomiale par une loi normale. On choisit la loi normale qui a la même espérance et le même écart-type. Donc X qui suit la loi $\mathcal{B}(140, 0.9)$ sera approchée par Y qui suit la loi $\mathcal{N}(126, 3.55)$.

Pour remplacer une loi discrète par une loi continue, il est préférable d'utiliser la correction de continuité,

$$p(X \geq 120) \simeq p(Y > 119.5).$$

On se ramène enfin à la loi normale centrée réduite. On pose $T = \frac{Y-126}{3.55}$, et

$$\begin{aligned} p(Y > 119.5) &= p(T > \frac{119.5 - 126}{3.55}) = p(T > -1.83) \\ &= p(T < 1.83) = 0.5 + \Phi(1.83) = 0.97. \end{aligned}$$

Conclusion : l'acheteur a 97 chances sur 100 de recevoir 120 bonnes pièces sur les 140 achetées.

Exercice 79 *Les statistiques antérieures d'une compagnie d'assurances permettent de prévoir qu'elle recevra en moyenne 300 réclamations durant l'année en cours. Quelle est la probabilité que la compagnie reçoive plus de 350 réclamations pendant l'année en cours ?*

Solution de l'exercice 79. *Réclamations.*

La variable X qui nous intéresse est le "nombre de réclamations reçues pendant une année". Il s'agit du nombre de réalisations d'un événement pendant un intervalle de temps donné. X suit donc une loi de Poisson. Le nombre moyen de réalisations dans une année est 300. Cette valeur moyenne est aussi le paramètre de la loi de Poisson. Donc X suit la loi $\mathcal{P}(300)$.

On cherche à déterminer $p(X > 350)$. Il n'y a pas de table de la loi de Poisson pour cette valeur du paramètre. Il nous faut donc approcher X qui suit la loi de Poisson $\mathcal{P}(300)$ par Y qui suit la loi normale de même espérance et de même écart-type, c'est-à-dire $\mathcal{N}(300, \sqrt{300})$.

Ici aussi, on remplace une loi discrète par une loi continue. Il faut donc appliquer la correction de continuité

$$p(X > 350) = p(X \geq 351) \simeq p(Y > 350.5).$$

On se ramène finalement à la loi normale centrée réduite. On pose $T = \frac{Y-300}{\sqrt{300}}$.

$$p(Y > 350.5) = p(T > \frac{350.5 - 300}{\sqrt{300}}) = p(T > 2.92) = 0.5 - \Phi(2.92) = 0.0017.$$

La compagnie d'assurances a donc 0.17% de chances de recevoir plus de 350 réclamations en un an.

Exercice 80 *Le nombre moyen de clients qui se présentent à la caisse d'un supermarché sur un intervalle de 5 minutes est de 10. Quelle est la probabilité qu'aucun client ne se présente à la caisse dans un intervalle de deux minutes (deux méthodes possibles) ?*

Solution de l'exercice 80. *Solution $n^0 1$.*

Considérons la variable aléatoire X = "nombre de clients se présentant à la caisse dans un intervalle de deux minutes". Nous reconnaissons une situation type et la variable X suit une loi de Poisson. Vu qu'en moyenne 10 clients se présentent en 5 mn, l'intensité α du processus est de 2 clients par minute, $\alpha = 2$. Or le paramètre de la loi de Poisson est αt_0, t_0 étant ici 2 minutes. D'où $\lambda = 4$.

On cherche à calculer $p(X = 0)$. D'après la formule du cours, $p(X = 0) = e^{-\lambda} = e^{-4} = 0.018$.

Solution de l'exercice 80. *Solution $n^0 2$.*

Considérons à présent la question sous un autre angle en s'intéressant au temps d'attente Y entre deux clients. Le cours nous dit que la loi suivie par une telle variable est une loi exponentielle. Son paramètre α est l'intensité du processus de Poisson soit ici $\alpha = 2$. Y suit donc la loi $Exp(2)$.

Sa fonction de densité est $2e^{-2x}$ pour $x > 0$ exprimé en minutes. On en déduit que

$$p(Y \geq 2) = \int_2^{+\infty} 2e^{-2x}\, dx = [-e^{-2x}]_2^{+\infty} = e^{-4} = 0.018.$$

3.8 Distributions dérivant du modèle gaussien

Les distributions que nous allons étudier sont importantes non pas pour représenter des modèles théoriques de séries statistiques comme les précédentes, mais en raison du rôle qu'elles jouent dans les problèmes d'estimation ou de tests que nous verrons par la suite. Pour l'instant leurs définitions peuvent sembler complexes, notamment parce que la notion de « degrés de liberté » n'a pas encore été précisée. Pour le moment, il importe simplement de connaître leur définition et de savoir lire les tables correspondantes.

3.8.1 La distribution du χ^2 de Pearson

Elle a été découverte en 1905 par le mathématicien britannique Karl Pearson (1857-1936) qui travailla également sur les problèmes de régression avec le généticien Sir Francis Galton. Cette distribution (qui se prononce khi-deux) est très importante pour tester l'ajustement d'une loi théorique à une distribution expérimentale (test du χ^2) et pour déterminer la loi de la variance d'un échantillon.

Définition 81 *Si $X_1, X_2, \ldots, X_n$ sont n variables aléatoires indépendantes qui suivent toute la loi normale centrée réduite, alors la quantité $X = X_1^2 + X_2^2 + \cdots + X_n^2$ est une variable aléatoire distribuée selon la loi du χ^2 à n degrés de liberté.*

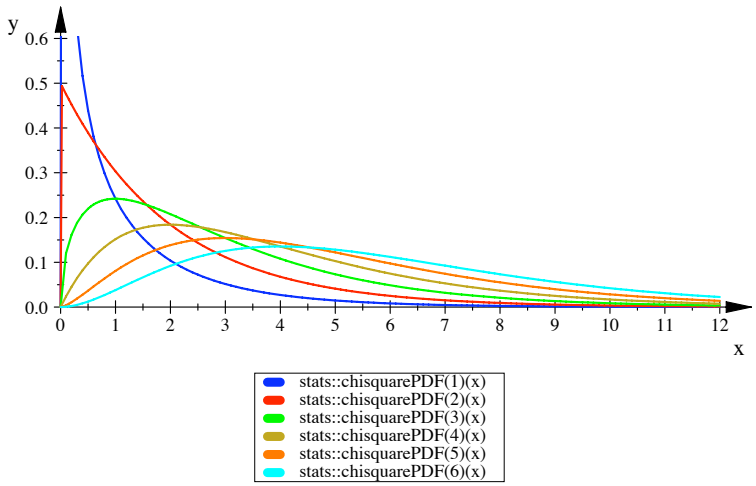
On note $X \rightharpoonup \chi_n^2$.

Forme de la distribution

L'expression de la densité de probabilités étant très compliquée et d'aucun intérêt pour nous, nous ne la donnons pas ici.

La distribution du χ^2 est continue à valeurs positives et présente un étalement sur le côté supérieur. Elle ne dépend que du nombre de degrés de liberté n.

Ci-dessous, densité de χ_n^2 pour $n = 1, \ldots, 6$.



Paramètres descriptifs

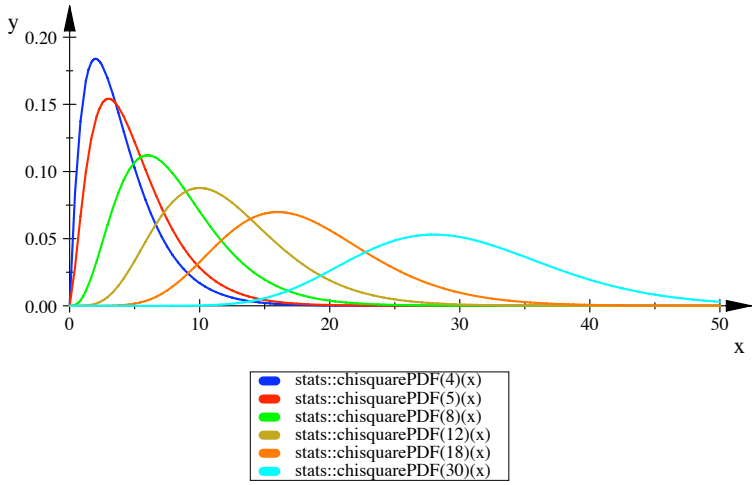
$$\mathrm{E(X)} = \mathrm{n}\ , \quad \mathrm{V(X)} = 2\mathrm{n}.$$

Somme de deux variables qui suivent une loi du χ^2

Si $X_1 \rightharpoonup \chi^2_{n_1}$ et $X_2 \rightharpoonup \chi^2_{n_2}$ sont indépendantes, alors $X_1 + X_2 \rightharpoonup \chi^2_{n_1+n_2}$.

Approximation par une loi normale

A mesure que n augmente, la loi du χ^2 tend vers la loi normale, comme on peut le constater sur le graphique ci-dessous.



Densité de χ^2_n pour $n = 4, 5, 8, 12, 18, 30$.

En pratique, on peut considérer que pour $n \geq 30$, on peut remplacer la loi du χ^2 à n degrés de liberté par la loi normale $\mathcal{N}(n, \sqrt{2n})$.

Utilisation de la table de Pearson

Pour des raisons de commodité, au lieu de donner la table des fonctions de répartition des variables aléatoires χ^2_n pour les différentes valeurs de n, on donne, en fonction de n (nombre de

degrés de liberté) et d'une probabilité α que l'on peut choisir, la valeur $\chi^2_{\alpha,n}$ définie par $P(\chi^2 > \chi^2_{\alpha,n}) = \alpha$. α est un seuil et a en fait une signification particulière dans les problèmes d'estimation et de tests. Il sera défini ultérieurement.

3.8.2 La distribution de Fisher-Snedecor

Cette distribution fut découverte par l'anglais Fisher en 1924 puis tabulée par Snédecor en 1934. Elle interviendra lors des comparaisons des variances de deux échantillons (test d'hypothèse F).

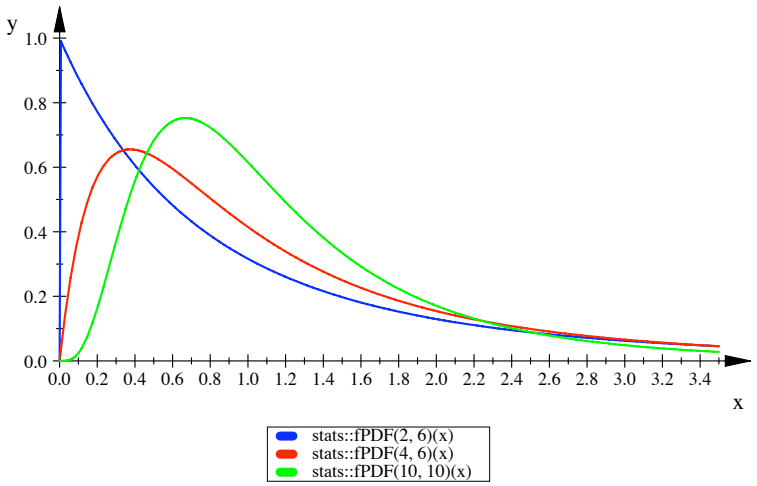
Définition 82 *Si X_1 et X_2 sont deux variables aléatoires indépendantes qui suivent toutes les deux une loi de khi-deux de degrés de liberté respectifs n_1 et n_2, alors la quantité $F = \dfrac{X_1/n_1}{X_2/n_2}$ est une variable aléatoire qui suit la loi de Fisher-Snedecor à n_1 et n_2 degrés de liberté.*

On note $F \rightharpoonup F_{n_1,n_2}$. Cette variable ne prend que des valeurs positives.

Forme de la distribution

On n'écrit pas ici l'expression de la fonction de densité, compliquée et inutile pour nous. Les formes de distribution dépendent de n_1 et n_2 et sont dissymétriques.

A mesure que les valeurs n_1 et n_2 augmentent, la loi de Fisher tend vers une loi normale.



Densité de F pour $(n_1, n_2) = (2,6)$, (4,6), (10,10).

Utilisation de la table de la distribution de Fisher

Les valeurs tabulées de la variable F dépendent d'un seuil α que l'on peut choisir et des nombres de degré de liberté n_1 et n_2. La table donne la valeur F_{α,n_1,n_2} définie par $P(F > F_{\alpha,n_1,n_2}) = \alpha$.

Remarque 83 *Il faut faire attention à l'ordre de n_1 et n_2. n_1 représente le nombre de degrés de liberté du* numérateur *et n_2 celui du* dénominateur *et ne peuvent être intervertis.*

Remarque 84 *Le nombre G_{α,n_1,n_2} tel que $P(F < G_{\alpha,n_1,n_2}) = \alpha$ est l'inverse de F_{α,n_2,n_1}.*

En effet, si une variable X suit la loi F_{n_1,n_2}, alors $1/X$ suit la loi F_{n_2,n_1}.

3.8.3 La distribution de Student

Student est le pseudonyme de V.S Gosset, 1908.

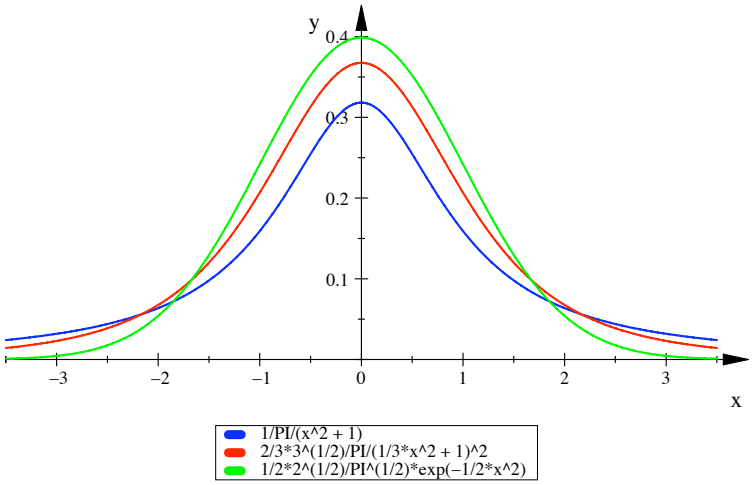
Définition 85 *Soient X et Y deux variables aléatoires indépendantes, la première étant distribuée selon une loi normale centrée réduite $\mathcal{N}(0,1)$ et la deuxième selon une loi de khi-deux à n degrés de liberté. La quantité $T = \dfrac{X\sqrt{n}}{\sqrt{Y}}$ est une variable aléatoire qui suit une* loi de Student à n degrés de liberté.

On écrit $T \rightharpoonup T_n$.

Allure de la distribution

La densité de probabilité est de forme complexe et nous n'en aurons jamais besoin.

La distribution est symétrique par rapport à l'origine et un peu plus aplatie que la distribution normale centrée réduite. Elle ne dépend que de la valeur n qui est son nombre de degrés de liberté.



Densité de T_n pour $n = 1$, 3 et densité de la loi normale standard.

Paramètres descriptifs

On a $E(T_n) = 0$ si $n > 1$ et $Var(T_n) = \frac{n}{n-2}$ si $n > 2$.

Approximation par la loi normale

A mesure que n augmente, la distribution de Student à n degrés de liberté se rapproche de plus en plus de celle de celle de la loi normale centrée réduite.

En pratique : si $T \rightharpoonup T_n$ pour $n \geq 30$, on pourra écrire que $T \rightharpoonup \mathcal{N}(0,1)$.

Tables de la loi de Student

Les valeurs tabulées de la variable T dépendent d'un seuil α que l'on peut choisir et du nombre de degré de liberté n. La table donne la valeur $t_{\alpha,n}$ définie par $P(|T| > t_{\alpha,n}) = \alpha$.

Chapitre 4

Distributions d'échantillonnage

4.1 Généralités sur la notion d'échantillonnage

4.1.1 Population et échantillon

On appelle population la totalité des unités de n'importe quel genre prises en considération par le statisticien. Elle peut être finie ou infinie.

Un échantillon est un sous-ensemble de la population étudiée.

Qu'il traite un échantillon ou une population, le statisticien décrit habituellement ces ensembles à l'aide de mesures telles que le nombre d'unités, la moyenne, l'écart-type et le pourcentage.

- Les mesures que l'on utilise pour décrire une population sont des paramètres. Un paramètre est une caractéristique de la population.
- Les mesures que l'on utilise pour décrire un échantillon sont appelées des statistiques. Une statistique est une caractéristique de l'échantillon.

Nous allons voir dans ce chapitre et dans le suivant comment les résultats obtenus sur un échantillon peuvent être utilisés pour décrire la population. On verra en particulier que les statistiques sont utilisées pour estimer les paramètres.

Afin de ne pas confondre les statistiques et les paramètres, on utilise des notations différentes, comme le présente le tableau récapitulatif suivant.

	POPULATION	ÉCHANTILLON
DÉFINITION	C'est l'ensemble des unités considérées par le statisticien.	C'est un sous-ensemble de la population choisie pour étude.
CARACTÉRISTIQUES	Ce sont les paramètres	Ce sont les statistiques
NOTATIONS	N = taille de la population (si elle est finie)	n = taille de l'échantillon
Si on étudie un caractère quantitatif	moyenne de la population $m = \frac{1}{N}\sum_{i=1}^{N} x_i$	moyenne de l'échantillon $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
	écart-type de la population $\sigma_{pop} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2}$	écart-type de l'échantillon $\sigma_{ech} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Si on étudie un qualitatif	proportion dans la population p	proportion dans l'échantillon f

4.1.2 L'échantillonage

Avantages de l'échantillonnage

- Coût moindre.
- Gain de temps.
- C'est la seule méthode qui donne des résultats dans le cas d'un test destructif.

