Chapitre 1

Rappels sur les propriétés de la loi Normale

1.1 Quelques propriétés de la loi Normale

Avant toute chose nous allons dans ce chapitre introductif rappeler les notions les plus importantes de la loi normale qui nous serviront en permanence par la suite.

Définition 1 *On appelle variable aléatoire normale ou gaussienne toute variable aléatoire absolument continue dont la densité de probabilité f est définie par*

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

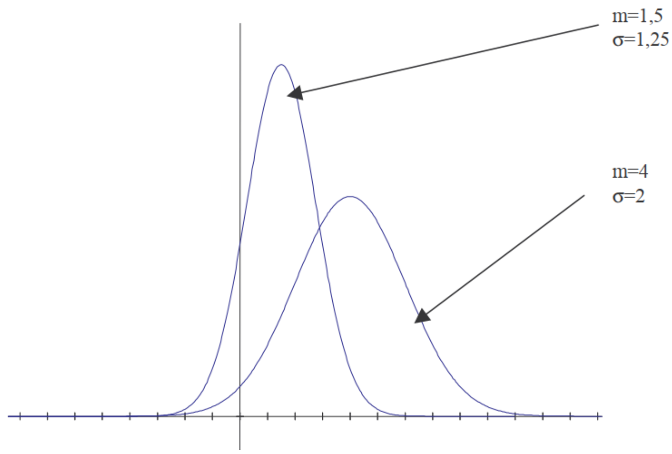


FIG. 1-1 – La courbe représentative de la loi $\mathcal{N}(m, \sigma)$

1.1.1 Fonction de répartition de la loi Normale $\mathcal{N}(m, \sigma)$

Soit Z une variable aléatoire qui suit une loi Normale de moyenne $m = \mu$ et d'écart type σ et dont la densité de probabilité est notée f et la fonction de répartition de la loi normale est donnée par :

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

1.1.2 Loi normale centrée réduite

On dit que X suit une loi normale centrée réduite et on note $X \sim \mathcal{N}(0, 1)$ si sa loi admet pour densité la fonction

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

La courbe représentative de la fonction f est donnée par la Figure 1.2. Sa densité de probabilité est la fonction f définie par

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

— Cette fonction f est paire, la courbe a un axe de symétrie qui est la droite des ordonnées.

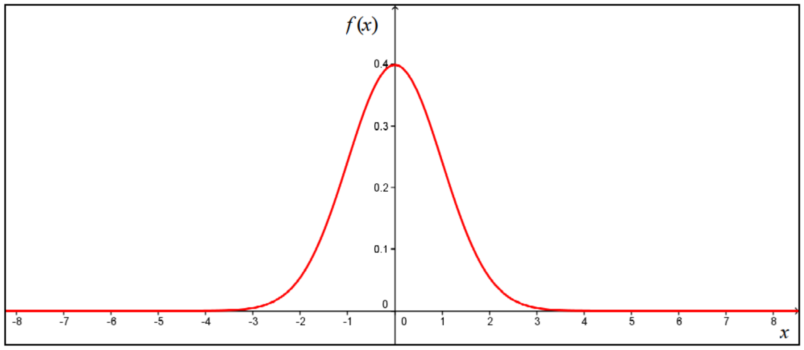


FIG. 1-2 – Représentation graphique de f

En $x = 0$, la fonction f vaut

$$f(0) = \frac{1}{\sqrt{2\pi}}.$$

— Les points d'inflexion de la fonction g se trouvent en $x = -1$ et $x = 1$.

— En général, les valeurs de $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ sont données à l'aide d'une table pour $x \geq 0$

Définition 2 *La fonction de répartition de la loi $\mathcal{N}(0,1)$ est souvent notée :*

$$\Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dt$$

Théorème 3 *Si $X \sim \mathcal{N}(m, \sigma)$ alors $Z = \dfrac{X - m}{\sigma} \sim \mathcal{N}(0,1)$*

On utilise la lettre Z pour désigner une loi normale centrée réduite.

