3.2 Estimation par intervalle de confiance

L'estimation ponctuelle nous a permis de construire des estimateurs du paramètre inconnu θ et par suite connaissant un échantillion d'observations on peut calculer une estimation de θ et donc une valeur approchée.

L'estimation par intervalle de confiance nous permettera de déterminer des intervalles contenant la vraie valeur de θ tout en évaluant la précision et la crédibilité de l'estimation. L'idée est de construite à partir d'un échantillion $X_1, ..., X_n$ de la variable $X \rightsquigarrow \mathcal{L}(\theta)$ deux statistiques $A(X_1, ..., X_n)$ et $B(X_1, ..., X_n)$ vérifiant :

$$\mathbb{P}(A(X_1, ..., X_n) \leq \theta \leq B(X_1, ..., X_n)) = 1 - \alpha$$

où α est un réel donné compris entre 0 et 1.

Définition 3.6
On dit que l'intervalle $[A(X_1, ..., X_n), B(X_1, ..., X_n)]$ est un ***intervalle de confiance de niveau*** $1-\alpha$ *ou au* ***seuil de risque*** α *si les deux statistiques $A(X_1, ..., X_n)$ et $B(X_1, ..., X_n)$ vérifient :*
i) $A(X_1, ..., X_n) \leq B(X_1, ..., X_n)$
ii) $\mathbb{P}(\theta \in [A(X_1, ..., X_n), B(X_1, ..., X_n)]) = 1 - \alpha$
ou encore : $\mathbb{P}(\theta \notin [A(X_1, ..., X_n), B(X_1, ..., X_n)]) = \alpha$.
Le nombre α est appelé ***seuil de risque*** *et la valeur $1-\alpha$ est appelé* ***niveau de confiance.***

Disposant d'un échantillion d'observations $(x_1, ..., x_n)$ réalisation de $(X_1, ..., X_n)$, notons :

$$a = A(x_1, ..., x_n) \quad \text{et} \quad b = B(x_1, ..., x_n)$$

l'intervale $[a, b]$ est appelé **fourchette d'estimation** au niveau de confiance $1-\alpha$ ou au niveau de risque α.

Remarque(s) 3.7
i) *Ainsi nous apprécions la crédibilité de notre estimation grâce au nombre $1-\alpha$ et sa précision par la mesure de l'erreur absolue laquelle est majorée par $b-a$.*
ii) *Pour n et α donnés, il est possible de construire une infinité d'intervalles de confiance, en effet il suffit de constater que si on a :*

$$\mathbb{P}(\theta \notin [A(X_1, ..., X_n), B(X_1, ..., X_n)]) = \alpha,$$

on peut choisir α_1 et α_2 telque $\alpha_1 + \alpha_2 = \alpha$ de telle sorte que :

$$\mathbb{P}(\theta < A(X_1, ..., X_n)) = \alpha_1 \quad \text{et} \quad \mathbb{P}(\theta > B(X_1, ..., X_n)) = \alpha_2.$$

Dans ce qui suit nous allons traiter quelques exemples, la méthodologie que nous allons suivre est de partir d'un estimateur de θ et de chercher une fonction algébrique de cet estimateur dont la loi est connue et ne dépend pas de θ ou du moins asymptotiquement connue. Construire un intervalle pour cette statistique pour ensuite le transformer en un intervalle de confiance pour le paramètre θ. Pour se faire rappelons la définition de quantile (fractile).

Définition 3.8 *Quantile*
Soient X une variable aléatoire réelle et F sa fonction de répartition. On appelle quantile (fractile) d'ordre $\alpha \in]0, 1[$ le nombre :

$$q_\alpha = \inf\{x \in \mathbb{R}, F(x) \geq \alpha\}.$$

Lorsque la fonction F est bijective, q_α n'est autre que $F^{-1}(\alpha)$.

3.2.1 Intervalle de confiance pour les paramètres d'une gaussienne

Soit X une variable aléatoire de loi normale de paramètre m et σ^2 et supposons qu'elle soit associée à un caractère quantitatif étudié dans une population dans laquelle on a tiré un n-échantillon $X_1, ..., X_n$. Nous allons construire, dans un premier temps, un intervalle de confiance pour m dans le cas : σ^2 connue et dans le cas σ^2 inconnue. Dans un second temps, on fera de même pour σ^2 dans le cas m connue et dans le cas m inconnue.

Intervalle de confiance pour m dans le cas σ^2 connue

Nous avons vu dans ce qui précède que l'estimateur $\overline{X}_n$ de m est sans biais, convergent et efficace. Par ailleurs, la variable $\overline{X}_n$ suit la loi normale $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$, il vient donc que :

$$\frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\overline{X}_n - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1).$$

Comme la fonction de répartition Φ de la loi $\mathcal{N}(0, 1)$ est symétrique pour tout réel a on a :

$$\mathbb{P}\left(-a \leq \sqrt{n}\frac{\overline{X}_n - m}{\sigma} \leq a\right) = 2\Phi(a) - 1,$$

ce qui permet d'écrire :

$$\mathbb{P}\left(-a\frac{\sigma}{\sqrt{n}} \leq \overline{X}_n - m \leq a\frac{\sigma}{\sqrt{n}}\right) = \mathbb{P}\left(\overline{X}_n - a\frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(a) - 1.$$

Etant donné un $\alpha \in [0, 1]$, on peut donc trouver a telque :

$$\mathbb{P}\left(\overline{X}_n - a\frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(a) - 1 = 1 - \alpha,$$

3.2 Estimation par intervalle de confiance

suite en h TD.

L'estimation ponctuelle nous a permis de construire des estimateurs du paramètre inconnu θ et par suite connaissant un échantillion d'observations on peut calculer une estimation de θ et donc une valeur approchée.

L'estimation par intervalle de confiance nous permettera de déterminer des intervalles contenant la vraie valeur de θ tout en évaluant la précision et la crédibilité de l'estimation. L'idée est de construite à partir d'un échantillion $X_1, ..., X_n$ de la variable $X \rightsquigarrow \mathcal{L}(\theta)$ deux statistiques $A(X_1, ..., X_n)$ et $B(X_1, ..., X_n)$ vérifiant :

$$\mathbb{P}(A(X_1, ..., X_n) \le \theta \le B(X_1, ..., X_n)) = \underbrace{1-\alpha}_{\text{seuil de confiance}}$$

où α est un réel donné compris entre 0 et 1.

Définition 3.6
*On dit que l'intervalle $[A(X_1, ..., X_n), B(X_1, ..., X_n)]$ est un **intervalle de confiance de niveau** $1-\alpha$ ou au **seuil de risque** α si les deux statistiques $A(X_1, ..., X_n)$ et $B(X_1, ..., X_n)$ vérifient :*
i) $A(X_1, ..., X_n) \le B(X_1, ..., X_n)$
ii) $\mathbb{P}(\theta \in [A(X_1, ..., X_n), B(X_1, ..., X_n)]) = 1-\alpha$
ou encore : $\mathbb{P}(\theta \notin [A(X_1, ..., X_n), B(X_1, ..., X_n)]) = \alpha$.
*Le nombre α est appelé **seuil de risque** et la valeur 1-α est appelé **niveau de confiance**.*

Disposant d'un échantillion d'obsevations $(x_1, ..., x_n)$ réalisation de $(X_1, ..., X_n)$, notons :

$$a = A(x_1, ..., x_n) \quad \text{et} \quad b = B(x_1, ..., x_n)$$

l'intervale $[a, b]$ est appelé **fourchette d'estimation** au niveau de confiance $1-\alpha$ ou au niveau de risque α.

Remarque(s) 3.7
i) *Ainsi nous apprécions la crédibilité de notre estimation grâce au nombre 1-α et sa précision par la mesure de l'erreur absolue laquelle est majorée par b-a.*
ii) *Pour n et α donnés, il est possible de construire une infinité d'intervalles de confiance, en effet il suffit de constater que si on a :*

$$\mathbb{P}(\theta \notin [A(X_1, ..., X_n), B(X_1, ..., X_n)]) = \alpha,$$

on peut choisir α_1 et α_2 telsque $\alpha_1 + \alpha_2 = \alpha$ de telle sorte que :

$$\mathbb{P}(\theta < A(X_1, ..., X_n)) = \alpha_1 \;\; et \;\; \mathbb{P}(\theta > B(X_1, ..., X_n)) = \alpha_2.$$

Dans ce qui suit nous allons traiter quelques exemples, la méthodologie que nous allons suivre est de partir d'un estimateur de θ et de chercher une fonction algébrique de cet estimateur dont la loi est connue et ne dépend pas de θ ou du moins asymptotiquement connue. Construire un intervalle pour cette statistique pour ensuite le transformer en un intervalle de confiance pour le paramètre θ. Pour se faire rappelons la définition de quantile (fractile).

Définition 3.8 Quantile
Soient X une variable aléatoire réelle et F sa fonction de répartition. On appelle quantile (fractile) d'ordre $\alpha \in]0,1[$ le nombre :

$$q_\alpha = \inf\{x \in \mathbb{R}, F(x) \ge \alpha\}.$$

Lorsque la fonction F est bijective, q_α n'est autre que $F^{-1}(\alpha)$.

3.2.1 Intervalle de confiance pour les paramètres d'une gaussienne

Théorème 1.28 chapitre 1

$\overline{X}_n \sim N(m, \sigma^2)$

$\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

Soit X une variable aléatoire de loi normale de paramètre m et σ^2 et supposons qu'elle soit associée à un caractère quantitatif étudié dans une population dans laquelle on a tiré un n-échantillon $X_1, ..., X_n$. Nous allons construire, dans un premier temps, un intervalle de confiance pour m dans le cas : σ^2 connue et dans le cas σ^2 inconnue. Dans un second temps, on fera de même pour σ^2 dans le cas m connue et dans le cas m inconnue.