Méthodes d'échantillonnage

- Échantillonnage sur la base du jugement (par exemple, dans les campagnes électorales certains districts électoraux sont des indicateurs fiables de l'opinion publique).
- Échantillonnage aléatoire simple. Tous les échantillons possibles de même taille ont la même probabilité d'être choisis et tous les éléments de la population ont une chance égale de faire partie de l'échantillon (On utilise souvent une table de nombres aléatoires pour s'assurer que le choix des éléments s'effectue vraiment au hasard).

Remarque 86 *Il existe deux autres méthodes d'échantillonnage aléatoire mais elles ne nous intéressent pas ici . Ce sont l'échantillonnage stratifié et l'échantillonnage par grappes.*

Bien entendu, seul l'échantillonnage aléatoire nous permettra de juger objectivement de la valeur des estimations faites sur les caractéristiques de la population.

Inconvénient de l'échantillonnage

L'échantillonnage a pour but de fournir suffisamment d'informations pour pouvoir faire des déductions sur les caractéristiques de la population. Mais bien entendu, les résultats obtenus d'un échantillon à l'autre vont être en général différents et différents également de la valeur de la caractéristique correspondante dans la population. On dit qu'il y a des fluctuations d'échantillonnage. Comment, dans ce cas, peut-on tirer des conclusions valables ? En déterminant les lois de probabilités qui régissent ces fluctuations. C'est l'objet de ce chapitre.

4.2 La variable aléatoire : moyenne d'échantillon

4.2.1 Introduction

Position du problème :

Si nous prélevons un échantillon de taille n dans une population donnée, la moyenne de l'échantillon nous donnera une idée approximative de la moyenne de la population. Seulement si nous prélevons un autre échantillon de même taille, nous obtiendrons une autre moyenne d'échantillon. Sur l'ensemble des échantillons possibles, on constatera que certains ont une moyenne proche de la moyenne de la population et que d'autres ont une moyenne qui s'en écarte davantage.

Comment traiter le problème ?

Un échantillon de taille n (appelé aussi un n-échantillon), obtenu par échantillonnage aléatoire, va être considéré comme le résultat d'une expérience aléatoire. A chaque échantillon de taille n on peut associer la valeur moyenne des éléments de l'échantillon. On a donc défini une variable aléatoire qui à chaque n-échantillon associe sa moyenne échantillonnale. On la note $\bar{X}$. Cette variable aléatoire possède bien entendu :

- Une distribution de probabilité.
- Une valeur moyenne (la moyenne des moyennes d'échantillons, vous suivez toujours ?).
- Un écart-type.

Le but de ce paragraphe est de déterminer ces trois éléments.

Avant de continuer, essayons de comprendre sur un exemple ce qui se passe.

Exemple 87 *Une population est constituée de 5 étudiants en statistique (le faible effectif n'est pas dû à un manque d'intérêt pour la matière de la part des étudiants mais au désir de ne pas multiplier inutilement les calculs qui vont suivre !). Leur professeur s'intéresse au temps hebdomadaire consacré à l'étude des statistiques par chaque étudiant.*

On a obtenu les résultats suivants.

Étudiant	Temps d'étude (en heures)
A	7
B	3
C	6
D	10
E	4
Total	30

La moyenne de la population est $m = 30/5 = 6$.

Si le professeur choisit un échantillon de taille 3, quelles sont les différentes valeurs possibles pour la moyenne de son échantillon ? Quelle relation existe-t-il entre cette moyenne d'échantillon et la véritable moyenne 6 de la population ?

Toutes les possibilités sont regroupées dans le tableau ci-dessous.

Numéro de l'échantillon	Échantillon	Valeurs du temps d'étude dans cet échantillon	Moyennes de l' échantillon
1	A, B, C	7,3,6	5.33
2	A, B, D	7,3,10	6.67
3	A, B, E	7,3,4	4.67
4	A, C, D	7,6,10	7.67
5	A, C, E	7,6,4	5.67
6	A, D, E	7,10,4	7.00
7	B, C, D	3,6,10	6.33
8	B, C, E	3,6,4	4.33
9	B, D, E	3,10,4	5.67
10	C, D, E	6,10,4	6.67
Total			60.00

On constate que :

- Il y a 10 échantillons ($C_5^3 = 10$).
- La moyenne des échantillons varie entre 4.33 et 7.67, ce qui signifie que la distribution des moyennes d'échantillon est moins dispersée que la distribution des temps d'étude des étudiants, située entre 3 et 10.
- Il est possible que deux échantillons aient la même moyenne. Dans cet exemple, aucun n'a la moyenne de la population ($m = 6$).
- La moyenne des moyennes d'échantillon est $E(\bar{X}) = 60/10 = 6$.

En fait, nous allons voir que le fait que l'espérance de $\bar{X}$ (c'est-à-dire la moyenne des moyennes d'échantillon) est égale à la moyenne de la population n'est pas vérifié seulement dans notre exemple. C'est une loi générale.

Bien, me direz-vous, mais pourquoi faire tout cela ? Dans la réalité, on ne choisit qu'un seul échantillon. Alors comment le professeur de statistique qui ne connaît qu'une seule moyenne d'échantillon pourra-t-il déduire quelque chose sur la moyenne de la population ? Tout simplement en examinant "jusqu'à quel point" la moyenne d'un échantillon unique s'approche de la moyenne de la population. Pour cela, il lui faut la distribution théorique de la variable aléatoire $\bar{X}$ ainsi que l'écart-type de cette distribution.

4.2.2 Etude de la variable : moyenne d'échantillon

Définition de la variable

On considère une population dont les éléments possèdent un caractère mesurable qui est la réalisation d'une variable aléatoire X qui suit une loi de probabilité d'espérance m et d'écart-type σ_{pop}. On suppose que la population est infinie ou si elle est finie que l'échantillonnage se fait avec remise.

- On prélève un échantillon aléatoire de taille n et on mesure les valeurs de X sur chaque élément de l'échantillon. On obtient une suite de valeurs $x_1, x_2, \ldots, x_n$.
- Si on prélève un deuxième échantillon toujours de taille n, la suite des valeurs obtenues est $x'_1, x'_2, \ldots, x'_n$, puis $x''_1, x''_2, \ldots, x''_n$... etc... pour des échantillons supplémentaires.

x_1, x'_1, x''_1,... peuvent être considérées comme les valeurs d'une variable aléatoire X_1 qui suit la loi de X. De même, x_2, x'_2, x''_2,... peuvent être considérées comme les valeurs d'une variable aléatoire X_2 qui suit aussi la loi de X, ... et x_n, x'_n, x''_n,... celles d'une variable aléatoire X_n qui suit encore et toujours la même loi, celle de X.

- X_1 pourrait se nommer "valeur du premier élément d'un échantillon". X_2 pourrait se nommer "valeur du deuxième élément d'un échantillon". ....X_n pourrait se nommer "valeur du n-ième élément d'un échantillon".
- L'hypothèse d'une population infinie ou d'un échantillonnage avec remise nous permet d'affirmer que ces n variables aléatoires sont indépendantes.

Rappel sur les notations : Par convention, on note toujours les variables aléatoires à l'aide de lettres majuscules (X_i) et les valeurs qu'elles prennent dans une réalisation à l'aide de lettres minuscules (x_i).

Si les valeurs prises par X dans un échantillon sont $x_1, x_2, \ldots, x_n$, la moyenne $\bar{x}$ de l'échantillon est donnée par $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$. Cette valeur n'est rien d'autre que la valeur prise dans cet échantillon de la variable aléatoire $\frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Définition 88 *On définit donc la* variable aléatoire *moyenne d'échantillon $\bar{X}$ par*

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Paramètres descriptifs de la distribution

On applique les propriétés de l'espérance et de la variance étudiées au chapitre 2.

- $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{nm}{n} = m$, car les variables suivent toutes la même loi d'espérance m.
- $Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2_{pop}}{n^2} = \frac{\sigma^2_{pop}}{n}$, car les variables suivent toutes la même loi de variance et sont indépendantes.

Proposition 89

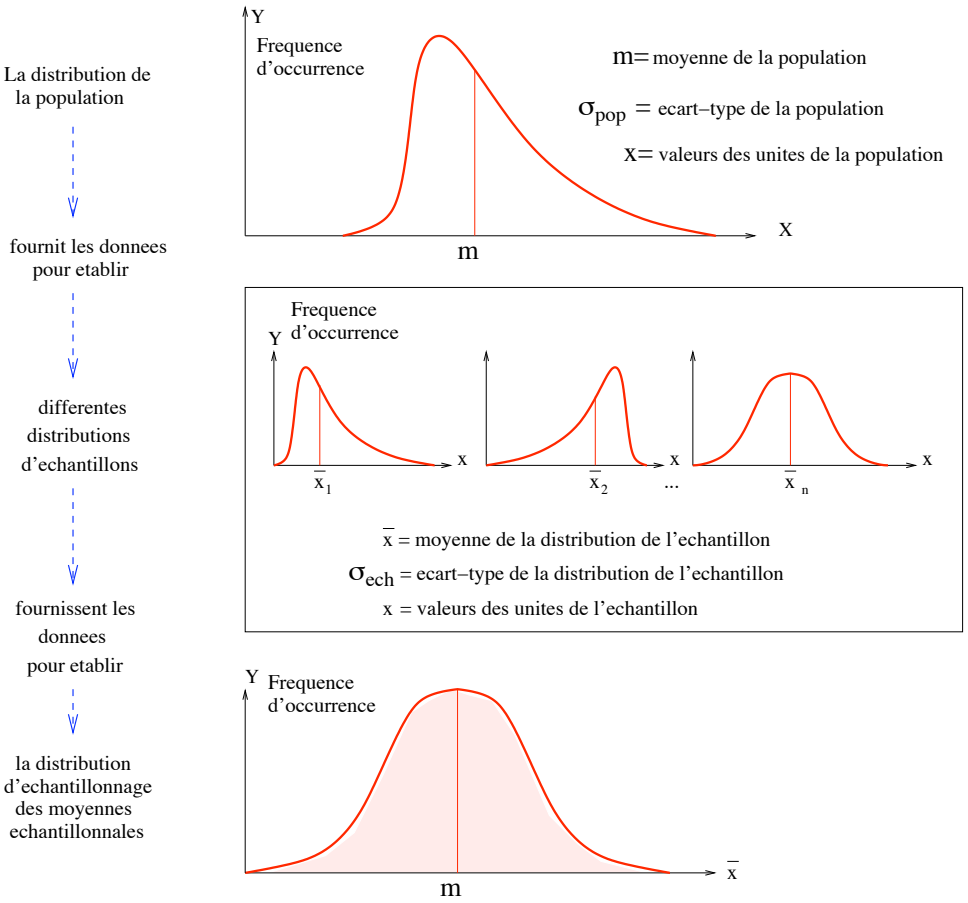
$$E(\bar{X}) = m, \quad Var(\bar{X}) = \frac{\sigma^2_{pop}}{n}.$$

Remarque 90 *1. Nous venons de démontrer ce que nous avions constaté sur notre exemple : la moyenne de la distribution d'échantillonnage des moyennes est égale à la moyenne de la population.*

2. On constate que plus n croît, plus $Var(\bar{X})$ décroît.

Dans l'exemple d'introduction, nous avions en effet constaté que la distribution des moyennes d'échantillon était moins dispersée que la distribution initiale. En effet, à mesure que la taille de l'échantillon augmente, nous avons accès à une plus grande quantité d'informations pour estimer la moyenne de la population. Par conséquent, la différence probable entre la vraie valeur de la moyenne de la population et la moyenne échantillonnale diminue. L'étendue des valeurs possibles de la moyenne échantillonnale diminue et le degré de dispersion de la distribution aussi. $\sigma(\bar{X})$ est aussi appelé l'*erreur-type* de la moyenne.

On peut schématiser le passage de la distribution de la variable aléatoire X à celle de la variable aléatoire $\bar{X}$ en passant par les différents échantillons par le graphique ci-après.



m= moyenne de la distribution d'echantillonnage des moyennes

$\sigma_{pop}/\sqrt{n}$ = ecart–type de la distribution d'echantillonnage des moyennes

$\bar{x}$ = valeurs des moyennes de tous les echantillons possibles

Mais connaître les paramètres descriptifs de la distribution de $\bar{X}$ ne suffit pas. Il faut connaître aussi sa distribution de probabilité. On se demande alors : dépend elle

1. de la distribution de X ?
2. de la taille n de l'échantillon ?

4.3 La variable aléatoire : variance d'échantillon

La variance $\sigma^2_{ech} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ d'un n-échantillon est la réalisation de la variable aléatoire $\Sigma^2_{ech} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. On peut se demander si cette variable possède la même propriété que la variable moyenne d'échantillon, c'est-à-dire si l'espérance de Σ^2_{ech} est égale à la variance de la population.

4.3.1 Espérance de la variable aléatoire Σ^2_{ech}

Autre expression de Σ^2_{ech}

$$\begin{aligned}\Sigma^2_{ech} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - m + m - \bar{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 + \frac{2}{n}\sum_{i=1}^{n}(X_i - m)(m - \bar{X}) + \frac{1}{n}\sum_{i=1}^{n}(m - \bar{X})^2 \\ &= A + B + C,\end{aligned}$$

Or, $B = \frac{2}{n}(m - \bar{X})\sum_{i=1}^{n}(X_i - m) = -2(m - \bar{X})^2$ et $C = (m - \bar{X})^2$. On en déduit que

$$\Sigma^2_{ech} = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 - (\bar{X} - m)^2.$$

Espérance de Σ^2_{ech}

Proposition 91

$$E(\Sigma^2_{ech}) = \frac{n-1}{n}\sigma^2_{pop}.$$

Preuve.

$$\begin{aligned}E(\Sigma^2_{ech}) &= \frac{1}{n}\sum_{i=1}^{n}E((X_i - m)^2) - E((\bar{X} - m)^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}Var(X_i) - Var(\bar{X}) \\ &= \sigma^2_{pop} - \frac{1}{n}\sigma^2_{pop} = \frac{n-1}{n}\sigma^2_{pop}.\end{aligned}$$

Conclusion. La moyenne des variances d'échantillon n'est pas la variance de la population, mais la variance de la population multipliée par $\frac{n-1}{n}$. Bien sûr, si n est très grand, ces deux nombres seront très proches l'un de l'autre.

4.3.2 La variable aléatoire S^2

Définition de S^2

Pour pouvoir déterminer une valeur approchée de σ^2_{pop} et savoir quelle erreur on commet en effectuant cette approximation, on veut disposer d'une variable dont l'espérance est la variance de la population. Nous allons donc considérer une nouvelle variable aléatoire S^2.

Définition 92 *On appelle* variance d'échantillon *la variable aléatoire S^2 définie par*

$$S^2 = \frac{n}{n-1}\Sigma^2_{ech} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

On a bien entendu $E(S^2) = \sigma^2_{pop}$.
Nous verrons plus tard que cela signifie que S^2 est un estimateur *sans biais* de σ^2_{pop}.

Distribution de S^2

Nous supposons ici que X suit une loi normale.

On considère la variable $Y = \frac{n\Sigma^2_{ech}}{\sigma^2_{pop}} = \sum_{i=1}^{n}(\frac{X_i - \bar{X}}{\sigma_{pop}})^2$.

Y est une somme d'écarts réduits relatifs à une variable normale. D'après ce que nous avons vu au chapitre 3, paragraphe 7.1, nous pouvons affirmer que Y suit une loi du χ^2 à $n-1$ degrés de liberté (on perd un degré de liberté car on a estimé le paramètre m par $\bar{X}$).

Proposition 93 *$Y = \frac{(n-1)S^2}{\sigma^2_{pop}}$ suit une loi χ^2_{n-1}.*

Remarque 94 *Encore une fois, on n'a pas directement la loi de S^2 mais celle de $\frac{(n-1)S^2}{\sigma^2_{pop}}$.*

Approximation de la distribution de S^2 dans le cas des grands échantillons : $n \geq 30$

Nous avons vu au chapitre 3 que lorsque n est grand ($n \geq 30$), on pouvait approcher la loi χ^2_ν par la loi $\mathcal{N}(\nu, \sqrt{2\nu})$. Donc Y suit approximativement une loi normale, $E(Y) \simeq n-1$ et $Var(Y) \simeq 2(n-1)$.

Proposition 95 *Si $n \geq 30$, S^2 suit une loi $\mathcal{N}(\sigma^2_{pop}, \sigma^2_{pop}\sqrt{\frac{2}{n-1}})$, en première approximation.*

Preuve. La loi de S^2 est alors approximativement normale, son espérance vaut σ^2_{pop} et sa variance approximativement

$$Var(S^2) = Var(\frac{\sigma^2_{pop}}{n-1}Y) = \frac{\sigma^4_{pop}}{(n-1)^2}Var(Y) \simeq \frac{2\sigma^4_{pop}}{n-1}.$$

4.4 Distribution de la moyenne d'échantillon

Nous allons distinguer deux cas : celui des grands échantillons ($n \geq 30$) et celui des petits échantillons ($n < 30$).