1.1.3 Calculs de probabilités

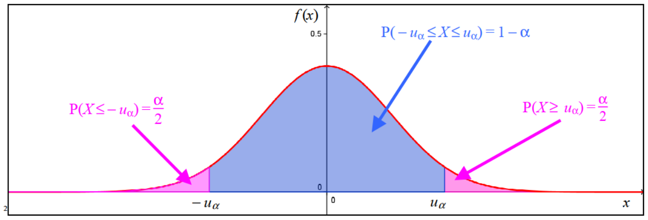
Théorème 4 *Si une variable aléatoire X suit une loi normale centrée réduite alors pour tous réels a et b tels que $a \leq b$, on a :*

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = \Phi(b) - \Phi(a)$$
$$P(X \geq a) = 1 - P(X \leq a) = 1 - \Phi(a)$$
$$P(X \leq -|a|) = 1 - \Phi(|a|)$$

On a :

$$\forall x \in \mathbb{R}_+ \ \Phi(x) = 1 - \Phi(-x)$$
$$P(|X| \leq x) = 2\Phi(x) - 1$$

- ▶ $P(X \in IR) = 1$, l'aire de la partie comprise entre l'axe des abscisses et la courbe représentative de f est égale à 1 unité d'aire.
- ▶ La symétrie de la courbe impose : $\begin{cases} P(X \leq 0) = P(X \geq 0) = 0,5 \\ P(X \leq -a) = P(X \geq a). \end{cases}$
- ▶ $(X > a)$ et $(X \leq a)$ étant des événements contraires : $P(X > a) = 1 - P(X \leq a)$



1.1.4 Probabilité d'intervalle centré en 0

Théorème 5 *X est une variable aléatoire qui suit une loi normale centrée réduite. Soit α un réel de l'intervalle $]0,1[$. Il existe un unique réel strictement positif u_α tel que :*

$$P(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha.$$

Exemple 6 *X est une variable aléatoire qui suit une loi normale centrée réduite.*

Déterminer l'intervalle I centré en 0 tel que $P(X \in I) = 0,8$.

On donnera les bornes de l'intervalle avec une précision de 10^{-2}.

Solution 7 *On a donc : $1 - \alpha = 0,8 \Leftrightarrow \alpha = 0,2$*

On doit donc avoir :

$$\Phi(u_\alpha) = 1 - \frac{\alpha}{2} = 0,9 \Leftrightarrow u_\alpha = \Phi^{-1}(0,9)$$

on trouve :

$$u_\alpha \simeq 1,28 \text{ donc } I = [-1,28, 1,28]$$

Exemple 8 *On suppose qu'une certaine variable $X \sim \mathcal{N}(0,1)$. Pour quelle proportion d'individus est-ce que $X \leq 1,56$?*

On cherche $P(X \leq 1,56)$ (rappel : on écrit aussi $\Phi(1,56)$. On cherche 1,56 dans la table

.	...	0.6...
1.5	...	***0.9406...***
.	...	...

Donc $P(X \leq 1,56) =$***0,9406.***

Exemple 9 *On suppose qu'une certaine variable* $X \sim \mathcal{N}(0,1)$*. Pour quelle proportion d'individus est-ce que* $X \geq 1,49$*?*

On cherche $P(X \geq 1,49)$*. On écrit d'abord*

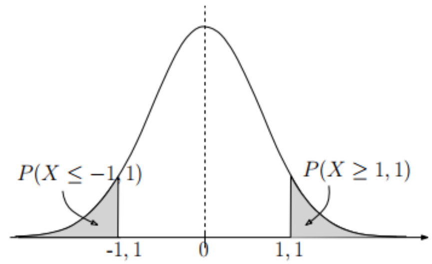
$$P\left(X \geq 1,49\right) = 1 - P\left(X \leq 1,49\right) = 1 - \Phi\left(1,49\right).$$

On cherche $\Phi\left(1,49\right)$ *dans la table.*

	...	...	0.09
	...	...	...
1.49	...	...	$\mathbf{0,9319}$
	...	...	....