Intervalle de confiance pour m dans le cas σ^2 connue

Nous avons vu dans ce qui précède que l'estimateur $\overline{X}_n$ de m est sans biais, convergent et efficace. Par ailleurs, la variable $\overline{X}_n$ suit la loi normale $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$, il vient donc que :

$$\frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\overline{X}_n - m}{\sigma} \rightsquigarrow \mathcal{N}(0,1).$$

Comme la fonction de répartition Φ de la loi $\mathcal{N}(0,1)$ est symétrique pour tout réel a on a :

$$\mathbb{P}\left(-a \le \sqrt{n}\frac{\overline{X}_n - m}{\sigma} \le a\right) = 2\Phi(a) - 1,$$

ce qui permet d'écrire :

$$\mathbb{P}\left(-a\frac{\sigma}{\sqrt{n}} \le \overline{X}_n - m \le a\frac{\sigma}{\sqrt{n}}\right) = \mathbb{P}\left(\overline{X}_n - a\frac{\sigma}{\sqrt{n}} \le m \le \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(a) - 1.$$

Etant donné un $\alpha \in [0,1]$, on peut donc trouver a telque :

$$\mathbb{P}\left(\overline{X}_n - a\frac{\sigma}{\sqrt{n}} \le m \le \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(a) - 1 = 1 - \alpha,$$

ou encore $\Phi(a) = 1 - \frac{\alpha}{2}$ c'est à dire que a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$. Ainsi nous avons donc construit un intervalle de confiance de niveau $1-\alpha$ pour m.

Propriétés 3.9 *Soit $X_1, ..., X_n$ un n-échantillon de la variable aléatoire X de loi normale d'esperance m inconnue et de variance σ^2 connue.*
L'intervalle $\left[\overline{X}_n - a\frac{\sigma}{\sqrt{n}};\ \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right]$, où a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, est un intervalle de confiance de niveau 1-α pour m. Par suite on a :

$$\mathbb{P}\left(m \in \left[\overline{X}_n - a\frac{\sigma}{\sqrt{n}};\ \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right]\right) = 1-\alpha.$$

Exemple(s) 3.10 *Des dosages de pH* dans les eaux d'un barrage ont donné les résultats $x_1, x_2, ..., x_{18}$ suivants :*

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
7,1	7,2	6,9	6,8	7,0	7,2	7,2	7,3	7,2

x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
6,9	7,0	7,1	7,4	7,4	7,9	6,4	7,2	7,1

.

On suppose que les valeurs observées x_i $(1 \le i \le 18)$ sont les valeurs prises par 18 variables aléatoires gaussiennes $X_1, X_2, ..., X_{18}$ indépendantes, de moyenne commune m inconnues et variance commune $\sigma^2 = 2$. A partir de l'échantillon observé, donner une estimation de m puis construire une fourchette d'estimation au niveau de confiance 0,95.

Une estimation de m est la moyenne $\overline{x}_{18}$ de cet échantillon

$$\widehat{m} = \overline{x}_{18} = \frac{1}{18}\sum_{i=1}^{18} x_i = 7.1278.$$

En appliquant le résultat précédent pour $1-\alpha = 0,95$ nous devons chercher le quantile de 0,975. Une lecture du tableau de la fonction de répartition de la loi normale $\mathcal{N}(0,1)$ nous donne que $a = 1,96$. L'intervalle de confiance de niveau $0,95$ pour m est donc :

$$\left[\overline{X}_{18} - 1,96\frac{\sqrt{2}}{\sqrt{18}};\ \overline{X}_{18} + 1,96\frac{\sqrt{2}}{\sqrt{18}}\right]$$

et par suite, une fourchette d'estimation au niveau de confiance $0,95$ est donnée par :

$$\left[\overline{x}_{18} - 1,96\frac{\sqrt{2}}{\sqrt{18}};\ \overline{x}_{18} + 1,96\frac{\sqrt{2}}{\sqrt{18}}\right] = \left[7.1278 - 1,96\frac{\sqrt{2}}{\sqrt{18}};\ 7.1278 + 1,96\frac{\sqrt{2}}{\sqrt{18}}\right]$$
$$= [6.474; 7.781].$$

Intervalle de confiance pour m dans le cas σ^2 inconnue

Dans ce cas nous allons remplacer dans la statistique précédente la variance σ^2 par son estimateur la variance empirique corrigée $S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$.

D'après le théorème 1.28, la variable :

$$\frac{\overline{X}_n - m}{S_n'/\sqrt{n}}$$

suit la loi de Student $t(n-1)$, sa fonction de répartition Ψ_{n-1} étant symétrique on a donc pour tout réel a :

$$\mathbb{P}\left(-a \le \sqrt{n}\frac{\overline{X}_n - m}{S_n'} \le a\right) = 2\Psi_{n-1}(a) - 1,$$

ce qui permet d'écrire :

$$\mathbb{P}\left(-a\frac{S_n'}{\sqrt{n}} \le \overline{X}_n - m \le a\frac{S_n'}{\sqrt{n}}\right) = \mathbb{P}\left(\overline{X}_n - a\frac{S_n'}{\sqrt{n}} \le m \le \overline{X}_n + a\frac{S_n'}{\sqrt{n}}\right) = 2\Psi_{n-1}(a) - 1.$$

Etant donné un $\alpha \in [0,1]$, on peut donc trouver a telque :

$$\mathbb{P}\left(\overline{X}_n - a\frac{S_n'}{\sqrt{n}} \le m \le \overline{X}_n + a\frac{S_n'}{\sqrt{n}}\right) = 2\Psi_{n-1}(a) - 1 = 1-\alpha,$$

ou encore $\Psi_{n-1}(a) = 1-\frac{\alpha}{2}$ c'est à dire que a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student $t(n-1)$. Ainsi nous avons donc construit un intervalle de confiance au niveau de confiance $1-\alpha$ pour m.

Propriétés 3.11 *Soit $X_1, ..., X_n$ un n-échantillon de la variable aléatoire X de loi normale d'espérance m inconnue et de variance σ^2 inconnue.*
L'intervalle $\left[\overline{X}_n - a\frac{S_n'}{\sqrt{n}};\ \overline{X}_n + a\frac{S_n'}{\sqrt{n}}\right]$, où a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student $t(n-1)$, est un intervalle de confiance au niveau de confiance 1-α pour m. Par suite, on a :

$$\mathbb{P}\left(m \in \left[\overline{X}_n - a\frac{S_n'}{\sqrt{n}};\ \overline{X}_n + a\frac{S_n'}{\sqrt{n}}\right]\right) = 1-\alpha.$$

Si nous disposons d'un échantillion d'obsevations $(x_1, ..., x_n)$ réalisation de $(X_1, ..., X_n)$, alors l'intervalle :

$$\left[\overline{x}_n - a\frac{s_n'}{\sqrt{n}};\ \overline{x}_n + a\frac{s_n'}{\sqrt{n}}\right]$$

est une fourchette d'estimation au niveau de confiance $1-\alpha$.

Remarque(s) 3.12 *Pour n assez grand ($n \geq 50$), on approchera la loi de Student $t(n)$ par la loi normale $\mathcal{N}(0,1)$ et par suite on lira les quantiles recherchés dans le tableau de la loi $\mathcal{N}(0,1)$.*

Exemple(s) 3.13 *Reprenons l'exemple pécédent et supposons la variance σ^2 inconnue, calculons dans ce cas la variance empirique corrigée des observations :*

$$s_n'^2 = \frac{1}{18-1}\sum_{i=1}^{18}(x_i - \overline{x}_{18})^2 = \frac{18}{17}\left(\frac{1}{18}\sum_{i=1}^{18} x_i^2 - \overline{x}_{18}^2\right) = \frac{18}{17}\left(50.893 - (7.1278)^2\right)$$
$$= \frac{18}{17} 8.724 \times 10^{-2} = 9.23 \times 10^{-2}.$$

Un intervalle de confiance au niveau de confiance 0,95 est donné par : $I = \left[\overline{X}_n \pm a\frac{S_n'}{\sqrt{n}}\right]$,
avec $a = \Psi_{17}^{-1}\left(1 - \frac{\alpha}{2}\right) = \Psi_{17}^{-1}(0.975) = 2.11$ *d'après le tableau de la loi de Student. En vue des observations, une fourchette d'estimation au niveau de confiance 0,95 est donnée par :*

$$I = \left[7.1278 \pm 2.11\sqrt{\frac{0.093}{18}}\right] = [6.97; 7.28].$$

Intervalle de confiance pour σ^2 dans le cas m connue

L'estimateur du maximum de vraisemblance de σ^2 dans ce cas est la statistique :

$$T_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2.$$

C'est un estimateur sans biais et convergent et de plus on a :

$$\frac{n}{\sigma^2}T_n = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - m)^2 \rightsquigarrow \mathcal{X}^2(n).$$

La fonction de répartition Λ_n de la loi Khi-deux n'étant pas symétrique, nous cherchons donc deux réels a et b telsque :

$$\mathbb{P}\left(a \leq \frac{n}{\sigma^2}T_n \leq b\right) = \Lambda_n(b) - \Lambda_n(a) = 1 - \alpha$$

pour un $\alpha \in [0,1]$ donné. En choisissant $\Lambda_n(a) = \frac{\alpha}{2}$ et $\Lambda_n(b) = 1 - \frac{\alpha}{2}$ autrement dit, a est le quantile d'ordre $\frac{\alpha}{2}$ et b est le quantile d'ordre $1 - \frac{\alpha}{2}$ relativement à la loi de Khi-deux à n degrés de liberté, on obtient :

$$\mathbb{P}\left(\Lambda_n^{-1}\left(\frac{\alpha}{2}\right) \leq \frac{n}{\sigma^2}T_n \leq \Lambda_n^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = \mathbb{P}\left(\frac{nT_n}{\Lambda_n^{-1}\left(1-\frac{\alpha}{2}\right)} \leq \sigma^2 \leq \frac{nT_n}{\Lambda_n^{-1}\left(\frac{\alpha}{2}\right)}\right) = 1 - \alpha.$$

Ainsi, l'intervalle $\left[\frac{nT_n}{\Lambda_n^{-1}(1-\frac{\alpha}{2})}; \frac{nT_n}{\Lambda_n^{-1}(\frac{\alpha}{2})}\right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour σ^2.

