

# **Simulation de Monte-Carlo**

**Présentation (simplifiée) de quelques notions de probabilité, d'estimation et de tests statistiques**

<b>1. LES PROBABILITES.</b>	<b>3</b>
1.1. INTRODUCTION	3
<i>Eléments de vocabulaire</i>	3
<i>Propriétés d'une probabilité</i>	4
1.2. VARIABLE ALEATOIRE DISCRETE, FONCTION DE DENSITE ET FONCTION DE REPARTITION.	5
<i>Variable aléatoire discrète</i>	5
<i>Loi de probabilité</i>	6
<i>Fonction de répartition</i>	7
1.3. MOMENTS, MOMENTS CENTRES, ESPERANCE, VARIANCE ET ECART-TYPE.	8
<i>Moments d'ordre k et espérance</i>	8
<i>Moments centrés d'ordre k, variance et écart-type</i>	9
1.4. COVARIANCE ET COEFFICIENT DE CORRELATION LINEAIRE	10
<i>Covariance</i>	10
<i>Coefficient de corrélation linéaire</i>	11
1.5. LOIS DE PROBABILITE DISCRETES.	11
<i>Permutations sans répétition (factorielle)</i>	12
<i>Arrangements</i>	12
<i>Combinaisons</i>	13
<i>Loi de Bernoulli</i>	13
<i>Loi Binomiale</i>	14
<i>Loi Hypergéométrique</i>	14
1.6. VARIABLE ALEATOIRE ET LOIS DE PROBABILITE CONTINUES.	15
<i>Variable aléatoire continue</i>	15
<i>Loi normale - Présentation</i>	16
<i>Loi normale - Exercice illustratif</i>	19
<i>Loi du <math>\chi^2</math></i>	20
<i>Loi de Student</i>	21
<b>2. ECHANTILLONNAGE, ESTIMATION ET INTERVALLE DE CONFIANCE.</b>	<b>23</b>
2.1. ECHANTILLONNAGE	23
2.2. ESTIMATION	24
2.3. ESTIMATION ET INTERVALLE DE CONFIANCE POUR UNE PROPORTION	24
<i>Estimation</i>	24
<i>Intervalle de confiance avec remise</i>	25
<i>Intervalle de confiance sans remise</i>	27
<i>Taille d'échantillon</i>	27
2.4. ESTIMATION ET INTERVALLE DE CONFIANCE POUR UNE MOYENNE	29
<i>Estimation</i>	29

<i>Intervalle de confiance</i> .....	29
<i>Taille d'échantillon</i> .....	30
2.5. ESTIMATION POUR UNE VARIANCE .....	30
<b>3. THEORIE DES TESTS STATISTIQUES .....</b>	<b>32</b>
3.1. INTRODUCTION .....	32
3.2. TEST DE CONFORMITE D'UNE PROPORTION .....	32
<i>Hypothèse nulle et alternative</i> .....	32
<i>Erreur de première et de seconde espèce</i> .....	33
<i>Procédure de test</i> .....	33
<i>Test bilatéral vs. unilatéral</i> .....	35
3.3. TEST DE CONFORMITE D'UNE MOYENNE .....	36
3.4. TEST DE COMPARAISON DE DEUX PROPORTIONS.....	37
3.5. TEST DE COMPARAISON D'UNE MOYENNE .....	38
<b>4. SYNTHESSES.....</b>	<b>39</b>
4.1. ESTIMATION DE PROPORTIONS .....	39
4.2. ESTIMATION DE MOYENNES .....	40
4.3. TESTS DE COMPARAISON .....	41

# 1. Les probabilités.

## 1.1. Introduction

### Eléments de vocabulaire.

Les probabilités fournissent une description mathématique (modèle) de l'incertain c'est-à-dire de ce qui n'est pas connu avec certitude. On parle dans ce cas d'**événements** « **aléatoires**<sup>1</sup> ». Un événement est dit « aléatoire » quand l'issue (le résultat) d'une **expérience** (exemple : jeter un dé, lancer une pièce de monnaie, tirer une carte, jouer à la roulette, observer un défaut dans un produit,...) n'est pas connue par avance c'est-à-dire qu'elle est soumise au *hasard*. Néanmoins, les phénomènes aléatoires respectent *a priori* certaines **propriétés sur le long terme** et ce sont ces propriétés que la théorie des probabilités va décrire au travers d'outils mathématiques.

Dans le cadre des probabilités, une **expérience aléatoire** (ou **épreuve**) est une procédure définie précisément et qui conduit à un résultat (résultat qui est donc par définition imprévisible). L'**ensemble fondamental** (noté  $\Omega$ ) des résultats (c'est-à-dire toutes les issues possibles – éventuellement en nombre infini) de l'expérience est connu **avant** l'expérience mais le **résultat** de l'expérience n'est connu qu'**après** la réalisation de l'expérience.

Chaque élément (noté  $\omega$ ) d'un ensemble fondamental est qualifié d'**événement élémentaire** (ou de **réalisation**). Ainsi, l'ensemble fondamental d'une expérience consistant à jeter un dé à six faces est :  $\Omega = [\text{Un}, \text{Deux}, \text{Trois}, \text{Quatre}, \text{Cinq}, \text{Six}]$ <sup>2</sup> tandis que celui associé au lancer d'une pièce est  $\Omega = [\text{Pile}, \text{Face}]$ . Dans le cas du dé, obtenir la valeur « Trois » est un exemple d'évènement élémentaire ou de réalisation de l'expérience aléatoire.

Un **événement** (ou **événement composé**) représente une partie  $E$  de  $\Omega$  et résulte d'une proposition logique relative au résultat de l'expérience. Pour illustrer ce point, prenons le cas d'une expérience aléatoire consistant à tirer au hasard une carte dans un jeu en contenant 32 ( $\text{card}(\Omega) = 32$ ) et considérons l'évènement composé  $E$  « obtenir une carte de ♥ » :

$$\Omega = [\text{Un} \clubsuit, \text{Un} \diamond, \text{Un} \heartsuit, \text{Un} \spadesuit, \text{Deux} \clubsuit, \text{Deux} \diamond, \text{Deux} \heartsuit, \text{Deux} \spadesuit, \dots, \text{R} \clubsuit, \text{R} \diamond, \text{R} \heartsuit, \text{R} \spadesuit]$$

$$\text{Et ici } E = [\text{Un} \heartsuit, \text{Deux} \heartsuit, \dots, \text{R} \heartsuit]$$

Une **probabilité**  $P$  est une fonction qui à un évènement associe une valeur comprise entre 0 et 1. Cette probabilité mesure les chances que cet évènement a de se produire. Par exemple, pour un

---

<sup>1</sup> Remarque : en latin, *alea* signifie « coup de dé ».

<sup>2</sup> Nous avons choisi de représenter les faces du dé à l'aide de lettres et non pas de chiffres pour insister sur la différence entre la notion d'évènement et celle de variable aléatoire définie à la suite.

dé parfaitement équilibré à six faces,  $\Omega = [\text{Un}, \text{Deux}, \text{Trois}, \text{Quatre}, \text{Cinq}, \text{Six}]$ . Le dé étant supposé équilibré, chaque face a autant de « chances » d'apparaître. En adoptant la définition<sup>3</sup> d'une probabilité comme étant le rapport entre le nombre de « cas favorables » sur le nombre de « cas possibles », on en déduit que :  $P(\text{Un}) = P(\text{Deux}) = \dots = P(\text{Six}) = 1/6$ . Selon la même logique, si on considère le cas d'une pièce de monnaie, on obtient :  $\Omega = [\text{Pile}, \text{Face}]$  et  $P(\text{Pile}) = P(\text{Face}) = 1/2$ .

### Propriétés d'une probabilité.

En nous basant sur la définition précédente d'une probabilité (on aura noté sur les deux exemples précédents que l'on raisonne sur un ensemble fondamental  $\Omega$  de taille finie), on tire un certain nombre de propriétés :

P1 : la probabilité qu'au moins l'un des événements de  $\Omega$  se produise est  $P(\Omega) = 1$

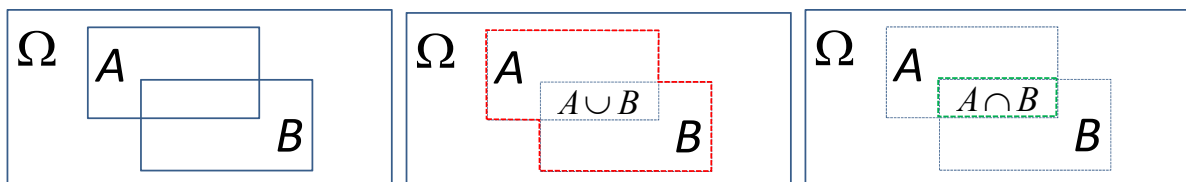
P2 : si  $\emptyset$  représente l'ensemble vide (ensemble ne contenant aucun élément) alors  $P(\emptyset) = 0$

P3 : pour tout événement  $A$  de  $\Omega$ , on a  $0 \leq P(A) \leq 1$

Le **complémentaire** de  $A$  par rapport à  $\Omega$  (noté  $A^c$  ou  $\bar{A}$ ) est constitué de l'ensemble des éléments de  $\Omega$  qui n'appartiennent pas à  $A$ . Par exemple, le complémentaire de l'événement obtenir un chiffre pair dans le cas du dé à six faces est l'événement obtenir un chiffre impair.

P4 : si  $A^c$  est le complémentaire de  $A$  par rapport à  $\Omega$  alors  $P(A^c) = 1 - P(A)$

Si on considère deux événements  $A$  et  $B$  de  $\Omega$  (graphique 1 ci-dessous), on appelle **réunion** de  $A$  et  $B$  (notée  $A \cup B$  représentée ci-dessous en rouge – graphique 2) l'ensemble des éléments qui appartiennent à  $A$  ou à  $B$  (donc soit à  $A$ , soit à  $B$  soit à  $A$  et à  $B$ ). On appelle **intersection** de  $A$  et  $B$  (notée  $A \cap B$  représentée ci-dessous en vert – graphique 3) l'ensemble des éléments qui appartiennent à la fois à  $A$  et à  $B$  :



P5<sup>4</sup> :  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ <sup>5</sup>

<sup>3</sup> Vision « fréquentiste » des probabilités.

<sup>4</sup> Cette propriété se généralise au cas où il y a plus de deux éléments.

<sup>5</sup> Dans le cas particulier où l'intersection de  $A$  et  $B$  est vide ( $A \cap B = \emptyset$  - événements disjoints ou incompatibles) on obtient donc de P5 que  $P(A \cup B) = P(A) + P(B)$ .

On considère toujours les deux mêmes événements  $A$  et  $B$ . On se pose maintenant la question de savoir quelle est la probabilité de  $B$  sachant que  $A$  est réalisé ( $A$  étant réalisé il est donc devenu certain). Cette valeur, appelée **probabilité conditionnelle**, est notée  $P(B|A)$  et se lit « probabilité de  $B$  sachant  $A$  » :

$$P6 : P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\text{Ou encore : } P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B)$$

Illustrons cette notion toujours en nous basant sur l'exemple du dé à six faces. On sait que la probabilité d'obtenir un résultat inférieur ou égal à trois vaut  $1/2$ . Si maintenant on sait que le résultat est impair, à combien est égale la probabilité d'obtenir un résultat inférieur ou égal à trois ? On cherche donc :  $P([\text{Un, Deux, Trois}] \cap [\text{Un, Trois, Cinq}])$

$$\begin{aligned} &= \frac{P([\text{Un, Deux, Trois}] \cap [\text{Un, Trois, Cinq}])}{P([\text{Un, Trois, Cinq}])} \\ &= \frac{P([\text{Un, Trois}])}{P([\text{Un, Trois, Cinq}])} \\ &= \frac{2/6}{3/6} = \frac{2}{3} \end{aligned}$$

De façon complémentaire, on obtient que deux événements  $A$  et  $B$  sont **indépendants** si la connaissance de l'un ne permet pas de modifier la probabilité de réalisation de l'autre. Dans ce cas :

$$P(B|A) = P(B) \text{ et } P(A|B) = P(A)$$

$$P7 : P(A \cap B) = P(A) * P(B)$$

$$\text{Et par ailleurs : } P(A \cup B) = P(A) + P(B) - P(A) * P(B)$$

## 1.2. Variable aléatoire discrète, fonction de densité et fonction de répartition.

### Variable aléatoire discrète.

Une **variable aléatoire**  $X$  ( $X: \Omega \rightarrow \mathbb{R}$ ) est une fonction à valeurs réelles ayant comme domaine de définition l'ensemble  $\Omega$ . La notion de variable aléatoire ne doit pas être confondue avec celle d'ensemble fondamental. Dans l'exemple précédent, obtenir un  $R \spadesuit$  n'est pas une variable aléatoire car  $R \spadesuit$  n'est pas une valeur numérique. En revanche, si on convient d'associer à chaque carte la valeur numérique  $X$  égale à sa valeur faciale quelle que soit sa couleur (donc 1 pour  $Un \clubsuit$ , 2

pour Deux ♦,...) et 10 points s'il s'agit d'une tête (Valet, Dame ou Roi et toujours quelle que soit sa couleur), alors  $X$  est une variable aléatoire.