On a $\Phi\left(1,49\right) = P\left(X \leq 1,49\right) = 0,9319$*, donc :*

$$P\left(X \geq 1,49\right) = 1 - 0,9319 = 0,0681$$

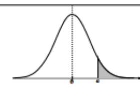
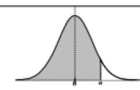
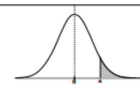
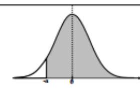
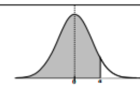


Exemple 10 *On suppose qu'une certaine variable* $X \sim \mathcal{N}(0,1)$*. Pour quelle proportion d'individus est-ce que* $X \leq -1,1$ *? On cherche* $P(X \leq -1,1)$*, c'est-à-dire*

$$\Phi\left(-1,1\right) = P(X \geq 1,1) = 1 - P(X \leq 1,1) = 1 - 0,8643 = 0,1357.$$

par exemple $\Phi\left(-1.1\right) = 1 - \Phi\left(1.1\right).$

Pour n'importe quel $a > 0$,

I	$\mathbb{P}(X \leq a)$				$\Rightarrow$ table
II	$\mathbb{P}(X \geq a)$		$= 1 -$		$\Rightarrow$ cas I
III	$\mathbb{P}(X \leq -a)$		$=$		$\Rightarrow$ cas II
IV	$\mathbb{P}(X \geq -a)$		$=$		$\Rightarrow$ cas I

Exemple 11 *On suppose qu'une certaine variable* $X \sim \mathcal{N}(11, 2)$. *Pour quelle proportion d'individus est-ce que* $X \leq 14$?

On cherche $P(X \leq 14)$.

- ▶ $P(X \leq 14) = P(\frac{X-11}{2} \leq \frac{14-11}{2}) = P(Z \leq 1.5)$.
- ▶ On cherche $1,5$ dans la table.

On trouve finalement $P\,(X \leq 14) = 0,9332$.

Chapitre 2

Échantillonnage

2.1 Introduction

La théorie de l'échantillonnage étudie les liens entre une population et des échantillons de cette population.

Dans cette partie, nous allons étudier comment se comporte un échantillon (éléments pris au hasard) dans une population dont on connaît les caractéristiques statistiques (lois,...) d'une variable considérée X. Dans ce cas, prendre un échantillon aléatoire de taille n consiste à considérer n réalisations de X ou encore considérer n variables aléatoires $X_1, ..., X_n$ indépendantes, de même loi que X.

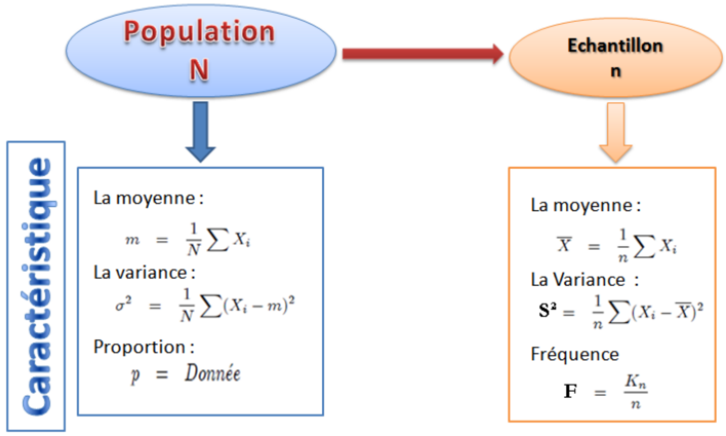
2.1.1 Avantages de l'échantillonnage

L'analyse d'un échantillon, par rapport à celle de la population, cout moindre, gain de temps et c'est la seule méthode qui donne des résultats dans le cas d'un test destructif.

2.2 Population - Echantillontion

2.2.1 Population

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



2.2.2 Echantillon

Soit $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$ une population de taille N. Soit X le caractère que l'on voudrait étudier sur cette population. Avec l'échantillon aléatoire simple : soit X_k le résultat aléatoire du $k^{ième}$ tirage, c'est une variable aléatoire qui suit la même loi que X. On note x_k le résultat du $k^{ième}$ tirage et on note $(X_1, X_2, ..., X_n)$ le résultat aléatoire de ces n tirages.

Donc les n variables aléatoires indépendantes $X_1, X_2, ..., X_n$ constituent un échantillon aléatoire simple de la variable X si et seulement si

$$\begin{aligned} E(X_1) &= E(X_2) = ... = E(X_n) = E(X) = m. \\ \sigma(X_1) &= \sigma(X_2) = ... = \sigma(X_n) = \sigma(X) = \sigma^2_{pop}. \end{aligned}$$

Définition 12 *$(X_1, X_2, ..., X_n)$ sont n v.a. indépendantes et de même loi (celle de X),(Par exemple la loi de Gauss) il est appelé n-échantillon ou échantillon de taille n de X.*