Propriétés 3.14 *Soit $X_1, ..., X_n$ un n-échantillon de la variable aléatoire X de loi normale d'espérance m connue et de variance σ^2 inconnue. Posons $T_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$, alors l'intervalle $\left[\frac{nT_n}{b}; \frac{nT_n}{a}\right]$, où a est le quantile d'ordre $\frac{\alpha}{2}$ de la loi de Khi-deux $\mathcal{X}^2(n)$ et b est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de de Khi-deux $\mathcal{X}^2(n)$, est un intervalle de confiance au niveau de confiance 1-α pour σ^2. Par suite on a :*

$$\mathbb{P}\left(\sigma^2 \in \left[\frac{nT_n}{b}; \frac{nT_n}{a}\right]\right) = 1 - \alpha.$$

Exemple(s) 3.15 *Afin d'étudier un caractère quantitatif donné dans une population on a tiré un échantillon $X_1, ..., X_{26}$ suivant la loi normale d'espérance 4 et de variance σ^2 inconnue. Les observations nous ont permis de calculer :*

$$\hat{\sigma} = \frac{1}{n}\sum_i^n (x_i - m)^2$$

$$\sum_{i=1}^{26} x_i = 104.5 \quad et \quad \sum_{i=1}^{26} x_i^2 = 601.4.$$

Dans le but de construire une fourchette d'estimation au niveau de confiance 0,90, nous calculons une estimation de σ^2

$$\widehat{\sigma^2} = \frac{1}{26}\sum_{i=1}^{26}(x_i - m)^2 = \frac{1}{26}\sum_{i=1}^{26} x_i^2 - \frac{2m}{26}\sum_{i=1}^{26} x_i + m^2 = 6.98.$$

Par ailleurs, vu que $\alpha = 0.1$ nous cherchons dans le tabeau de la loi Khi-deux à 26 degrés de liberté les quantiles a et b d'ordre 0.05 et d'ordre 0.95 et on a :

$$a = 15.379 \quad et \quad b = 38.885.$$

En conclusion, l'intervalle :

$$\left[\frac{n\widehat{\sigma^2}}{b}; \frac{n\widehat{\sigma^2}}{a}\right] = \left[\frac{26 \times 6.98}{38.885}; \frac{26 \times 6.98}{15.379}\right] = [4.66; 11.8]$$

est une fourchette d'estimation au niveau de confiance 0,90.

Intervalle de confiance pour σ^2 dans le cas m inconnue

Considérons la variance empirique corrigée $S_n'^2$ c"est un estimateur de σ^2 sans biais et convergent et d'après le théorème 1.28 on a :

$$\frac{n-1}{\sigma^2}S_n'^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \rightsquigarrow \mathcal{X}^2(n-1).$$

Etant donné un $\alpha \in [0,1]$, en choisissant $\Lambda_{n-1}(a) = \frac{\alpha}{2}$ et $\Lambda_{n-1}(b) = 1 - \frac{\alpha}{2}$ autrement dit, a est le quantile d'ordre $\frac{\alpha}{2}$ et b est le quantile d'ordre $1 - \frac{\alpha}{2}$ relativement à la loi de Khi-deux à $n-1$ degrés de liberté, on obtient :

$$\mathbb{P}\left(\Lambda_{n-1}^{-1}\left(\frac{\alpha}{2}\right) \leq \frac{n-1}{\sigma^2}S_n'^2 \leq \Lambda_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\right) = \mathbb{P}\left(\frac{(n-1)\,S_n'^2}{\Lambda_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)} \leq \sigma^2 \leq \frac{(n-1)\,S_n'^2}{\Lambda_{n-1}^{-1}\left(\frac{\alpha}{2}\right)}\right)$$
$$= 1-\alpha.$$

Ainsi, l'intervalle $\left[\frac{(n-1)S_n'^2}{\Lambda_n^{-1}(1-\frac{\alpha}{2})};\ \frac{(n-1)S_n'^2}{\Lambda_n^{-1}(\frac{\alpha}{2})}\right]$ est un intervalle de confiance au niveau de confiance $1-\alpha$ pour σ^2.

Propriétés 3.16 *Soit $X_1,...,X_n$ un n-échantillon de la variable aléatoire X de loi normale d'espérance m inconnue et de variance σ^2 inconnue. Soit $S_n'^2$ la variance empirique corrigée, alors l'intervalle $\left[\frac{(n-1)S_n'^2}{b};\ \frac{(n-1)S_n'^2}{a}\right]$, où a et b sont respectivement les quantiles d'ordre $\frac{\alpha}{2}$ et $1-\frac{\alpha}{2}$ de la loi de de Khi-deux $\mathcal{X}^2(n-1)$, est un intervalle de confiance au niveau de confiance 1-α pour σ^2. Par suite, on a :*

$$\mathbb{P}\left(\sigma^2 \in \left[\frac{(n-1)\,S_n'^2}{b};\ \frac{(n-1)\,S_n'^2}{a}\right]\right) = 1-\alpha.$$

Exemple(s) 3.17 *A partir des observations d'un échantillon, de taille 20, tirées d'une population normale on a obtenu les résultats suivant :*

$$\sum_{i=1}^{20} x_i = 54.8 \quad et \quad \sum_{i=1}^{20} x_i^2 = 302.4.$$

Afin de construire une fourchette d'estimation au niveau de confiance 0,95, *nous allons commencer par calculer* $(20-1)\,s_{20}'^2$:

$$\begin{aligned}(20-1)\,s_{20}'^2 &= \sum_{i=1}^{20}(x_i-\overline{x}_{20})^2 = \sum_{i=1}^{20}x_i^2 - 2\overline{x}_{20}\sum_{i=1}^{20}x_i + 20\overline{x}_{20}\\ &= \sum_{i=1}^{20}x_i^2 - 2\frac{1}{20}\sum_{i=1}^{20}x_i + \sum_{i=1}^{20}x_i\\ &= 302.4 - \frac{1}{10}54.8 + 54.8 = 351.72.\end{aligned}$$

Après une lecture dans le tableau de la loi Khi-deux, on a donc a et b sont respectivement les quantiles d'ordre 0.025 *et* 0.975 *de la loi de de Khi-deux* $\mathcal{X}^2(19)$:

$$a = 8.907 \quad et \quad b = 32.852,$$

il vient dont que l'intervalle recherché est :

$$\left[\frac{351.72}{32.852};\frac{351.72}{8.907}\right] = [10.706; 28.261].$$

3.2.2 Intervalle de confiance de la moyenne d'une loi quelconque

Dans ce paragraphe on suppose que $X_1,...,X_n$ est un n-échantillon de la variable aléatoire X de loi quelconque d'espérance m inconnue et de variance σ^2. Nous allons construire, dans un premier temps, un intervalle de confiance pour m dans le cas : σ^2 connue et dans un second temps, le cas σ^2 inconnue, et ceci en appliquant le théorème de la limite centrale.

Intervalle de confiance de la moyenne d'une loi quelconque, dont la variance est connue

D'après le TLC on a :

$$\sqrt{n}\left(\frac{\overline{X}_n-m}{\sigma}\right) \xrightarrow[n\to\infty]{\mathcal{L}} Y \sim \mathcal{N}(0,1),$$

ce qui permet d'écrire que pour tout réel a :

$$\lim_{n\to\infty}\mathbb{P}\left(-a \leq \sqrt{n}\frac{\overline{X}_n-m}{\sigma} \leq a\right) = \mathbb{P}(-a \leq Y \leq a) = 2\Phi(a)-1.$$

En choisissant a le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, on obtient :

$$\lim_{n\to\infty}\mathbb{P}\left(-a \leq \sqrt{n}\frac{\overline{X}_n-m}{\sigma} \leq a\right) = 1-\alpha.$$

Propriétés 3.18 *Soit $X_1,...,X_n$ un n-échantillon de la variable aléatoire X de loi quelconque d'esperance m inconnue et de variance σ^2 connue.*
L'intervalle $\left[\overline{X}_n - a\frac{\sigma}{\sqrt{n}};\ \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right]$, où a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, est un intervalle de confiance au niveau ***asymptotique*** *1-α pour m.*
Par suite on a :

$$\lim_{n\to\infty}\mathbb{P}\left(m \in \left[\overline{X}_n - a\frac{\sigma}{\sqrt{n}};\ \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right]\right) = 1-\alpha.$$

Intervalle de confiance de la moyenne d'une loi quelconque, dont la variance est inconnue

Partant du résultat de l'extension du TLC, le théorème 1.20, on a :

$$\sqrt{n-1}\frac{\overline{X}_n-m}{S_n} = \sqrt{n}\frac{\overline{X}_n-m}{S_n'} \xrightarrow[n\to\infty]{\mathcal{L}} Y \sim \mathcal{N}(0,1).$$