Une variable aléatoire est **discrète** si elle ne prend qu'un nombre fini ou dénombrable de valeurs (habituellement entières). Selon cette définition, on peut donc repérer (numéroter) chacun de ses éléments par des indices eux-mêmes entiers. Pour illustrer ce point, reprenons l'exemple classique du dé à 6 faces et définissons  $S$  comme étant le résultat de la somme des lancers de deux dés (un Vert et un Rouge par exemple). Pour chacun des dés, on a  $\Omega = \{\text{Un, Deux, Trois, Quatre, Cinq, Six}\}$ . Si on note  $v$  le résultat du dé Vert et  $r$  celui du dé Rouge, les événements élémentaires de cette expérience sont  $\omega = (v, r)$ <sup>6</sup> et on peut définir  $S(\omega) = v + r$  dont les valeurs sont comprises entre 2 (les deux dés sont tombés sur la face un) et 12 (ils sont tombés sur la face six). On obtient dès lors :

$$\Omega' = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

En observant l'ensemble fondamental  $\Omega'$ , on constate que le nombre de valeurs possibles est de 11 et que chacune de ces valeurs est entière. Comme indiqué précédemment, on pourrait repérer chaque valeur possible par des indices entiers et écrire :

$$\Omega' = \{\omega'_1, \omega'_2, \omega'_3, \dots, \omega'_{11}\} \text{ où } \omega'_1 = 1, \omega'_2 = 2, \omega'_3 = 3, \dots, \omega'_{11} = 11$$

En résumé, à chaque réalisation d'une expérience on pourra donc associer une valeur unique c'est-à-dire une variable aléatoire, avec une certaine probabilité (loi de probabilité ou fonction de masse) que celle-ci soit obtenue.

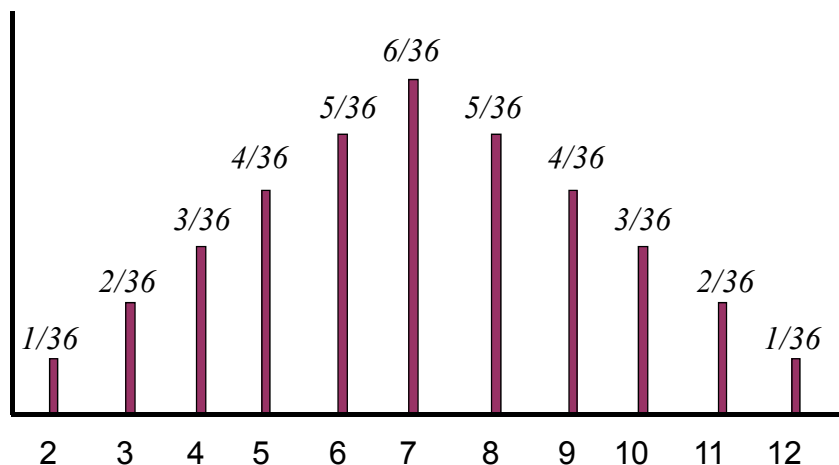
### Loi de probabilité.

Dans l'exemple précédent, il n'y a qu'un seul couple  $\omega = (v, r) = (1, 1)$  dont la somme soit égale à  $S = 2$ . En revanche il y en a deux,  $\omega = (v, r) = (1, 2)$  et  $\omega = (v, r) = (2, 1)$  qui donne  $S = 3$ . Comme il y a 36 valeurs possibles,  $P(S = 1) = \frac{1}{36}$  et  $P(S = 2) = \frac{1+1}{36}$ . La loi de probabilité de  $S$  peut être représentée sous la forme d'un tableau<sup>7</sup> et d'un graphique associé (diagramme en bâtons) :

$s$	2	3	4	5	6	7	...	12
$P(S=s)$	1/36	2/36	3/36	4/36	5/36	6/36	...	1/36

<sup>6</sup>  $\Omega = \{(1,1); (1,2); \dots; (6,6)\}$  et chaque élément  $\omega = (v, r)$  a une probabilité de  $\frac{1}{36}$  d'être réalisé.

<sup>7</sup> On verra par la suite qu'une loi de probabilité peut également se présenter sous la forme d'une fonction mathématique.



De façon générale, on associe donc à une variable aléatoire discrète  $X$  l'un des deux tableaux suivants :

Le tableau suivant si  $X$  est une variable aléatoire discrète finie dont  $\text{card}(\Omega) = n$  alors  $p_i \geq 0$  et

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

$x$	$x_1$	$x_2$	$\dots$	$x_i$	$\dots$	$x_n$
$P(X=x)$	$p_1$	$p_2$	$\dots$	$p_i$	$\dots$	$p_n$

Ou le tableau équivalent si  $X$  est une variable aléatoire discrète dénombrable dont  $\text{card}(\Omega) = \infty$  alors  $p_i \geq 0$  et

$$\sum_{i=1}^{\infty} p_i = 1$$

$x$	$x_1$	$x_2$	$\dots$	$x_i$	$\dots \dots \dots$
$P(X=x)$	$p_1$	$p_2$	$\dots$	$p_i$	$\dots \dots \dots$

### Fonction de répartition.

La **fonction de répartition** (ou probabilité cumulée) est une fonction  $F$  qui prend ses valeurs entre 0 et 1 (elle est définie dans  $[0;1]$ ) et qui a une valeur  $x$  associe la valeur  $P(X \leq x)$  (ou  $P(X < x)$  chez certains) :

$$\mathbb{R} \rightarrow [0; 1]$$

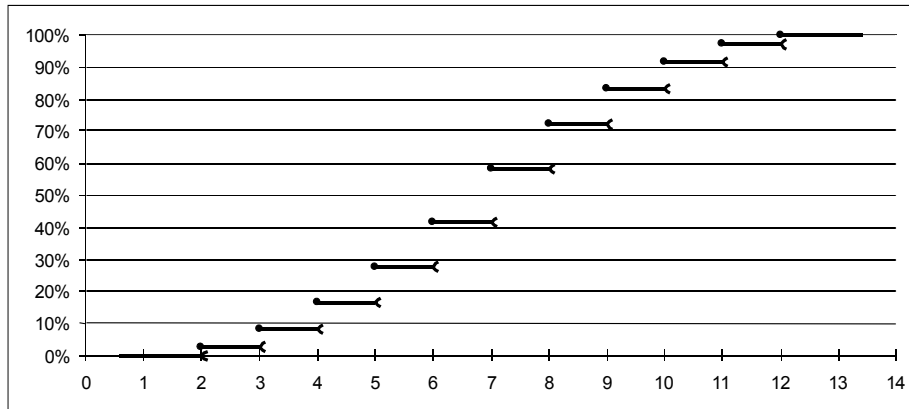
$$x \mapsto P(X \leq x)$$



Si on reprend l'exemple de la somme (notée  $X$ ) des deux dés à six faces, on obtient la fonction de répartition et la représentation graphique suivantes :

$x$	2	3	4	5	6	7	...	12
$P(X \leq x)$	1/36	3/36	6/36	10/36	15/36	21/36	...	36/36

$$P(X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{1+2+3}{36} = \frac{1}{6}$$



### 1.3. Moments, moments centrés, espérance, variance et écart-type.

#### Moments d'ordre $k$ et espérance.

On définit la notion de « **moment d'ordre  $k$**  » d'une variable aléatoire finie<sup>8</sup>  $X$  ainsi :

$$m_k(X) = \sum_{i=1}^n p_i x_i^k$$

Cette expression va nous permettre en particulier de définir l'**espérance mathématique** d'une variable aléatoire comme étant égale au moment d'ordre 1. De façon plus explicite, l'espérance mathématique<sup>9</sup> est équivalente à la **moyenne arithmétique** en statistique descriptive<sup>10</sup> :

$$m_1(X) = E(X) = \sum_{i=1}^n p_i x_i$$

A partir de l'exemple de la somme des deux dés, on obtient aisément :

$$E(X) = 2 * \frac{1}{36} + 3 * \frac{2}{36} + \dots + 12 * \frac{1}{36} = 7$$

<sup>8</sup> Si elle est dénombrable, on remplace  $n$  par  $\infty$  dans le signe somme.

<sup>9</sup> L'espérance mathématique n'est pas une valeur aléatoire, c'est un nombre réel.

<sup>10</sup> En statistique descriptive, la moyenne se calcule ainsi :  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Propriétés de l'espérance :

$$E(aX + b) = aE(X) + b$$

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

$$E(XY) \neq E(X)E(Y) \text{ sauf si } X \text{ et } Y \text{ sont indépendantes}$$

### Moments centrés d'ordre $k$ , variance et écart-type.

On va maintenant définir la notion de « **moment centré d'ordre  $k$**  » d'une variable aléatoire finie  $X$ . Pour cela, définissons  $X' = X - E(X)$  et calculons :

$$mc_k(X) = m_k(X') = \sum_{i=1}^n p_i [x_i - E(X)]^k$$

Au sein de ces moments centrés, on s'intéresse particulièrement à celui d'ordre 2 qui permet de définir la **variance** d'une variable aléatoire<sup>11</sup> :

$$Var(X) = mc_2(X) = \sum_{i=1}^n p_i [x_i - E(X)]^2$$

L'**écart-type** se calcule lui comme la racine carrée de la variance :

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

Le théorème de Koenig nous donne une autre manière de calculer une variance :

$$Var(X) = \sigma^2(X) = m_2(X) - [m_1(X)]^2 = E(X^2) - E(X)^2$$

Propriétés de la variance :

$$V(X) \geq 0$$

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

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

où  $Cov(X, Y)$  est la covariance entre la variable  $X$  et la variable  $Y$ .

---

<sup>11</sup> En statistique descriptive, on utilisera deux formules pour calculer la variance (cf. partie sur l'estimation) suivant que l'on dispose de la population ou d'un échantillon. Dans le premier cas :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right) - \mu^2 \quad \text{Variance d'une population}$$

Dans le second cas :

$$s'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2 \quad \text{Variance d'un échantillon}$$

$$s^2 = \frac{n}{n-1} s'^2 \quad \text{Estimation de la variance d'une population à partir de celle d'un échantillon}$$

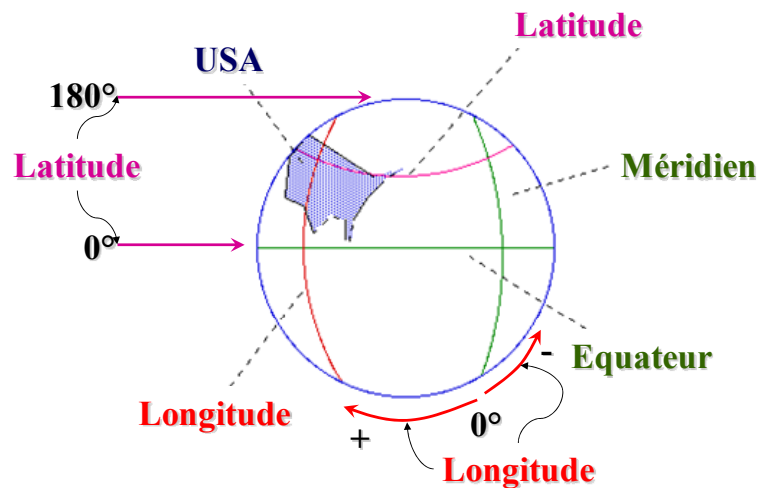
## 1.4. Covariance et coefficient de corrélation linéaire

### Covariance

La **covariance**<sup>12</sup> mesure la relation qui peut exister entre deux variables. Une covariance positive traduit l'idée que les deux variables varient dans le même sens c'est-à-dire que lorsque l'une augmente, l'autre en fait autant. Une covariance négative indique en revanche une relation inverse. Finalement, une covariance proche de 0 signifie qu'il n'y a pas de relation. D'ailleurs, deux variables aléatoires indépendantes (cf. propriétés de la covariance plus loin) ont par définition une covariance nulle (attention, la réciproque n'est pas vérifiée).

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

Pour illustrer ce point, définissons trois variables, la latitude, la longitude et la température moyenne minimale en Janvier (exprimée en degrés Fahrenheit) pour une série (individus) de villes des USA. La latitude permet de repérer les villes du nord au sud et la longitude d'est en ouest comme indiqué ci-dessous :

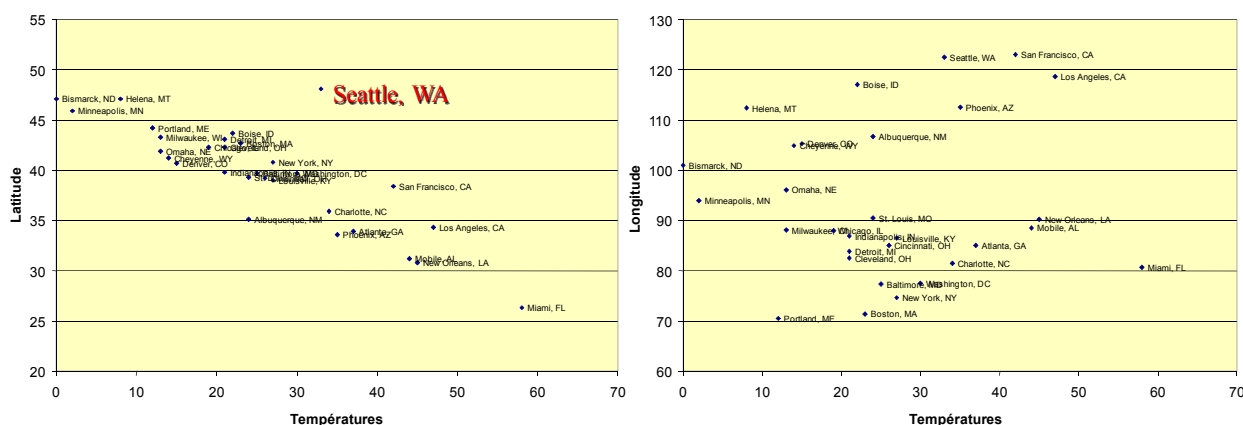


En admettant que dans l'hémisphère nord les villes les plus au sud ont les températures les plus élevées, on peut supposer qu'il existe une relation entre ces deux variables. Une latitude proche de 0° indiquant une ville se situant à proximité de l'équateur avec donc une température élevée, la covariance entre ces deux variables est négative. On observe (graphique de gauche ci-dessous) un nuage de points étirés le long d'une droite avec un individu qui semble obéir à une logique différente, Seattle. En revanche, la position est-ouest est « sans influence » sur la température et donc la covariance est proche de 0. Sur le graphique de droite ci-dessous, les points sont répartis

<sup>12</sup> En statistique descriptive, la covariance se calcule ainsi :

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}$$

« aléatoirement » dans le plan c'est-à-dire sans qu'il ne soit possible d'observer a priori une forme spécifique.