4.4.1 Cas des grands échantillons : $n \geq 30$.

On peut appliquer le théorème centrale-limite.

1. Nous sommes en présence de n variables aléatoires indépendantes.
2. Elles suivent la même loi d'espérance m et de variance σ^2_{pop}, donc aucune n'est prépondérante.

Conclusion. Lorsque n devient très grand, la distribution de $S = X_1 + \cdots + X_n$ se rapproche de celle de la loi normale d'espérance nm et de variance $n\sigma^2_{pop}$, S suit approximativement $\mathcal{N}(nm, n\sigma^2_{pop})$.

Par conséquent, pour n assez grand, la distribution de $\bar{X} = S/n$ se rapproche de celle de la loi normale d'espérance m et de variance σ^2_{pop}/n c'est-à-dire $\mathcal{N}(m, \frac{\sigma_{pop}}{\sqrt{n}})$. On peut donc considérer que $\frac{\bar{X} - m}{\sigma_{pop}/\sqrt{n}}$ suit la loi $\mathcal{N}(0,1)$.

Proposition 96 *Si $n \geq 30$, $\bar{X}$ suit approximativement $\mathcal{N}(m, \frac{\sigma_{pop}}{\sqrt{n}})$.*

Remarque 97 – *En pratique, on considère que cela est vrai à partir de $n \geq 30$ et que lorsque la forme de la distribution de X est pratiquement symétrique, $n \geq 15$ est convenable.*

- *Ce théorème est très puissant car il n'impose aucune restriction sur la distribution de X dans la population.*
- *Si la variance est inconnue, un grand échantillon ($n \geq 30$) permet de déduire une valeur fiable pour σ^2_{pop} en calculant la variance de l'échantillon σ^2_{ech} et en posant*

$$\sigma^2_{pop} = \frac{n}{n-1}\sigma^2_{ech} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

comme on l'a vu au paragraphe précédent.

4.4.2 Cas des petits échantillons : $n < 30$

Nous nous plaçons alors exclusivement dans le cas où X suit une loi normale dans la population. Nous allons encore distinguer deux cas : celui où σ_{pop} est connu et celui où σ_{pop} est inconnu.

Cas où σ_{pop} est connu

X suit une loi normale $\mathcal{N}(m, \sigma_{pop})$ donc les variables X_i suivent toutes la même loi $\mathcal{N}(m, \sigma_{pop})$. De plus elles sont indépendantes. D'après la propriété vue au chapitre 3 sur la somme de lois normales indépendantes, $S = X_1 + \cdots + X_n$ a une distribution normale et la variable $\bar{X} = S/n$ suit aussi une loi normale, la loi $\mathcal{N}(m, \frac{\sigma_{pop}}{\sqrt{n}})$. Donc $\dfrac{\bar{X} - m}{\sigma_{pop}/\sqrt{n}}$ suit la loi $\mathcal{N}(0, 1)$.

$$\boxed{\text{Si } \begin{cases} n < 30 \\ \sigma_{pop} \text{ connu} \end{cases}, \bar{X} \text{ suit } \mathcal{N}(m, \tfrac{\sigma_{pop}}{\sqrt{n}}).}$$

Cas où σ_{pop} est inconnu

Lorsque l'échantillonnage s'effectue à partir d'une population normale de variance inconnue et que la taille de l'échantillon est petite ($n < 30$), l'estimation de la variance effectuée au paragraphe précédent n'est plus fiable. On ne peut plus écrire $\sigma^2_{pop} \simeq \frac{n}{n-1}\sigma^2_{ech}$ car σ^2_{ech} varie trop d'échantillon en échantillon.

L'écart-type de la distribution de $\bar{X}$ n'est donc plus une constante $\frac{\sigma_{pop}}{\sqrt{n}}$ connue approximativement grâce à $\frac{\sigma_{pop}}{\sqrt{n}} \simeq \frac{\sigma_{ech}}{\sqrt{n-1}}$. On va alors considérer que l'écart-type de $\bar{X}$ sera donné dans chaque échantillon par une valeur différente de $\frac{\sigma_{ech}}{\sqrt{n-1}}$.

Nous devons donc considérer σ_{ech} comme la réalisation d'une variable aléatoire, la variable écart-type d'échantillon, notée Σ_{ech} et définie par $\Sigma^2_{ech} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

La variable aléatoire $T = \dfrac{\bar{X} - m}{\Sigma_{ech}/\sqrt{n-1}} = \dfrac{\sqrt{n-1}(\bar{X} - m)}{\Sigma_{ech}}$ ne suit plus alors une loi normale car le dénominateur n'est pas une constante.

En divisant numérateur et dénominateur par σ_{pop}, on écrit T sous la forme

$$T = \frac{\sqrt{n-1}(\bar{X} - m)}{\Sigma_{ech}} = \frac{\sqrt{n-1}\frac{\bar{X}-m}{\sigma_{pop}/\sqrt{n}}}{\sqrt{\sum_{i=1}^{n}(\frac{X_i - \bar{X}}{\sigma_{pop}})^2}}.$$

On reconnaît au numérateur une variable aléatoire qui suit une loi $\mathcal{N}(0, 1)$, multipliée par un facteur $\sqrt{n-1}$, et au dénominateur une somme de carrés de variables suivant aussi la loi $\mathcal{N}(0, 1)$. Le carré du dénominateur suit donc une loi du χ^2. Mais quel est son nombre de degrés de liberté ?

Le concept de degré de liberté

- Pourquoi chercher le nombre de degrés de liberté ?
  Pour pouvoir utiliser correctement les tables de lois de probabilité qui dépendent d'un nombre de degrés de liberté (en particulier pour la distribution de Student et celle du χ^2).
- Que représente le nombre de degrés de liberté ?
  C'est une quantité qui est toujours associée à une somme de carrés et qui représente le nombre de carrés indépendants dans cette somme.
- Comment calculer le nombre de degrés de liberté ?
  Il y a deux façons de procéder.
  1. Soit on effectue la différence entre le nombre total de carrés et le nombre de relations qui lient les différents éléments de la somme.
  2. Soit on effectue la différence entre le nombre total de carrés et le nombre de paramètres que l'on doit estimer pour effectuer le calcul.

Dans le cas de notre somme $\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma_{pop}})^2$, envisageons les deux façons de compter le nombre de degrés de liberté.

1. Le nombre de carrés dans la somme est n. Il y a une relation entre les variables $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$. Le nombre de degrés de liberté est donc $n-1$.
2. Le nombre de carrés dans la somme est n. Lorsqu'on dit que $\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma_{pop}})^2$ est une somme de carrés de variables normales centrées réduites, on remplace m par $\bar{X}$. On a donc estimé un paramètre. On trouve encore que le nombre de degrés de liberté est $n-1$.

Proposition 98 *Si* $\begin{cases} n < 30 \\ \sigma_{pop} \text{ connu} \end{cases}$ *, la variable* $T = \dfrac{\bar{X} - m}{\Sigma_{ech}/\sqrt{n-1}}$ *suit une loi de Student à* $n-1$ *degrés de liberté, notée* T_{n-1}.

Revoir éventuellement la définition de la loi de Student dans le chapitre précédent.

Remarque 99 *Dans ce dernier cas (petits échantillons et* X *suit une loi normale de variance inconnue), on ne trouve pas directement la loi suivie par mais celle suivie par* $T = \dfrac{\bar{X} - m}{\Sigma_{ech}/\sqrt{n-1}}$.

Exercice 100 *Le responsable d'une entreprise a accumulé depuis des années les résultats à un test d'aptitude à effectuer un certain travail. Il semble plausible de supposer que les résultats au test d'aptitude sont distribués suivant une loi normale de moyenne* $m = 150$ *et de variance* $\sigma^2_{pop} = 100$. *On fait passer le test à 25 individus de l'entreprise. Quelle est la probabilité que la moyenne de l'échantillon soit entre 146 et 154 ?*

Solution de l'exercice 100. *Test d'aptitude.*

On considère la variable aléatoire $\bar{X}$ moyenne d'échantillon pour les échantillons de taille $n = 25$. On cherche à déterminer $P(146 < \bar{X} < 154)$.

Pour cela, il nous faut connaître la loi suivie par $\bar{X}$. Examinons la situation. Nous sommes en présence d'un petit échantillon ($n < 30$) et heureusement dans le cas où la variable X (résultat au test d'aptitude) suit une loi normale. De plus, σ_{pop} est connu. Donc $\bar{X}$ suit $\mathcal{N}(m, \frac{\sigma_{pop}}{\sqrt{n}}) = \mathcal{N}(150, 10/5)$. On en déduit que $T = \frac{\bar{X}-150}{2}$ suit $\mathcal{N}(0,1)$.

La table donne

$$\begin{aligned} P(146 < \bar{X} < 154) &= P(\frac{146-150}{2} < T < \frac{154-150}{2}) = P(-2 < T < 2) \\ &= 2P(0 < T < 2) = 2 \times 0.4772 = 0.9544. \end{aligned}$$

4.5 Distribution de la variable proportion d'échantillon

Il arrive fréquemment que nous ayions à estimer dans une population une proportion p d'individus possédant un caractère qualitatif donné.

Bien sûr, cette proportion p sera estimée à l'aide des résultats obtenus sur un n-échantillon. La proportion f obtenue dans un n-échantillon est la valeur observée d'une variable aléatoire F, fréquence d'apparition de ce caractère dans un échantillon de taille n, appelée proportion d'échantillon. On se pose une troisième fois la question. La moyenne des fréquences d'observation du caractère sur l'ensemble de tous les échantillons de taille n est-elle égale à la proportion p de la population ?

4.5.1 Paramètres descriptifs de la distribution de F

F est la fréquence d'apparition du caractère dans un échantillon de taille n. Donc $F = X/n$ où X est le nombre de fois où le caractère apparaît dans le n-échantillon.

Par définition X suit $\mathcal{B}(n, p)$. Donc $E(F) = np$ et $Var(F) = npq$.

Il en résulte que

$$E(X) = p \quad \text{et} \quad Var(X) = \sqrt{\frac{pq}{n}}.$$

Conséquences.

1. La réponse à la question que nous nous posions est oui : l'espérance de la fréquence d'échantillon est égale à la probabilité théorique d'apparition dans la population.
2. Lorsque la taille de l'échantillon augmente, la variance de F diminue, ce qui est logique : plus on a d'informations, plus il est probable que la proportion observée dans l'échantillon soit proche de la proportion de la population.

4.5.2 Distribution de la proportion d'échantillon dans le cas des grands échantillons

On sait que si $n \geq 30$, $np \geq 15$ et $nq \geq 15$, on peut approcher la loi binomiale par la loi normale de même espérance et de même écart-type. Donc F suit approximativement $\mathcal{N}(p, \sqrt{\frac{pq}{n}})$, et la variable $T = \dfrac{F - p}{\sqrt{\frac{pq}{n}}}$ suit alors approximativement la loi $\mathcal{N}(0, 1)$.

Exercice 101 *Selon une étude sur le comportement du consommateur, 25% d'entre eux sont influencés par la marque, lors de l'achat d'un bien. Si on interroge 100 consommateurs pris au hasard, quelle est la probabilité pour qu'au moins 35 d'entre eux se déclarent influencés par la marque ?*

Solution de l'exercice 101. *Influence de la marque.*

Appelons F la variable aléatoire :"proportion d'échantillon dans un échantillon de taille 100". Il s'agit ici de la proportion de consommateurs dans l'échantillon qui se déclarent influencés par la marque. On cherche à calculer $P(F > 0.35)$.

Il nous faut donc déterminer la loi de F. Or $np = 100 \times 0.25 = 25$ et $nq = 100 \times 0.75 = 75$. Ces deux quantités étant supérieures à 15, on peut considérer que F suit $\mathcal{N}(p, \sqrt{\frac{pq}{n}}) = \mathcal{N}(0.25, 0.0433)$.

On utilise la variable $T = \frac{F - 0.25}{0.0433}$ qui suit la loi $\mathcal{N}(0, 1)$. Il vient

$$P(F > 0.35) = P(T > 2.31) = 0.5 - P(0 < T < 2.31) = 0.5 - 0.4896 = 0.0104.$$

Conclusion. Il y a environ une chance sur 100 pour que plus de 35 consommateurs dans un 100 - échantillon se disent influencés par la marque lorsque l'ensemble de la population contient 25% de tels consommateurs.

En pratique, il est peu fréquent de connaître p : on doit plutôt l'estimer à partir d'un échantillon. Comment faire ? C'est ce que nous traiterons dans le prochain chapitre.

Chapitre 5

Estimation

5.1 Introduction

Estimer ne coûte presque rien,
Estimer incorrectement coûte cher.

Vieux proverbe chinois.

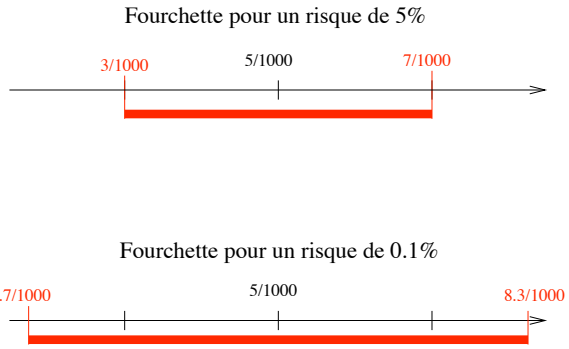
Dans de nombreux domaines (scientifiques, économiques, épidémiologiques...), on a besoin de connaître certaines caractéristiques d'une population. Mais, en règle générale, on ne peut pas les évaluer facilement du fait de l'effectif trop important des populations concernées. La solution consiste alors à estimer le paramètre cherché à partir de celui observé sur un échantillon plus petit.

L'idée de décrire une population à partir d'un échantillon réduit, à l'aide d'un "multiplicateur", n'a été imaginée que dans la seconde moitié du XVIIIème siècle, notamment par l'école arithmétique politique anglaise. Elle engendra une véritable révolution : l'observation d'échantillons permettait d'éviter des recensements d'une lourdeur et d'un prix exorbitants. Toutefois, on s'aperçut rapidement que les résultats manquaient d'exactitude. Nous savons maintenant pourquoi : on ne prenait en considération ni la *représentativité* de l'échantillon, ni les *fluctuations* d'échantillonnage. C'est là que le hasard intervient.

La première précaution à prendre est donc d'obtenir un échantillon représentatif. Nous pourrons en obtenir un par tirage au sort (voir le chapitre précédent sur l'échantillonnage aléatoire simple) : le hasard participe donc au travail du statisticien qui l'utilise pour pouvoir le maîtriser ! Mais, même tiré au sort, un échantillon n'est pas l'image exacte de la population, en raison des fluctuations d'échantillonnage. Lorsque, par exemple, on tire au sort des échantillons dans un urne contenant 20 % de boules blanches, on obtient des échantillons où la proportion de boules blanches fluctue autour de 20 %. Ces fluctuations sont imprévisibles : le hasard peut produire n'importe quel écart par rapport à la proportion de la population (20 %). Cependant, on s'en doute, tous les écarts ne sont pas également vraisemblables : les très grands écarts sont très peu probables. Au moyen du calcul des probabilités, le statisticien définit un intervalle autour du taux observé, intervalle qui contient probablement le vrai taux : c'est "l'intervalle de confiance" ou, plus couramment, la "fourchette".

Si l'on ne peut connaître le vrai taux par échantillonnage, peut-on au moins le situer avec certitude dans la fourchette ? Non. Le hasard étant capable de tous les caprices, on ne peut raisonner qu'en termes de probabilités, et la fourchette n'a de signification qu'assortie d'un certain risque d'erreur. On adopte souvent un risque de 5 % : cinq fois sur cent, le taux mesuré sur l'échantillon n'est pas le bon, le vrai taux étant en dehors de la fourchette. On peut diminuer le risque d'erreur mais alors la fourchette grandit et perd de son intérêt. Bien entendu, il existe une infinité de fourchettes, une pour chaque risque d'erreur adopté. On doit trouver un compromis entre le risque acceptable et le souci de précision.

Exemple 102 *Mesure du taux de séropositifs pour le sida dans une population.*



On a observé 25 séropositifs sur un échantillon de 5000 sujets, soit un taux de 0.5%. Ce taux observé n'a de signification qu'assorti d'une fourchette : le risque que le vrai taux sorte d'une fourchette comprise entre 0.3% et 0.7% est acceptable (figure du haut). On peut diminuer ce risque, mais alors la fourchette est plus large, et devient moins intéressante (figure du bas).

Dans ce cours, nous allons apprendre à estimer à l'aide d'un échantillon :

– Dans le cas d'un caractère quantitatif la moyenne m et l'écart-type d'une population.
– Dans le cas d'un caractère qualitatif, la proportion p de la population.

Ces estimations peuvent s'exprimer par une seule valeur (estimation ponctuelle), soit par un intervalle (estimation par intervalle de confiance). Bien sûr, comme l'échantillon ne donne qu'une information partielle, ces estimations seront accompagnées d'une certaine marge d'erreur.

5.2 L'estimation ponctuelle

5.2.1 Définition

Estimer un paramètre, c'est en chercher une valeur approchée en se basant sur les résultats obtenus dans un échantillon. Lorsqu'un paramètre est estimé par un seul nombre, déduit des résultats de l'échantillon, ce nombre est appelé *estimation ponctuelle* du paramètre.