Par suite, pour tout réel a, on peut écrire :

$$\lim_{n\to\infty}\mathbb{P}\left(-a \leq \sqrt{n-1}\frac{\overline{X}_n-m}{S_n} \leq a\right) = \lim_{n\to\infty}\mathbb{P}\left(-a \leq \sqrt{n}\frac{\overline{X}_n-m}{S_n'} \leq a\right) = 2\Phi(a)-1$$

et en choisissant a le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, on obtient :

$$\lim_{n\to\infty}\mathbb{P}\left(-a\leq\sqrt{n-1}\frac{\overline{X}_n-m}{S_n}\leq a\right)=\lim_{n\to\infty}\mathbb{P}\left(-a\leq\sqrt{n}\frac{\overline{X}_n-m}{S'_n}\leq a\right)=1-\alpha.$$

Propriétés 3.19 *Soit $X_1,...,X_n$ un n-échantillon de la variable aléatoire X de loi quelconque d'espérance m inconnue et de variance σ^2 inconnue.*
*L'intervalle $\left[\overline{X}_n-a\frac{S'_n}{\sqrt{n}};\ \overline{X}_n+a\frac{S'_n}{\sqrt{n}}\right]$, où a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, est un intervalle de confiance au niveau **asymptotique** de confiance 1-α pour m. Par suite, on a :*

$$\lim_{n\to\infty}\mathbb{P}\left(m\in\left[\overline{X}_n-a\frac{S'_n}{\sqrt{n}};\ \overline{X}_n+a\frac{S'_n}{\sqrt{n}}\right]\right)=1-\alpha.$$

Remarque(s) 3.20 *Dans la pratique, pour n assez grand : $n\geq 50$ dans le cas de la variance connue et $n\geq 100$ dans le cas de variance inconnue, les intervalles de confiance asymptotiques ci-dessus seront considérés comme intervalles de confiance approchés au niveau de confiance $1-\alpha$.*

3.2.3 Intervalle de confiance d'une proportion

On s'intéresse à un caractère qualitatif $\mathfrak{X}$ à deux modalités codées par 1 ou 0. Notons p la proportion des individus de la population qui répondent à la modalité codée 1. La loi de la variable X associée au caractère $\mathfrak{X}$ est donc une Bernoulli de paramètre p. Soit $(X_1,...,X_n)$ un échantillon de X.

La variable aléatoire fréquence empirique :

$$F_n=\frac{X_1+...+X_n}{n}$$

représente la proportion des individus dans l'échantillon qui répondent à la modalité codée 1 et elle vérifie : $\mathbb{E}(F_n)=p$ et $\mathbb{V}(F_n)=\frac{p(1-p)}{n}$.D'une part, les lois des grands nombres permettent d'écrire :

$$F_n=\frac{X_1+...+X_n}{n}\underset{n\to\infty}{\longrightarrow}p\ \text{ p.s. et en probabilité,}$$

de plus en appliquant le théorème de continuité, on obtient :

$$\sqrt{F_n(1-F_n)}\xrightarrow[n\to\infty]{\mathbb{P}}\sqrt{p(1-p)},$$

vu que la fonction g définie par : $g(t)=\sqrt{t(1-t)}$ est continue. D'autre part, le théorème de la limite centale donne :

$$\sqrt{n}\left(\frac{F_n-p}{\sqrt{p(1-p)}}\right)\xrightarrow[n\to\infty]{\mathcal{L}}Y\sim\mathcal{N}(0,1).$$

En combinant les deux dernières propriétés et par le théorème de Slutsky, on obtient que :

$$\sqrt{n}\left(\frac{F_n-p}{\sqrt{F_n(1-F_n)}}\right)\xrightarrow[n\to\infty]{\mathcal{L}}Y\sim\mathcal{N}(0,1).$$

Par suite, pour tout réel a, on peut écrire :

$$\lim_{n\to\infty}\mathbb{P}\left(-a\leq\sqrt{n}\left(\frac{F_n-p}{\sqrt{F_n(1-F_n)}}\right)\leq a\right)=2\Phi(a)-1$$

et en choisissant a le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, on obtient :

$$\lim_{n\to\infty}\mathbb{P}\left(-a\leq\sqrt{n}\left(\frac{F_n-p}{\sqrt{F_n(1-F_n)}}\right)\leq a\right)=1-\alpha$$

ou encore :

$$\lim_{n\to\infty}\mathbb{P}\left(F_n-a\sqrt{\frac{F_n(1-F_n)}{n}}\leq p\leq F_n+a\sqrt{\frac{F_n(1-F_n)}{n}}\right)=1-\alpha.$$

Propriétés 3.21 *Soit $X_1,...,X_n$ un n-échantillon de la variable aléatoire X de loi de Bernouilli de paramètre p.*
*L'intervalle $\left[F_n-a\sqrt{\frac{F_n(1-F_n)}{n}};\ F_n+a\sqrt{\frac{F_n(1-F_n)}{n}}\right]$, où a est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$, est un intervalle de confiance au niveau **asymptotique** de confiance 1-α pour p. Par suite on a :*

$$\lim_{n\to\infty}\mathbb{P}\left(p\in\left[F_n-a\sqrt{\frac{F_n(1-F_n)}{n}};\ F_n+a\sqrt{\frac{F_n(1-F_n)}{n}}\right]\right)=1-\alpha.$$

Exemple(s) 3.22 *Après une enquête au prés de 900 téléspectateurs pris au hassard, 200 parmi eux affirment regarder les informations de 20h sur la chaine nationale. On est donc en présence d'un échantillon de Bernouilli, et les données permettent de calculer une estimation de la proportion de téléspectateurs qui regardent les informations de 20h sur la chaine nationale. On a donc :*

$$\widehat{p}=\frac{200}{900}=0.222.$$

Afin de construire un intervalle de confiance approché de niveau de confiance 0.95, on lit sur le tableau de la normale la valeur du quantile de 0.975 et on a : $\Phi(1.96)=0.975$ et par suite, l'intervalle

$$\left[0.222-1.96\sqrt{\frac{0.222(1-0.222)}{900}};\quad 0.222+1.96\sqrt{\frac{0.222(1-0.222)}{900}}\right]=[0.194;0.249]$$

est une fourchette d'estimation au niveau de confiance 0,95 pour la proportion p.

3.2.4 Exercices

Exercice 1.

1. Dans un hôpital, on mesure le temps écoulé (en minutes) entre l'admission d'un malade et l'instant où il est reçu par le médecin. Ces mesures $x_1, x_2, \cdots, x_{26}$ sont effectuées sur un échantillon de 26 malades et donnent les résultats suivants :

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
43	41	38	40	38	25	43	50	41	41	42	39	42

x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}
38	41	41	44	42	40	40	39	38	26	49	40	39

Calculer la moyenne $\overline{x}$ de cet échantillon et sa variance empirique s^2.

2. On suppose que les valeurs observées x_i $(1 \leq i \leq 26)$ sont les valeurs prises par 26 variables aléatoires gaussiennes $X_1, X_2, ..., X_{26}$ indépendantes, de même loi $\mathcal{N}(m, \sigma^2)$. Donner un intervalle fourchette de confiance pour m à partir de l'échantillon observé au niveau de confiance $0,95$.

   (a) En supposant σ^2 connue : $\sigma^2 = 25$.

   (b) En supposant σ^2 inconnue.

Exercice 2.

Certains types de maux de tête sont traités avec un médicament M. Une étude statistique a montré que le temps (donné en minute) de disparition de la douleur chez les malades traités avec ce médicament est une variable aléatoire de loi normale $\mathcal{N}(30; \sigma^2)$, σ^2 étant inconnue.

1. Déterminer par la méthode du maximum de vraisemblance un estimateur T_n de σ^2 et donner ses propriétés statistiques (biais, convergence et efficacité).
   On rappelle que si X est de loi $\mathcal{N}(m; \sigma^2)$ alors $\mathrm{V}(X^2) = 2\sigma^4 + 4m\sigma^2$.
2. Déterminer par la méthode des moments un estimateur U_n de σ^2 et donner ses propriétés statistiques (biais, convergence et efficacité).
3. Comparer les deux estimateurs.
4. Au vue d'observations relevées auprès de 20 patients traités avec le médicament M nous avons calculé les valeurs suivantes

$$\sum_{i=1}^{20} x_i = 620 \quad \text{et} \quad \sum_{i=1}^{20} x_i^2 = 19520$$

où x_i est le temps de disparition de la douleur chez le $i^{ième}$ patient. Donner une estimation de la vraie valeur de la variance σ^2, puis déterminer une fourchette de confiance pour σ^2 au niveau de confiance $0,90$.

Exercice 3.

Un télescope est programmé pour s'ouvrir tous les soirs afin de couvrir une certaine partie du ciel. L'erreur commise par son mécanisme d'ouverture (en degré) est modélisée par une variable aléatoire X de densité :

$$f(x) = \frac{1}{2\theta} \exp\left(-\frac{|x|}{\theta}\right), \ x \in \mathbb{R} \text{ et } \theta > 0$$

et on suppose que les erreurs commises d'un jour à l'autre sont indépendantes.

1. Déterminer par la méthode du maximum de vraisemblance un estimateur T_n de θ et donner ses propriétés statistiques (biais, convergence et efficacité).
2. Montrer que
$$\sqrt{n}\frac{T_n - \theta}{T_n} \underset{n \to \infty}{\longrightarrow} \mathcal{N}(0; 1) \quad \textit{en loi.}$$
3. Soit $\alpha \in]0, 1[$ fixé, construire un intervalle de confiance au niveau asymptotique de confiance 1-α pour θ.
4. Le directeur de l'observatoire a décidé d'envoyer une équipe de techniciens pour effectuer des réparations si la moyenne de l'erreur absolue dépasse 5 degrés. Suite à des observations effectuées sur une période d'un mois (30 jours) afin de mesurer l'erreur, on a relevé les données suivantes

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
2.5	−1.2	5.8	3.2	−6.1	2.3	0.1	−0.8	2.9	2.4	3.3	−3.5	2.1	0	1.9

x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
2.9	1.8	3.1	−3.2	−3	2.8	−2.7	5.8	4.2	6	2.9	5.4	−2	−0.3	2.8

a- Donner une estimation de θ et un intervalle fourchette de confiance au niveau asymptotique de confiance 0,95 pour θ. 0,95

b- Quelle serait alors la décision du directeur de l'observatoire au vu de ces résultats.