Définition 13 *La réalisation unique $(x_1, ..., x_n)$ de l'échantillon $(X_1, X_2, ..., X_n)$ est l'ensemble des valeurs observées.*

Exemple 14 *On fait l'hypothsèe que la taille (en cm) des 4000 étudiants masculins d'une école de génie est une variable aléatoire X distribuée normalement, c'est-à-dire que $X \sim \mathcal{N}(\mu, \sigma)$. Un échantillon aléatoire de taille 50 de cette population est une suite de 50 variables aléatoires $X_i \sim \mathcal{N}(\mu, \sigma)$, $i = 1, 2, ..., 50$.*

Définition 15 (Définition d'une statistique) *Une statistique* Y *sur un échantillon* $(X_1, X_2, ..., X_n)$ *est une v.a., fonction mesurable des* X_k; $Y = f(X_1, X_2, ..., X_n)$. *Après réalisation, la v.a.* Y *(statistique) prend la valeur* $f(x_1, ..., x_n)$.

Exemples de statistiques :

1. La moyenne échantionnale $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. La variance échantionnale $S^2_{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

Dans notre cours, nous allons travailler sur l'échantillonnage aléatoire simple, avec deux cas :

$a)$ **Non Exaustif : Avec Remise** (car la taille de la population est grande).

$b)$ **Exaustif** : **Sans Remise** : (car la taille de population est finie)

2.3 Les distributions d'échantillonnage.

Soit dans une population mère de taille N, une variable aléatoire X pour laquelle l'espérance mathématique m, la proportion P et l'écart-type σ_{pop} sont connues. De cette population sont issus k échantillons $E_1, E_2, ..., E_k$ de taille n qui auront des moyennes et des écarts-types différents. La notion de distribution d'échantillonnage peut être résumé et schématisée :

Population mère : Ω	**Echantillon** 1	**Echantillon** 2	......	**Echantillon** k
Taille : N	Taille : n	Taille : n	......	Taille : n
Moyenne : $m = \mu$ (connue)	Moyenne : $\bar{X}_1$	Moyenne : $\bar{X}_2$	......	Moyenne : $\bar{X}_k$
Proportion : P (connue)	proportion : f_1	proportion : f_2	......	proportion : f_k
Ecart-type : σ^2_{pop} (connue)	Ecart-type : σ_1	Ecart-type : σ_2	......	Ecart-type : σ_k

2.4 Distribution échantillonnale de la moyenne $\bar{\mathbf{X}}$

Dans une population mère de taille N, on peut tirer plusieurs échantillons de taille n :

Pour chaque échantillon, on peut calculer une moyenne :

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

et une variance

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

La valeur de l'espérance mathématique $E\left(\bar{X}\right)$ et de la variance $S^2_{\bar{X}}$ varient d'un échantillon à l'autre.

C'est cette variation qui donne naissance à la distribution des variables aléatoires :

– **Echantillonnage de la moyenne ou moyenne d'échantillon** $\bar{X}$, caractérisée par :

$E(\bar{X})$: l'espérance mathématique des moyennes calculées sur tous les échantillons de taille n.

$S_{\bar{X}}$: l'écart type de la distribution d'échantillonnage, qui représente la dispersion de l'ensemble des moyennes d'échantillons de taille n autour de $E(\bar{X})$

– **Variance d'échantillon** $S^2_{\mathbf{c}}$ définie par

$$S^2_{\mathbf{c}} = \frac{n}{n-1}S^2_{\bar{X}}.$$

On a bien entendu $E(S^2_{\mathbf{c}}) = \sigma^2_{pop}$.

Nous verrons plus tard que cela signifie que $S^2_{\mathbf{c}}$ (variance corrigée de l'échantillon) est un estimateur sans biais de σ^2

1) Cas : moyenne m et écart-type σ_{pop} de la population connus

A) Si la population est infinie ou si l'´echantillonnage est **non exhaustif** (tirage avec remise)

– L'espérance mathématique de la variable aléatiore $\bar{X}$ est égale à celle de la population mère :

$$E\left(\bar{X}\right) = \mu_{\bar{X}} = m$$

– Pour la variance $V(\bar{X})$

$$\begin{aligned} V(\bar{X}) &= \sigma^2_{\bar{X}} = \frac{\sigma^2_{pop}}{n}. \\ \sigma_{\bar{X}} &= \frac{\sigma_{pop}}{\sqrt{n}} \end{aligned}$$

Alors dans ce cas $\bar{X} \sim \mathcal{N}\left(m, \frac{\sigma_{pop}}{\sqrt{n}}\right)$, ou bien $\bar{X}$ suit approximativement $\mathcal{N}\left(m, \frac{\sigma_{pop}}{\sqrt{n}}\right)$

(en pratique $n > 30$)

Exemple 16 *Une machine effectue l'ensachage d'un produit.*

On sait que les sacs ont un poids moyen de $250g$ *avec un écart-type de* $25g$.