L'estimation ponctuelle se fait à l'aide d'un *estimateur*, qui est une variable aléatoire d'échantillon. L'estimation est la valeur que prend la variable aléatoire dans l'échantillon observé.

5.2.2 Propriétés des estimateurs ponctuels

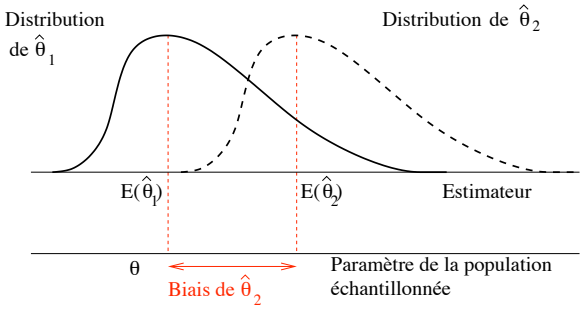
Lorsqu'on utilise fréquemment des estimateurs ponctuels on souhaite qu'ils possèdent certaines propriétés. Ces propriétés sont importantes pour choisir le meilleur estimateur du paramètre correspondant, c'est-à-dire celui qui s'approche le plus possible du paramètre à estimer. Un paramètre inconnu peut avoir plusieurs estimateurs. Par exemple, pour estimer le paramètre m, moyenne d'une population, on pourrait se servir de la moyenne arithmétique, de la médiane ou du mode. Les qualités que doit posséder un estimateur pour fournir de bonnes estimations sont décrites ci-après.

Définition 103 Estimateur non biaisé. *On note θ le paramètre de valeur inconnue, $\hat{\theta}$ l'estimateur de θ.*

Un estimateur est sans biais *si la moyenne de sa distribution d'échantillonnage est égale à la valeur θ du paramètre de la population à estimer, c'est-à-dire si $E(\hat{\theta}) = \theta$.*

Si l'estimateur est biaisé, son biais est mesuré par l'écart suivant : BIAIS $= E(\hat{\theta}) - \theta$.

La figure suivante représente les distributions d'échantillonnage d'un estimateur sans biais $\hat{\theta}_1$ et d'un estimateur biaisé $\hat{\theta}_2$.



Exemple 104 *On a vu au chapitre 4 que $E(\bar{X}) = m$. Donc la moyenne d'échantillon $\bar{X}$ est un estimateur sans biais du paramètre m, moyenne de la population.*

En revanche, la médiane d'échantillon M_e est un estimateur biaisé lorsque la population échantillonnée est asymétrique.

Exemple 105 *Nous avons vu également que $E(\Sigma^2_{ech}) = \frac{n-1}{n}\sigma^2_{pop}$. Donc Σ^2_{ech} est un estimateur biaisé du paramètre σ^2_{pop}, variance de la population.*

C'est pour cette raison que l'on a introduit la variance d'échantillon $S^2 = \frac{n}{n-1}\Sigma^2_{ech}$ qui est un estimateur sans biais de σ^2_{pop}, puisque $E(S^2) = \sigma^2_{pop}$.

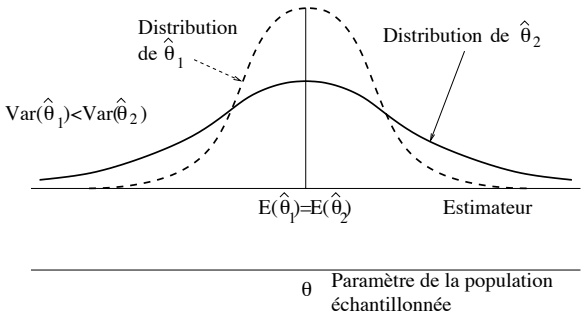
L'absence de biais, à elle toute seule, ne garantit pas que nous avons un bon estimateur. En effet, certains paramètres peuvent avoir plusieurs estimateurs sans biais. Le choix parmi les estimateurs sans biais s'effectue en comparant les variances des estimateurs. En effet, un estimateur sans biais mais à variance élevée peut fournir des estimations très éloignées de la vraie valeur du paramètre.

Définition 106 Estimateur efficace. *Un estimateur sans biais est efficace si sa variance est la plus faible parmi les variances des autres estimateurs sans biais.*

Ainsi, si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs sans biais du paramètre θ, l'estimateur $\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$ si

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta \quad \text{et } V(\hat{\theta}_1) < V(\hat{\theta}_2).$$

La notion d'estimateur efficace peut s'illustrer de la façon suivante :



Définition 107 Estimateur convergent. *Un estimateur $\hat{\theta}$ est convergent si sa distribution tend à se concentrer autour de la valeur inconnue à estimer, θ, à mesure que la taille d'échantillon augmente, c'est-à-dire si $n \to \infty$.*

Par exemple, $\bar{X}$ est un estimateur convergent puisque $V(\bar{X}) = \frac{\sigma^2_{pop}}{n}$ tend vers 0.

Remarque 108 *Un estimateur sans biais et convergent est dit absolument correct.*

Ces trois propriétés sont les principales qualités que nous recherchons pour un estimateur. Nous n'insisterons pas sur les propriétés mathématiques que doivent posséder les estimateurs.

Conséquences : L'étude du chapitre 4 nous a appris que

$$E(\bar{X}) = m \quad \text{et} \quad V(\bar{X}) = \frac{\sigma^2_{pop}}{n},$$
$$E(S^2) = \sigma^2_{pop} \quad \text{et} \quad V(S^2) = \frac{2\sigma^4_{pop}}{n-1},$$
$$E(F) = p \quad \text{et} \quad V(F) = \frac{pq}{n}.$$

On peut donc affirmer que :
- $\bar{X}$ est un estimateur absolument correct de la moyenne m pour un caractère quantitatif.
- S^2 est un estimateur absolument correct de la variance pour un caractère quantitatif.
- F est un estimateur absolument correct de la proportion p pour un caractère qualitatif.

Nous pourrons donc estimer m par $\bar{X}$, σ^2_{pop} par S^2, p par F.

Mais les estimations ponctuelles bien qu'utiles, ne fournissent aucune information concernant la précision des estimations, c'est-à-dire qu'elles ne tiennent pas compte de l'erreur possible dans l'estimation, erreur attribuable aux fluctuations d'échantillonnage. Quelle confiance avons-nous dans une valeur unique ? On ne peut répondre à cette question en considérant uniquement l'estimation ponctuelle obtenue des résultats de l'échantillon. Il faut lui associer un intervalle qui permet d'englober avec une certaine fiabilité, la vraie valeur du paramètre correspondant.

5.3 Estimation par intervalle de confiance

5.3.1 Définition

L'estimation par intervalle d'un paramètre inconnu θ consiste à calculer, à partir d'un estimateur choisi $\hat{\theta}$, un intervalle dans lequel il est vraisemblable que la valeur correspondante du paramètre s'y trouve. L'intervalle de confiance est défini par deux limites LI et LS auxquelles est associée une certaine probabilité, fixée à l'avance et aussi élevée qu'on le désire, de contenir la valeur vraie du paramètre. La probabilité associée à l'intervalle de confiance et exprimée en pourcentage est égale à S où S est le seuil de confiance ou niveau de confiance de l'intervalle, exprimé également en pourcentage. Autrement dit,

$$P(LI \leq \theta \leq LS) = S,$$

où
- LI est la limite inférieure de l'intervalle de confiance.
- LS est la limite supérieure de l'intervalle de confiance.
- S est la probabilité associée à l'intervalle d'encadrer la vraie valeur du paramètre.

LI et LS sont appelées les *limites de confiance* de l'intervalle et sont des quantités qui tiennent compte des fluctuations d'échantillonnage, de l'estimateur $\hat{\theta}$ et du seuil de confiance S. La quantité 1 - S est égale à la probabilité, exprimée en pourcentage, que l'intervalle n'encadre pas la vraie valeur du paramètre. On note $\alpha = 1 - S$. α s'appelle le risque ou le *seuil de signification* de l'intervalle.

A quoi correspond l'intervalle de confiance ?

Si nous répétons l'expérience un grand nombre de fois (prélever un grand nombre de fois un échantillon de taille n de la même population), dans 100S cas sur 100 les intervalles obtenus (différents à chaque réalisation de l'expérience) recouvrent la vraie valeur du paramètre.

Remarques :
- L'intervalle ainsi défini est un intervalle aléatoire puisqu'avant l'expérience, les limites de l'intervalle sont des variables aléatoires (elles sont fonctions des observations de l'échantillon).
- Le niveau de confiance est toujours associé à l'intervalle et non au paramètre inconnu θ. θ n'est pas une variable aléatoire : il est ou n'est pas dans l'intervalle $[LI, LS]$.

– Le niveau de confiance doit être choisi *avant* que ne s'effectue l'estimation par intervalle. Il arrive souvent que le chercheur non averti calcule plusieurs intervalles d'estimation à des niveaux de confiance différents et choisisse par la suite l'intervalle qui lui semble le plus approprié. Une telle approche constitue en réalité une interprétation inacceptable des données en ce qu'elle fait dire aux résultats échantillonnaux ce que l'on veut bien entendre.
– Il y a une infinité de solutions possibles pour déterminer l'intervalle $[LI, LS]$. On choisira de prendre des risques symétriques, c'est-à-dire de choisir LI et LS tels que

$$P(\theta \leq LI) = P(\theta \geq LS) = \frac{1-S}{2}.$$

Pour calculer l'intervalle de confiance, on doit connaître la distribution d'échantillonnage (distribution de probabilité) de l'estimateur correspondant, c'est-à-dire connaître de quelle façon sont distribuées toutes les valeurs possibles de l'estimateur obtenues à partir de tous les échantillons possibles de même taille prélevés de la même population. Ce travail a été effectué au chapitre précédent. Nous allons voir à présent comment déduire des distributions d'échantillonnage la construction des intervalles de confiance.

5.3.2 Estimation d'une moyenne par intervalle de confiance

On se propose d'estimer, par intervalle de confiance, la moyenne m d'un caractère mesurable d'une population. Il s'agit donc de calculer, à partir de la moyenne $\bar{x}$ (valeur prise par l'estimateur $\bar{X}$) de l'échantillon, un intervalle dans lequel il est vraisemblable que la vraie valeur de m se trouve. Cet intervalle se définit d'après l'équation $P(A \leq m \leq B) = S$. Les limites A et B de cet intervalle sont des quantités aléatoires et prendront, après avoir prélevé l'échantillon et calculé l'estimation $\bar{x}$, la forme $LI \leq m \leq LS$. Nous allons déterminer LI et LS en utilisant la distribution d'échantillonnage de $\bar{X}$. L'étude du chapitre 4 nous amène donc à distinguer deux cas, suivant la taille de l'échantillon.

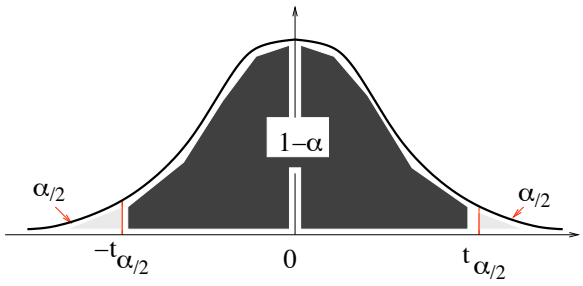
a. On dispose d'un grand échantillon ($n \geq 30$) ou d'un petit échantillon ($n < 30$) dont la distribution est normale d'écart-type connu σ_{pop}

Dans ces conditions on considère que la variable aléatoire $\bar{X}$ suit une loi normale,

$$\bar{X} \rightharpoonup \mathcal{N}(m, \frac{\sigma_{pop}}{\sqrt{n}}).$$

Donc $T = \dfrac{\bar{X} - m}{\frac{\sigma_{pop}}{\sqrt{n}}}$ suit la loi $\mathcal{N}(0,1)$.

On cherche à déterminer A et B tels que $P(A \leq m \leq B) = S$.



Puisqu'on choisit des risques symétriques, on va déterminer dans la table de la loi normale centrée réduite la valeur $t_{\alpha/2}$ telle que $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = S$, ce qui peut s'écrire

$$P(\bar{X} - t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}) = S,$$

qui est bien de la forme cherchée en posant

$$A = \bar{X} - t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}, \quad B = \bar{X} + t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}.$$

Signification. Avant toute expérience, la probabilité que l'intervalle aléatoire $[\bar{X} - t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}]$ contienne la vraie valeur de m est S. Ces deux limites sont des variables aléatoires qui prendront des valeurs numériques particulières une fois que l'échantillon est choisi et qu'on a obtenu la valeur de $\bar{x}$ (réalisation de la variable aléatoire $\bar{X}$). On en déduit par la suite un intervalle d'extrémités fixes (et non plus un intervalle aléatoire) qui s'écrit $[\bar{x} - t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}]$, et on lui attribue, non pas une probabilité, mais un niveau de confiance de α de contenir la vraie valeur de m.

Conclusion. A partir d'un échantillon de grande taille ($n \geq 30$) ou à partir d'un échantillon de petite taille ($n < 30$), prélevé à partir d'une population normale de moyenne m (inconnue) et de variance σ^2_{pop} connue, on définit un intervalle de confiance ayant un niveau de confiance S de contenir la vraie valeur de m par

$$[\bar{x} - t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{\sigma_{pop}}{\sqrt{n}}].$$

Remarque 109 *Dans le cas d'un grand échantillon, si la variance σ^2_{pop} de la population est inconnue, on peut l'estimer sans problème par la variance d'échantillon $s^2 = \frac{n}{n-1}\sigma^2_{ech}$ (voir chapitre 4).*

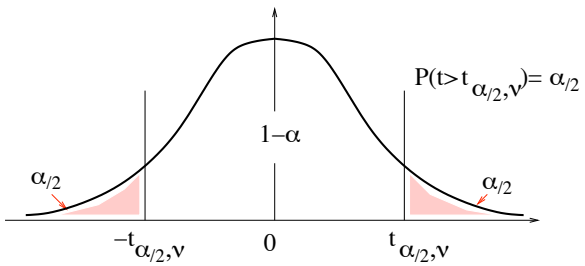
b. On dispose d'un petit échantillon ($n < 30$) et la distribution de X est normale d'écart-type inconnu

Dans ces conditions, l'étude du chapitre 4 nous a appris que nous ne disposions pas directement de la loi de $\bar{X}$ mais de celle de

$$T = \frac{\bar{X} - m}{\Sigma_{ech}/\sqrt{n-1}}.$$

T suit une loi de Student à $n-1$ degrés de liberté : $T \rightharpoonup T_{n-1}$.

Pour trouver l'intervalle de confiance de m au risque α, nous allons procéder comme dans le cas précédent.



On détermine dans la table de la loi de Student la valeur $t_{\alpha/2,\nu}$ (où $\nu = n-1$) telle que $P(-t_{\alpha/2,\nu} \leq T \leq t_{\alpha/2,\nu}) = S$, ce qui peut s'écrire

$$P(\bar{X} - t_{\alpha/2,\nu}\frac{\Sigma_{ech}}{\sqrt{n-1}}, \bar{X} + t_{\alpha/2,\nu}\frac{\Sigma_{ech}}{\sqrt{n-1}}) = S.$$

Après avoir choisi l'échantillon, $\bar{X}$ a pris la valeur $\bar{x}$ et Σ_{ech} la valeur σ_{ech}. On en déduit par la suite un intervalle d'extrémités fixes (et non plus un intervalle aléatoire) qui s'écrit $[\bar{x} - t_{\alpha/2,\nu}\frac{\sigma_{ech}}{\sqrt{n-1}}, \bar{x} + t_{\alpha/2,\nu}\frac{\sigma_{ech}}{\sqrt{n-1}}]$ et on lui attribue, non pas une probabilité, mais un niveau de confiance de α de contenir la vraie valeur de m.

Conclusion. A partir d'un échantillon de petite taille ($n < 30$), prélevé à partir d'une population normale de moyenne m (inconnue) et de variance σ^2_{pop} inconnue, on définit un intervalle de confiance ayant un niveau de confiance S de contenir la vraie valeur de m par

$$[\bar{x} - t_{\alpha/2,\nu}\frac{\sigma_{ech}}{\sqrt{n-1}}, \bar{x} + t_{\alpha/2,\nu}\frac{\sigma_{ech}}{\sqrt{n-1}}],$$

où $\nu = n - 1$ est le nombre de degrés de liberté de la distribution de Student.

On pourra bien sûr remplacer $\frac{\sigma_{ech}}{\sqrt{n-1}}$ par $\frac{s}{\sqrt{n}}$.

5.3.3 Remarques

1. L'intervalle de confiance pourra être numériquement différent chaque fois qu'on prélève un échantillon de même taille de la population puisque l'intervalle est centré sur la moyenne de l'échantillon qui varie de prélèvement en prélèvement.

2. Le niveau de confiance est associé à l'intervalle et non au paramètre m. Il ne faut pas dire que la vraie valeur de m a, disons 95 chances sur 100, de se trouver dans l'intervalle mais plutôt que l'intervalle de confiance a 95 chances sur 100 de contenir la vraie valeur de m ou encore que 95 fois sur 100, l'intervalle déterminé contiendra la vraie valeur de m. Une fois que l'intervalle est calculé, m est ou n'est pas dans l'intervalle (pour une population donnée, m est une constante et non une variable aléatoire).