Exercice 4.

Une usine de fabrication de pièces mécaniques fait des contrôles de qualité. On s'intéresse à la proportion de pièces défectueuses fabriquées par jour. Afin d'estimer cette proportion p on a effectué un test de qualité sur des échantillons de 15 lots choisis au hasard de 100 pièces chacun.

Ce contrôle effectué sur les 15 lots de la même façon a donné la statistique suivante :

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
6	3	5	4	4	2	7	7	5	6	6	5	7	5	3

Grâce à l'échantillon observé, donner une estimation de la vraie valeur de la proportion p puis une fourchette de confiance de cette proportion au niveau asymptotique de confiance $0,95$.

Exercice 5.

Etant donnés deux échantillons $X_1, ..., X_n$ et $Y_1, ..., Y_n$ indépendants et telles que les variables $X_1, ..., X_n$ sont i.i.d et distribuées selon la loi normale $\mathcal{N}(\nu; \sigma^2)$ et les variables $Y_1, ..., Y$ sont i.i.d et distribuées selon la loi normale $\mathcal{N}(\mu; \sigma^2)$, avec ν, μ et σ^2 inconnues.

1. Justifier les assertions suivantes :

   (a) La variable $\overline{X}_n - \overline{Y}_n$ suit la loi $\mathcal{N}\left(\nu - \mu; \frac{2}{n}\sigma^2\right)$.

   (b) La variable $\frac{n}{\sigma^2}\left(S_{1,n}^2 + S_{2,n}^2\right)$ suit la loi Khi-deux à $(2n-2)$ degrés de liberté, avec les variables $S_{1,n}^2$ et $S_{2,n}^2$ sont respectivement les variances empiriques de $X_1, ..., X_n$ et $Y_1, ..., Y_n$.

   (c) Si on note $d = \nu - \mu$ et $\overline{D}_n = \overline{X}_n - \overline{Y}_n$, alors la variable

$$\sqrt{n-1}\frac{\overline{D}_n - d}{\sqrt{S_{1,n}^2 + S_{2,n}^2}}$$

   suit la loi de Student $t(2n-2)$.

2. Soit $\alpha \in]0,1[$ fixé et a le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student $t(2n-2)$. Montrer que l'intervalle :

$$I = \left[\overline{D}_n - a\frac{\sqrt{S_{1,n}^2 + S_{2,n}^2}}{\sqrt{n-1}}; \overline{D}_n + a\frac{\sqrt{S_{1,n}^2 + S_{2,n}^2}}{\sqrt{n-1}}\right]$$

   est un intervalle de confiance pour la différence des moyennes $d = \nu - \mu$ au niveau de confiance $1 - \alpha$.

3. **Application :**

   On desire comparer les consomations moyennes en litre et par jour de carburant avant et après une augmentation des prix. On suppose que les consommations avant l'augmentation des prix sont distribuées selon une loi normale $\mathcal{N}(\nu; \sigma^2)$ et que les consommations après l'augmentation des prix sont distribuées selon une loi normale $\mathcal{N}(\mu; \sigma^2)$.

   i) Avant l'augmentation et sur un éhantillon de taille $n = 50$ les observations nous ont donné les résultats suivants :

$$\sum_{i=1}^{50} x_i = 285 \quad \text{et} \quad \sum_{i=1}^{50} x_i^2 = 1624,58$$

   ii) Après l'augmentation et sur un éhantillon indépendant du premier et de même taille, les observations nous ont donné les résultats suivants :

$$\sum_{i=1}^{50} y_i = 255 \quad \text{et} \quad \sum_{i=1}^{50} y_i^2 = 1300,62.$$

   (a) Donner un intervalle fourchette de confiance pour la différence des moyennes $d = \nu - \mu$ au niveau de confiance $0,95$.

   (b) Que peut-on dire sur l'incidence de l'augmentation des prix sur la consommation moyenne de carburant.

TD

(1)

Exercice 1 :

1- $\bar{x} = \frac{1}{26}\sum_{1}^{26} x_i = \frac{1}{26}\sum_{1}^{26} x_i$

$= 40$

$S^2 = \frac{1}{n}\sum_{1}^{n} (X_i - \bar{X}_n)^2$

$= \frac{1}{26}\sum_{1}^{26} X_i^2 - \bar{X}_{26}^2$

$= \frac{1}{26}\sum_{1}^{26} X_i^2 - 40^2$

$= 1625,84 - 40^2 = 25,84$

2- $X \sim \mathcal{N}(m, \sigma^2)$

a) $\sigma^2 = 25$

$\left[\bar{X}_n - a\frac{\sigma}{\sqrt{n}} ; \bar{X}_n + a\frac{\sigma}{\sqrt{n}}\right]$ IC du niveau de confiance $1-\alpha$

a = le quantile d'ordre $1-\frac{\alpha}{2}$

on a $\phi^{-1}(1-\frac{\alpha}{2}) = a \Rightarrow \phi^{-1}(0,975) = a = 1,96$

$\Rightarrow \left[40 - \frac{a\times 5}{\sqrt{26}} ; 40 + \frac{5a}{\sqrt{26}}\right]$ fourchette d'estimation au niveau de confiance $1-\alpha$.

$[38,078 ; 41,921]$ IC

b— σ^2 inconnu. ②

$$S_n'^2 = \frac{n}{n-1} S_n^2 = \frac{26}{25} \times 25,84$$

$$= 26,87$$

$$\left[\overline{X}_n - a \frac{S_n'}{\sqrt{n}} ; \overline{X}_n + a \frac{S_n'}{\sqrt{n}} \right]$$ fourchette d'estimation au niveau de confiance $1-\alpha$

$$\left[40 - \varphi_{25}^{-1}\left(1-\frac{\alpha}{2}\right) \times \frac{\sqrt{26,87}}{\sqrt{26}} ; 40 + \varphi_{25}^{-1}\left(1-\frac{\alpha}{2}\right) \frac{\sqrt{26,87}}{\sqrt{26}} \right]$$

~~[39,64 ; 50,328] IC au niveau de confiance $1-\alpha$.~~

(Tableau de Student)

$$\varphi_{25}^{-1}\left(1-\frac{\alpha}{2}\right) = 2,060$$

$k = n-1$
$= 25$
$\alpha = 0,05$
$\frac{\alpha}{2} = 0,025$

~~[23,13 ; 56,86]~~ IC ~~au niveau de~~

~~[30,95 ; 49,8]~~

~~[38,26 ; 41,73]~~

~~[38,26 ; 42,05]~~

[37,50 ; 42,05]

Exercice 2 :

$X \sim \mathcal{N}(30, \sigma^2)$; σ^2 inconnue.

On pose $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2$

$= \frac{1}{n} \sum_{1}^{n} (X_i^2 - 2mX_i + m^2)$

$= \frac{1}{n} \left(\sum_{1}^{n} X_i^2 - \sum_{1}^{n} 2m X_i + \sum_{1}^{n} m^2 \right)$

$= \frac{1}{20} \left(\cancel{620}\ 15520 - 2 \times 30 \times 620 + 20 \times 30^2 \right)$

$= 16$

FC $\left[\frac{n T_n}{b} ; \frac{n T_n}{a} \right]$, a = quantile d'ordre $\frac{\alpha}{2}$ et b = quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi khi-deux $X^2(20)$

on a $1 - \alpha = 0,9 \Rightarrow \alpha = 0,1$

$a = \Lambda_{20}^{-1}\left(\frac{\alpha}{2}\right) = \Lambda_{20}^{-1}(0,05) = 10,85$ | $1-\gamma = 0,05$, $\gamma = 0,95$

$b = \Lambda_{20}^{-1}\left(1 - \frac{\alpha}{2}\right) = \Lambda_{20}^{-1}(0,95) = 31,41$

$\Rightarrow \left[\frac{20 \times 16}{31,41} ; \frac{20 \times 16}{10,85} \right]$

F.C $[10,187 ; 29,49]$

Exercice 3

(1)

1/ $f(x) = \frac{1}{2\theta} \exp\left(-\frac{|x|}{\theta}\right) \quad , \quad x \in \mathbb{R}, \ \theta > 0$

$$L(x_1, \dots, x_n, \theta) = \prod_{1}^{n} f_\theta(x_i)$$
$$= \prod_{1}^{n} \left(\frac{1}{2\theta} \exp\left(-\frac{|x_i|}{\theta}\right)\right)$$
$$= \left(\frac{1}{2\theta}\right)^n \exp \sum_{1}^{n} \left(-\frac{|x_i|}{\theta}\right)$$

$$\ln(L(x_1 \dots x_n, \theta)) = \ln\left(\frac{1}{(2\theta)^n} \exp\left(\sum_{1}^{n} \left(\frac{-|x_i|}{\theta}\right)\right)\right)$$
$$= \cancel{\ln} - n \ln(2\theta) + \sum_{1}^{n} \frac{-|x_i|}{\theta}$$
$$= -n \ln 2 - n \ln(\theta) + \sum_{1}^{n} \left(\frac{-|x_i|}{\theta}\right)$$
$$= -n \ln 2 - n \ln(\theta) - \frac{1}{\theta} \sum_{1}^{n} |x_i|$$

$$\frac{\partial}{\partial\theta} \ln(L) = \cancel{+}\frac{n}{\theta} - \sum_{1}^{n} |x_i| \left(\frac{-1}{\theta^2}\right)$$
$$= \cancel{+}\frac{n}{\theta} \cancel{+} \frac{1}{\theta^2} \sum_{1}^{n} |x_i|$$