Quelles sont les caractéristiques de la moyenne des poids d'un échantillon de 100 *sacs ?*

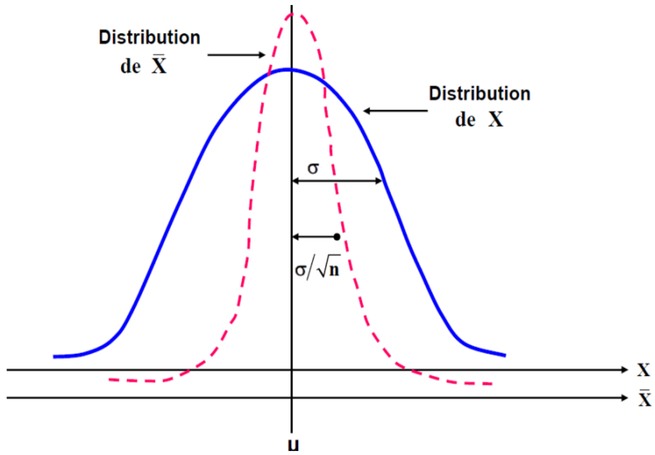
Solution 17 $(P) : m = \mu = 250,\ \sigma = 2,5;\ (E) : n = 100 > 30$

$\bar{X}$ *suit la loi normale de paramètres* $m = 250$ *et* $\frac{\sigma_{pop}}{\sqrt{n}} = \frac{25}{10} = 2,5.$

B) Si l'échantillonnage est **exhaustif** (tirage **sans remise**) dans une population finie : (Taille N sera donnée)

$$\begin{aligned} V(\bar{X}) &= \sigma_{\bar{X}}^2 = \frac{\sigma_{pop}^2}{n}.\frac{N-n}{N-1} \\ \sigma_{\bar{X}} &= \frac{\sigma_{pop}}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} \end{aligned}$$

Donc dans ce cas $\bar{X} \sim \mathcal{N}\left(m, \frac{\sigma_{pop}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$,



Remaque 18 $\sigma_{\bar{X}} = \dfrac{\sigma_{pop}}{\sqrt{n}}$ *est aussi appelé l'erreur-type de la moyenne.*

Remaque 19 *Si les échantillons sont issus d'une population mère finie et sont constituée sans remise. L'espérance mathématique de $\bar{X}$ est toujours égale à m, mais l'écart-type est corrigé par le facteur d'exhaustivité (facteur de correction)*

$\sigma_{\bar{X}} = \dfrac{\sigma_{pop}}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} \approx \dfrac{\sigma_{pop}}{\sqrt{n}}.\sqrt{1-\frac{n}{N}}$ *tel que* $\dfrac{n}{N}$ *représente le taux de sondage.*

Exemple 20 *Dans une usine textile, on utilise une machine automatique pour couper des morceaux de tissu. Lorsque la machine est correctement ajustée, la longueur des morceaux de tissu est en moyenne de 90cm avec un écart type de 0.60 cm.*

Pour contrôler la longueur des morceaux de tissu, on tire dans la production d'une journée un échantillon aléatoire de 200 morceaux.

a) Si l'on suppose que la longueur X des morceaux de tissu suit une loi normale, calculer la probabilité que la moyenne de l'échantillon soit au plus égale à 89.90 cm, ceci dans 2 cas :

—Production de la journée : 10000 morceaux

—Production de la journée : 2000 morceaux.

b) Déterminer la même probabilité sans faire l'hypothèse que X soit distribuée normalement.

Solution 21 *a) Production journalière* $= N = 10000$*, Taille de l'échantillon* $= n = 200$, $\dfrac{n}{N} = 0.02$

Même si l'échantillonnage est exhaustif, ce n'est pas la peine de tenir compte du coefficient d'exhaustivité.