3. Plus le niveau de confiance est élevé, plus l'amplitude de l'intervalle est grande. Pour la même taille d'échantillon, on perd de la précision en gagnant une plus grande confiance.

4. Dans le cas où la variance de la population est inconnue, des échantillonnages successifs de la population peuvent conduire pour une même taille d'échantillon et le même niveau de confiance, à des intervalles de diverses amplitudes parce que l'écart-type s variera d'échantillon en échantillon.

5.3.4 Estimation d'une variance par intervalle de confiance

On se propose d'estimer, par intervalle de confiance, la variance σ^2_{pop} d'un caractère mesurable d'une population. Il s'agit donc de déterminer, à partir de la variance de l'échantillon σ^2_{ech}, un intervalle dans lequel il est vraisemblable que la vraie valeur de σ^2_{pop} se trouve.

On cherche un intervalle $[A, B]$ vérifiant $P(A \leq \sigma^2_{pop} \leq B) = S$. Les limites de cet intervalle prendront, après avoir prélevé l'échantillon et calculé l'estimation les valeurs prises par les deux quantités aléatoires A et B, la forme $a \leq \sigma^2_{pop} \leq b$.

Nous allons déterminer A et B en utilisant la distribution d'échantillonnage de la variance d'échantillon S^2.

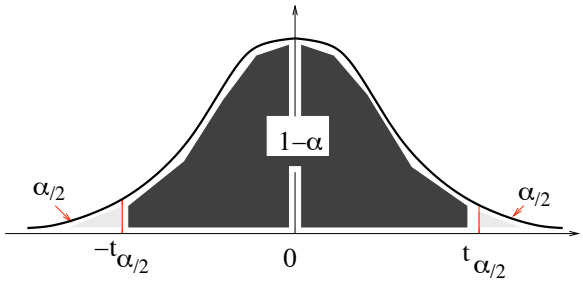
Nous supposerons par la suite que la population est "normale", c'est-à-dire que le caractère X suit une loi normale. L'étude du chapitre 4 nous amène donc à distinguer deux cas.

a. La population est "normale" et on dispose d'un grand échantillon ($n \geq 30$)

La variance d'échantillon $S^2 = \frac{n}{n-1}\Sigma^2_{ech}$ suit approximativement une loi normale (voir chapitre 4), $S^2 \rightharpoonup \mathcal{N}(\sigma^2_{pop}, \sigma^2_{pop}\sqrt{\frac{2}{n-1}})$, donc

$$T = \frac{S^2 - \sigma^2_{pop}}{\sigma^2_{pop}\sqrt{\frac{2}{n-1}}}$$

suit une loi normale centrée réduite.



On peut déterminer dans la table de la loi normale centrée réduite la valeur $t_{\alpha/2}$ telle que $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = S$, ce qui peut s'écrire

$$P(-t_{\alpha/2} \leq \frac{S^2 - \sigma^2_{pop}}{\sigma^2_{pop}\sqrt{\frac{2}{n-1}}} \leq t_{\alpha/2}) = 1 - \alpha.$$

Comme on a un grand échantillon, on peut estimer σ^2_{pop} par $s^2 = \frac{n}{n-1}\sigma^2_{ech}$. Soit encore

$$P(S^2 - t_{\alpha/2}s^2\sqrt{\frac{2}{n-1}} \leq \sigma^2_{pop} \leq S^2 + t_{\alpha/2}s^2\sqrt{\frac{2}{n-1}}) = 1 - \alpha,$$

qui est bien de la forme cherchée.

Ces deux limites sont des variables aléatoires qui prendront des valeurs numériques particulières une fois que l'échantillon est choisi et qu'on a obtenu la valeur de s^2 (réalisation de la variable aléatoire S^2). On en déduit par la suite un intervalle d'extrémités fixes (et non plus un intervalle aléatoire) qui s'écrit $[s^2 - t_{\alpha/2}s^2\sqrt{\frac{2}{n-1}} \leq \sigma^2_{pop} \leq s^2 + t_{\alpha/2}s^2\sqrt{\frac{2}{n-1}}]$, et on lui attribue un niveau de confiance S de contenir la vraie valeur de σ^2_{pop}.

Conclusion. A partir d'un échantillon de grande taille ($n \geq 30$), prélevé à partir d'une population normale de variance σ^2_{pop} inconnue, on définit un intervalle de confiance ayant un niveau de confiance $1 - \alpha$ de contenir la vraie valeur de σ^2_{pop} par

$$[s^2 - t_{\alpha/2}s^2\sqrt{\frac{2}{n-1}} \leq \sigma^2_{pop} \leq s^2 + t_{\alpha/2}s^2\sqrt{\frac{2}{n-1}}].$$

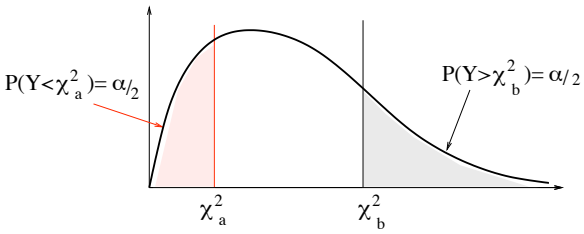
b. La population est "normale" et on dispose d'un petit échantillon ($n < 30$)

La variable

$$Y = \frac{n\Sigma^2_{ech}}{\sigma^2_{pop}} = \frac{(n-1)S^2}{\sigma^2_{pop}}$$

suit une loi du χ^2 à $n-1$ degrés de liberté (voir chapitre 4), $Y \rightharpoonup \chi^2_{n-1}$.

Nous allons chercher un intervalle $[\chi^2_a, \chi^2_b]$ de valeurs telles que $P(\chi^2_a \leq Y \leq \chi^2_b) = S$.



On choisit un intervalle correspondant à des risques symétriques, c'est-à-dire tel que

$$P(Y < \chi^2_a) = P(\chi^2_b < Y) = \frac{1 - S}{2} = \frac{\alpha}{2}.$$

Les deux valeurs χ_a^2 et χ_b^2 se déterminent à l'aide des tables. On peut alors écrire que $P(\chi_a^2 \leq \frac{n\Sigma_{ech}^2}{\sigma_{pop}^2} \leq \chi_b^2) = S$ et donc que

$$P(\frac{n\Sigma_{ech}^2}{\chi_b^2} \leq \sigma_{pop}^2 \leq \frac{n\Sigma_{ech}^2}{\chi_a^2}) = S.$$

Ces deux limites sont des variables aléatoires qui prendront des valeurs numériques particulières une fois que l'échantillon est choisi et qu'on a obtenu la valeur de s^2 (réalisation de la variable aléatoire S^2). On en déduit par la suite un intervalle d'extrémités fixes qui s'écrit $[\frac{n\sigma_{ech}^2}{\chi_b^2}, \frac{n\sigma_{ech}^2}{\chi_a^2}]$ et on lui attribue un niveau de confiance S de contenir la vraie valeur de σ_{pop}^2.

Conclusion. A partir d'un échantillon de petite taille ($n < 30$), prélevé à partir d'une population normale de variance σ_{pop}^2 inconnue, on définit un intervalle de confiance ayant un niveau de confiance S de contenir la vraie valeur de σ_{pop}^2 par

$$[\frac{n\sigma_{ech}^2}{\chi_b^2}, \frac{n\sigma_{ech}^2}{\chi_a^2}].$$

5.3.5 Estimation d'une proportion par intervalle de confiance

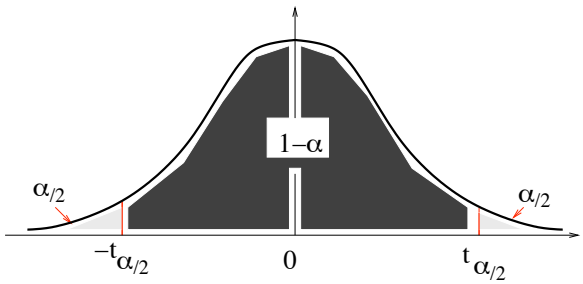
On se propose d'estimer, par intervalle de confiance, la proportion p d'un caractère quantitatif d'une population. Il s'agit donc de déterminer, à partir de la proportion de l'échantillon f, un intervalle dans lequel il est vraisemblable que la vraie valeur de p s'y trouve. On cherche un intervalle $[A, B]$ vérifiant $P(A \leq p \leq B) = S$. Les limites de cet intervalle prendront, après avoir prélevé l'échantillon et calculé les valeurs prises par les deux quantités aléatoires A et B, la forme $LI \leq p \leq LS$.

Nous allons déterminer A et B en utilisant la distribution d'échantillonnage de la proportion d'échantillon F.

Nous supposons que nous sommes en présence d'un *grand échantillon* ($n \geq 30$) et que p (que nous devons estimer) n'est pas trop petit ($np \geq 15$ et $nq \geq 15$). La fréquence d'échantillon F suit approximativement une loi normale (voir chapitre 4), $F \rightharpoonup \mathcal{N}(p, \sqrt{\frac{pq}{n}})$. Donc

$$T = \frac{F - p}{\sqrt{\frac{pq}{n}}}$$

suit approximativement une loi normale centrée réduite.



On peut déterminer dans la table de la loi normale centrée réduite la valeur $t_{\alpha/2}$ telle que $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = S$. Ce qui peut s'écrire :

$$P(-t_{\alpha/2} \leq \frac{F - p}{\sqrt{\frac{pq}{n}}} \leq t_{\alpha/2}) = S.$$

Le problème est qu'on ignore la valeur de p et qu'elle intervient dans l'écart-type. Comme n est grand, il est correct d'estimer p par la valeur f (prise par l'estimateur F) trouvée dans l'échantillon.

En effet, la grande taille de l'échantillon garantit que f ne fluctue pas trop d'échantillon en échantillon. Soit encore

$$P(F - t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \leq p \leq F + t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}) = S.$$

qui est bien de la forme cherchée.

Ces deux limites sont des variables aléatoires qui prendront des valeurs numériques particulières une fois que l'échantillon est choisi et qu'on a obtenu la valeur de f (réalisation de la variable aléatoire F). On en déduit par la suite un intervalle d'extrémités fixes qui s'écrit $[f - t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}, f + t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}]$ et on lui attribue un niveau de confiance S de contenir la vraie valeur de p.

Conclusion. A partir d'un échantillon de grande taille ($n \geq 30$), prélevé à partir d'une population dont la proportion p d'un caractère qualitatif est inconnue mais pas trop petite, on définit un intervalle de confiance ayant un niveau de confiance S de contenir la vraie valeur de p par

$$[f - t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}, f + t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}].$$

5.3.6 Comment contrôler l'erreur ?

Il arrive souvent que la précision de l'estimation soit spécifiée avant même que l'échantillon ne soit prélevé. Par exemple, vous voulez vérifier un lot de pièces de machinerie : ces pièces doivent avoir un certain diamètre et l'erreur tolérée dans la fabrication doit être très petite, sinon plusieurs d'entre elles seront inutilisables. Pour vérifier le lot, vous prélevez un échantillon, mais vous voulez que l'estimation se fasse avec la plus petite erreur d'échantillonnage possible : vous voulez une estimation précise. D'une trop grande erreur d'échantillonnage résulte une longueur d'intervalle trop grande et cela rend souvent inutile l'intervalle de confiance construit.

Nous pouvons contrôler l'erreur d'échantillonnage en choisissant une taille d'échantillon appropriée. L'erreur d'échantillonnage survient lorsque l'échantillon ne prend pas en considération la population dans sa totalité. Chaque fois qu'un échantillon est prélevé, nous perdons une certaine partie de l'information concernant la population, ce qui entraîne immanquablement une erreur dans l'estimation. Par conséquent, si nous voulons un très haut niveau de précision, nous devons prélever un échantillon dont la taille permet d'extraire de la population l'information suffisante pour réaliser l'estimation avec la précision désirée.

Nous verrons en travaux dirigés sur des exemples comment procéder.

Chapitre 6

Les tests d'hypothèse

6.1 Généralités

6.1.1 Principe d'un test d'hypothèses

Les tests d'hypothèse constituent un autre aspect important de l'inférence statistique. Le principe général d'un test d'hypothèse peut s'énoncer comme suit :

- On étudie une population dont les éléments possèdent un caractère (mesurable ou qualitatif) et dont la valeur du paramètre relative au caractère étudié est inconnue.
- Une hypothèse est formulée sur la valeur du paramètre : cette formulation résulte de considérations théoriques, pratiques ou encore elle est simplement basée sur un pressentiment.
- On veut porter un jugement sur la base des résultats d'un échantillon prélevé de cette population.

Il est bien évident que la statistique (c'est-à-dire la variable d'échantillonnage) servant d'estimateur au paramètre de la population ne prendra pas une valeur rigoureusement égale à la valeur théorique proposée dans l'hypothèse. Cette variable aléatoire comporte des fluctuations d'échantillonnage qui sont régies par des distributions connues.

Pour décider si l'hypothèse formulée est supportée ou non par les observations, il faut une méthode qui permettra de conclure si l'écart observé entre la valeur de la statistique obtenue dans l'échantillon et celle du paramètre spécifiée dans l'hypothèse est trop important pour être uniquement imputable au hasard de l'échantillonnage.

La construction d'un test d'hypothèse consiste en fait à déterminer entre quelles valeurs peut varier la variable aléatoire, en supposant l'hypothèse vraie, sur la seule considération du hasard de l'échantillonnage.

Les distributions d'échantillonnage d'une moyenne, d'une variance et d'une proportion que nous avons traitées dans un chapitre précédent vont être particulièrement utiles dans l'élaboration des tests statistiques.

6.1.2 Définition des concepts utiles à l'élaboration des tests d'hypothèse

Hypothèse statistique.

Une *hypothèse statistique* est un énoncé (une affirmation) concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations) d'une population.

Test d'hypothèse.

Un *test d'hypothèse* (ou test statistique) est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques.

Hypothèse nulle (H_0) et hypothèse alternative (H_1).

L'hypothèse selon laquelle on fixe à priori un paramètre de la population à une valeur particulière s'appelle l'*hypothèse nulle* et est notée H_0. N'importe quelle autre hypothèse qui diffère de l'hypothèse H_0 s'appelle l'*hypothèse alternative* (ou contre-hypothèse) et est notée H_1.

C'est l'hypothèse nulle qui est soumise au test et toute la démarche du test s'effectue en considérant cette hypothèse comme vraie.

Dans notre démarche, nous allons établir des règles de décision qui vont nous conduire à l'acceptation ou au rejet de l'hypothèse nulle H_0. Toutefois cette décision est fondée sur une information partielle, les résultats d'un échantillon. Il est donc statistiquement impossible de prendre la bonne décision à coup sûr. En pratique, on met en oeuvre une démarche qui nous permettrait, à long terme de rejeter à tort une hypothèse nulle vraie dans une faible proportion de cas. La conclusion qui sera déduite des résultats de l'échantillon aura un caractère probabiliste : on ne pourra prendre une décision qu'en ayant conscience qu'il y a un certain risque qu'elle soit erronée. Ce risque nous est donné par le seuil de signification du test.

Seuil de signification du test

Le risque, consenti à l'avance et que nous notons α, de rejeter à tort l'hypothèse nulle H_0 alors qu'elle est vraie, s'appelle le *seuil de signification* du test et s'énonce en probabilité ainsi,

$$\alpha = P(\text{rejeter } H_0 \mid H_0 \text{ vraie}).$$

A ce seuil de signification, on fait correspondre sur la distribution d'échantillonnage de la statistique une *région de rejet* de l'hypothèse nulle (appelée également région critique). L'aire de cette région correspond à la probabilité α. Si par exemple on choisit $\alpha = 0.05$, cela signifie que l'on admet d'avance que la variable d'échantillonnage peut prendre, dans 5% des cas, une valeur se situant dans la zone de rejet de H_0, bien que H_0 soit vraie et ceci uniquement d'après le hasard de l'échantillonnage. Sur la distribution d'échantillonnage correspondra aussi une région complémentaire, dite *région d'acceptation* de H_0 (ou région de non-rejet) de probabilité $1 - \alpha$.

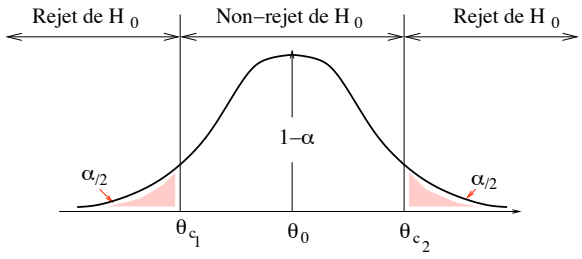
Remarque 110 *1. Les seuils de signification les plus utilisés sont $\alpha = 0.05$ et $\alpha = 0.01$, dépendant des conséquences de rejeter à tort l'hypothèse H_0.*

2. La statistique qui convient pour le test est donc une variable aléatoire dont la valeur observée sera utilisée pour décider du « rejet » ou du « non-rejet » de H_0. La distribution d'échantillonnage de cette statistique sera déterminée en supposant que l'hypothèse H_0 est vraie.