$\frac{1}{\hat\theta_{MV}} \left(-n + \frac{1}{\hat\theta_{MV}} a\right) = 0$

$-n + \frac{1}{\hat\theta_{MV}} a = 0$

$= \frac{n}{a}$

$\frac{a}{n}$

$$\frac{\partial}{\partial\theta} \ln(L) \Big|_{\theta = \hat\theta_{MV}} = 0$$

$$\cancel{+}\frac{n}{\hat\theta_{MV}} \cancel{+} \frac{1}{\hat\theta_{MV}^2} \sum_{1}^{n} |x_i| = 0$$

$$\Rightarrow \boxed{\hat\theta_{MV} = \frac{1}{n} \sum_{1}^{n} |x_i|}$$

$$E(W_n) = \frac{1}{n}\sum_{1}^{n} E(|x_i|)$$

$$E(|x|) = \int_0^{+\infty} |x| f_\theta(x)\, dx$$

$$= \cancel{\int_{-\infty}^{+\infty} \frac{|x|}{2\theta} \exp\left(-\frac{|x|}{\theta}\right) dx}$$

$$= \int_0^{\infty} x f_\theta(x)\, dx$$

$$= \int_0^{\infty} \frac{x}{\theta} \exp\left(-\frac{|x|}{\theta}\right) dx = \theta$$

$$E(|X|) = \theta$$

$$E(|X^2|) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = 2\theta^2$$

$$= 2\int_0^{\infty} x^2 f(x)\, dx$$

$$V(|X|) = E(X^2) - (E(|X|))^2$$

$$= \theta^2$$

$$E(W_n) = E(|X|) = \theta \text{ sans biais}$$

$$V(W_n) = \frac{n}{n^2} V(|X|) = \frac{\theta^2}{n} \to 0$$

$\Rightarrow$ CV.

(3)

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$$

$$= -E\left(\frac{n}{\theta^2} - \frac{3}{\theta^2}\sum |X_i|\right) = \frac{n}{\theta^2}$$

$$V(W_n) = \frac{1}{I_n(\theta)} \Rightarrow W_n \text{ st efficace.}$$

2/ 1) $W_n \xrightarrow{P} E(W_n) = \theta$

2) TLC : $\sqrt{n}\,\dfrac{W_n - \theta}{\sqrt{\theta^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

$\left.\right\}$ (+) slutsky

$$\Rightarrow \sqrt{n}\,\frac{W_n - \theta}{W_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

3) $P\left(-a < \sqrt{n}\,\dfrac{W_n - \theta}{W_n} < a\right) \longrightarrow 2\Phi(a) - 1$

$$\Rightarrow a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$\Rightarrow P\left(W_n - a\frac{W_n}{\sqrt{n}} \leq \theta \leq W_n + a\frac{W_n}{\sqrt{n}}\right)$$

$$\Rightarrow IC = \left[W_n \pm \frac{a W_n}{\sqrt{n}}\right]$$

4

4) a- $w_n = \frac{1}{n} \sum_{1}^{n} |x_i|$

$= \frac{1}{30} \sum_{1}^{30} |x_i| = 2{,}82.$

$\cancel{\Phi^{-1}\left(1-\frac{\alpha}{2}\right) = 0{,}99} \Rightarrow \cancel{2\alpha = 0{,}01}$

a le quantile d'ordre $1-\frac{\alpha}{2}$

$\Rightarrow \Phi^{-1}(\underbrace{0{,}55}_{1-\frac{\alpha}{2}}) = 0{,}33$

IC $\left[2{,}5 \pm 1{,}23\right] = \left[1{,}67 , 4{,}13\right]$

b) La moyenne de l'erreur ne dépasse pas 5 degrés

→ Il est pas nécessaire d'envoyer l'équipe des techniciens.

6

Exercice 1

$X \sim \mathcal{N}(\nu, \sigma^2)$

$Y \sim \mathcal{N}(\mu, \sigma^2)$

1) a – $\bar{D}_n = \bar{X}_n - \bar{Y}_n$

$\bar{X}_n \sim \mathcal{N}\left(\nu, \frac{\sigma^2}{n}\right)$

$\bar{Y}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ cours

$\Rightarrow \bar{D}_n \sim \mathcal{N}\left(\nu - \mu, \frac{2\sigma^2}{n}\right)$

3 –

b – $\frac{n}{\sigma^2} S_{1,n}^2 \sim \chi^2(n-1)$ cours

$\frac{n}{\sigma^2} S_{2,n}^2 \sim \chi^2(n-1)$ indépendantes

$\frac{n}{\sigma^2}\left(S_{1,n}^2 + S_{2,n}^2\right) \sim \chi^2(2n-2)$

c – $\sqrt{n-1}\,\frac{\bar{D}_n - d}{\sqrt{S_{1,n}^2 + S_{2,n}^2}} \sim t(2n-2)$ Théo 1, 28 chapitre 1.

2)

$$\mathbb{P}\left(-a \leq \sqrt{n-1}\,\frac{\overline{D}_n - d}{\sqrt{S_{1,n}^2 + S_{2,n}^2}} \leq a\right) = 2\,\Psi_{2n-2}(a) - 1 \quad (6)$$

$$\Rightarrow \mathbb{P}\left(\overline{D}_n - a\,\frac{\sqrt{S_{1,n}^2 + S_{2,n}^2}}{\sqrt{n-1}} \leq d \leq \overline{D}_n + a\,\frac{\sqrt{S_{1,n}^2 + S_{2,n}^2}}{\sqrt{n-1}}\right)$$

$$= 2\,\Psi_{2n-2}(a) - 1$$

$$\Rightarrow a = \Psi_{2n-2}^{-1}\left(1-\frac{\alpha}{2}\right)$$

$$I = \left[\overline{D}_n \pm a\,\frac{\sqrt{S_{1,n}^2 + S_{2,n}^2}}{\sqrt{n-1}}\right]$$

$9{,}03 \cdot 10^{-3}$

$\hat{v} = \overline{X}_n = 5{,}7$

$\hat{\mu} = \overline{Y}_n = 5{,}1$

3 – Application

$$\hat{d} = \overline{D}_n = \overline{X}_n - \overline{Y}_n$$

$$= \frac{1}{50}\sum_{1}^{50} x_i - \frac{1}{50}\sum_{1}^{n} y_i$$

$$= 0{,}6$$

$1-\frac{\alpha}{2} = 0{,}95$

$\Rightarrow \alpha = 0{,}1$

$$a = \Psi_{2n-2}^{-1}\left(1-\frac{\alpha}{2}\right) = \Psi_{2n-2}^{-1}(0{,}95)$$

$$= \Psi_{\boxed{98}}^{-1}(0{,}95)$$

$$= \Phi^{-1}(0{,}975)$$

$$= 1{,}96$$

7

$$S_{1,n}^2 = \frac{1}{50} \sum_{1}^{50} (X_i^2) - \overline{X}_n^2$$

$$= 1,6 \cdot 10^{-3} = 0,0016$$

$$S_{2,n}^2 = \frac{1}{50} \sum_{1}^{50} (Y_i)^2 - \overline{Y}_n^2$$

$$= 2,4 \cdot 10^{-3} = 0,0024$$

$$I = \left[0,6 \pm 1,96 \times (9,03 \cdot 10^{-3}) \right]$$

$$= [0,582 \, , \, 0,618] \subset \mathbb{R}_+^*$$

$\to d > 0$

$\mu - \mu > 0 \Rightarrow \nu > \mu$.

avant après

~~différence positive~~

La moyenne de consommation a baissé avec une probabilité 0,95.

d

Chapitre 4

Tests d'hypothèses

4.1 Problème de décision

La prise de décision concerne tous les domaines : industrie, recherche, médecine, jeux, justice etc...

La statistique ne se limite pas à l'estimation de paramètres, mais aussi à formuler et à tester des hypothèses sur les valeurs des paramètres inconnus ou encore sur les lois de probabilités des variables choisies. Cette prise de décision est toujours motivée par l'observation de phénomènes donnés. Citons par exemple :

Industrie : contrôle de qualité. Prendre la décision de remplacer une machine en tenant compte du nombre de pièces défectueuses produites par cette machine. Cette décision engendre des dépenses et donc un manque à gagner pour l'entreprise.

Médecine : Décider de l'efficacité d'un médicament par rapport à un autre.

Jeux : Décider si un dé est pipé ou pas, et donc le changer ou le garder.

4.1.1 Test d'hypothèses

Le problème de décision consiste donc à choisir en tenant compte des observations entre deux hypothèses, une appelée **hypothèse nulle**, notée H_0 et une autre appelée **hypothèse alternative** et notée H_1.

Un **test d'hypothèses** est une procédure qui permet de choisir entre deux hypothèses en se basant sur les observations d'un échantillon. On distingue deux catégories de tests :

i) **Tests paramétriques** : ce sont des tests sur des hypothèses portant sur les valeurs numériques des paramètres de la loi étudiée (proportion, moyenne, variance...).

ii) **Tests non paramétriques** ce sont des tests sur les hypothèses portant sur la loi de la variable à étudier ou tout autre type ne portant pas sur les paramètres de loi donnés.

On ne peut pas parler de décision sans parler des erreurs possibles faite lors de la prise de décision. En justice par exemple, on peut condamner un innocent ou acquitter un coupable. Se pose alors la question de la gravité de l'erreur suite à la prise de décision.