Propriétés<sup>13</sup> de la covariance :

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

$$Cov(X, X) = Var(X)$$

$$Cov(a_1X_1 + a_2X_2, Y) = a_1Cov(X_1, Y) + a_2Cov(X_2, Y)$$

$$Cov(X, Y) = 0 \text{ si } X \text{ et } Y \text{ sont indépendants}$$

### Coefficient de corrélation linéaire

La covariance est un indicateur statistique qui prend ses valeurs sur  $\mathbb{R}$  et dont l'échelle de mesure dépend de celles des variables  $X$  et  $Y$  dont elle évalue la relation. En complément de la covariance, on définit donc le **coefficient de corrélation linéaire** qui, comme son nom l'indique, va mesurer l'intensité de la relation linéaire qui lie les variables  $X$  et  $Y$ . Ce coefficient est égal au rapport entre la covariance entre  $X$  et  $Y$  et le produit des écarts-types de  $X$  et  $Y$  :

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

Sachant que  $Cov(X, X) = Var(X)$  et que  $Cov(X, -X) = -Var(X)$ , on vérifiera aisément que le coefficient de corrélation linéaire est borné :  $-1 \leq r_{XY} \leq 1$ . Comme pour la covariance, si  $X$  et  $Y$  sont indépendantes alors  $r_{XY} = 0$  mais la réciproque n'est pas vraie.

### 1.5. Lois de probabilité discrètes.

Nous allons maintenant introduire trois lois de probabilité discrète<sup>14</sup>, les lois de Bernoulli, binomiales et hypergéométriques. Ces trois lois partagent le fait qu'elles permettent de modéliser

<sup>13</sup> Les propriétés de l'espérance (moyenne), de la variance et de la covariance que nous venons de voir seront reprises dans le document sur la régression linéaire.

des phénomènes qui se traduisent systématiquement par deux issues (exclusives et exhaustives). Avant cela, nous devons tout d'abord introduire les notions de « factorielle », « d'arrangements » et de « combinaisons ».

### Permutations sans répétition (factorielle).

Considérons  $n$  objets qui doivent être placés dans un certain ordre sachant que chaque objet ne peut être utilisé qu'une seule fois et posons-nous la question de savoir combien cela représente de possibilités (permutations) ? Nous pouvons choisir le premier objet parmi les  $n$  objets disponibles. Pour le deuxième, il n'en reste plus que  $n-1$ , puis  $n-2$  pour le troisième :

Rang de l'objet	Objet 1	Objet 2	Objet 3	...	Objet $n-1$	Objet $n$
Nombre d'objets disponibles	$n$ choix	$(n-1)$ choix	$(n-2)$ choix	...	2 choix	1 choix

Au final, le nombre de permutations (sans répétition) de  $n$  objets est égal à :

$$P_n = n * (n - 1) * (n - 2) ... * 2 * 1 = n!$$

On nomme cette expression « **factorielle**<sup>15</sup> ».

### Arrangements

$n$  objets doivent être placés **dans un certain ordre** et sans répétition sachant que l'on n'en retient que  $p$  ( $p \leq n$ ). Selon la même logique que précédemment, nous pouvons choisir le premier objet parmi les  $n$  objets disponibles. Pour le deuxième, il n'en reste plus que  $n-1$ , puis  $n-2$  pour le troisième... Pour le  $p^{\text{ième}}$  objet, il n'en restera plus que  $n-p+1$  :

Rang de l'objet	Objet 1	Objet 2	Objet 3	...	Objet $p$
Nombre d'objets disponibles	$n$ choix	$(n-1)$ choix	$(n-2)$ choix	...	$(n-p+1)$ choix

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

---

<sup>14</sup> Il existe bien d'autres lois discrètes que vous aborderez notamment dans le cadre du cours de simulation mais qui ne seront pas présentées ici.

<sup>15</sup> On pose  $0!=1$

Exemple : choisir  $p=2$  lettres parmi  $n=4$  (A, B, C et D)

<b>AB</b>	<b>AC</b>	<b>AD</b>
<b>BA</b>	<b>BC</b>	<b>BD</b>
<b>CA</b>	<b>CB</b>	<b>CD</b>
<b>DA</b>	<b>DB</b>	<b>DC</b>

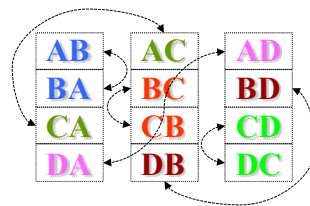
On notera que  $A_n^n = n!$

### Combinaisons

$n$  objets doivent être placés **dans n'importe quel ordre** et sans répétition sachant que l'on n'en retient que  $p$  ( $p \leq n$ ). La situation est donc comparable au cas précédent sauf que si l'on reprend l'exemple des lettres, la combinaison AB devient identique à la combinaison BA puisqu'on ne tient pas compte de l'ordre d'apparition des lettres :

$$C_n^p = \frac{n!}{p!(n-p)!}$$

Exemple : choisir  $p=2$  lettres parmi  $n=4$  (A, B, C et D)



On notera que  $C_n^0 = C_n^n = 1$

### Loi de Bernoulli.

Supposons qu'une expérience aléatoire se traduise par deux événements mutuellement exclusifs et exhaustifs que l'on nommera par commodité « Succès » et « Echec ». Soit  $p$  la probabilité d'un succès et  $q=(1-p)$  celle d'un échec. En définissant la variable aléatoire  $X=1$  pour un succès et 0 sinon (échec), sa loi de probabilité est telle que :

$$P(X = 0) = 1 - p \text{ et } P(X = 1) = p$$

Ce qui peut aussi s'écrire :

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & \text{si } x = 0 \text{ ou } 1 \\ 0 & \text{sinon} \end{cases}$$

L'espérance et la variance d'une variable aléatoire qui suit une loi de Bernoulli sont donc respectivement égales à :

$$E(X) = (1-p).0 + p.1 = p$$

$$V(X) = p(1-p) = pq$$

### Loi Binomiale.

Généralisons la distribution de Bernoulli en supposant que l'on répète cette expérience  $n$  fois et que le résultat d'une expérience n'influence pas celui d'une autre (les réalisations sont indépendantes – on dit également que les tirages sont effectués avec remise). Le nombre de succès  $X$  résultant de ces  $n$  essais peut donc varier entre 0 et  $n$ . Si on s'intéresse à la probabilité d'obtenir exactement  $X=k$  succès en  $n$  essais, alors  $X$  suit une loi binomiale de paramètres  $n$  et  $p$  ce que l'on note  $X \sim \mathcal{B}(n, p)$ . On montre que :

$$P(X = k) = C_n^k p^k (1 - p)^{n-k} = C_n^k p^k q^{n-k}$$

L'espérance et la variance d'une variable aléatoire qui suit une loi binomiale sont :

$$E(X) = np$$

$$V(X) = np(1 - p) = npq$$

### Loi Hypergéométrique.

Si l'hypothèse selon laquelle le résultat d'une expérience n'influence pas celui d'une autre n'est pas respectée (c'est-à-dire que l'on effectue un **tirage sans remise**) alors on remplacera la distribution binomiale par la distribution hypergéométrique. Soient  $N$  la taille de la population totale qui comprend  $M$  individus<sup>16</sup> « positifs »,  $p$  ( $p=M/N$ ) la probabilité d'un succès (ou proportion de positifs) et  $n$  le nombre d'essais, le nombre de succès  $X$  résultant de ces  $n$  essais suit une loi hypergéométrique de paramètres  $N$ ,  $n$  et  $p$  ce que l'on note  $X \sim \mathcal{H}(N, n, p)$ . L'espérance et la variance d'une variable aléatoire qui suit une loi hypergéométrique sont :

$$E(X) = np$$

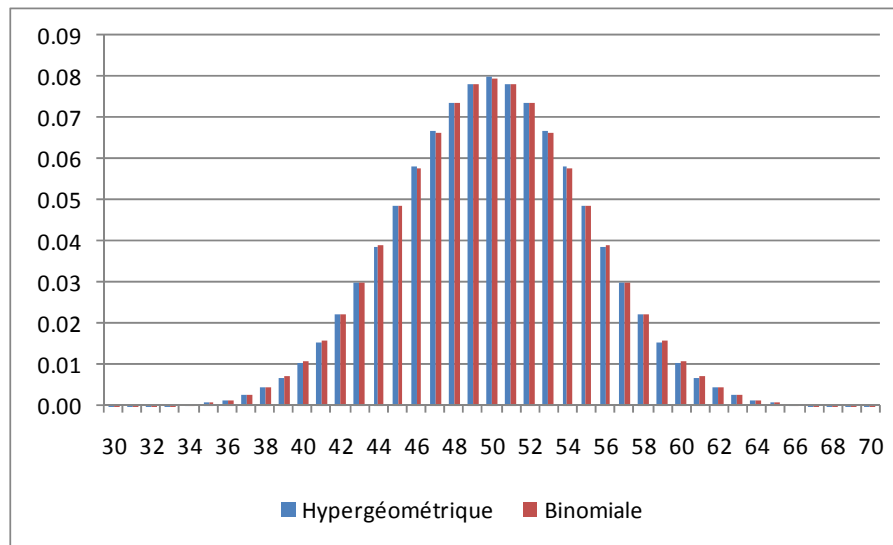
$$V(X) = \frac{N - n}{N - 1} npq$$

Remarque : à condition que  $N$  et  $M$  soient grands, la loi hypergéométrique  $\mathcal{H}(N, n, p)$  peut être approximée par une loi binomiale  $\mathcal{B}(n, p)$ . En pratique cette approximation est possible dès lors que le taux de sondage  $\frac{n}{N}$  est inférieur à 10 %.

---

<sup>16</sup> Et donc  $N-M$  individus « négatifs »

Dans le graphique ci-dessous, nous avons illustré cette approximation en prenant le cas où  $N = 10\,000$ ,  $M = 5\,000$  et  $n = 100$  :



## 1.6. Variable aléatoire et lois de probabilité continues.

### Variable aléatoire continue

Une variable aléatoire  $X$  est **continue**<sup>17</sup> si il existe une fonction  $f$  (définie sur  $\mathbb{R}$ ) qui possède les trois propriétés suivantes :

- 1) pour tout  $x \in \mathbb{R}$ ,  $f(x) \geq 0$
- 2)  $f$  est continue sur  $\mathbb{R}$
- 3) pour tout  $x \in \mathbb{R}$ ,  $F(x) = \int_{-\infty}^x f(t)dt$  et en particulier  $\int_{-\infty}^{\infty} f(t)dt = 1$

Dans ce cas,  $f$  est appelée « **densité de probabilité** » de  $X$  tandis que  $F$  est sa « **fonction de répartition** ». D'un point de vue mathématique,  $F$  est la fonction primitive de  $f$ . A l'aide de cette définition, on peut donc calculer la probabilité qu'une variable aléatoire  $X$  soit inférieure à une valeur  $a$  ainsi :

$$P(X \leq a) = \int_{-\infty}^a f(t)dt$$

Cette définition nous permet de poser que :

$$\text{pour tout couple } a, b \in \mathbb{R}, P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a) = \int_a^b f(t)dt$$

$$\text{ainsi pour tout } x \in \mathbb{R}, P(X = x) = F(x) - F(x) = 0 \text{ et donc } P(X < x) = P(X \leq x)$$

<sup>17</sup> De façon plus intuitive, et par analogie avec une variable aléatoire discrète, on dira qu'une variable aléatoire est continue si elle prend un nombre infini de valeurs (et donc des valeurs décimales) entre deux bornes.