Exemple 111 *Supposons que nous affirmions que la valeur d'un paramètre θ d'une population est égale à la valeur θ_0. On s'intéresse au changement possible du paramètre θ dans l'une ou l'autre direction (soit $\theta > \theta_0$, soit $\theta < \theta_0$). On effectue un test bilatéral.*

Les hypothèses H_0 et H_1 sont alors $\begin{cases} H_0 & \theta = \theta_0 \\ H_1 & \theta \neq \theta_0. \end{cases}$

On peut schématiser les régions de rejet et de non-rejet de H_0 comme suit :



Si, suite aux résultats de l'échantillon, la valeur de la statistique utilisée se situe dans l'intervalle $[\theta_{c_1}, \theta_{c_2}]$, on acceptera H_0 au seuil de signification choisi. Si, au contraire, la valeur obtenue est supérieure à c_2 ou inférieure à c_1, on rejette H_0 et on accepte H_1.

Remarque 112 *Si on s'intéresse au changement du paramètre dans une seule direction, on opte pour un test unilatéral, en choisissant comme hypothèse H_1 soit $\theta > \theta_0$, soit $\theta < \theta_0$. La région critique est alors localisée uniquement à droite ou uniquement à gauche de la région d'acceptation.*

Dans un souci de simplification, nous nous intéresserons dans ce cours essentiellement aux tests bilatéraux.

6.2 Tests permettant de déterminer si un échantillon appartient à une population donnée

6.2.1 Test sur une moyenne : comparaison d'une moyenne expérimentale à une moyenne théorique dans le cas d'un caractère quantitatif

Nous voulons déterminer si l'échantillon de taille n dont nous disposons appartient à une population de moyenne m_0 au seuil de signification α. Nous allons dans tous les tests travailler de la même façon, en procédant en quatre étapes.

1ère étape : Formulation des hypothèses.

L'échantillon dont nous disposons provient d'une population de moyenne m. Nous voulons savoir si m $= m_0$. On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 : $\begin{cases} H_0 & m = m_0 \\ H_1 & m \neq m_0. \end{cases}$.

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

- On détermine la statistique qui convient pour ce test. Ici, l'estimateur de la moyenne m, c'est-à-dire $\bar{X}$, semble tout indiqué.
- On détermine la loi de probabilité de $\bar{X}$ en se plaçant sous l'hypothèse H_0. Deux cas peuvent se produire.

Premier cas : L'échantillon est de grande taille (ou bien la population est normale de variance σ^2_{pop} connue).

$\bar{X}$ suit alors une loi normale de moyenne m_0 (puisqu'on se place sous H_0) et d'écart-type $\frac{\sigma_{pop}}{\sqrt{n}}$, $\bar{X} \rightharpoonup \mathcal{N}(m_0, \frac{\sigma_{pop}}{\sqrt{n}})$. On pose

$$T = \frac{\bar{X} - m_0}{\frac{\sigma_{pop}}{\sqrt{n}}}.$$

T mesure un écart réduit. T est aussi appelée *fonction discriminante du test*. $T \rightharpoonup \mathcal{N}(0, 1)$.

Deuxième cas : L'échantillon est de petite taille (prélevé au hasard d'une population normale de variance σ^2_{pop} inconnue).

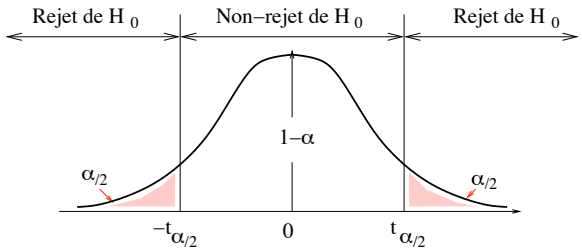
Dans ce cas la fonction discriminante du test sera

$$T = \frac{\bar{X} - m_0}{\frac{\Sigma_{ech}}{\sqrt{n-1}}}.$$

Ici $T \rightharpoonup T_{n-1}$ (loi de Student à $n-1$ degrés de liberté).

3ème étape : Détermination des valeurs critiques de Tdélimitant les zones d'acceptation et de rejet.

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.



Il nous faut donc déterminer dans la table la valeur maximale $t_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire vérifiant $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$.

4ème étape : Calcul de la valeur de T prise dans l'échantillon et conclusion du test.
On calcule la valeur t_0 prise par T dans l'échantillon.
- Si la valeur t_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est statistiquement significatif au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.
- Si la valeur t_0 se trouve dans la zone d'acceptation, on dira que l'écart-réduit observé n'est pas significatif au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

6.2.2 Tests sur une proportion

Nous nous proposons de tester si la proportion p d'éléments dans la population présentant un certain caractère qualitatif peut être ou non considérée comme égale à une valeur hypothétique p_0. Nous disposons pour ce faire de la proportion d'éléments possédant ce caractère dans un échantillon de taille n. Nous allons procéder comme au paragraphe précédent, en quatre étapes.

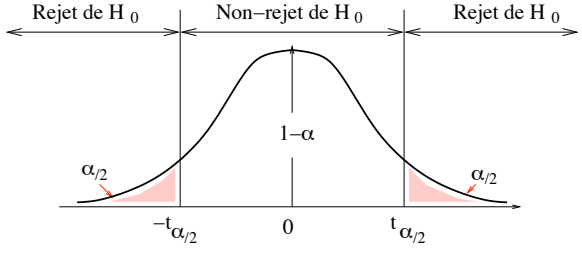
1ère étape : Formulation des hypothèses.
L'échantillon dont nous disposons provient d'une population dont la proportion d'éléments présentant le caractère qualitatif est p. Nous voulons savoir si $p = p_0$. On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 : $\begin{cases} H_0 & p = p_0 \\ H_1 & p \neq p_0. \end{cases}$.

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.
On détermine la statistique qui convient pour ce test. Ici, l'estimateur de la proportion p, c'est-à-dire F, semble tout indiquée.
On détermine la loi de probabilité de F *en se plaçant sous l'hypothèse* H_0. On suppose que l'on dispose d'un grand échantillon (et que "p n'est pas trop petit" (de manière que l'on ait $np \geq 15$ et $n(1-p) \geq 15$). F suit alors une loi normale de moyenne p_0 (puisqu'on se place sous H_0) et d'écart-type $\sqrt{\frac{p_0(1-p_0)}{n}}$, $F \rightharpoonup \mathcal{N}(p_0, \sqrt{\frac{p_0(1-p_0)}{n}})$.
On pose

$$T = \frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

T mesure un écart réduit. T est aussi appelée fonction discriminante du test. $T \rightharpoonup \mathcal{N}(0, 1)$.

3ème étape : Détermination des valeurs critiques de T délimitant les zones d'acceptation et de rejet.
On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.



Il nous faut donc déterminer dans la table la valeur maximale $t_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire vérifiant $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$.

4ème étape : Calcul de la valeur de T prise dans l'échantillon et conclusion du test.
On calcule la valeur t_0 prise par T dans l'échantillon.

- Si la valeur t_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est statistiquement significatif au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.
- Si la valeur t_0 se trouve dans la zone d'acceptation, on dira que l'écart-réduit observé n'est pas significatif au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

Nous étudierons ces sortes de tests sur des exemples en travaux dirigés.

6.3 Risques de première et de deuxième espèce

6.3.1 Définitions

Tous les règles de décision que nous avons déterminées acceptaient un risque α qui était le risque de rejeter à tort l'hypothèse H_0, c'est-à-dire le risque de rejeter l'hypothèse H_0, alors que H_0 est vraie. Ce risque s'appelle aussi le *risque de première espèce.*

La règle de décision du test comporte également un deuxième risque, à savoir de celui de ne pas rejeter l'hypothèse nulle H_0 alors que c'est l'hypothèse H_1 qui est vraie. C'est le *risque de deuxième espèce.*

Les deux risques peuvent se définir ainsi :

α = $P(\text{rejeter}H_0 \mid H_0\text{vraie})$ = probabilité de commettre une erreur de première espèce.
β = $P(\text{ne pas rejeter}H_0 \mid H_1\text{vraie})$ = probabilité de commettre une erreur de deuxième espèce.

Le risque de première espèce α est choisi à priori. Toutefois le risque de deuxième espèce β dépend de l'hypothèse alternative H_1 et on ne peut le calculer que si on spécifie des valeurs particulières du paramètre dans l'hypothèse H_1 que l'on suppose vraie.

Les risques liés aux tests d'hypothèses peuvent se résumer ainsi :

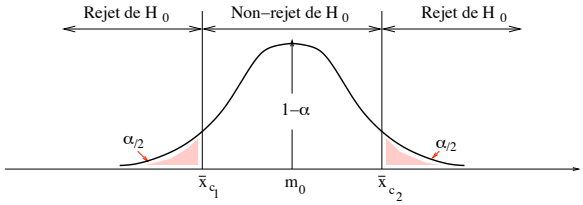
		Conclusion	**du test**
		Accepter H_0	**Rejeter** H_0
H_0 est vraie	La décision est	**Bonne**	**Fausse**
	probabilité de prendre cette décision avant l'expérience	$1 - \alpha$	α
H_1 est vraie	La décision est	**Fausse**	**Bonne**
	probabilité de prendre cette décision avant l'expérience	β	$1 - \beta$

Remarque 113 *La probabilité complémentaire du risque de deuxième espèce $(1 - \beta)$ définit la* puissance du test *à l'égard de la valeur du paramètre dans l'hypothèse alternative H_1. La puissance du test représente la probabilité de rejeter l'hypothèse nulle H_0 lorsque l'hypothèse vraie est H_1. Plus β est petit, plus le test est puissant.*

6.3.2 Schématisation des deux risques d'erreur sur la distribution d'échantillonnage

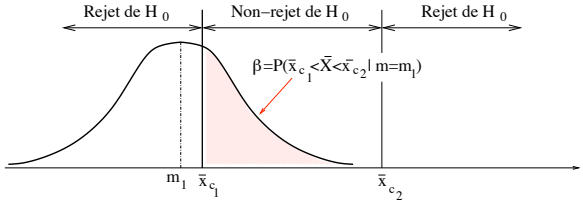
A titre d'exemple, regardons ce qu'il se passe à propos d'un test sur la moyenne. On peut visualiser sur la distribution d'échantillonnage de la moyenne comment sont reliés les deux risques d'erreur associés aux tests d'hypothèses.

Les zones d'acceptation de H_0 $(m = m_0)$ et de rejet de H_0 se visualisent ainsi :

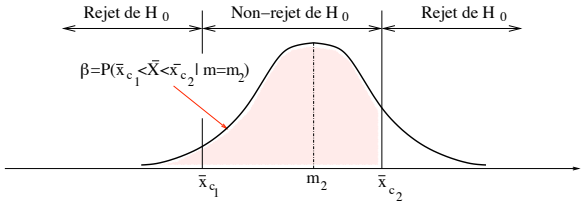


Donnons diverses valeurs à m (autres que m_0) que l'on suppose vraie et schématisons le risque de deuxième espèce β.

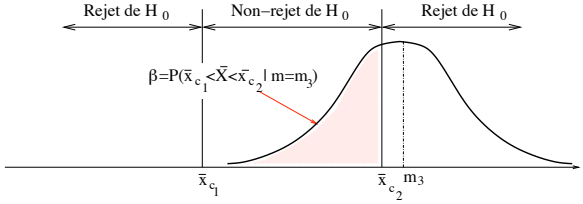
Hypothèse vraie : $m = m_1$ $(m_1 < m_0)$ La distribution d'échantillonnage de $\bar{X}$ en supposant vraie $m = m_1$ est illustrée en pointillé et l'aire hachurée sur cette figure correspond à la région de non-rejet de H_0. Cette aire représente β par rapport à la valeur m_1.



Hypothèse vraie : $m = m_2$ $(m_2 > m_0)$



Hypothèse vraie : $m = m_3$ $(m_3 > m_0)$



Cette schématisation permet d'énoncer quelques propriétés importantes concernant les deux risques d'erreur.

1. Pour un même risque α et une même taille d'échantillon, on constate que, si l'écart entre la valeur du paramètre posée en H_0 et celle supposée dans l'hypothèse vraie H_1 augmente, le risque β diminue.
2. Une réduction du risque de première espèce (de $\alpha = 0.05$ à $\alpha = 0.01$ par exemple) élargit la zone d'acceptation de H_0. Toutefois, le test est accompagné d'une augmentation du risque de deuxième espèce β. On ne peut donc diminuer l'un des risques qu'en consentant à augmenter l'autre.
3. Pour une valeur fixe de α et un σ déterminé, l'augmentation de la taille d'échantillon aura pour effet de donner une meilleure précision puisque $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ diminue. La zone d'acceptation de H_0 sera alors plus restreinte, conduisant à une diminution du risque β. Le test est alors plus puissant.

6.4 Tests permettant de déterminer si deux échantillons appartiennent à la même population

6.4.1 Introduction

Il existe de nombreuses applications qui consistent, par exemple, à comparer deux groupes d'individus en regard d'un caractère quantitatif particulier (poids, taille, rendement scolaire, quotient intellectuel,....) ou à comparer deux procédés de fabrication selon une caractéristique quantitative particulière (résistance à la rupture, poids, diamètre, longueur,...) ou encore de comparer les proportions d'apparition d'un caractère qualitatif de deux populations (proportion de défectueux, proportion de gens favorisant un parti politique,...). Les variables aléatoires qui sont alors utilisées pour effectuer des tests d'hypothèses (ou aussi calculer des intervalles de confiance) sont la *différence des moyennes* d'échantillon, le *quotient des variances* d'échantillon ou la *différence des proportions* d'échantillon.

6.4.2 Comparaison de deux moyennes d'échantillon : "test T"

Nous nous proposons de tester si la moyenne de la première population (m_1) peut être ou non considérée comme égale à la moyenne de la deuxième population (m_2). Nous allons alors comparer les deux moyennes d'échantillon $\bar{x}_1$ et $\bar{x}_2$. Il est évident que si $\bar{x}_1$ et $\bar{x}_2$ diffèrent beaucoup, les deux échantillons n'appartiennent pas la même population. Mais si $\bar{x}_1$ et $\bar{x}_2$ diffèrent peu, il se pose la question de savoir si l'écart $d = \bar{x}_1 - \bar{x}_2$ peut être attribué aux hasards de l'échantillonnage. Afin de donner une réponse rigoureuse à cette question, nous procéderons encore en quatre étapes.

1ère étape : Formulation des hypothèses.

Le premier échantillon dont nous disposons provient d'une population dont la moyenne est m_1. Le deuxième échantillon dont nous disposons provient d'une population dont la moyenne est m_2.

Nous voulons savoir si il s'agit de la même population en ce qui concerne les moyennes, c'est-à-dire si $m_1 = m_2$. On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 : $\begin{cases} H_0 & m_1 = m_2 \\ H_1 & m_1 \neq m_2. \end{cases}$.

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

On détermine la statistique qui convient pour ce test. Ici, la différence $D = \bar{X}_1 - \bar{X}_2$ des deux moyennes d'échantillon, semble tout indiquée.

On détermine la loi de probabilité de D *en se plaçant sous l'hypothèse* H_0. On suppose que l'on dispose de grands échantillons ($n_1 \geq 30$ et $n_2 \geq 30$). $\bar{X}_1$ suit alors une loi normale de moyenne m_1 et d'écart-type $\frac{\sigma_{pop1}}{\sqrt{n_1}}$ que l'on peut sans problème estimer par $\frac{\sigma_{ech1}}{\sqrt{n_1 - 1}}$ (car $n_1 \geq 30$). I.e. $\bar{X}_1 \rightharpoonup \mathcal{N}(m_1, \frac{\sigma_{ech1}}{\sqrt{n_1 - 1}})$.

De même $\bar{X}_2$ suit alors une loi normale de moyenne m_2 et d'écart-type $\frac{\sigma_{pop2}}{\sqrt{n_2}}$ que l'on peut sans problème estimer par $\frac{\sigma_{ech2}}{\sqrt{n_2 - 1}}$ (car $n_2 \geq 30$). I.e. $\bar{X}_2 \rightharpoonup \mathcal{N}(m_2, \frac{\sigma_{ech2}}{\sqrt{n_2 - 1}})$.

On en déduit, puisque $\bar{X}_1$ et $\bar{X}_2$ sont indépendantes que D suit également une loi normale.

$E(D) = E(\bar{X}_1) - E(\bar{X}_2) = m_1 - m_2 = 0$ puisqu'on se place sous H_0.

$E(D) = V(\bar{X}_1) + V(\bar{X}_2) = \dfrac{\sigma^2_{ech1}}{n_1 - 1} + \dfrac{\sigma^2_{ech2}}{n_2 - 1}$ puisque les variables sont indépendantes.

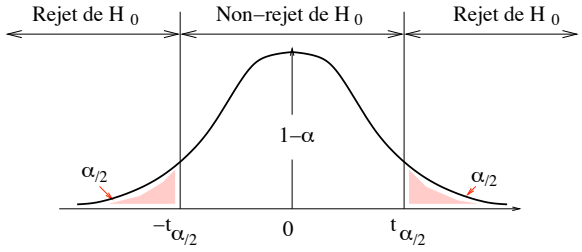
On pose

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma^2_{ech1}}{n_1 - 1} + \frac{\sigma^2_{ech2}}{n_2 - 1}}}.$$

T mesure un écart réduit. T est la fonction discriminante du test. $T \rightharpoonup \mathcal{N}(0, 1)$.

3ème étape : Détermination des valeurs critiques de T délimitant les zones d'acceptation et de rejet.

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.



Il nous faut donc déterminer dans la table la valeur maximale $t_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire vérifiant $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$.