On distingue deux types d'erreurs :

- **Erreur de première espèce :** décider que H_1 est vraie alors que H_0 est vraie.
- **Erreur de seconde espèce :** décider que H_0 est vraie alors que H_1 est vraie.

On s'intéressera donc au calcul de la probabilité de décider juste et celle de se tromper.

Dans le cas d'erreur de première espèce, la probabilité de rejeter à tort H_0 est notée α **seuil de signification du test** (ou risque fournisseur) et le nombre $1-\alpha$ est appelé **niveau de confiance du test.**

Dans le cas d'erreur de seconde espèce la probabilité de choisir H_0 alors que c'est H_1 qui est vraie est notée β et appelée **risque client** et $1-\beta$ est alors la probabilité de décider H_1 ou de rejeter H_0 à raison est appelée **puissance du test.**

Réalité → ↓ *Décision*	H_0 est vraie	H_0 est fausse
Accepter H_0	Décision correcte $\mathbb{P}(\text{accepter } H_0 \ / \ H_0 \text{ vraie}) = 1-\alpha$	Erreur de seconde espèce $\mathbb{P}(\text{accepter } H_0 \ / \ H_0 \text{ fausse}) = \beta$
Rejeter H_0	Erreur de première espèce $\mathbb{P}(\text{rejeter } H_0 \ / \ H_0 \text{ vraie}) = \alpha$	Décision correcte $\mathbb{P}(\text{rejeter } H_0 \ / \ H_0 \text{ fausse}) = 1-\beta$

Diminuer les deux risques en même temps est malheureusement pas possible, d'autant plus que la procédure de diminuer l'un engendre l'augmentation de l'autre. Par suite, il y a donc un choix à faire, ce choix revient au décideur de privilégié l'une des deux alternatives.

Reprenons l'exemple de justice :

$$H_0 : \textit{la personne est innocente}; \quad H_1 : \textit{la personne est coupable}.$$

Le dilemme c'est de choisir entre privilégié l'erreur « Condamner un innocent » ou « acquitter un coupable ».

La pratique et l'éthique fait que nous cherchons à minimiser l'erreur de condamner un innocent, en d'autres termes minimiser α.

En pratique, on choisit H_0 et H_1 de telle sorte que l'on cherche à minimiser l'erreur de première espèce.

4.1.2 Variable de décision et région critique

Étant donné un échantillon $X_1, \ldots, X_n$, une fonction $T(X_1, \ldots, X_n)$ est appelé **variable de décision** si elle est porteuse d'informations sur le choix entre les deux hypothèses et si elle est distribuée ou asymptotiquement distribuée selon une loi connue (normale, khideux, Student, Fisher).

Au vue des observations, cette variable permettra de choisir entre H_0 et H_1 sous la contrainte que la probabilité de rejeter à tort H_0 est égale à α fixé.

Donc la variable de décision permet de déterminer la région dans laquelle on rejette H_0.

Définition 4.1
*On appelle **région critique** du test, l'ensemble des échantillons pour les quels on rejette H_0 en faveur de H_1*

$$W = \{(X_1, \ldots, X_n);\ H_0 \text{ est rejetée}\}$$

telle que :

$$\mathbb{P}(W\ /\ H_0 \text{ est vraie}) = \alpha.$$

Remarque(s) 4.2
L'ensemble complémentaire de W s'appelle région d'acceptation de l'hypothèse H_0,

$$\overline{W} = \{(X_1, \ldots, X_n);\ H_0 \text{ est acceptée}\}$$

telle que

$$\mathbb{P}\left(\overline{W}\ /\ H_0 \text{ est vraie}\right) = 1 - \alpha.$$

Pour récapituler, la démarche à suivre afin d'effectuer un test d'hypothèses et atteindre une décision est la suivante :

1. Choisir les hypothèses H_0 et H_1 de sorte à priviliger le rejet de H_0.
2. Déterminer la variable de décision.
3. Se fixer α et déterminer la région critique.
4. Regarder si les observations se trouvent ou non dans W.
5. Prendre la décision de rejeter ou non l'hypothèse H_0.

4.2 Tests paramétriques

Étant donné un échantillon $X_1, \ldots, X_n$ d'une variable aléatoire X de loi $\mathcal{L}(\theta)$ dépendant d'un paramètre inconnu $\theta \in \mathbb{R}$. On suppose que nous disposons des observations $x_1, \ldots, x_n$ réalisation des variables $X_1, \ldots, X_n$.

Définition 4.3
*Une hypothèse est **simple** si elle est du type "$\theta = \theta_0$", avec θ_0 est un réel fixé.*
*Une hypothèse est **composite** si elle est du type "$\theta \in A$", avec A est une partie de $\mathbb{R}$ donnée non réduite à un singleton.*

4.2.1 Tests d'hypothèses simples

On dit qu'un test d'hypothèses est un **test d'hypothèses simples** s'il est du type H_0 :"$\theta = \theta_0$" contre H_1 :"$\theta = \theta_1$". Son seuil est la probabilité de rejeter à tort H_0 :

$$\alpha = \mathbb{P}\left((X_1, \ldots, X_n) \in W\ /\ H_0 \text{ est vraie}\right).$$

Ce test permet uniquement de dire au vu des observations laquelle des deux hypothèses est la plus vraisemblable, d'autant plus que nous ne connaissons pas la vraie valeur de θ.

4.2.2 Tests d'hypothèses composites

On dit qu'un test d'hypothèses est un **test d'hypothèses composites** si au moin l'une des hypothèses est composite. On distingue deux types :

test bilatéral : H_0 :"$\theta = \theta_0$" contre H_1 :"$\theta \neq \theta_0$", seule H_1 est composite.

test unilatéral : H_0 :"$\theta \leq \theta_0$" contre H_1 :"$\theta > \theta_0$", ou H_0 :"$\theta \geq \theta_0$" contre H_1 :"$\theta < \theta_0$", H_0 et H_1 sont composites.

Ce n'est pas les seules varriantes possibles, on peut aussi considérer les tests suivants : Soit $I \subset \mathbb{R}$ un intervalle et considérons le test : H_0 :"$\theta \in I$" contre H_1 :"$\theta \notin I$".

La puissance d'un test doit dépendre de la vraie valeur de θ, d'ou les définitions suivantes :

Définition 4.4
La fonction β de θ définie par :

$$\begin{aligned} \beta :\ \mathbb{R} &\longrightarrow [0,1] \\ \theta &\longmapsto \beta(\theta) = \mathbb{P}\left((X_1, \ldots, X_n) \in W\ /\ \theta\right), \end{aligned}$$

*est la **puissance** du test portant sur la valeur de θ, c'est la probabilité de rejeter H_0 quand la vraie valeur du paramètre est θ. Le **seuil du test** est :*

$$\alpha = \sup_{H_0} \beta(\theta),$$

c'est la probabilté maximale de rejeter H_0 alors que H_0 est vraie.

Exemple(s) 4.5 *Certains types de maux de tête sont traités avec un médicament M. Une étude statistique a montré que le temps (donné en minute) de disparition de la douleur chez les malades traités avec ce médicament est une variable aléatoire de loi normale $\mathcal{N}(\theta_0 = 30; 25)$. Une équipe de bio-pharmaciens a conçu un nouveau médicament M' et décide de le tester. En considérant que que le temps (donné en minute) de disparition de la douleur chez les malades traités avec ce nouveau médicament est une variable aléatoire de loi normale $\mathcal{N}(\theta; \sigma^2)$.*
En considérant que l'efficacité du médicament se mesure grâce à la durée moyenne de disparition de la douleur et que pour lancer la commercialisation du médicament M', on choisira de conserver un médicament moins performant que le nouveau plutôt que commercialiser un médicament moins performant que l'ancien. Pour se faire, on emet les hypothèses suivantes :

"$\theta = \theta_0$" : le médicament M' a en moyenne le même effet que le médicament M.
"$\theta < \theta_0$" : le médicament M' est en moyenne plus efficace que le médicament M.
"$\theta > \theta_0$" : le médicament M' est en moyenne moins efficace que le médicament M.

L'hypothèse à retenir dont le but de commercialiser le médicament M' est donc l'hypothèse "$\theta < \theta_0$" et elle correspond au rejet de l'hypothèse H_0. En conclusion, nous allons donc tester

$$H_0 : "\theta \geq \theta_0" \text{ contre } H_1 : "\theta < \theta_0"$$

et chercher la région critique du test de seuil α donné et au vu de n réalisations $x_1, \ldots, x_n$ d'un échantillon $X_1, \ldots, X_n$ de loi $\mathcal{N}(\theta; \sigma^2)$.
La région critique est donc de la forme :

$$W = \{(X_1, \ldots, X_n);\ H_0 \text{ est rejetée}\} = \{(X_1, \ldots, X_n);\ \overline{X}_n < r_\alpha\},$$

reste alors à calculer r_α en utilisant le fait que :

$$\alpha = \sup_{H_0} \beta(\theta) = \sup_{\theta \geq \theta_0} \mathbb{P}\left(\overline{X}_n < r_\alpha;\ \theta\right).$$

Par ailleurs, on sait que la variable $\sqrt{n}\frac{\overline{X}_n - \theta}{S'_n}$ suit la loi de Student $t(n-1)$ de fonction de répartition Ψ_{n-1}, il vient donc :

$$\alpha = \sup_{\theta \geq \theta_0} \mathbb{P}\left(\sqrt{n}\frac{\overline{X}_n - \theta}{S'_n} < \sqrt{n}\frac{r_\alpha - \theta}{S'_n};\ \theta\right) = \sup_{\theta \geq \theta_0} \Psi_{n-1}\left(\sqrt{n}\frac{r_\alpha - \theta}{S'_n}\right) = \Psi_{n-1}\left(\sqrt{n}\frac{r_\alpha - \theta_0}{S'_n}\right).$$

On a donc $\sqrt{n}\frac{r_\alpha - \theta_0}{S'_n} = \Psi_{n-1}^{-1}(\alpha)$ ou encore : $r_\alpha = \theta_0 - \frac{S'_n}{\sqrt{n}}\Psi_{n-1}^{-1}(\alpha)$.