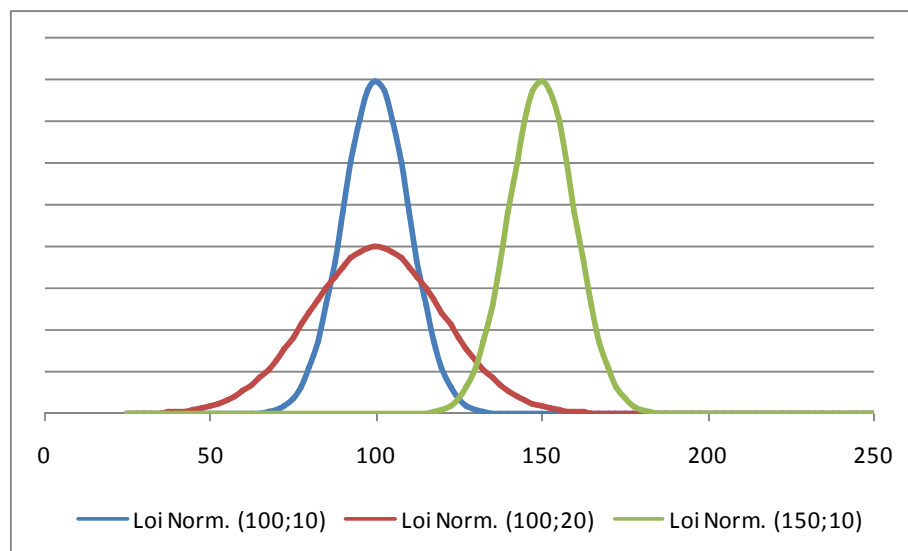


## Loi normale - Présentation

Une variable aléatoire  $X$  suit une **loi normale**<sup>18</sup> (ou loi normale gaussienne ou loi de Laplace-Gauss) de moyenne  $\mu$  et d'écart-type  $\sigma$  – ce que l'on note  $X \sim \mathcal{N}(\mu, \sigma)$  – si elle admet pour densité de probabilité la fonction

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

La moyenne  $\mu$  est un indicateur de tendance centrale tandis que l'écart-type  $\sigma$  mesure la dispersion des données autour de la tendance centrale. Sur le graphique ci-dessous, nous avons donc retenu trois couples « moyenne – écart-type » qui permettent de voir comment se comporte la loi suivant les valeurs respectives de ces paramètres. Pour une moyenne donnée ( $\mu = 100$ ), une augmentation de l'écart-type conduit à un plus grand étalement de la densité (courbe bleue vs. courbe rouge). Pour un écart-type donné ( $\sigma = 10$ ), une augmentation de la moyenne se traduit par un déplacement de la densité vers la droite (courbe bleue vs. courbe verte) :



Quelles que soient les valeurs de la moyenne  $\mu$  et de l'écart-type  $\sigma$ , on constate que la loi normale est symétrique autour de sa moyenne qui est donc également sa valeur médiane et son mode.

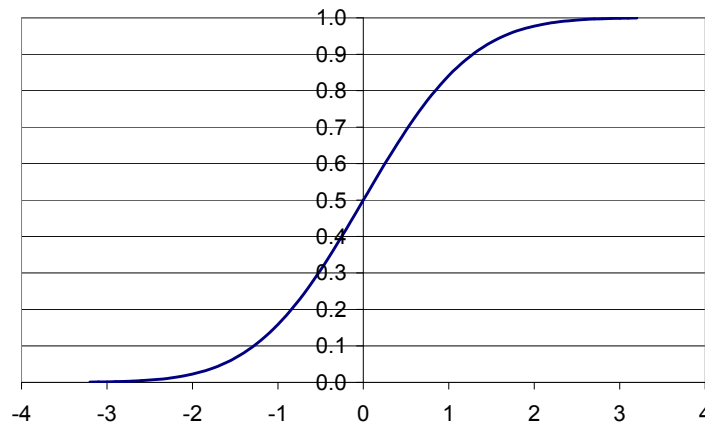
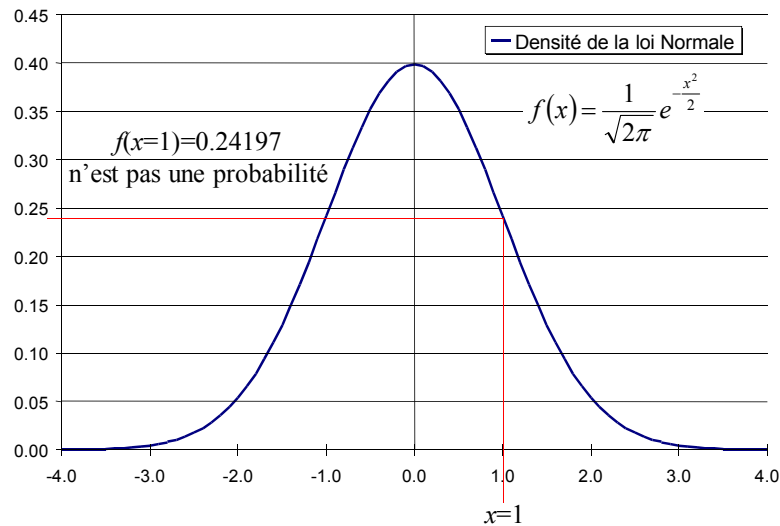
Si  $X \sim \mathcal{N}(\mu, \sigma)$  alors  $T = \frac{X-\mu}{\sigma}$  suit une loi normale centrée (le fait de soustraire la moyenne) et réduite (le fait de diviser par l'écart-type). En utilisant les propriétés de l'espérance et de la

<sup>18</sup> La loi normale est, de loin, la loi la plus « importante » en probabilité et ceux pour au moins deux raisons : elle permet de modéliser énormément de phénomènes (taille, poids d'individus) et de nombreuses autres lois statistiques peuvent être approchées par la loi normale (notamment dans le cas des grands échantillons). Elle sera donc particulièrement utilisée en théorie des tests.

variance, on en déduit que  $\mu_T = 0$  et  $\sigma_T^2 = \sigma_T = 1$  et donc que  $T \sim \mathcal{N}(0,1)$  admet pour fonction de densité :

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

Dans les deux graphiques ci-dessous, nous avons fait figurer la représentation graphique de la densité et de la répartition de la loi normale centrée-réduite :



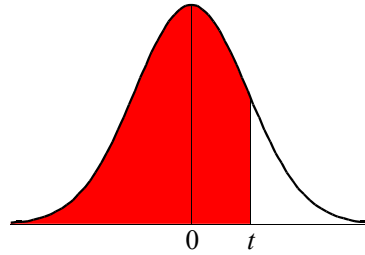
La fonction de densité de la loi normale n'admettant pas de primitive parmi les fonctions usuelles, on ne peut donc pas calculer  $F(\cdot)$  aisément c'est-à-dire sans l'aide d'un ordinateur. C'est pour cela que les valeurs de la fonction de répartition de la loi normale centrée réduite sont reprises dans des « tables<sup>19</sup> ». En définissant :

$$\pi(t) = P(T < t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

<sup>19</sup> Dans le cours, nous verrons comment utiliser Excel pour calculer les probabilités directement.

$\pi(t)$  donne la probabilité qu'une variable aléatoire  $T$  soit inférieure à une valeur  $t$  quelconque.

Graphiquement cela correspond en fait à l'aire représentée en rouge dans le graphique ci-dessous :



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Si on souhaite connaître la probabilité que  $T$  soit inférieure à un fractile  $t=1.96$ , on commence par décomposer  $t$  en notant que  $t=1.9+0.06$ . A l'intersection de la ligne 1.9 et de la colonne 0.06 (dont la somme vaut donc 1.96) on lit la valeur 0.975 ce qui signifie que :

$$\pi(1.96) = P(T < 1.96) = 0.975 = 97.5 \%$$

Valeurs remarquables de la loi normale :

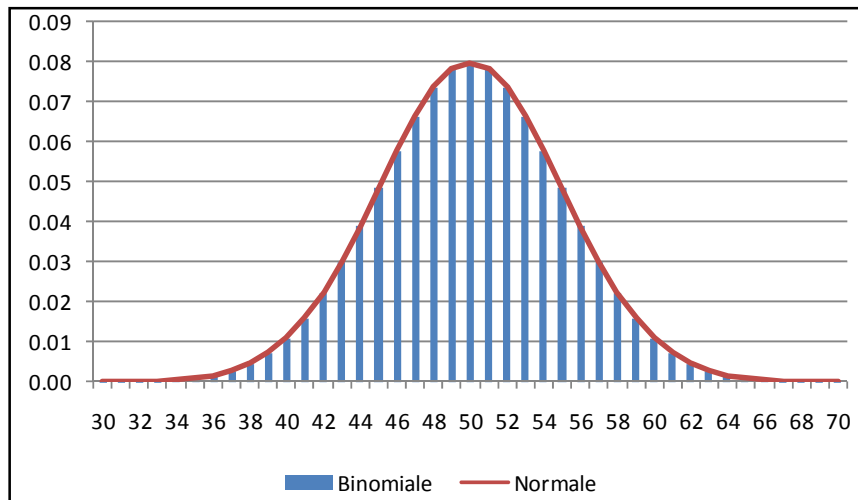
$$P\left(\mu - \frac{2}{3}\sigma < X < \mu + \frac{2}{3}\sigma\right) \approx 50 \%$$

$$P(\mu - 1\sigma < X < \mu + 1\sigma) \approx 68 \%$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 95 \%$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 99.7 \%$$

Remarque : la loi normale est une loi continue. Néanmoins, sous certaines conditions, elle peut être utilisée pour approximer la loi binomiale qui elle est discrète. Ainsi, à condition que  $n$  soit grand,  $p$  pas trop proche de 0 ou 1 et  $npq > 5$ , la loi binomiale  $\mathcal{B}(n, p)$  peut être approximée par une loi normale  $\mathcal{N}(np, \sqrt{npq})$ . Dans le graphique ci-dessous, nous avons illustré cette approximation en prenant le cas où  $n = 100$ ,  $p = 0.5$  :



### Loi normale - Exercice illustratif

Une étude récente a été réalisée sur les mensurations des hommes et des femmes du pays d'Utopie afin d'adapter les tailles des vêtements. Pour simplifier le problème, on ne s'intéressera qu'aux hommes et on supposera que la variable taille suit une loi normale de moyenne  $\mu = 175 \text{ cm}$  et d'écart-type  $\sigma = 10 \text{ cm}$  (la variable poids sera ici ignorée). En précisant explicitement les calculs, dans combien de cas (en %) observe-t-on un homme mesurant :

- Moins de 1,85 mètre ?

$$P(X \leq 185) = P\left(T = \frac{X - 175}{10} \leq \frac{185 - 175}{10}\right) = P(T \leq 1) = \pi(1) = 0.8413 = 84.13 \%$$

- Moins de 1,60 mètre ?

$$P(X \leq 160) = P\left(T = \frac{X - 175}{10} \leq \frac{160 - 175}{10}\right) = P(T \leq -1.5) = 1 - \pi(1.5) = 1 - 0.9332 = 6.68 \%$$

- Plus de 2 mètres ?

$$P(X \geq 200) = P\left(T = \frac{X - 175}{10} \geq \frac{200 - 175}{10}\right) = P(T \geq 2.5) = 1 - \pi(2.5) = 1 - 0.9938 = 0.62 \%$$

- Entre 1,65 mètre et 1,85 mètre ?

$$P(165 \leq X \leq 185) = P\left(\frac{165 - 175}{10} \leq T = \frac{X - 175}{10} \leq \frac{185 - 175}{10}\right) = P(-1 \leq T \leq 1)$$

$$P(T \leq -1) = 1 - \pi(1) = 1 - 0.8413 = 15.87 \%$$

$$P(T \leq 1) = \pi(1) = 0.8413 = 84.13 \%$$

$$P(-1.5 \leq T \leq 1) = 84.13 \% - 15.87 \% = 68.26 \%$$

Pour simplifier de nouveau le problème, on suppose que l'on cherche à définir simplement trois intervalles de tailles (moins de  $x_1$  cm, entre  $x_1$  et  $x_2$  cm et plus de  $x_2$  cm) de façon que chaque intervalle concerne exactement 1/3 de la population.

- déterminez les deux tailles  $x_1$  et  $x_2$  en explicitant vos calculs.

Posons  $t_2 = \frac{x_2 - 175}{10}$ , on cherche donc la valeur  $t_2$  telle que  $\pi(t_2) = \frac{2}{3}$ . Par **lecture inverse** de la table de la loi normale, on trouve :

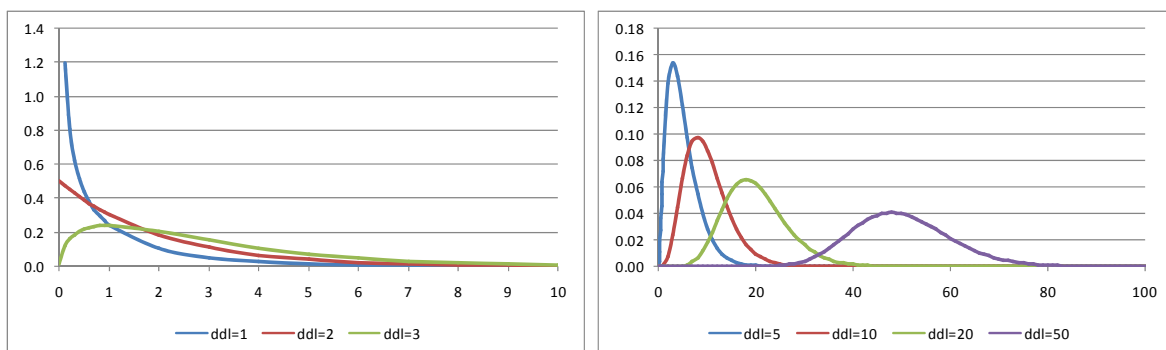
$$\pi^{-1}\left(\frac{2}{3}\right) = 0.4307 = t_2$$

Et par conséquent :  $x_2 = 175 + 10 \times t_2 = 179.31$ . En tirant parti de la symétrie de la loi normale, on en déduit que :  $x_1 = 170.69$ .