Dans ce cas $E\left(\bar{X}\right) = 90$ *cm et* $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.6}{\sqrt{200}} = 0.042.$

Comme $X \sim \mathcal{N}(90, 0.6) \longrightarrow \bar{X} \sim \mathcal{N}((90, 0.042))$

$P(\bar{X} \leq 89.9) = P\left(T \leq \dfrac{89.9 - 90}{0.042}\right) = P(T \leq -2.38) = 1 - \Phi(2.38) = 0.0087 \longrightarrow 0.87\%$

Production journaliére $= N = 2000 \longrightarrow \dfrac{n}{N} = 0.1 \longrightarrow$ *on doit tenir compte du coefficient d'exhaustivité*

$s_{\bar{X}} = \dfrac{\sigma^2_{pop}}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}} = \dfrac{0.6}{\sqrt{200}}\sqrt{\frac{2000-200}{2000-1}} = 0.04$

$\bar{X} \sim \mathcal{N}(90, 0.04)$

$P(\bar{X} \leq 89.9) = P\left(T \leq \dfrac{89.9 - 90}{0.04}\right) = P(T \leq -2.5) = 1 - \Phi(2.5) = 0.0062 \longrightarrow 0.62\%$

Même si l'on ne fait plus l'hypothèse que X *soit une variable normale, comme* $n = 200 > 30$, *le théorème central limite permet de dire quel* $\bar{X} \backsim \mathcal{N}(90, 0.042)$ *pour* $N = 10000$. *On trouvera donc la même probabilité*

$P(\bar{X} \leq 89.9) = 0.0087 \longrightarrow 0.87\%.$

Définition 22 *On appelle variance empirique de l'échantillon* $(X_1, X_2, ..., X_n)$ *de* X *, la statistique*

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

Sa réalisation est $s^2_{\bar{X}} = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$ (qui est la variance de l'échantillon), aussi appelée variance observée.

$$E\left(S^2_{\bar{X}}\right) = \frac{n-1}{n}\sigma^2_{pop}.$$

Calculons $E\left(S^2_{\bar{X}}\right)$

$$\begin{aligned}
E\left(S^2_{\bar{X}}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}^2\right) = \frac{1}{n}\sum E\left(X_i^2\right) - E\left(\bar{X}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[V(X_i) + (E(X_i))^2\right] - \left[V(\bar{X}) + \left(E(\bar{X})\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[V(X) + (E(X))^2\right] - \frac{1}{n}\sigma^2_{pop} - m^2 \\
&= V(X) + (E(X))^2 - \frac{1}{n}\sigma^2_{pop} - m^2 = \sigma^2 + m^2 - \frac{1}{n}\sigma^2_{pop} - m^2 \\
&= \left(1 - \frac{1}{n}\right)\sigma^2_{pop} = \frac{n-1}{n}\sigma^2_{pop}
\end{aligned}$$

Conclusion 23 *La moyenne des variances d'échantillon n'est pas la variance de la population, mais la variance de la population multipliée par* $\dfrac{n-1}{n}$*. Bien sûr, si* n *est très grand, ces deux nombres seront très proches l'un de l'autre.*

2) Cas écart-type σ de la population inconnu

A) Cas des grands échantillon ($n \geq 30$)

-Si la variance est **inconnue**, un grand échantillon permet de déduire une valeur fiable pour σ^2_{pop} en calculant la variance de l'échantillon S^2 et en posant

$$\sigma^2_{pop} = \frac{n}{n-1} S^2_{\bar{X}} = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

$$\text{Donc si } \begin{cases} n \geq 30 \\ \sigma_{pop} \text{ inconnue} \end{cases}, \quad \bar{X} \text{ suit } \mathcal{N}\left(m, \sqrt{\frac{n}{n-1}} S\right)$$

B) Cas des petits échantillons **:** ($n < 30$)

On considère exclusivement le cas où X suit une loi normale dans la population.

$$\text{Si : } \begin{cases} n < 30 \\ \sigma_{pop} \text{ inconnue} \end{cases}, \quad T = \frac{\bar{X} - m}{\frac{S}{\sqrt{n-1}}} \text{ suit une loi de Student à } n-1 \text{ degrés de liberté, notée } T_{n-1}.$$

