

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



N° Réf :.....

Université Abd Elhafid Boussouf Mila

Institut des Sciences et Technologie

Département de Mathématiques et Informatique

## Mémoire préparé en vue de l'obtention du diplôme de Master

EN: Mathématiques

Spécialité : Mathématiques appliquées

### Test séquentiel du rapport des probabilités appliqué sur COVID-19

Préparé par

- SIFOUR HAYAT
- BOUCHEKOUT IMANE

Devant le jury

Laouira Widad (MCB)

C.U.Abd Elhafid Boussouf

Président

Zerari Amel (MAA)

C.U.Abd Elhafid Boussouf

Rapporteur

Boularouk Yaakoub (MCB)

C.U.Abd Elhafid Boussouf

Examineur

Année Universitaire : 2019/2020

## Remerciement

*En premier lieu, nous remercions Dieu qui nous a procuré de la patience et de courage à fin d'achever ce travail et d'atteindre cette réussite.*

*Nous tenons tout d'abord à exprimer notre profonde gratitude a madame Z'ERARI AM'EL qui nous a dirigé ce travail et avoir su nous apporter l'aide. Nécessaire à sa réalisation, pour son aide, sa patience et ses conseils tout au long de cette recherche.*

*Nous tenons à remercier toutes les personnes qui ont participé à différents degrés pour réaliser ce modeste travail.*

*Sans oublier aussi de remercier infiniment tous les enseignants du département de mathématique, sur leurs aides, efforts et conseils qui ont contribué à notre réussite.*

*Nous voudrions exprimer notre reconnaissance envers les amis et collègues qui m'ont apporté leur soutien moral et intellectuel tout au long de notre démarche.*

# Dédicace

Je dédie ce mémoire de master

A mon père "ABDELMADJIDE"

Qui m'a toujours transmis l'amour du travail et le sens du perfectionnisme et Qui m'a toujours encadré avec beaucoup d'amour et d'attention, Que dieu lui réserve bonne santé pour continuer ce long chemin avec ma petite sœur;

A ma mère chérie "AICHA"

Ma raison d'être, ma raison de vivre, la lanterne qui éclaire mon chemin et m'illumine de douceur et d'amour, qui par ses sacrifices consentis et son affection profonde m'a toujours guidé sur la voie du succès, Que dieu la réserve bonne santé;

A ma grand-mère **AKILA** que dieu la protège;

A mes grands-pères **MAHMOUD** et **AMAR**, ma grand-mère **LOUIZA**, mes oncles **MEKKI** et **ABDELHAMID** que DIEU les accueille en son paradis;

A mes chères frères **MOHAMED**, **OUSSAMA** et leurs femmes;

A mes chères sœurs **KAOUTHER**, **MARWA** et son mari;

A mes chères neveux **YOUCEF**, **ABDESSAMAD**, **NAZIM** et ma chère nièce **MARIA ISRAA**;

A mes tantes et mon oncle **ABDELNAAIM** et leurs enfants;

A mes amies et à tous ceux qui m'aiment qu'ils trouvent dans ce travail l'expression de mon affection et ma grande admiration.

*Imane*

# Dédicace

Je dédie ce mémoire de master

A mon père "OMAR"

Qui m'a toujours transmis l'amour du travail et le sens du perfectionnisme et Qui m'a toujours encadré avec beaucoup d'amour et d'attention, Que dieu lui réserve bonne santé pour continuer ce long chemin avec ma petite sœur;

A ma mère chérie "HADDA"

Ma raison d'être, ma raison de vivre, la lanterne qui éclaire mon chemin et m'illumine de douceur et d'amour, qui par ses sacrifices consentis et son affection profonde m'a toujours guidé sur la voie du succès, Que dieu la réserve bonne santé;

A mes chère frères AISSAM et RAYAD et MOSTAPHA et  
ABD NOUR;

A ma chère sœur YOUSRA ;

A mes tantes et leurs enfants;

A mes amies et à tous ceux qui m'aiment qu'ils trouvent dans ce travail l'expression de mon affection et ma grande admiration.

**HAYAT**

# Liste de notations

$H_0$	Hypothèse nulle
$H_1$	Hypothèse alternative
$\theta$	Paramètre inconnu
$\theta_0$	Valeur numérique
$\theta_1$	Valeur numérique
$\Theta$	Espace paramétrique
$\Theta_0$	Sous ensemble non vide de $\Theta$
$\Theta_1$	Sous ensemble non vide de $\Theta$
$X$	Variable aléatoire
$\alpha$	Risque de premier espèce
$\beta$	Risque de deuxième espèce
$1 - \alpha$	Niveau de confiance
$1 - \beta$	Puissance d'un test
$n$	nombre de réalisation
$\omega$	Région critique
$\varpi$	Région d'acceptation
TSRP	Test séquentiel du rapport des probabilités
$f_{\theta}(x_i)$	Fonction de densité de $x$
$\varpi(\theta)$	Fonction d'efficacité
ASN	Average sample number (nombre d'échantillon moyen)
$N$	Taille d'échantillon

## Résumé

Ce travail comporte l'analyse séquentielle appliquée sur les essais cliniques. Nous étudions les tests classiques et les tests séquentiels et en particulier le test séquentiel de probabilité dans le domaine statistique.

Nous intéressons aux tests de Pocock et d'O'Brien et Fleming et pour l'erreur de 1<sup>ère</sup> espèce fixée  $\alpha = 0.05$  à fin de diminuer le nombre de la taille d'échantillon dans le but de trouver l'efficacité de traitement.