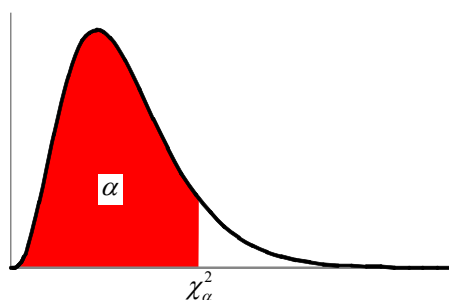
A la suite de la loi normale, nous allons maintenant présenter deux lois continues particulièrement utiles dans le cadre de la théorie des tests pour la première (loi du  $\chi^2$  - prononcer Chi-deux ou Chi-carré, le terme « Chi » étant parfois écrit « Khi ») et de l'estimation et de la théorie des tests pour la seconde (loi de Student). Notez qu'il existe un très grand nombre de lois continues (uniforme, triangulaire, log-normale, exponentielle, de Cauchy, de Weibull,...) mais ce sont essentiellement les deux à venir que nous allons rencontrer dans le cadre de ce cours.

### Loi du $\chi^2$

Soient  $n$  variables aléatoires indépendantes suivant toutes une loi normale centrée-réduite, alors la somme de leur carré,  $X = X_1^2 + X_2^2 + \dots + X_n^2$  suit une loi du Chi-deux à  $n$  degrés de liberté ( $n$  entier supérieur ou égal à 1) que l'on note  $\chi^2(n)$ . La fonction de densité de cette loi n'est pas étudiée ici. En revanche, sur le graphique ci-dessous nous l'avons représentée pour différents degrés de liberté (notés ddl) :



Comme dans le cas de la loi normale, le calcul des probabilités n'est pas aisé. Cette loi est donc tabulée :



$\alpha \backslash n$	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995
1	0.708	1.074	1.642	2.706	3.841	5.024	6.635	7.879
2	1.833	2.408	3.219	4.605	5.991	7.378	9.210	10.597
3	2.946	3.665	4.642	6.251	7.815	9.348	11.345	12.838
4	4.045	4.878	5.989	7.779	9.488	11.143	13.277	14.860
5	5.132	6.064	7.289	9.236	11.070	12.833	15.086	16.750
6	6.211	7.231	8.558	10.645	12.592	14.449	16.812	18.548
7	7.283	8.383	9.803	12.017	14.067	16.013	18.475	20.278
8	8.351	9.524	11.030	13.362	15.507	17.535	20.090	21.955
9	9.414	10.656	12.242	14.684	16.919	19.023	21.666	23.589
10	10.473	11.781	13.442	15.987	18.307	20.483	23.209	25.188
11	11.530	12.899	14.631	17.275	19.675	21.920	24.725	26.757
12	12.584	14.011	15.812	18.549	21.026	23.337	26.217	28.300
13	13.636	15.119	16.985	19.812	22.362	24.736	27.688	29.819
14	14.685	16.222	18.151	21.064	23.685	26.119	29.141	31.319
15	15.733	17.322	19.311	22.307	24.996	27.488	30.578	32.801
16	16.780	18.418	20.465	23.542	26.296	28.845	32.000	34.267
17	17.824	19.511	21.615	24.769	27.587	30.191	33.409	35.718
18	18.868	20.601	22.760	25.989	28.869	31.526	34.805	37.156
19	19.910	21.689	23.900	27.204	30.144	32.852	36.191	38.582
20	20.951	22.775	25.038	28.412	31.410	34.170	37.566	39.997
21	21.991	23.858	26.171	29.615	32.671	35.479	38.932	41.401
22	23.031	24.939	27.301	30.813	33.924	36.781	40.289	42.796
23	24.069	26.018	28.429	32.007	35.172	38.076	41.638	44.181
24	25.106	27.096	29.553	33.196	36.415	39.364	42.980	45.559
25	26.143	28.172	30.675	34.382	37.652	40.646	44.314	46.928
26	27.179	29.246	31.795	35.563	38.885	41.923	45.642	48.290
27	28.214	30.319	32.912	36.741	40.113	43.195	46.963	49.645
28	29.249	31.391	34.027	37.916	41.337	44.461	48.278	50.993
29	30.283	32.461	35.139	39.087	42.557	45.722	49.588	52.336
30	31.316	33.530	36.250	40.256	43.773	46.979	50.892	53.672
40	41.622	44.165	47.269	51.805	55.758	59.342	63.691	66.766
50	51.892	54.723	58.164	63.167	67.505	71.420	76.154	79.490
60	62.135	65.227	68.972	74.397	79.082	83.298	88.379	91.952
70	72.358	75.689	79.715	85.527	90.531	95.023	100.425	104.215
80	82.566	86.120	90.405	96.578	101.879	106.629	112.329	116.321
90	92.761	96.524	101.054	107.565	113.145	118.136	124.116	128.299
100	102.946	106.906	111.667	118.498	124.342	129.561	135.807	140.169

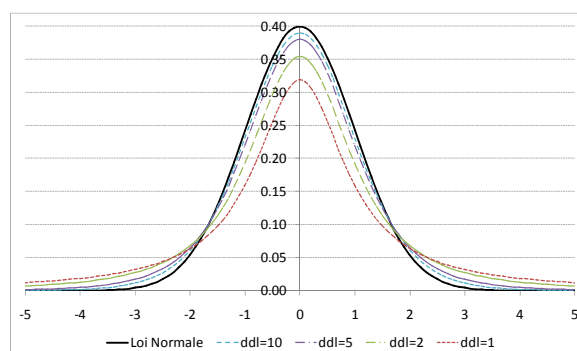
## Loi de Student

Très utilisée en estimation et en théorie des tests, la loi de Student a été développée, comme son nom ne l'indique pas, par William Sealy Gosset (13 juin 1876 – 16 octobre 1937), statisticien anglais qui était employé par la brasserie Guinness pour stabiliser le goût de la bière. Elle est définie pour une variable aléatoire  $T$  égale au rapport entre une variable aléatoire  $X$  qui suit une loi normale centrée-réduite et la racine carrée d'une variable aléatoire  $Y$  qui suit une loi du  $\chi^2$  avec  $n$  degrés de liberté :

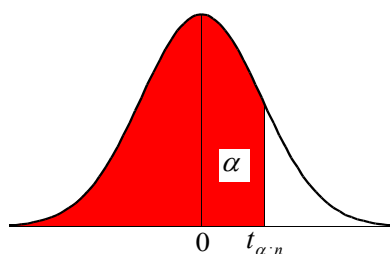
$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

Comme pour loi du  $\chi^2$ , la loi de Student possède un argument qui est le nombre de degrés de liberté. Lorsque ce nombre de degrés de liberté devient important ( $n > 30$ ), la loi de Student peut être

approximée par la loi normale centrée-réduite. Nous avons représenté ci-dessous la fonction de densité de la loi de Student avec, à titre de comparaison, la loi normale représentée en noir épais :



Comme dans le cas de la loi normale, le calcul des probabilités n'est pas aisé. Cette loi est donc tabulée :



$\alpha \backslash n$	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.995
1	0.325	1.000	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.684	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.684	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.683	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.681	0.851	1.303	1.684	2.021	2.423	2.704
50	0.255	0.679	0.849	1.299	1.676	2.009	2.403	2.678
60	0.254	0.679	0.848	1.296	1.671	2.000	2.390	2.660
70	0.254	0.678	0.847	1.294	1.667	1.994	2.381	2.648
80	0.254	0.678	0.846	1.292	1.664	1.990	2.374	2.639
90	0.254	0.677	0.846	1.291	1.662	1.987	2.368	2.632
100	0.254	0.677	0.845	1.290	1.660	1.984	2.364	2.626
200	0.254	0.676	0.843	1.286	1.653	1.972	2.345	2.601
500	0.253	0.675	0.842	1.283	1.648	1.965	2.334	2.586
Infini	0.253	0.674	0.842	1.282	1.645	1.960	2.326	2.576

## 2. Échantillonnage, estimation et intervalle de confiance.

### 2.1. Échantillonnage.

L'échantillonnage consiste à prélever des échantillons (liste de  $n$  individus,  $n$  étant la taille de l'échantillon) au sein d'une population de taille  $N$ . Le rapport  $\frac{n}{N}$  définit ce que l'on appelle le taux de sondage. La base de sondage est quant à elle une liste (souvent informatique) exhaustive (ou au moins quasi-exhaustive) de tous les individus qui composent la population dans son ensemble. Comme l'indique le site « Statistique Canada » : « *L'échantillonnage permet aux statisticiens de tirer des conclusions au sujet d'un tout en y examinant une partie. Il nous permet d'estimer des caractéristiques d'une population en observant directement une partie de l'ensemble de la population. Les chercheurs ne s'intéressent pas à l'échantillon lui-même, mais à ce qu'il est possible d'apprendre à partir de l'enquête et à la façon dont on peut appliquer cette information à l'ensemble de la population.* » Dans bien des cas, avoir accès à l'ensemble de la population est tout simplement impossible pour des raisons de coûts. Par exemple, que penser d'un constructeur automobile qui passerait tous ses véhicules en sortie de chaîne à un « crash-test »... En présence de contrôle qualité dits destructifs, le recours à un échantillon s'impose de lui-même.

Il existe deux types de méthodes d'échantillonnage : l'échantillonnage aléatoire (ou probabiliste) et l'échantillonnage raisonné (ou non probabiliste). Dans le premier cas, on laisse le hasard sélectionner les individus, chaque individu ayant une certaine (souvent la même) probabilité d'être prélevé. Dans le second cas, on « choisit » les individus afin d'imposer à l'échantillon d'avoir des caractéristiques comparables à la population dont il est issu. C'est notamment le cas de la méthode des « quotas ». Une présentation très claire et complète de ces méthodes est disponible à l'adresse suivante<sup>20</sup> : <http://www.statcan.gc.ca/edu/power-pouvoir/ch13/5214895-fra.htm>.

Par ailleurs, un échantillon peut être prélevé avec (échantillon non-exhaustif) ou sans remise (échantillon exhaustif). Dans le premier cas, un même individu peut donc apparaître plusieurs fois dans l'échantillon alors qu'il ne sera présent (le cas échéant) qu'une seule fois dans le second mode de tirage. De façon pragmatique, lorsque le taux de sondage  $\frac{n}{N}$  est faible (en pratique 10 %), la différence entre les deux méthodes est négligeable.

### 2.2.

---

<sup>20</sup> L'adresse [Hhttp://www.statcan.gc.ca/edu/power-pouvoir/toc-tdm/5214718-fra.htm](http://www.statcan.gc.ca/edu/power-pouvoir/toc-tdm/5214718-fra.htm)H donne des informations générales sur le « pouvoir des données ».



## Estimation

Dans le cadre qui nous intéresse, lorsqu'un **paramètre** (moyenne, variance, proportion,...) d'une population est inconnu, on va chercher à l'estimer. Pour cela, on définit un **estimateur**, qui, calculé sur un échantillon, nous fournit une **estimation** du paramètre en question. Un **estimateur** est donc une **variable aléatoire** qui peut prendre différentes valeurs en fonction des échantillons prélevés. Une **estimation** est quant à elle une valeur numérique. En effet, c'est la valeur prise par un estimateur pour un échantillon donné.

Le fait que l'estimation soit liée à l'échantillon nous conduit à penser qu'en prélevant un autre échantillon il est quasi certain qu'on obtienne une autre valeur. On demande donc à un estimateur d'être **sans biais**, c'est à dire d'être égal « en moyenne », sur l'ensemble des échantillons, à la valeur du paramètre qu'il estime. Par ailleurs, la **qualité** d'un estimateur se traduit par le fait qu'il ne doit pas trop fluctuer selon l'échantillon prélevé. En d'autres termes, la qualité d'un estimateur est mesurée par sa variance qui doit donc être la plus faible possible.

Le meilleur estimateur sera celui dont la variance est la plus faible (cette variance est limitée par une borne inférieure, appelée **Borne de Cramer-Rao**). Un estimateur sans biais et de variance minimale est qualifié d'estimateur **efficace**.

### 2.3. Estimation et intervalle de confiance pour une proportion

#### Estimation

On considère une population de taille  $N$  (en général importante) composée d'individus positifs en proportion  $p$  et d'individus négatifs en proportion  $q=1-p$ . Bien évidemment cette proportion  $p$  de la population est inconnue et il n'est pas envisageable de réaliser un recensement. Il faut donc trouver un moyen pour **estimer**  $p$ . La première idée qui vient à l'esprit consiste à prélever dans la population totale un échantillon (petite partie) composé de  $n$  individus et de regarder la proportion d'individus positifs dans cet échantillon (qu'on notera  $f$ ) et on proposera  $f$  comme **estimation** de  $p$ . Suivant la manière dont on prélèvera l'échantillon (c'est-à-dire avec ou sans remise), on retrouve ici le cas binomiale ou hypergéométrique ce qui nous conduira à envisager deux cas différents.

Comme indiqué précédemment, si on prélève un autre échantillon, on trouvera vraisemblablement une autre estimation notée par exemple  $f'$ . En fait, si l'on appelle  $F^{21}$  la proportion d'individus positifs sur un échantillon (de manière générale et sans préciser lequel), on peut considérer  $F$  comme une **variable aléatoire** (qui prendra respectivement les valeurs  $f$  et  $f'$  pour

---

<sup>21</sup>  $F$  ne doit pas ici être confondu avec la fonction de répartition d'une variable aléatoire.

les deux échantillons envisagés). On dira donc que  $F$  est un **estimateur** de  $p$  (c'est-à-dire une façon de calculer une estimation). Si on note  $X$  le **nombre** d'individus positifs sur un échantillon de taille  $n$ , on aura :

$$F = \frac{X}{n}$$

La loi suivie par  $X$  dépend de la façon dont l'échantillon est constitué : s'il est prélevé avec remise alors  $X \sim \mathcal{B}(n, p)$  et s'il est prélevé sans remise,  $X \sim \mathcal{H}(N, n, p)$ .