4.2.3 Test de Student sur les paramétres d'une loi normale

En généralisant ce qui précéde, on obtient le résultat connu sous le nom de test de Student et qui regroupe le test bilatéral et les deux tests unilatéraux portant sur la moyenne de la loi normale :

Soit X une variable aléatoire de loi normale de paramètre m et σ^2 et supponsons qu'elle soit associée à un caractère quantitatif étudié dans une population dans laquelle on a tiré un n-échantillon $X_1, ..., X_n$ et réaliser des observations $x_1, ..., x_n$.

Nous allons construire, dans un premier temps, un test d'hypothèses sur la moyenne m dans le cas : σ^2 inconnue :

Proposition 4.6 *Test de Student sur la moyenne d'une loi normale de variance inconnue :*
***Test de** : "$\theta \leq \theta_0$" contre : "$\theta > \theta_0$" : $W = \left\{\sqrt{n}\frac{\overline{x}_n - \theta_0}{s'_n} > \Psi_{n-1}^{-1}(1-\alpha)\right\}$*
***Test de** : "$\theta \geq \theta_0$" contre : "$\theta < \theta_0$" : $W = \left\{\sqrt{n}\frac{\overline{x}_n - \theta_0}{s'_n} > \Psi_{n-1}^{-1}(\alpha)\right\}$*
***Test de** : "$\theta = \theta_0$" contre : "$\theta \neq \theta_0$" : $W = \left\{\sqrt{n}\left|\frac{\overline{x}_n - \theta_0}{s'_n}\right| > \Psi_{n-1}^{-1}(1 - \frac{\alpha}{2})\right\}$.*

Pour les autres cas : test sur la moyenne m avec σ^2 connue, test sur σ^2 dans le cas m connue et dans le cas m inconnue se traitent en suivant la même démarche que précédement.

4.2.4 Tests sur une proportion

On s'intéresse à un caractère qualitatif $\mathfrak{X}$ à deux modalités codées par 1 ou 0. Notons p la proportion des individus de la population qui répondent à la modalité codée 1. La loi de la variable X associée au caractère $\mathfrak{X}$ est donc une Bernoulli de paramètre p. Soit $(X_1, ..., X_n)$ un échantillon de X et $x_1, ..., x_n$ des réalisation des variables $X_1, ..., X_n$.

La variable aléatoire fréquence empirique :

$$F_n = \frac{X_1 + ... + X_n}{n}$$

représente la proportion des individus dans l'échantillon qui répondent à la modalité codée 1 et elle vérifie le théorème de la limite centale, ce qui donne que la variable $\sqrt{n}\left(\frac{F_n - p}{\sqrt{p(1-p)}}\right)$ est approximativement de loi $\mathcal{N}(0, 1)$.

Proposition 4.7 *Tests asymptotiques sur une proportion*

Test de : "$p \le p_0$" *contre :* "$p > p_0$" : $W = \left\{ \sqrt{n} \dfrac{\overline{x}_n - p_0}{\sqrt{p_0(1-p_0)}} > \Phi^{-1}(1-\alpha) \right\}$

Test de : "$p \ge p_0$" *contre :* "$p < p_0$" : $W = \left\{ \sqrt{n} \dfrac{\overline{x}_n - p_0}{\sqrt{p_0(1-p_0)}} > -\Phi^{-1}(1-\alpha) \right\}$

Test de : "$p = p_0$" *contre :* "$p \ne p_0$" : $W = \left\{ \sqrt{n} \left| \dfrac{\overline{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \right| > \Phi^{-1}(1-\frac{\alpha}{2}) \right\}$.

4.2.5 Remarques

1. Le rejet ou non de l'hypotnèse H_0 dépend étroitement du choix du seuil de risque α. Plus le seuil est petit, moins on aura tendance à rejeter l'hypothèse H_0.
2. Il existe un seuil critique α_C telque pour tout seuil α inférieur à α_C on ne rejetera pas H_0, autrement dit, pour tout seuil α supérieur à α_C on rejetera pas H_0. Cette valeur est appelée **p-valeur**, il existe des logicels de statistiques qui permettent de la calculer.
3. On a vu que pour construire un test paramétrique pour un paramétre θ, nous avons utilisé la même technique que pour la construction d'un intervalle de confiance. Le plus classique est donc de partir d'un estimateur $\widehat{\theta}_n$ de θ et de procéder comme suit :

   Test de : "$\theta \le \theta_0$" contre : "$\theta > \theta_0$" on cherche r_α tel que $W = \left\{ \widehat{\theta}_n > r_\alpha \right\}$,

   Test de : "$\theta \ge \theta_0$" contre : "$\theta < \theta_0$" : on cherche r_α tel que $W = \left\{ \widehat{\theta}_n < r_\alpha \right\}$,

   Test de : "$\theta = \theta_0$" contre : "$\theta \ne \theta_0$" : on cherche r_α tel que

   $$W = \left\{ \widehat{\theta}_n < r_{1,\alpha} \text{ ou } \widehat{\theta}_n > r_{2,\alpha} \right\}.$$

   Revenir à la difinition de $\alpha = \sup\limits_{H_0} \beta(\theta) = \sup\limits_{H_0} \mathbb{P}(X_1, \dots, X_n) \in W \ / \theta)$ permet donc de calculer les différents cas $r_\alpha, r_{1,\alpha}, r_{2,\alpha}$.
4. Cette méthode ne fonctionne pas toujours et même dans les cas où elle fonctionne elle ne donne pas nécessairment un test optimal. Des méthodes statistiques plus puissantes et plus performantes existes, leures étudent dépasse le cadre de ce cours, un des résultats le plus important est le théorème de Neyman-Pearson.

4.2.6 Exercices

Exercice 1.

On reprend l'exemple 4.5 et on suppose que l'équipe de bio-pharmaciens qui a conçu le nouveau médicament M′ a testé ce médicament sur 12 malades et les observations de la durée de disparition de la douleur sont les suivantes :

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
25	28	20	32	17	24	41	28	25	30	27	24

1. On décide de ne commercialiser le nouveau médicament que si on est sûr à 95% qu'il est plus efficace. Au vu des observations, qu'elle sera la décision de l'équipe de bio-pharmaciens qui a conçu le nouveau médicament M′.
2. Qu'elle sera cette décision si le seuil de risque était de 1%.

Exercice 2.

Un candiat à une éléction lance un sondage afin de savoir s'il va être élu ou pas. Sur les 800 personnes intérrogé, 420 d'entre elles ont déclaré vouloir voter pour lui. Peut-on affirmer que ce candidat sera élu avec moins de 5% de chances de se tromper.

Exercice 3.

Une étude des salaires pour un secteur donné a montré qu'il est distribué selon une loi normale de moyenne 700dt et de variance σ^2. Suite aux observations récoltées à l'essu d'un tirage d'un échantillon de 30 salaires, nous avons relevé que :

$$\sum_{i=1}^{30} (x_i - 700)^2 = 2013.560.$$

Au seuil de 5% peut-on affirmer que $\sigma^2 = 120$.

Exercice 3

①

~~note: $1-\alpha = 0{,}95$~~

~~$\alpha = 0{,}05$~~

~~$\psi^{-1}(0{,}05) =$~~

$\bar{x}_n = 26{,}75$

$s_n' = 6{,}08$

$r_\alpha = \,?$ $\qquad \alpha = 0{,}05$

$\psi^{-1}(1-\alpha) = -\psi^{-1}(\alpha)$

on a $\sqrt{12}\,\dfrac{r_\alpha - 30}{6{,}08} = \psi_{11}^{-1}(0{,}05) = -1{,}796$.

~~donc~~ $r_\alpha = \dfrac{6{,}08 \times (-1{,}796)}{\sqrt{12}} + 30$

$= 26{,}84$

$\bar{X}_n < r_\alpha$

Donc les observations sont dans la region critique. Donc on rejette H_0 donc, on conclut que n'est plus efficace que n avec moins de 5% de ~~confiance~~ chance de tromper.

pour $\alpha = 0{,}01$, $\psi^{-1}(0{,}01) = -2{,}718$

$r_\alpha = 25{,}22 < \bar{X}_n$

Si on veut le seuil de telle sorte qu'il soit 1% on commercialise pas le nouveau medicament.

Exercice 2 :

$$\omega = \{ (X_1 - x_n) \; ; \; H_0 \text{ rejeté} \}$$

$$P(W / H_0 \text{ est vraie}) = \alpha$$

$P_0 = 50\,\% = 0,5$

Test de $H_0 : p < \frac{1}{2}$ contre $H_1 : p \geq \frac{1}{2}$

$$W = \left\{ \sqrt{n} \, \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \geq \phi^{-1}(1-\alpha) \right\}$$

$\cancel{\sqrt{0,05}}$ $\alpha = 0,05 \rightarrow 1-\alpha = 0,95$

$\phi^{-1}(0,95) = 1,645$.

Au vue des observation, on a $n = 800$, $\bar{x}_n = \frac{420}{800} = 0,525$

et donc $\sqrt{n} \, \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{2} = 1,41 \leq 1,645$

On n'est pas dans la région critique, donc on ne rejette pas H_0 donc on ne peut pas ~~...~~ affirmer que le candidat sera élu avec moins de 5% de chance.