Exercise 24 *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* $\sigma^2 = 100$: *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 25 *On considére la variable aléatoire* $\bar{X}$ *moyenne d'échantillon pour les échantillons de taille* $n = 25$: *On cherche à déterminer* $P(146 < \bar{X} < 154)$

Nous sommes en présence d'un petit échantillon ($n < 30$) *et heureusement dans le cas où la variable* X *suit une loi normale. De plus,* σ_{pop} *est connu. Donc* $\bar{X}$ *suit* $\mathcal{N}\left(m, \dfrac{\sigma_{pop}}{\sqrt{n}}\right) = \mathcal{N}\left(150, \frac{10}{5}\right)$*. On en déduit que* $Z = \dfrac{\bar{X} - 150}{2}$ *suit* $\mathcal{N}(0,1)$.

La table donne

$$\begin{aligned} P(146 &< \bar{X} < 154) = P\left(\frac{146-150}{2} < Z < \frac{154-150}{2}\right) = P\left(-2 < Z < 2\right). \\ &= 2P\left(0 < Z < 2\right) = 2 \times \left(P\left(Z < 2\right) - P\left(Z < 0\right)\right) = 2 \times \left(0,9772 - 0,5\right). \\ &= 2 \times 0,4772 = 0,9544. \end{aligned}$$

2.4.1 Distribution de la variance d'échantillon $S_{\bar{X}}^2$

Supposons que X suit une loi normale.

On considère la variable

$$Y = \frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} = \frac{n}{\sigma_{pop}^2}\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma_{pop}^2}\right)^2.$$

D'aprés la décomposition de $S_{\bar{X}}^2$:

$$\sum_{i=1}^{n}(X_i - m)^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - m\right)^2.$$

En divisant par σ^2 :

$$\begin{aligned} \sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma_{pop}^2}\right)^2 &= \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma_{pop}^2}\right)^2 + \frac{n}{\sigma_{pop}^2}\left(\bar{X} - m\right)^2. \\ &= \frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} + \left(\frac{\bar{X} - m}{\frac{\sigma_{pop}}{\sqrt{n}}}\right)^2. \end{aligned}$$

On a : $\dfrac{X_i - m}{\sigma_{pop}} \sim \mathcal{N}(0,1) \Rightarrow \left(\dfrac{X_i - m}{\sigma_{pop}}\right)^2 \sim \chi_1^2$ (Khi-deux 1 degret de libérté).

D'ou $\displaystyle\sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma_{pop}}\right)^2 \sim \chi_n^2$ (Comme somme de n carrés de variables aléatoires indépendantes normales centrées réduite)

$$\frac{\bar{X} - m}{\frac{\sigma_{pop}}{\sqrt{n}}} \sim \mathcal{N}(0,1) \Rightarrow \left(\frac{\bar{X} - m}{\frac{\sigma_{pop}}{\sqrt{n}}}\right)^2 \sim \chi_1^2.$$

D'où, on en déduit :

$$\begin{aligned} Y &= \frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} \sim \chi_{n-1}^2 \\ \text{c'est à dire } S_{\bar{X}}^2 &\sim \frac{\sigma_{pop}^2}{n}\chi_{n-1}^2 \quad \chi_{n-1}^2 \text{ (Khi-deux } n-1 \text{ deg ret de libérté)} \end{aligned}$$

$$E(S_{\bar{X}}^2) = \frac{\sigma_{pop}^2}{n}(n-1) \text{ et } Var\left(S_{\bar{X}}^2\right) = 2(n-1)\frac{\sigma_{pop}^4}{n^2}.$$

Exercise 26 *On prélève 25 pièces dans une production industrielle. Une étude préalable a montré que le diamètre de ces pièces suivait une loi gaussienne de moyenne 10mm et d'écart-type 2mm. Entre quelles valeurs a-t-on 85% de chances de trouver l'écart-type de ces pièces?*

Solution 27 *Pour commencer, il faut déterminer α et β t.q.*

$$0.85 = P(\alpha < \frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} < \beta) = P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} < \beta) - P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} < \alpha)$$

$$\begin{aligned} &= 1 - P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} > \beta) - [1 - P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} > \alpha)] \\ &= P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} > \alpha) - P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} > \beta). \end{aligned}$$