### Intervalle de confiance avec remise

On a trouvé une estimation de  $p$  par une valeur notée  $f$ . On qualifie cette estimation de **ponctuelle** c'est-à-dire que l'on associe **une seule valeur** à l'estimateur à partir des données observées dans un échantillon. Le problème est qu'on ne sait pas comment cette valeur dépend de l'échantillon qui permet de la calculer. On va donc chercher à donner une estimation « moins fine » de cette valeur sous la forme d'une « fourchette » dont on pourra en revanche préciser la validité c'est-à-dire connaître la probabilité que la valeur inconnue se situe dans la fourchette.

On appelle cette fourchette un **intervalle de confiance** auquel on associe un **niveau de confiance** qui représente lui la probabilité d'avoir raison (et sera noté  $1-\alpha$ ). Un Intervalle de Confiance au niveau  $1-\alpha$  (noté  $IC_{1-\alpha}$ ) sera donc un intervalle dans lequel la valeur à estimer a une probabilité  $1-\alpha$  de se trouver.

Nous allons maintenant déterminer un IC pour  $p$  **dans le cas où  $n$  est supérieur à 100 et  $npq$  supérieur à 5** avec un échantillon prélevé **avec** remise. Dans ce cas, comme on vient de le préciser,  $X$  le nombre d'individus positifs dans l'échantillon suit une loi binomiale de paramètres  $n$  et  $p$ . Puisque nous sommes dans les conditions d'approximation d'une loi binomiale par une loi normale, on peut écrire que :

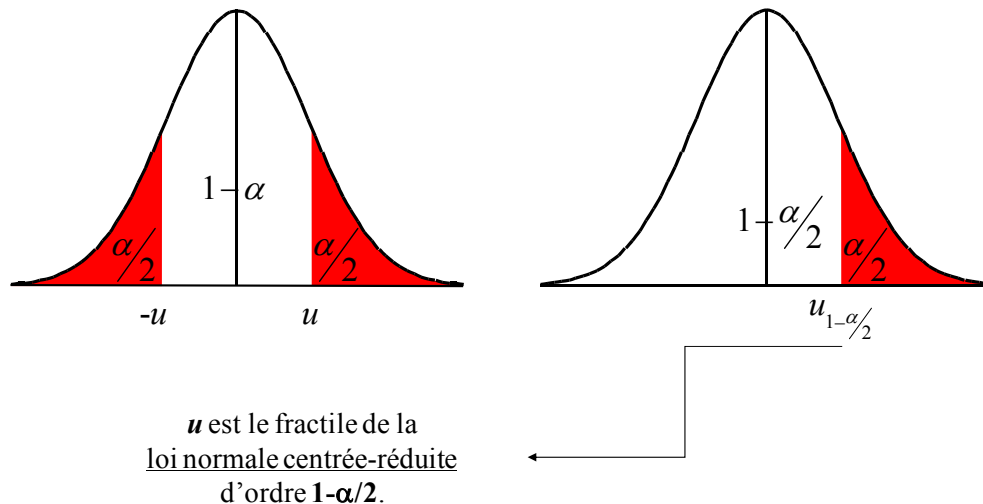
$$X \sim \mathcal{N}(np, \sqrt{npq})$$

$$F = \frac{X}{n} \sim \mathcal{N}\left(p, \sqrt{\frac{pq}{n}}\right)$$

$$\frac{F - p}{\sqrt{\frac{pq}{n}}} \sim \mathcal{N}(0,1)$$

Il ne reste plus qu'à trouver la valeur  $u$  telle que :  $P\left(-u < \frac{F-p}{\sqrt{\frac{pq}{n}}} < u\right) = 1 - \alpha$ .

Puisque la surface comprise entre  $-u$  et  $+u$  doit être égale à  $1-\alpha$ , la surface à l'extérieur doit être égale à  $\alpha$ , donc, pour des raisons de symétrie de la loi normale, elle doit être de  $\alpha/2$  de chaque côté. La surface après  $u$  doit donc être égale à  $\alpha/2$ , la surface avant  $u$  doit donc être égale à  $1-\alpha/2$  comme indiqué sur le graphique suivant.



Comme  $u$  est le fractile d'ordre  $1 - \alpha/2$  de la loi normale centrée-réduite, on le note  $u_{1-\alpha/2}$  et on en déduit que :

$$P\left(F - u_{1-\frac{\alpha}{2}} * \sqrt{\frac{pq}{n}} < p < F + u_{1-\frac{\alpha}{2}} * \sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

On a ainsi défini un IC pour  $p$ , c'est-à-dire un intervalle dans lequel  $p$  a une probabilité  $1-\alpha$  de se situer. Le problème est que les bornes de cet intervalle ne peuvent être calculées que si  $p$  est connu ce qui justement n'est pas le cas puisque  $p$  est inconnu. On peut alors penser à estimer les bornes de cet intervalle en remplaçant  $p$  par son estimation  $f$ , mais en fait, c'est  $p(1-p)$  qu'il faut estimer et  $f(1-f)$  n'est pas une estimation **sans biais** de  $p(1-p)$ . Pour avoir une estimation sans biais<sup>22</sup> de  $p(1-p)$ , il faut prendre  $\frac{n}{n-1} f(1-f)$  ce qui donne comme intervalle de confiance **estimé** :

$$f \pm u_{1-\frac{\alpha}{2}} * \sqrt{\frac{f(1-f)}{n-1}}$$

Pour illustrer notre propos, considérons un lot de production qui est de taille  $n=1\,000$  et supposons (ce n'est pas rassurant) qu'on observe  $X=200$  produits défectueux. On estime la

<sup>22</sup> Comme nous sommes dans le cas où  $n$  est grand ( $> 100$ ), la différence entre  $n$  et  $n-1$  est très peu sensible et donc souvent omise.

proportion de produits défectueux dans l'ensemble de la production grâce à un intervalle de

confiance de 95 % ainsi : 
$$\frac{200}{1000} \pm 1.96 * \sqrt{\frac{\frac{200}{1000}(1-\frac{200}{1000})}{1000-1}}.$$

On peut aussi suivre une autre démarche : puisqu'on ne peut pas connaître les vraies bornes de l'intervalle de confiance, on va en chercher un systématiquement plus grand, on va **majorer** l'intervalle. Pour cela, il faut majorer le rayon de l'intervalle de confiance. Il est aisé de remarquer que  $p(1-p)$  pour  $p$  compris entre 0 et 1 passe par un maximum pour  $p = 1/2$ . Dans ce cas,  $p(1-p)$  vaut donc au maximum  $1/4$ . En conséquence, l'intervalle de confiance **majoré**<sup>23</sup> vaut :

$$f \pm u_{1-\frac{\alpha}{2}} * \frac{1}{2\sqrt{n}}$$

Remarque : cette majoration est la seule possible si on n'a aucune information sur  $p$ . Mais si l'on sait, par exemple, que  $p$  est inférieure à une valeur donnée (notée  $p_0$  et elle-même inférieure à 50 %), on pourra majorer  $p(1-p)$  par  $p_0(1-p_0) = p_0q_0$  et non plus par  $1/4$ . On obtiendra alors la formule suivante pour la majoration de l'IC :

$$f \pm u_{1-\frac{\alpha}{2}} * \frac{\sqrt{p_0q_0}}{n}$$

Cette formule reste valable si on a une information du type :  $p \geq p_0 \geq 1/2$ .

### Intervalle de confiance sans remise

Pour ce qui concerne les échantillons prélevés **sans remise**, la formule pour l'IC tient compte (en partie) du facteur d'exhaustivité<sup>24</sup> et est égale à :

$$f \pm u_{1-\frac{\alpha}{2}} * \sqrt{\frac{f(1-f)}{n-1} * \frac{N-n}{N}}$$

### Taille d'échantillon

Supposons qu'on veuille atteindre, pour notre estimation, une certaine précision; ce qu'on voudra, en fait, c'est réduire la taille de la fourchette. La précision de l'estimation est donc mesurée par le « rayon » de l'IC. Estimer une proportion à 1 % près, par exemple, c'est fixer un rayon de 1 % pour l'IC. La valeur 1 % s'appellera ici la **précision absolue** (de l'estimation) à atteindre, on la notera  $\Delta$ .

<sup>23</sup> C'est en se basant sur cette formule que l'on peut définir la taille  $n$  d'un échantillon.

<sup>24</sup> Si le taux de sondage est faible on voit que  $\frac{N-n}{N} \rightarrow 1$  et donc les deux situations deviennent comparables.

La question qu'on va se poser est la suivante : quelle est la taille d'échantillon nécessaire pour atteindre une précision absolue fixée ?

### *Taille d'échantillon dans le cas avec remise*

Il faut que le rayon de l'IC soit inférieur à  $\Delta$ , mais on ne connaît pas la « vraie » valeur de ce rayon. En revanche, si l'on rend inférieur à  $\Delta$  quelque chose de plus grand que ce rayon, on est sûr d'avoir réussi. Or, on dispose déjà d'un majorant pour ce rayon, donc on écrit :

$$u_{1-\frac{\alpha}{2}} \times \frac{1}{2\sqrt{n}} \leq \Delta \Rightarrow n \geq \frac{1}{4} \left( \frac{u_{1-\frac{\alpha}{2}}}{\Delta} \right)^2$$

**Exemple :** dans le cas d'un IC à 95 %, la taille d'échantillon nécessaire pour estimer une proportion à 1 % près sera de  $\frac{1}{4} \left( \frac{1.96}{0.01} \right)^2 = 9\,604$  pour un échantillon prélevé avec remise. En supposant que le nombre de « positifs » soit de 4 802, on vérifiera que  $f = \frac{4\,802}{9\,604} = \frac{1}{2} = 0.5$  et que les bornes de l'intervalle sont :  $IC_{95\%} = [0.49; 0.51]$ . On voit donc que la proportion est bien estimée à  $\pm 1\%$ . C'est la taille minimale d'échantillon avec remise nécessaire pour atteindre la précision souhaitée, on la notera  $n_0$ .

Remarque : dans le cas où  $p$  est inférieure à une valeur donnée  $p_0$ , la formule précédente est remplacée par une formule avec  $p_0 q_0$  à la place de  $\frac{1}{4}$  :

$$n \geq p_0 q_0 \left( \frac{u_{1-\frac{\alpha}{2}}}{\Delta} \right)^2$$

### *Taille d'échantillon dans le cas sans remise*

Dans le cas d'un échantillon sans remise, le problème est le même, l'équation à résoudre devant juste tenir compte du facteur d'exhaustivité :

$$u_{1-\frac{\alpha}{2}} \times \frac{1}{2\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} \leq \Delta$$

$$u_{1-\frac{\alpha}{2}} \times \frac{1}{2 \sqrt{\frac{n}{\frac{N-n}{N-1}}}}} \leq \Delta$$

On se rend compte que cette équation se résout comme la précédente, on trouve :

$$n = n_0 \times \frac{N-n}{N-1} \Rightarrow n \geq n_0 \times \frac{N}{N+n_0-1}$$

**Exemple :** si la population étudiée est celle d'une ville de 30 000 habitants, dans le cas d'un IC à 95 %, la taille d'échantillon, prélevé sans remise, nécessaire pour estimer une proportion à 1 % près sera de :  $9\,604 \times \frac{30\,000}{30\,000 + 9\,604 - 1} = 7\,275.21 \approx 7\,276$ .

Il va sans dire que si la remarque s'applique (majoration de  $p$  par une valeur inférieure à 1/2 ou minoration par une valeur supérieure à 1/2), la valeur de  $n_0$  sera celle indiquée au paragraphe précédent (avec remise).

## 2.4. Estimation et intervalle de confiance pour une moyenne

### Estimation

Soit  $\mu$  la moyenne d'une variable  $X$  dans une population donnée. Un estimateur sans biais de  $\mu$  obtenu à partir d'un échantillon de taille  $n$  est tel que<sup>25</sup> :

$$\bar{x} = m = \frac{\sum x_i}{n}$$

### Intervalle de confiance

Si la variable  $X$  suit une loi normale de moyenne  $\mu$  et d'écart-type  $\sigma$  alors l'estimateur précédent suit également une loi normale de moyenne  $\mu$  (estimateur sans biais) et d'écart-type  $\frac{\sigma}{\sqrt{n}}$ . L'expression  $\frac{\sigma}{\sqrt{n}}$  montre que la « précision » de l'estimation croît avec la taille de l'échantillon.

Si on ne connaît pas la loi de  $X$  mais que l'on a un échantillon de « grande » taille (en pratique  $n \geq 30$ ) alors l'estimateur suit également une loi normale de moyenne  $\mu$  et d'écart-type  $\frac{\sigma}{\sqrt{n}}$ .

Si l'écart-type de la population est connu (cas théorique car pour connaître  $\sigma$  il faut connaître  $\mu$  et le problème de son estimation ne se pose donc plus) :

$$IC = \left[ m - u_{1-\frac{\alpha}{2}} * \sqrt{\frac{\sigma}{n}} ; m + u_{1-\frac{\alpha}{2}} * \sqrt{\frac{\sigma}{n}} \right]$$

où  $u_{1-\frac{\alpha}{2}}$  est le quantile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale centrée-réduite<sup>26</sup>.