4ème étape : Calcul de la valeur de T prise dans l'échantillon et conclusion du test.
On calcule la valeur t_0 prise par T dans l'échantillon.
- Si la valeur t_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est *statistiquement significatif* au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.
- Si la valeur t_0 se trouve dans la zone d'acceptation, on dira que l'écart-réduit observé *n'est pas significatif* au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

Remarque 114 *Si on travaille sur de petits échantillons, si la loi suivie par la grandeur est une loi normale et si on ignore les écarts-type des populations, on doit utiliser la loi de Student.*

6.4.3 Comparaison de deux variances d'échantillon : "test F"

1ère étape : Formulation des hypothèses.
Le premier échantillon dont nous disposons provient d'une population dont l'écart-type est σ_{pop1}. Le deuxième échantillon dont nous disposons provient d'une population dont l'écart-type est σ_{pop2}. Nous voulons savoir si il s'agit de la même population en ce qui concerne les écarts-type, c'est-à-dire si $\sigma_{pop1} = \sigma_{pop2}$. On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 :
$\begin{cases} H_0 & \sigma_{pop1} = \sigma_{pop2} \\ H_1 & \sigma_{pop1} \neq \sigma_{pop2}. \end{cases}$.

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.
On détermine la statistique qui convient pour ce test. Ici, la variable aléatoire dont on connaît la loi est le rapport $F = \dfrac{S_1^2}{S_2^2}$ où S_1^2 et S_2^2 sont les variables aléatoires variances d'échantillon.
On détermine la loi de probabilité de F *en se plaçant sous l'hypothèse* H_0.
On suppose ici que les deux populations dont nous avons tiré les échantillons sont normales. Il en découle que
- $\dfrac{(n_1 - 1)S_1^2}{\sigma_{pop1}^2}$ suit la loi du khi-deux à $n_1 - 1$ degrés de liberté.
- De même, $\dfrac{(n_2 - 1)S_2^2}{\sigma_{pop2}^2}$ suit la loi du khi-deux à $n_2 - 1$ degrés de liberté.

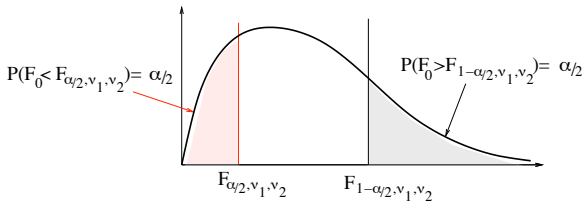
On considère alors le quotient

$$F_0 = \frac{\frac{S_1^2}{\sigma_{pop1}^2}}{\frac{S_2^2}{\sigma_{pop2}^2}}$$

qui est distribué suivant la loi de Fisher avec $\nu_1 = n_1 - 1$ et $\nu_2 = n_2 - 1$ degrés de liberté. Lorsqu'on se place sous l'hypothèse H_0, c'est le rapport $F_0 = \dfrac{S_1^2}{S_2^2}$ qui suit la loi de Fisher avec ν_1 et ν_2 degrés de liberté puisque $\sigma_{pop1} = \sigma_{pop2}$. Ici la *fonction discriminante du test* est F_0.

3ème étape : Détermination des valeurs critiques de F_0 délimitant les zones d'acceptation et de rejet.

On impose maintenant à la zone d'acceptation de H_0 concernant le quotient des deux variances d'échantillon d'être centrée autour de 1.



On détermine dans les tables les deux valeurs $F_{\alpha/2,\nu_1,\nu_2}$ et $F_{1-\alpha/2,\nu_1,\nu_2}$ telles que $P(F_{\alpha/2,\nu_1,\nu_2} < F_0 < F_{1-\alpha/2,\nu_1,\nu_2} = 1 - \alpha$.

On rejettera H_0 si la valeur f_0 prise par F_0 dans l'échantillon se trouve à l'extérieur de l'intervalle $[F_{\alpha/2,\nu_1,\nu_2}, F_{1-\alpha/2,\nu_1,\nu_2}]$.

Remarque 115 *On notera que pour obtenir la valeur critique inférieure de F_0, on doit utiliser la relation*

$$F_{1-\alpha/2,\nu_1,\nu_2} = \frac{1}{F_{\alpha/2,\nu_2,\nu_1}}.$$

4ème étape : Calcul de la valeur de F_0 prise dans l'échantillon et conclusion du test.

On calcule la valeur f_0 prise par F_0 dans l'échantillon.

- Si la valeur F_0 se trouve dans la zone de rejet, on dira que la valeur observée pour F est *statistiquement significative* au seuil α. Ce quotient est éloigné de 1 et ne permet pas d'accepter H_0. On rejette H_0.
- Si la valeur F_0 se trouve dans la zone d'acceptation, on dira que la valeur observée pour F *n'est pas significative* au seuil α. L'écart constaté par rapport à la valeur 1 attendue est imputable aux fluctuations d'échantillonnage. On accepte H_0.

6.4.4 Comparaison de deux proportions d'échantillon

Il y a de nombreuses applications (échéances électorales, expérimentations médicales...) où nous devons décider si l'écart observé entre deux proportions échantillonnales est significatif où s'il est attribuable au hasard de l'échantillonnage. Pour répondre à cette question, nous procéderons comme d'habitude en quatre étapes.

1ère étape : Formulation des hypothèses.

Le premier échantillon dont nous disposons provient d'une population 1 dont les éléments possèdent un caractère qualitatif dans une proportion inconnue p_1. Le deuxième échantillon dont nous disposons provient d'une population 2 dont les éléments possèdent le même caractère qualitatif dans une proportion inconnue p_2.

Nous voulons savoir si il s'agit de la même population en ce qui concerne les proportions, c'est-à-dire si $p_1 = p_2$. On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 : $\begin{cases} H_0 & p_1 = p_2 \\ H_1 & p_1 \neq p_2. \end{cases}$.

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

Nous traiterons uniquement le cas où nous sommes en présence de grands échantillons.

On détermine la statistique qui convient pour ce test. Ici, la différence $D = F_1 - F_2$ des deux proportions d'échantillon, semble tout indiquée, puisque F_1 est un estimateur sans biais de p_1 et F_2 un estimateur sans biais de p_2..

On détermine la loi de probabilité de D *en se plaçant sous l'hypothèse* H_0. F_1 suit alors une loi normale de moyenne p_1 et d'écart-type $\sqrt{\frac{p_1(1-p_1)}{n_1}}$.

De même, F_2 suit alors une loi normale de moyenne p_2 et d'écart-type $\sqrt{\frac{p_2(1-p_2)}{n_2}}$.

On en déduit, puisque F_1 et F_2 sont indépendantes que D suit également une loi normale.

$E(D) = E(F_1) - E(F_2) = p_1 - p_2 = 0$ puisqu'on se place sous H_0.

$V(D) = V(F_1) + V(F_2) = \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}$ puisque les variables sont indépendantes. Ici, on a posé $p_1 = p_2 = p$ puisque l'on se place sous H_0.

Mais comment trouver p puisque c'est justement sur p que porte le test ? Puisque nous raisonnons en supposant l'hypothèse H_0 vraie, on peut considérer que les valeurs de F_1 et F_2 obtenues sur nos échantillons sont des approximations de p. De plus, plus la taille de l'échantillon est grande, meilleure est l'approximation (revoir le chapitre sur les intervalles de confiance). Nous allons donc pondérer les valeurs observées dans nos échantillons par la taille respective de ces échantillons. On approchera p dans notre calcul par $\hat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$.

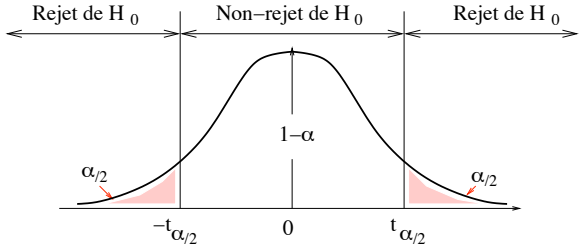
On pose

$$T = \frac{D}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}.$$

T mesure un écart réduit. T est la *fonction discriminante du test.* $T \rightharpoonup \mathcal{N}(0,1)$.

3ème étape : Détermination des valeurs critiques de T délimitant les zones d'acceptation et de rejet

On impose toujours à la zone d'acceptation de H_0 concernant l'écart réduit d'être centrée autour de 0.



Il nous faut donc déterminer dans la table la valeur maximale $t_{\alpha/2}$ de l'écart réduit imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire vérifiant $P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$.

4ème étape : Calcul de la valeur de T prise dans l'échantillon et conclusion du test

On calcule la valeur t_0 prise par T dans l'échantillon.

- Si la valeur t_0 se trouve dans la zone de rejet, on dira que l'écart-réduit observé est *statistiquement significatif* au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.
- Si la valeur t_0 se trouve dans la zone d'acceptation, on dira que l'écart-réduit observé *n'est pas significatif* au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

6.5 Test d'ajustement de deux distributions : "test du khi-deux"

6.5.1 Introduction

Dans le chapitre 1 de ce cours, nous avons traité de diverses distributions expérimentales dans lesquelles on présentait la répartition des fréquences (absolues ou relatives) pour divers caractères. Lorsque nous avons accumulé suffisamment de données sur une variable statistique, on peut alors examiner si la distribution des observations semble s'apparenter à une distribution théorique connue (comme une loi binomiale, de Poisson, normale...). Un outil statistique qui permet de vérifier la concordance entre une distribution expérimentale et une distribution théorique est le *test de Pearson*, appelé aussi le *test du khi-deux*.

On cherche donc à déterminer si un modèle théorique est susceptible de représenter adéquatement le comportement probabiliste de la variable observée, comportement fondé sur les fréquences des résultats obtenus sur l'échantillon.

Comment procéder ?

Répartitions expérimentales

On répartit les observations suivant k classes (si le caractère est continu) ou k valeurs (si le caractère est discret). On dispose alors des effectifs des k classes : $n_1, n_2, \ldots, n_k$. On a bien sûr la relation

$$\sum_{i=1}^{k} n_i = N,$$

où N est le nombre total d'observations effectuées.

Remarque 116 *Dans la pratique, on se placera dans le cas où $N \geq 50$ et où chaque n_i est supérieur ou égal à 5. Si cette condition n'est pas satisfaite, il y a lieu de regrouper deux ou plusieurs classes adjacentes. Il arrive fréquemment que ce regroupement s'effectue sur les classes aux extrémités de la distribution. k représente donc le nombre de classes après regroupement.*

Répartitions théoriques

En admettant comme plausible une distribution théorique particulière, on peut construire une répartition idéale des observations de l'échantillon de taille N en ayant recours aux probabilités tabulées (ou calculées) du modèle théorique : $p_1, p_2, \ldots, p_k$. On obtient alors les effectifs théoriques $n_{t,i}$ en écrivant $n_{t,i} = Np_i$. On dispose automatiquement de la relation $\sum_{i=1}^{k} n_{t,i} = N$.

Définition de l'écart entre les deux distributions

Pour évaluer l'écart entre les effectifs observés n_i et les effectifs théoriques $n_{t,i}$, on utilise la somme des écarts normalisés entre les deux distributions, à savoir

$$\chi^2 = \frac{(n_1 - n_{t,1})^2}{n_{t,1}} + \frac{(n_2 - n_{t,2})^2}{n_{t,2}} + \cdots + \frac{(n_k - n_{t,k})^2}{n_{t,k}}.$$

Plus le nombre χ^2 ainsi calculé est grand, plus la distribution étudiée différer de la distribution théorique.

Quelques considérations théoriques à propos de cet écart

Le nombre d'observations n_i parmi l'échantillon de taille N susceptible d'appartenir à la classe i est la réalisation d'une variable binomiale N_i de paramètres N et p_i (chacune des N observations appartient ou n'appartient pas à la classe i avec une probabilité p_i). Si N est suffisamment grand (on se place dans le cas d'échantillons de taille 50 minimum) et p_i pas trop petit (on a effectué des regroupements de classes pour qu'il en soit ainsi), on peut approcher la loi binomiale par la loi normale, c'est-à-dire $\mathcal{B}(N, p_i)$ par $\mathcal{N}(Np_i, \sqrt{Np_i(1-p_i)})$. Pour simplifier, on approxime $Np_i(1-p_i)$ par Np_i. Donc $\frac{N_i - Np_i}{Np_i}$ suit la loi $\mathcal{N}(0,1)$. Lorsqu'on élève au carré toutes ces quantités et qu'on en

fait la somme, on obtient une somme de k lois normales centrées réduites (presque) indépendantes. Nous avons vu au chapitre 3 que cette somme suivait une loi du khi-deux.

Mais quel est le nombre de degrés de liberté de cette variable du khi-deux ?

Il y a k carrés, donc à priori k degrés de liberté. Mais on perd toujours un degré de liberté car on a fixé l'effectif total de l'échantillon,

$$\sum_{i=1}^{k} N_i = N.$$

On peut perdre d'autres degrés de liberté si certains paramètres de la loi théorique doivent être estimés à partir de l'échantillon.

1. Si la distribution théorique est entièrement spécifiée, c'est-à-dire si on cherche à déterminer si la distribution observée suit une loi dont les paramètres sont connus avant même de choisir l'échantillon, on a $k-1$ degrés de liberté (k carrés indépendants moins une relation entre les variables).
2. S'il faut d'abord estimer r paramètres de la loi à partir des observations de l'échantillon (par exemple on cherche si la distribution est normale mais on ne connaît d'avance ni sa moyenne ni son écart-type), il n'y a plus que $k-1-r$ degrés de liberté.

Dans le cas général, on dira que la loi du khi-deux suivie par l'écart entre les deux distributions a $k-1-r$ degrés de liberté lorsqu'on a estimé r paramètres de la loi théorique à partir des observations de l'échantillon (avec la possibilité pour r de valoir 0).

6.5.2 Le test d'ajustement de Pearson

Il nous faut maintenant décider, à l'aide de cet indicateur qu'est le χ^2, si les écarts entre les effectifs théoriques et ceux qui résultent des observations sont significatifs d'une différence de distribution ou si ils sont dus aux fluctuations d'échantillonnage. Nous procéderons comme d'habitude en quatre étapes.

1ère étape : Formulation des hypothèses.

On va donc tester l'hypothèse H_0 contre l'hypothèse H_1 :

$$\begin{cases} H_0 & \text{Les observations suivent la distribution théorique spécifiée,} \\ H_1 & \text{Les observations ne suivent pas la distribution théorique spécifiée.} \end{cases}$$

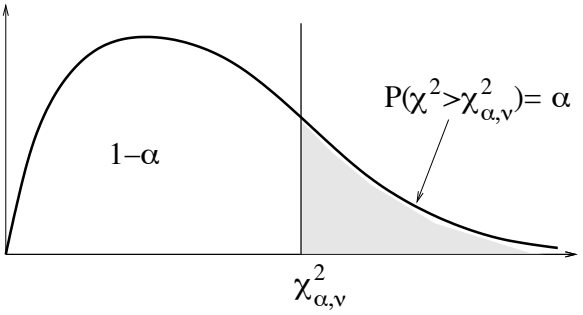
2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

On utilise la variable aléatoire

$$\chi^2 = \frac{(N_1 - n_{t,1})^2}{n_{t,1}} + \frac{(N_2 - n_{t,2})^2}{n_{t,2}} + \cdots + \frac{(N_k - n_{t,k})^2}{n_{t,k}}.$$

3ème étape : Détermination des valeurs critiques de χ^2 délimitant les zones d'acceptation et de rejet.

On impose à la zone d'acceptation de H_0 concernant la valeur du χ^2 d'être un intervalle dont 0 est la borne inférieure (car un χ^2 est toujours positif).



Il nous faut donc déterminer dans la table la valeur maximale $\chi^2_{\alpha,\nu}$ de l'écart entre les deux distributions imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire vérifiant $P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$. $\chi^2_{\alpha,\nu}$ représente donc la valeur critique pour un test sur la concordance entre deux distributions et le test sera toujours unilatéral à droite.

4ème étape : Calcul de la valeur de χ^2 prise dans l'échantillon et conclusion du test.
On calcule la valeur χ^2_0 prise par χ^2 dans l'échantillon.
- Si la valeur χ^2_0 se trouve dans la zone de rejet, on dira que l'écart observé entre les deux distributions est *statistiquement significatif* au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.
- Si la valeur χ^2_0 se trouve dans la zone d'acceptation, on dira que l'écart-réduit observé *n'est pas significatif* au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

6.6 Test d'homogénéïté de plusieurs populations

6.6.1 Introduction

On prélève au hasard k échantillons de tailles $n_1, n_2, \ldots, n_k$ de k populations. Les résultats du caractère observé dans chaque population sont ensuite classés selon r modalités. Dans ce cas, les totaux marginaux (les n_i) associés aux k échantillons sont fixés et ne dépendent pas du sondage. Il s'agit de savoir comparer les k populations entre elles et de savoir si elles ont un comportement semblable en regard du caractère étudié (qualitatif ou quantitatif). On rassemble les données dans un tableau à double entrée appelé *tableau de contingence*.