On sait que $\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} \sim \mathcal{X}_{25-1}^2 = \mathcal{X}_{24}^2$ et alors on cherche dans la table du $\mathcal{X}^2$ à 24 degrés de liberté les valeurs α et β comme suit :

$$\begin{cases} P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} > \alpha) = 0.90 \\ P(\frac{nS_{\bar{X}}^2}{\sigma_{pop}^2} > \beta) = 0.05 \end{cases} \quad \textit{d'après la table de } \mathcal{X}^2$$

on trouve $\alpha = 15,659$ et $\beta = 36,415$

et alors :

$$P\left(15.659 < \frac{25S_{\bar{X}}^2}{2^2} < 36.415\right) = 0.85.$$

$$P(2.5054 < S^2 < 5.8264) = 0.85.$$

$$P(1.58 < S < 2.41) = 0.85.$$

2.4.2 Distribution d'échantillonnage d'une proportion F

Soit une population comportant deux modalités A et B. Soit p la proportion d'individus de la population possédant la modalité A. $1-p$ est donc la proportion des individus de la population possédant la modalité B.

On extrait de la population un échantillon de taille n. Soit X la v.a qui représente le nombre d'individus dans l'échantillon ayant la modalité A.

Définition 28 *La v.a.* $F = \dfrac{X}{n}$ *s'appelle fréquence empirique. Sa réalisation* f *est la proportion d'individus dans l'échantillon ayant la modalité* A*. Où* X *est le nombre de fois où le caractère apparaît dans le* $n-$*échantillon.*

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

i) **Si la population est infinie ou si l'échantillonnage est non exhaustif (tirage avec remise)**, on montre que :

$$\begin{cases} E(F) = p \\ V(F) = S_F^2 = \dfrac{p\,(1-p)}{n}, \ S_F = \sqrt{\dfrac{p\,(1-p)}{n}} \end{cases}$$

– **Loi de probabilité pour** F

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

$$\text{Et on écrit } F \sim \mathcal{N}\left(p, \sqrt{\frac{p\,(1-p)}{n}}\right)$$

ii) Si l'échantillonnage est exhaustif (tirage sans remise) dans une population finie $(n > 0,05N)$

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

Exercise 29 *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 d´eclarent influencés par la marque ?*

Solution 30 *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 d´eclarent 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*

$$\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = \mathcal{N}(0.25, 0.0433).$$

On utilise la variable $T = \dfrac{F - 0.25}{0.0433}$ *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 31 *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.*

Exercise 32 *Le directeur financier d'une société sait par expérience que* 12% *des factures émises ne sont pas réglées dans les* 10 *jours ouvrables suivant l'échéance. Il fait prélever un échantillon aléatoire de* 500 *factures.*

Exemple 33 *Quelle est la probabilité qu'au moins* 70 *factures ne sont pas réglées dans le délais, sachant quel'ensemble des factures pouvant être étudiées est de plusieurs dizaines de milliers*

Solution 34 *Soit $F =$ "proportion d'échantillon dans un échantillon de taille* 500*".*

$P\left(F \geq \frac{500}{70}\right)$ =?- *Distribution d'échantillonnage d'une proportion F, échantillonnage exhaustif (tirage sans remise) dans une population finie,*

mais $n < 0,05N$, donc il ne faut pas tenir compte du facteur d'exhaustivité.

Ici $p = 0,12$, $q = 1 - p = 1 - 0,12 = 0,88$.

Comme $n = 500 > 30$, $np = 500 \times 0;12 = 60 > 15$, $nq = 500 \times 0,88 = 440 > 15 \Rightarrow$ approximation de la loi binomiale par la loi normale

$$\begin{aligned}
1)\ F &\sim \mathcal{N}\left(p, \sqrt{\frac{p\,(1-p)}{n}}\right) = \mathcal{N}\left(0,12, \sqrt{\frac{0,12 \times 0,88}{500}}\right) = \mathcal{N}\,(0,12, 0,015) \\
2)\ P\left(F \geq \frac{500}{70}\right) &= P\left(Z > \frac{0,139 - 0,12}{0,015}\right) \\
&= 1 - P\left(Z < \frac{0,019}{0,015}\right) = 1 - P(Z < 1,27) = 1 - \Phi(1,27) = 1 - 0,8997 \approx 0,1 \\
&\approx 10\%\ \textit{de chances pour que plus de } 70 \textit{ factures dans un } 500 \textit{ échantillon soient} \\
&\quad \textit{non réglées dans le délais.}
\end{aligned}$$