Si l'écart-type de la population est inconnu, on obtient alors :

$$IC = \left[ m - t_{1-\frac{\alpha}{2}; n-1} * \sqrt{\frac{s}{n-1}} ; m + t_{1-\frac{\alpha}{2}; n-1} * \sqrt{\frac{s}{n-1}} \right]$$

<sup>25</sup> Les notations  $\bar{x}$  et  $m$  seront utilisées indifféremment.

<sup>26</sup> Nous sommes ici dans le cas avec remise. Dans le cas sans remise il faut multiplier l'écart-type par le facteur d'exhaustivité  $\sqrt{\frac{N-n}{N-1}}$ .

où  $t_{1-\frac{\alpha}{2}, n-1}$  est le quantile d'ordre  $1 - \frac{\alpha}{2}$  avec  $n - 1$  degrés de liberté de la loi de Student et  $s$  l'estimateur sans biais de  $\sigma$  comme indiqué ci-dessous.

### Taille d'échantillon

Selon la même logique que pour une proportion, il est possible de calculer la taille minimale d'un échantillon permettant, pour un niveau de confiance donné, d'atteindre une certaine précision  $\Delta$ . Notez que cette fois la précision s'exprime non pas sous la forme d'un pourcentage comme pour une proportion mais dans la même unité de mesure que la variable étudiée.

Etablir la taille d'un échantillon dans le cas d'une moyenne, nécessite de connaître la valeur de l'écart-type  $\sigma$  de la variable dans la population. Ce faisant, on montre que la taille  $n$  doit être :

$$n \geq \left( \frac{u_{1-\frac{\alpha}{2}} \times \sigma}{\Delta} \right)^2$$

**Exemple :** dans le cas d'un IC à 95 %, la taille d'échantillon prélevé avec remise nécessaire pour estimer avec une précision absolue (amplitude) de 100 la moyenne d'une variable dont l'écart-type  $\sigma$  est égal à 1 000 sera de  $\left( \frac{1.96 \times 1\,000}{100} \right)^2 = 384.15 \approx 385$ .

Dans le cas d'un échantillon prélevé sans remise, on procède de nouveau en deux étapes : on commence par définir la taille avec remise  $n_0$  puis on applique la correction suivante :

$$n \geq n_0 \times \frac{N}{N + n_0 - 1}$$

Si la population étudiée est celle d'une ville de 30 000 habitants, dans ce cas on vérifiera que la taille minimale de l'échantillon est de 380.

## 2.5. Estimation pour une variance

Dans le cadre d'une population de taille  $N$ , la variance  $\sigma^2$  et l'écart-type  $\sigma$  d'une variable  $X$  sont égaux à :

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N} \text{ et } \sigma = \sqrt{\sigma^2}$$

A partir de ces définitions, on serait tenter d'utiliser comme estimateur de  $\sigma^2$  une expression faisant intervenir  $m$  (estimation de la moyenne) à la place de  $\mu$  et  $n$  (taille de l'échantillon) à la place de  $N$  (taille de la population) :

$$s'^2 = \frac{\sum (x_i - m)^2}{n}$$

Comme on l'a vu précédemment, un estimateur doit être sans biais. Dans le cadre de l'estimation de la variance, on remplacera  $s'^2$  par  $s^2$  qui est tel que :

$$s^2 = \frac{\sum (x_i - m)^2}{n - 1}$$

La différence entre les deux estimateurs est liée au dénominateur. On passe aisément de l'un à l'autre en multipliant (ou divisant) par un facteur  $\frac{n}{n-1}$  dont on voit aisément que la limite lorsque  $n \rightarrow \infty$  est égale à 1.



### 3. Théorie des tests statistiques

#### 3.1. Introduction

Dans la première partie de cette note de cours nous avons plusieurs fois pris pour exemple des dés ou des pièces en précisant à chaque fois qu'on les supposait « parfaitement équilibrés ». Nous allons maintenant inverser cette logique et considérer une pièce de monnaie (nous délaissions les dés) et nous demander si elle est équilibrée.

Pour répondre à cette question, on pourrait faire appel à un physicien et lui demander de « mesurer » la pièce pour vérifier cette hypothèse. Mais en tant que statisticiens nous allons procéder autrement. Dans cette optique, nous allons lancer la pièce et observer ce qui se passe. Bien sûr, effectuer un seul lancer ne servirait à rien et nous allons donc répéter cette procédure un « grand nombre » de fois (disons 100 fois pour fixer les idées). A l'issue de cette expérience, on observe que la pièce est tombée 54 fois sur Pile (et donc 46 fois sur Face). A l'aide de ces informations, peut-on conclure qu'elle est équilibrée ? En fait, on ne peut toujours pas répondre à la question. Intuitivement, on serait tenté de dire qu'elle est équilibrée si le nombre de fois où elle tombe sur Pile est « proche » de 50 mais encore faut-il préciser cette notion.

#### 3.2. Test de conformité d'une proportion

##### Hypothèse nulle et alternative

Au vu des données observées dans un échantillon<sup>27</sup>, on est donc amené à **tester** une hypothèse (« La pièce est équilibrée »), c'est-à-dire à décider, au vu des résultats d'une **expérience aléatoire** (les 100 lancers) si on accepte<sup>28</sup> cette hypothèse ou si on la refuse. Il convient dès lors de définir les **valeurs critiques** à partir desquelles on change d'avis.

Dans le cadre de la théorie des tests statistiques, l'hypothèse testée (ici « La pièce est équilibrée ») est appelée **hypothèse nulle**. Par convention, on la note  $H_0$ . On nomme **hypothèse alternative** (notée  $H_A$  ou  $H_1$ ) l'hypothèse concurrente (à savoir ici que : « La pièce n'est pas équilibrée »).

Quand on accepte  $H_0$ , on refuse  $H_A$  et réciproquement c'est-à-dire que lorsqu'on refuse  $H_0$ , c'est au profit de  $H_A$ . Par conséquent, les hypothèses doivent être mutuellement exclusives (leur intersection est vide) et exhaustives (leur réunion couvre l'ensemble des cas possibles).

---

<sup>27</sup> Cette approche est donc « inférentielle » puisque l'on part de données observées sur un échantillon pour en conclure quelque chose sur la population entière.

<sup>28</sup> Plutôt que « accepter » une hypothèse on préférera ne pas la « rejeter ». Cette subtilité de vocabulaire sera précisée plus tard.

### Erreur de première et de seconde espèce

Il faut maintenant préciser quand accepter  $H_0$  et quand la refuser, c'est-à-dire définir une **règle de décision**. Cette règle devra être telle qu'on commette le moins d'erreurs possible. Or il y a deux types d'erreurs :

- refuser  $H_0$  alors qu'elle est vraie – erreur dite de **1<sup>ère</sup> Espèce** que l'on convient de noter  $\alpha$ . La valeur  $\alpha$  est appelé **seuil** du test.
- ne pas rejeter (accepter)  $H_0$  alors qu'elle est fausse – erreur dite de **2<sup>de</sup> Espèce** que l'on convient de noter  $\beta$ . La valeur  $1 - \beta$  est appelée **puissance du test**.

Le mieux serait de choisir une règle de décision rendant  $\alpha$  et  $\beta$  nulles. Pour rendre  $\alpha$  nulle, il suffit de décider d'accepter toujours  $H_0$  mais alors  $\beta$  sera maximum. De même, pour rendre  $\beta$  nulle, il suffit de toujours refuser  $H_0$  mais alors dans ce cas c'est  $\alpha$  qui sera maximum. Nous pouvons illustrer ce point à l'aide de l'analogie suivante : un délit a été commis et une personne est jugée. Soit  $H_0$  : « la personne est innocente ». Dans ce cadre, l'erreur de première espèce  $\alpha$  consisterait à rejeter à tort l'hypothèse nulle c'est-à-dire à condamner un innocent. L'erreur de seconde espèce  $\beta$  consisterait elle à ne pas rejeter l'hypothèse nulle alors qu'elle est fausse c'est-à-dire à acquitter un coupable :

		Réalité	
		Innocent	Coupable
Procès	Acquitté	OK	$\beta$
	Condamné	$\alpha$	OK

Ainsi, quand on veut faire baisser  $\alpha$  on fait augmenter  $\beta$  (pour être sûr de ne condamner aucun innocent il faudrait acquitter tout le monde et notamment tous les coupables) et réciproquement. Ces deux objectifs sont contradictoires et il faut donc choisir entre les deux puisqu'il n'y a aucun moyen de les satisfaire simultanément (sauf peut-être à allonger la durée de l'instruction pour collecter plus d'informations).

### Procédure de test

La procédure de test que l'on va définir maintenant (et qui est due à Neyman et Pearson) consiste à fixer le seuil  $\alpha$  du test puis, parmi toutes les règles assurant cette valeur de  $\alpha$ , à choisir celle rendant  $\beta$  minimale. Les seuils les plus communément utilisés seront de 5 % (0,05) et de 1 % (0,01). On peut bien sûr choisir a priori n'importe quelle valeur pour le seuil d'un test mais seules des valeurs faibles (proches de 0) auront une utilisation pratique puisqu'elles représentent une probabilité d'erreur. Fixer un faible  $\alpha$  risque de rejet à tort, revient à avoir une attitude prudente ou

conservatrice. Cela se comprend lorsque  $H_0$  est une situation de référence, qu'on n'est pas désireux d'abandonner à moins d'une forte évidence contre elle.

Si on définit  $X$  comme le nombre de Pile obtenus en 100 lancers, on sait que si  $H_0$  est vraie alors  $X$  suit une loi binomiale<sup>29</sup> de paramètres  $n=100$  et  $p=1/2$  dont l'espérance vaut  $E(X) = np = 50$  et la variance  $V(X) = npq = 25$  (son écart-type étant donc égale à  $\sqrt{25} = 5$ ). On pourra dès lors approximer cette loi binomiale par une loi normale de paramètres  $\mu=50$  et  $\sigma=5$  ( $X \sim \mathcal{N}(50,5)$ ). En supposant toujours  $H_0$  vraie, on obtient en centrant et en réduisant :

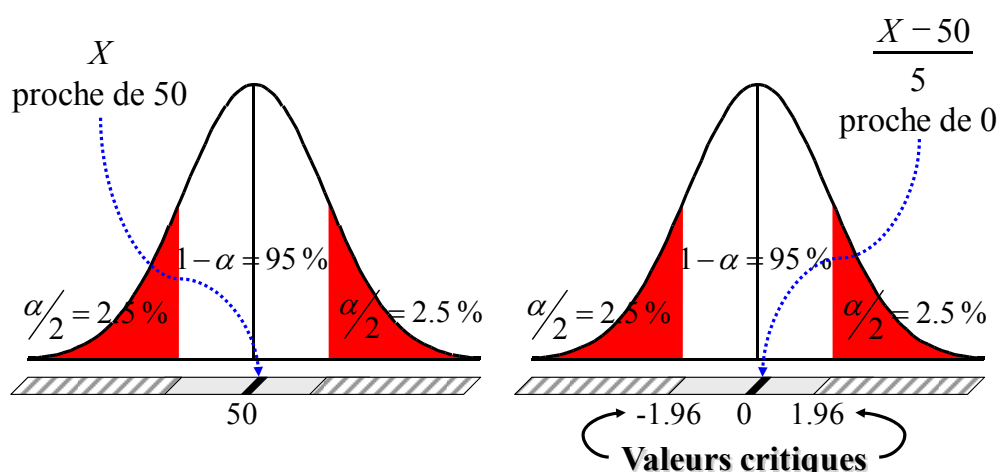
$$T = \frac{X - 50}{5} \sim \mathcal{N}(0,1)$$

On a déjà remarqué qu'une règle de décision « intuitivement raisonnable » consisterait à accepter  $H_0$  à la condition que  $X$  soit proche de 50 (où  $T = \frac{X-50}{5}$  proche de 0) et à refuser  $H_0$  sinon. On acceptera donc  $H_0$  si  $\left| \frac{X-50}{5} \right| \leq u$ ,  $u$  étant la **valeur critique** du test. Il ne nous reste donc plus qu'à déterminer la valeur de  $u$  pour que la probabilité de rejeter une hypothèse vraie soit égale à  $\alpha$  :

$$P\left(-u \leq \frac{X - 50}{5} \leq u\right) = 1 - \alpha$$

$u$  est donc le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale centrée réduite. En résumé, on accepte  $H_0$  si  $\left| \frac{X-50}{5} \right| \leq u_{1-\frac{\alpha}{2}}$  et on la refuse sinon.

Dans notre exemple, si on retient un seuil  $\alpha = 5\%$ , on sait que  $u_{0.975} = 1.96$ . Graphiquement, la situation peut alors être résumée ainsi :



<sup>29</sup> On vérifie bien qu'il n'y a que deux issues possibles (Pile ou Face) et on peut admettre que les résultats des lancers successifs sont sans influence les uns sur les autres (indépendance).

A la suite de notre expérience, on a obtenu  $X=54$  fois Pile. La valeur critique du test vaut donc  $u = \frac{54-50}{5} = \frac{4}{5}$  et comme  $\frac{4}{5} \in [-1.96; 1.96]$  on accepte  $H_0$  à savoir qu'au vu des résultats de notre expérience l'hypothèse selon laquelle la pièce est équilibrée n'est pas remise en cause.