		Populations échantillonnées					
		$j=1$	$j=2$	...	j	...	$j=k$
Caractère observé selon r modalités	$i=1$	n_{11}	n_{12}		n_{1j}		n_{1k}
	$i=2$	n_{21}	n_{22}		n_{2j}		n_{2k}
	...						
	i	n_{i1}	n_{i2}		n_{ij}		n_{ik}
	...						
	$i=r$	n_{r1}	n_{r2}		n_{rj}		n_{rk}
		$n_1 = \sum_{i=1}^{r} n_{i1}$	$n_2 = \sum_{i=1}^{r} n_{i2}$		$n_j = \sum_{i=1}^{r} n_{ij}$		$n_k = \sum_{i=1}^{r} n_{ik}$

6.6.2 Test d'homogénéïté

Il s'agit de comparer les effectifs observés pour chaque modalité du caractère avec les effectifs théoriques sous l'hypothèse d'une répartition équivalente entre les k populations et ceci pour chaque modalité du caractère. Si nous notons p_{ij} la probabilité théorique pour qu'une unité statistique choisie au hasard dans la population j présente la modalité i du caractère étudié, on peut alors préciser les hypothèses de la façon suivante :

1ère étape : Formulation des hypothèses.
H_0 : $p_{i1} = p_{i2} = \cdots = p_{ik}$ pour $i = 1, 2, \ldots, r$. Soit encore : les proportions d'individus présentant chaque modalité du caractère sont les mêmes dans les k populations.
H_1 : $p_{ij_1} \neq p_{ij_2}$ pour au moins un i parmi $1, 2, \ldots, r$ et pour au moins deux j_1 et j_2 différents choisis parmi $1, 2, \ldots, k$. Soit encore : les proportions d'individus présentant chaque modalité du caractère ne sont pas identiques pour toutes les populations pour au moins une modalité du caractère.

2ème étape : Détermination de la fonction discriminante du test et de sa distribution de probabilité.

Sous l'hypothèse d'homogénéité des populations, on doit comparer les effectifs observés aux effectifs théoriques. Pour calculer les effectifs théoriques, il nous faut déterminer p_i, la proportion d'individus associée à la modalité i et que l'on suppose identique dans les k populations. On obtiendra une estimation de cette proportion en utilisant l'ensemble des données collectées. On choisit donc

$$p_i = \frac{\sum_{j=1}^{k} n_{ij}}{\sum_{j=1}^{k} n_j}.$$

On en déduit les effectifs théoriques de chaque classe grâce à la relation

$$n_{t,ij} = p_i n_j.$$

Pour comparer les écarts entre ce qu'on observe et ce qui se passe sous l'hypothèse H_0, on considère la somme des écarts réduits de chaque classe, à savoir la quantité

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(N_{ij} - n_{t,ij})^2}{n_{t,ij}}.$$

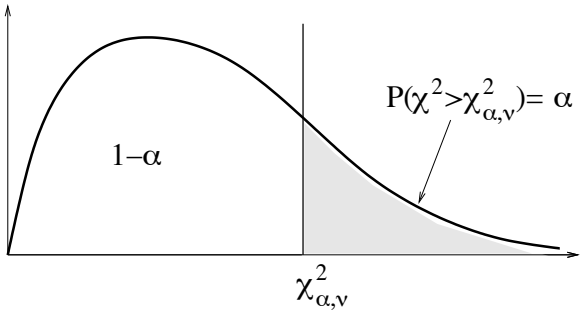
Cette variable aléatoire suit une loi du khi-deux (voir paragraphe précédent), mais quel est donc son nombre de degrés de liberté ?

Calcul du nombre de degrés de liberté du khi-deux.

- A priori, on a kr cases dans notre tableau donc kr degrés de liberté. Mais il faut retirer à cette valeur, le nombre de paramètres estimés ainsi que le nombre de relations entre les différents éléments des cases.
- On a estimé r probabilités théoriques à l'aide des valeurs du tableau$(p_1,\ p_2, \ldots, p_r)$, mais seulement $r-1$ sont indépendantes, puisqu'on impose la restriction $\sum_{i=1}^{r} p_i = 1$. Par ces estimations, on a donc supprimé $r-1$ degrés de liberté.
- Les effectifs de chaque colonne sont toujours liés par les relations $\sum_{i=1}^{r} N_{ij} = n_j$ (puisque les n_j sont imposés par l'expérience) et ces relations sont au nombre de k.
- Finalement, le nombre de degrés de liberté du khi-deux est $kr - (r-1) - k = (k-1)(r-1)$.

3ème étape : Détermination des valeurs critiques de délimitant les zones d'acceptation et de rejet.

On impose à la zone d'acceptation de H_0 concernant la valeur du χ^2 d'être un intervalle dont 0 est la borne inférieure (car un χ^2 est toujours positif).



Il nous faut donc déterminer dans la table la valeur maximale $\chi^2_{\alpha,\nu}$ de l'écart entre les deux distributions imputable aux variations d'échantillonnage au seuil de signification α, c'est-à-dire vérifiant $P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$.

4ème étape : Calcul de la valeur de χ^2 prise dans l'échantillon et conclusion du test.

On calcule la valeur χ^2_0 prise par χ^2 dans l'échantillon.

- Si la valeur χ^2_0 se trouve dans la zone de rejet, on dira que l'écart observé entre les deux distributions est *statistiquement significatif* au seuil α. Cet écart est anormalement élevé et ne permet pas d'accepter H_0. On rejette H_0.

– Si la valeur χ_0^2 se trouve dans la zone d'acceptation, on dira que l'écart-réduit observé *n'est pas significatif* au seuil α. Cet écart est imputable aux fluctuations d'échantillonnage. On accepte H_0.

CONCLUSION : Nous avons appris à effectuer un certain nombre de tests. Il en existe d'autres. Tous fonctionnent sur le même principe. Si vous avez compris ce qui précède, vous serez capables de les appréhender correctement lorsque vous les rencontrerez : suivez le modèle.

LOI DE POISSON ($0.1 \leq \lambda \leq 6$)

k	$\lambda = 0.1$		$\lambda = 0.2$		$\lambda = 0.3$		$\lambda = 0.4$		$\lambda = 0.5$	
	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$
0	0.9048	0.9048	0.8187	0.8187	0.7408	0.7408	0.6703	0.6703	0.6065	0.6065
1	0.0905	0.9953	0.1637	0.9825	0.2222	0.9631	0.2681	0.9384	0.3033	0.9098
2	0.0045	0.9998	0.0164	0.9989	0.0333	0.9964	0.0536	0.9921	0.0758	0.9856
3	0.0002	1	0.0011	0.9999	0.0033	0.9997	0.0072	0.9992	0.0126	0.9982
4	0	1	0.0001	1	0.0003	1	0.0007	0.9999	0.0016	0.9998
5			0	1	0	1	0.0001	1	0.0002	1
	$\lambda = 0.6$		$\lambda = 0.7$		$\lambda = 0.8$		$\lambda = 0.9$		$\lambda = 1$	
	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$
0	0.5488	0.5488	0.4966	0.4966	0.4493	0.4493	0.4066	0.4066	0.3679	0.3679
1	0.3293	0.8781	0.3476	0.8442	0.3595	0.8088	0.3659	0.7725	0.3679	0.7358
2	0.0988	0.9769	0.1217	0.9659	0.1438	0.9526	0.1647	0.9371	0.1839	0.9197
3	0.0198	0.9966	0.0284	0.9942	0.0383	0.9909	0.0494	0.9865	0.0613	0.9810
4	0.0030	0.9996	0.0050	0.9992	0.0077	0.9986	0.0111	0.9977	0.0153	0.9963
5	0.0004	1	0.0007	0.9999	0.0012	0.9998	0.0020	0.9997	0.0031	0.9994
6	0	1	0.0001	1	0.0002	1	0.0003	1	0.0005	0.9999
7	0	1	0	1	0	1	0	1	0.0001	1
	$\lambda = 2$		$\lambda = 3$		$\lambda = 4$		$\lambda = 5$		$\lambda = 6$	
	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$
0	0.1353	0.1353	0.0498	0.0498	0.0183	0.0183	0.0067	0.0067	0.0025	0.0025
1	0.2707	0.4060	0.1494	0.1991	0.0733	0.0916	0.0337	0.0404	0.0149	0.0174
2	0.2707	0.6767	0.2240	0.4232	0.1465	0.2381	0.0842	0.1247	0.0446	0.0620
3	0.1804	0.8571	0.2240	0.6472	0.1954	0.4335	0.1404	0.2650	0.0892	0.1512
4	0.0902	0.9473	0.1680	0.8153	0.1954	0.6288	0.1755	0.4405	0.1339	0.2851
5	0.0361	0.9834	0.1008	0.9161	0.1563	0.7851	0.1755	0.6160	0.1606	0.4457
6	0.0120	0.9955	0.0504	0.9665	0.1042	0.8893	0.1462	0.7622	0.1606	0.6063
7	0.0034	0.9989	0.0216	0.9881	0.0595	0.9489	0.1044	0.8666	0.1377	0.7440
8	0.0009	0.9998	0.0081	0.9962	0.0298	0.9786	0.0653	0.9319	0.1033	0.8472
9	0.0002	1	0.0027	0.9989	0.0132	0.9919	0.0363	0.9682	0.0688	0.9161
10	0	1	0.0008	0.9997	0.0053	0.9972	0.0181	0.9863	0.0413	0.9574
11	0	1	0.0002	0.9999	0.0019	0.9991	0.0082	0.9945	0.0225	0.9799
12	0	1	0.0001	1	0.0006	0.9997	0.0034	0.9980	0.0113	0.9912
13	0	0	0	1	0.0002	0.9999	0.0013	0.9993	0.0052	0.9964
14	0	0	0	1	0.0001	1	0.0005	0.9998	0.0022	0.9986
15	0	0	0	0	0	1	0.0002	0.9999	0.0009	0.9995
16	0	0	0	0	0	1	0	1	0.0003	0.9998
17	0	0	0	0	0	1	0	1	0.0001	0.9999
18	0	0	0	0	0	0	0	1	0	1

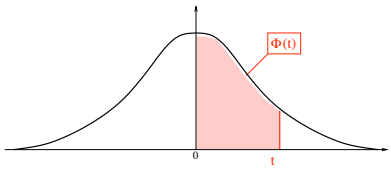
LOI DE POISSON ($7 \leq \lambda \leq 11$)

k	$\lambda = 7$		$\lambda = 8$		$\lambda = 9$		$\lambda = 10$		$\lambda = 11$	
	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$
0	0.0009	0.0009	0.0003	0.0003	0.0001	0.0001	0	0	0	0
1	0.0064	0.0073	0.0027	0.0030	0.0011	0.0012	0.0005	0.0005	0.0002	0.0002
2	0.0223	0.0296	0.0107	0.0138	0.0050	0.0062	0.0023	0.0028	0.0010	0.0012
3	0.0521	0.0818	0.0286	0.0424	0.0150	0.0212	0.0076	0.0103	0.0037	0.0049
4	0.0912	0.1730	0.0573	0.0996	0.0337	0.0550	0.0189	0.0293	0.0102	0.0151
5	0.1277	0.3007	0.0916	0.1912	0.0607	0.1157	0.0378	0.0671	0.0224	0.0375
6	0.1490	0.4497	0.1221	0.3134	0.0911	0.2068	0.0631	0.1301	0.0411	0.0786
7	0.1490	0.5987	0.1396	0.4530	0.1171	0.3239	0.0901	0.2202	0.0646	0.1432
8	0.1304	0.7291	0.1396	0.5925	0.1318	0.4557	0.1126	0.3328	0.0888	0.2320
9	0.1014	0.8305	0.1241	0.7166	0.1318	0.5874	0.1251	0.4579	0.1085	0.3405
10	0.0710	0.9015	0.0993	0.8159	0.1186	0.7060	0.1251	0.5830	0.1194	0.4599
11	0.0452	0.9467	0.0722	0.8881	0.0970	0.8030	0.1137	0.6968	0.1194	0.5793
12	0.0263	0.9730	0.0481	0.9362	0.0728	0.8758	0.0948	0.7916	0.1094	0.6887
13	0.0142	0.9872	0.0296	0.9658	0.0504	0.9261	0.0729	0.8645	0.0926	0.7813
14	0.0071	0.9943	0.0169	0.9827	0.0324	0.9585	0.0521	0.9165	0.0728	0.8540
15	0.0033	0.9976	0.0090	0.9918	0.0194	0.9780	0.0347	0.9513	0.0534	0.9074
16	0.0014	0.9990	0.0045	0.9963	0.0109	0.9889	0.0217	0.9730	0.0367	0.9441
17	0.0006	0.9996	0.0021	0.9984	0.0058	0.9947	0.0128	0.9857	0.0237	0.9678
18	0.0002	0.9999	0.0009	0.9993	0.0029	0.9976	0.0071	0.9928	0.0145	0.9823
19	0.0001	1	0.0004	0.9997	0.0014	0.9989	0.0037	0.9965	0.0084	0.9907
20	0	1	0.0002	0.9999	0.0006	0.9996	0.0019	0.9984	0.0046	0.9953
21	0	1	0.0001	1	0.0003	0.9998	0.0009	0.9993	0.0024	0.9977
22	0	1	0	1	0.0001	0.9999	0.0004	0.9997	0.0012	0.9990
23	0	1	0	1	0	1	0.0002	0.9999	0.0006	0.9995
24	0	1	0	1	0	1	0.0001	1	0.0003	0.9998
25	0	1	0	1	0	1	0	1	0.0001	0.9999
26	0	1	0	1	0	1	0	1	0	1

LOI DE POISSON ($12 \leq \lambda \leq 16$)

	$\lambda = 12$		$\lambda = 13$		$\lambda = 14$		$\lambda = 15$		$\lambda = 16$	
	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$	p_k	$\sum p_k$
0	0	0	0	0	0	0	0	0	0	0
1	0.0001	0.0001	0	0	0	0	0	0	0	0
2	0.0004	0.0005	0.0002	0.0002	0.0001	0.0001	0	0	0	0
3	0.0018	0.0023	0.0008	0.0011	0.0004	0.0005	0.0002	0.0002	0.0001	0.0001
4	0.0053	0.0076	0.0027	0.0037	0.0013	0.0018	0.0006	0.0009	0.0003	0.0004
5	0.0127	0.0203	0.0070	0.0107	0.0037	0.0055	0.0019	0.0028	0.0010	0.0014
6	0.0255	0.0458	0.0152	0.0259	0.0087	0.0142	0.0048	0.0076	0.0026	0.0040
7	0.0437	0.0895	0.0281	0.0540	0.0174	0.0316	0.0104	0.0180	0.0060	0.0100
8	0.0655	0.1550	0.0457	0.0998	0.0304	0.0621	0.0194	0.0374	0.0120	0.0220
9	0.0874	0.2424	0.0661	0.1658	0.0473	0.1094	0.0324	0.0699	0.0213	0.0433
10	0.1048	0.3472	0.0859	0.2517	0.0663	0.1757	0.0486	0.1185	0.0341	0.0774
11	0.1144	0.4616	0.1015	0.3532	0.0844	0.2600	0.0663	0.1848	0.0496	0.1270
12	0.1144	0.5760	0.1099	0.4631	0.0984	0.3585	0.0829	0.2676	0.0661	0.1931
13	0.1056	0.6815	0.1099	0.5730	0.1060	0.4644	0.0956	0.3632	0.0814	0.2745
14	0.0905	0.7720	0.1021	0.6751	0.1060	0.5704	0.1024	0.4657	0.0930	0.3675
15	0.0724	0.8444	0.0885	0.7636	0.0989	0.6694	0.1024	0.5681	0.0992	0.4667
16	0.0543	0.8987	0.0719	0.8355	0.0866	0.7559	0.0960	0.6641	0.0992	0.5660
17	0.0383	0.9370	0.0550	0.8905	0.0713	0.8272	0.0847	0.7489	0.0934	0.6593
18	0.0255	0.9626	0.0397	0.9302	0.0554	0.8826	0.0706	0.8195	0.0830	0.7423
19	0.0161	0.9787	0.0272	0.9573	0.0409	0.9235	0.0557	0.8752	0.0699	0.8122
20	0.0097	0.9884	0.0177	0.9750	0.0286	0.9521	0.0418	0.9170	0.0559	0.8682
21	0.0055	0.9939	0.0109	0.9859	0.0191	0.9712	0.0299	0.9469	0.0426	0.9108
22	0.0030	0.9970	0.0065	0.9924	0.0121	0.9833	0.0204	0.9673	0.0310	0.9418
23	0.0016	0.9985	0.0037	0.9960	0.0074	0.9907	0.0133	0.9805	0.0216	0.9633
24	0.0008	0.9993	0.0020	0.9980	0.0043	0.9950	0.0083	0.9888	0.0144	0.9777
25	0.0004	0.9997	0.0010	0.9990	0.0024	0.9974	0.0050	0.9938	0.0092	0.9869
26	0.0002	0.9999	0.0005	0.9995	0.0013	0.9987	0.0029	0.9967	0.0057	0.9925
27	0.0001	0.9999	0.0002	0.9998	0.0007	0.9994	0.0016	0.9983	0.0034	0.9959
28	0	1	0.0001	0.9999	0.0003	0.9997	0.0009	0.9991	0.0019	0.9978
29	0	1	0.0001	1	0.0002	0.9999	0.0004	0.9996	0.0011	0.9989
30	0	1	0	1	0.0001	0.9999	0.0002	0.9998	0.0006	0.9994
31	0	1	0	1	0	1	0.0001	0.9999	0.0003	0.9997
32	0	1	0	1	0	1	0	1	0.0001	0.9999
33	0	1	0	1	0	1	0	1	0.0001	0.9999

Loi de Laplace-Gauss



t	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4.0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000