Pour rejeter  $H_0$  au seuil  $\alpha = 5 \%$ , il aurait fallu que  $X \geq 60$  (ou symétriquement que  $X \leq 40$ ). En revanche, pour un seuil  $\alpha = 1 \%$ , on sait que  $u_{0.995} = 2.5758$  et dans ce cas le nombre de piles nécessaire pour rejeter  $H_0$  devrait être  $X \geq 63$  (ou symétriquement  $X \leq 37$ ). En clair, si on est plus exigeant sur le risque de déclarer à tort que la pièce n'est pas équilibrée (rejet de  $H_0$ ), on demande à avoir plus d'observations attestant ce fait<sup>30</sup>.

### Test bilatéral vs. unilatéral

L'hypothèse  $H_0$  selon laquelle «  $H_0$  : la pièce est équilibrée » revient à dire que la proportion de Pile est égale à 50 %. L'hypothèse alternative  $H_A$  traduit l'idée que «  $H_A$  : la pièce n'est pas équilibrée » c'est-à-dire que la proportion de pile est différente de 50 %. Soit  $p$  la proportion de pile, on peut donc écrire les deux hypothèses<sup>31</sup> du test ainsi :

$$\begin{cases} H_0 : p = 0.5 \\ H_A : p \neq 0.5 \end{cases}$$

Le test que nous venons d'étudier appartient donc à la catégorie des **tests de conformité à une norme**. De plus, il s'agit d'un test **bilatéral**. On fait un test bilatéral quand on veut savoir si **deux valeurs sont égales ou différentes**. On fait un test **unilatéral** quand on veut savoir si **l'une est au-dessus ou en dessous de l'autre**. Quand on fait un test unilatéral, il faut choisir l'hypothèse alternative ( $H_A : < \text{ou} >$ ).

Dans le chapitre sur les intervalles de confiance, on a pris l'exemple d'un lot de production de taille  $n=1\,000$  dans lequel on a observé  $X=200$  produits défectueux soit une proportion de 20 %. On cherche à savoir maintenant si cette proportion dans l'ensemble de la production (c'est-à-dire la population) est supérieure à 25 % ?

Dans la mesure où la proportion trouvée sur l'échantillon (qui est une **estimation** de la proportion sur la population restant inconnue) est elle-même inférieure à 25 %, il est inutile de mener le test  $H_0 : p = 25 \%$  contre  $H_A : p > 25 \%$  car on en connaît par avance le résultat :  $H_0$  sera acceptée car aucun test ne conduira à dire que la proportion sur la population totale est supérieure à 25 % quand elle est inférieure sur l'échantillon. C'est donc l'autre test unilatéral qu'il faut mener,

---

<sup>30</sup> Plutôt que de comparer une valeur critique à des fractiles, on verra lors de la présentation des tests dans le cadre de la régression linéaire que les logiciels renvoient ce que l'on nomme une **probabilité critique**, probabilité critique que l'on comparera toujours aux seuils usuels de 1 ou 5 %.

<sup>31</sup> Ou en le basant sur  $X : \begin{cases} H_0 : X = 50 \\ H_A : X \neq 50 \end{cases}$

c'est-à-dire  $H_0 : p = 25 \%$  contre  $H_A : p < 25 \%$ . Si ce test conduit à refuser  $H_0$ , cela voudra dire que l'on pense que  $p$  est effectivement inférieure à 25 %. Mais si ce test conduit à accepter  $H_0$ , cela signifiera concrètement que l'on ne sait pas si  $p$  est inférieure ou supérieure à 25 %. Tout se passe alors comme si une proportion de 20 % trouvée dans l'échantillon n'était pas jugée suffisamment inférieure à 25 % pour qu'on puisse affirmer que la proportion  $p$  inconnue (dont elle n'est qu'une estimation) est elle-même inférieure à 25 %. En résumé, on choisit comme hypothèse alternative d'un test unilatéral l'hypothèse vérifiée par les données de l'échantillon.

### 3.3. Test de conformité d'une moyenne

Etudions maintenant le cas du test de conformité d'une moyenne. On dispose d'un échantillon de taille  $n$  sur une variable  $X$ , variable dont on suppose qu'elle suit une loi normale de moyenne  $\mu$  et d'écart-type  $\sigma$ . On souhaite tester :

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_A : \mu \neq \mu_0 \end{cases}$$

On commence par calculer  $m$  la moyenne sur l'échantillon. Comme dans le cas d'un intervalle de confiance, on distingue ensuite le cas où  $\sigma$  est connu (peu probable) de celui où il ne l'est pas.

Si  $\sigma$  est connu : pour un risque  $\alpha$  donné, on rejette  $H_0$  si

$$m < \mu_0 - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

ou

$$m > \mu_0 + t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

où  $t_{1-\frac{\alpha}{2}}$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale centrée réduite. De manière équivalente, on peut calculer la valeur critique du test :

$$z = \frac{m - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Et on rejette  $H_0$  si :

$$z \notin \left[ -t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}} \right]$$

Si  $\sigma$  est inconnu : on estime l'écart-type à partir des données de l'échantillon :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n - 1}}$$

et pour un risque  $\alpha$  donné, on rejette  $H_0$  dans l'un des deux cas suivants :

$$m < \mu_0 - t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}}$$

$$m > \mu_0 + t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}}$$

où  $t_{1-\frac{\alpha}{2}; n-1}$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi de Student à  $n-1$  degrés de liberté (on rappelle par ailleurs que pour  $n$  grand ( $n > 30$ ) :  $t_{1-\frac{\alpha}{2}; n-1} \approx t_{1-\frac{\alpha}{2}}$ ). De manière équivalente, on peut calculer la valeur critique du test :

$$z = \frac{m - \mu_0}{\frac{s}{\sqrt{n}}}$$

Et on rejette  $H_0$  si :

$$z \notin \left[ -t_{1-\frac{\alpha}{2}; n-1}; t_{1-\frac{\alpha}{2}; n-1} \right]$$

où  $t_{1-\frac{\alpha}{2}; n-1}$  est toujours le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi de Student à  $n-1$  degrés de liberté.

### 3.4. Test de comparaison de deux proportions

Soient deux populations distinctes possédant respectivement une proportion  $p_1$  et  $p_2$  d'individus positifs. On se pose la question de savoir si « ces deux proportions sont identiques » vs. « ces deux proportions sont différentes ». Pour cela, de ces deux populations, on extrait deux échantillons (de taille  $n_1$  pour le premier et  $n_2$  pour le second) et pour un risque  $\alpha$  donné on teste :

$$\begin{cases} H_0 : p_1 = p_2 \\ H_A : p_1 \neq p_2 \end{cases}$$

Pour obtenir la valeur critique du test on commence par calculer  $\hat{p}$  la proportion globale de positifs sur les deux échantillons puis la valeur  $z$  :

$$\hat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

$$z = \frac{f_1 - f_2}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

On rejette  $H_0$  si :

$$z \notin \left[ -t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}} \right]$$

où  $\left[ -t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}} \right]$  où  $t_{1-\frac{\alpha}{2}}$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale.

### 3.5. Test de comparaison d'une moyenne

Dans le cas où on souhaite effectuer un test de comparaison entre deux moyennes, il faut de nouveau se poser la question de savoir si les écarts-types ( $\sigma_1$  et  $\sigma_2$ ) sont connus (cas peu fréquent) ou pas. Dans le cadre de cette note, nous n'allons traiter que le cas où les écarts-types sont inconnus. Par ailleurs, nous allons supposer que nous disposons de deux échantillons de grande taille à savoir que  $n_1$  et  $n_2$  sont tous les deux supposés être supérieurs à 30. Dans ce cas, on calcule la valeur critique du test :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Et on rejette  $H_0$  si :

$$z \notin \left[ -t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}} \right]$$

où  $\left[ -t_{1-\frac{\alpha}{2}}; t_{1-\frac{\alpha}{2}} \right]$  où  $t_{1-\frac{\alpha}{2}}$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi normale

## 4. Synthèses.

### 4.1. Estimation de proportions

- (1)  $\left[ f \pm u_{1-\frac{\alpha}{2}} \times \sqrt{\frac{f(1-f)}{n-1}} \right]$  : Intervalle de confiance au niveau  $1-\alpha$  pour une proportion inconnue (échantillon **avec remise**)
- (2)  $\left[ f \pm u_{1-\frac{\alpha}{2}} \times \sqrt{\frac{f(1-f)}{n-1}} \times \frac{N-n}{N} \right]$  : idem pour un échantillon prélevé **sans remise**
- (3)  $\frac{1}{4} \left( \frac{u_{1-\frac{\alpha}{2}}}{\Delta} \right)^2$  : Taille d'échantillon **avec remise** nécessaire pour atteindre une précision absolue  $\Delta$
- (4)  $p_0 q_0 \left( \frac{u_{1-\frac{\alpha}{2}}}{\Delta} \right)^2$  : idem si l'on a une information sur  $p$  du type :  $p \leq p_0 \leq 1/2$  ou  $p \geq p_0 \geq 1/2$
- (5)  $n_0 \times \frac{N}{N+n_0-1}$  : Taille d'échantillon **sans remise** nécessaire pour atteindre une précision absolue  $\Delta$  en notant  $n_0$  la valeur trouvée en (3) ou (4)

**f** = Proportion observée sur l'échantillon

**u** = Fractile de la loi normale centrée réduite

**1-α** = Niveau de confiance de l'intervalle

**n** = Taille de l'échantillon

**N** = Taille de la population totale

**Δ** = Précision absolue à atteindre (rayon que doit avoir l'intervalle de confiance)

**p<sub>0</sub>** = Majorant (inférieur à 50 %) ou Minorant (supérieur à 50 %) pour la proportion inconnue  $p$

**q<sub>0</sub>** =  $1-p_0$



## 4.2. Estimation de moyennes

(6)  $\left[ \bar{X} \pm u_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \right]$  : Intervalle de confiance au niveau  $1-\alpha$  pour une moyenne inconnue si l'échantillon est prélevé avec remise et  $\sigma$  connu

(7)  $\left[ \bar{X} \pm u_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} \right]$  : idem si l'échantillon est prélevé sans remise

(8)  $\left[ \bar{X} \pm t_{1-\frac{\alpha}{2}}(n-1) \times \frac{s}{\sqrt{n}} \right]$  ou  $\left[ \bar{X} \pm t_{1-\frac{\alpha}{2}}(n-1) \times \frac{s'}{\sqrt{n-1}} \right]$  : idem que (6) si  $\sigma$  inconnu

(9)  $\left( \frac{u_{1-\frac{\alpha}{2}} \times \sigma}{\Delta} \right)^2$  : Taille minimale d'échantillon avec remise nécessaire pour atteindre une précision absolue  $\Delta$

(10)  $n_0 \times \frac{N}{N+n_0-1}$  : idem si l'échantillon est prélevé sans remise en notant  $n_0$  la valeur trouvée en (9)

$\bar{X}$  = Moyenne observée sur l'échantillon

$n$  = Taille de l'échantillon

$s'$  = Ecart-type observé sur l'échantillon

$\Delta$  = Précision absolue (Rayon de l'intervalle de confiance) à atteindre pour l'estimation de la moyenne inconnue sur la population totale

$\sigma$  = Ecart-type sur la population totale

$N$  = Taille de la population totale

$s$  = Estimation de  $\sigma = s' \times \sqrt{\frac{n}{n-1}}$

### 4.3. Tests de comparaison

Test Bilatéral :	$H_0 : "="$	contre	$H_a : " \neq "$	Règle :	Accepter $H_0$ si $ vc  < z_{1-\alpha/2}$ .	Refuser sinon
Tests Unilatéraux :	$H_0 : "="$	contre	$H_a : ">"$	Règle :	Accepter $H_0$ si $vc < z_{1-\alpha}$ .	Refuser sinon
	$H_0 : "="$	contre	$H_a : "<"$	Règle :	Accepter $H_0$ si $vc > -z_{1-\alpha}$ .	Refuser sinon

**Dans un test unilatéral, on choisit comme hypothèse alternative, l'hypothèse vérifiée sur l'échantillon**

On note **vc** la valeur calculée du Test qui prendra différentes valeurs suivant les cas  
**z** représente la valeur tabulée ou critique ; c'est le fractile d'une loi à déterminer suivant les cas

**Cas 1 :** Comparaison d'une proportion inconnue **p** à une norme **p<sub>0</sub>**

**Cas 2 :** Comparaison de deux proportions inconnues **p<sub>1</sub>** et **p<sub>2</sub>**

**Cas 3 :** Comparaison d'une moyenne inconnue **m** à une norme **m<sub>0</sub>** si  $\sigma$  est connu

**Cas 4 :** Comparaison d'une moyenne inconnue **m** à une norme **m<sub>0</sub>** si  $\sigma$  est inconnu

**Cas 5 :** Comparaison de deux moyennes inconnues **m<sub>1</sub>** et **m<sub>2</sub>** si  $\sigma_1$  et  $\sigma_2$  connus

**Cas 6 :** Comparaison de deux moyennes inconnues **m<sub>1</sub>** et **m<sub>2</sub>** si  $\sigma_1$  ou  $\sigma_2$  inconnus et  $n_1$  et  $n_2 > 30$

	<b>Cas 1</b>	<b>Cas 2</b>	<b>Cas 3</b>
VC	$\frac{f - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$	$\frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \text{ avec } f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$	$\frac{\bar{X} - m_0}{\sigma / \sqrt{n}}$
VT	u	u	u
	<b>Cas 4</b>	<b>Cas 5</b>	<b>Cas 6</b>
VC	$\frac{\bar{X} - m_0}{s / \sqrt{n}} \text{ ou } \frac{\bar{X} - m_0}{s' / \sqrt{n-1}}$	$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
VT	t(n-1)	u	u