## Les mots clés

La fonction de l'effectif moyen, l'efficacité de traitement, les tests statistiques, les tests séquentiels, la taille de l'échantillon, COVID-19.

## **Abstract**

This work involves sequential analysis in clinical trials. We study classical tests and sequential tests and in particular the sequential probability test in statistics.

We are interested in the Pocock test and the O'Brien and Fleming test and for the 1re species error fixed  $\alpha=0.05$  in order to decrease the number of the sample size in the goal of finding treatment efficiency.

## **Key words**

The function of the average workforce, treatment efficiency, statistical tests, sequential testing, sample size, COVID-19.

## ملخص

يشمل هذا العمل التحليل المتسلسل المطبق على التجارب السريرية. نحن ندرس الاختبارات التقليدية والتتابعية وخاصة اختبار الاحتمال التسلسلي في المجال الإحصائي.

نحن مهتمون في اختبارات بوكوك وأوبراين وفليمغ ولخطأ الأنواع الأولى التي تم تعيينها في  $\alpha = 0.05$  من أجل تقليل عدد حجم العينة من أجل العثور على فعالية العلاج.

## الكلمات المفتاحية

وظيفة متوسط قوة العمل، كفاءة المعالجة، الاختبارات الاحصائية، الاختبارات المتسلسلة، حجم العينة، كوفيد-19.



# Table des matières

<b>Introduction Générale</b>	<b>13</b>
<b>1 Test d'hypothèse</b>	<b>16</b>
1.1 Le test d'hypothèse . . . . .	16
1.1.1 Définitions . . . . .	16
1.1.2 Test unilatéral et test bilatéral . . . . .	18
1.1.3 La région critique, La région d'acceptation . . . . .	20
1.1.4 La statistique de test : variable de décision . . . . .	22
1.1.5 P-valeur (ou la probabilité critique) . . . . .	23
1.1.6 La décision . . . . .	24
1.1.7 Puissance du test . . . . .	24
1.2 Mise en oeuvre pratique . . . . .	25
1.2.1 La démarche d'un test . . . . .	25
<b>2 Test séquentiel</b>	<b>32</b>
2.1 Expériences séquentielles . . . . .	33
2.1.1 Définition du test séquentiel du rapport des probabilités . . . . .	33
2.1.2 Détermination des constantes A et B du TSRP . . . . .	36
2.1.3 Mise en oeuvre du test séquentiel sur un exemple . . . . .	41
2.2 Propriétés importantes du TSRP . . . . .	44
2.2.1 La courbe d'efficacité . . . . .	44
2.2.2 La fonction de l'effectif moyen de l'échantillon (ASN) . . . . .	48
2.3 Comparaison entre tests classiques et tests séquentiels . . . . .	51

*TABLE DES MATIÈRES*

---

2.3.1	Avantages des tests séquentiels . . . . .	51
2.3.2	Inconvénients des tests séquentiels . . . . .	51
2.3.3	Approche décisionnelle entre le test classique et le test séquentiel . . . . .	51
2.3.4	Gain de la méthode séquentielle . . . . .	52
2.3.5	Contrôle par échantillonnage . . . . .	57
2.3.6	Plan d'échantillonnage . . . . .	57
2.3.7	Contrôle tronqué . . . . .	58
2.3.8	Les propriétés du plan tronqués . . . . .	59
2.4	Tests bilatéraux . . . . .	64
2.4.1	Tests bilatéraux pour comparer deux traitements avec une réponse normale de la variance connue : . . . . .	64
2.4.2	Un test d'échantillonnage fixe . . . . .	65
2.4.3	Tests séquentiels de groupe . . . . .	66
2.5	Test de Pocock . . . . .	67
2.5.1	L'exigence de puissance . . . . .	68
2.6	Test d'O'Brien et Fleming . . . . .	69
2.6.1	L'exigence de puissance . . . . .	70
<b>3</b>	<b>Application numérique</b>	<b>72</b>
3.1	le COVID-19 . . . . .	73
3.1.1	Appliquées des donnée réelle dans le COVID-19 . . . . .	74
3.1.2	Propriétés des tests Pocock et O'Brien et Fleming : . . . . .	78
	<b>Conclusion</b>	<b>87</b>

# Table des figures

1.1	Test bilatéral . . . . .	19
1.2	Un test unilatéral à droite . . . . .	19
1.3	Un test unilatéral à gauche . . . . .	19
2.1	Augmentation du nombre d'observations . . . . .	41
2.2	Représentation du rejet de $H_0(\theta = 1)$ . . . . .	43
2.3	Représentation de l'acceptation de $H_0(\theta = 2)$ . . . . .	44
2.4	Approximation de la fonction d'efficacité . . . . .	47
2.5	Approximation de La fonction de l'effectif moyen de l'échantillon (ASN) . . . . .	50
2.6	Plan d'échantillonnage simple . . . . .	58
2.7	Plan tronqué . . . . .	59
2.8	Courbe d'efficacité . . . . .	60
2.9	Courbe d'efficacité idéale pour $[\theta_0, \theta_1]$ . . . . .	60
2.10	La fonction de l'effectif de moyen de l'échantillon dans le cas de plan tronqué . . . . .	61
2.11	Frontières fermées . . . . .	62
2.12	Frontières ouvertes . . . . .	63
3.1	Test de Pocock avec 5 groupe d'observation . . . . .	76
3.2	Test O'Brien et Fleming avec 5 groupe d'observation . . . . .	77

# Liste des tableaux

1.1	Le tableau représente les risques de première et deuxième espèce . . . . .	21
3.1	Caractéristiques de la population étudiée . . . . .	75
3.2	Test de Pocock avec 5 groupe d'observation . . . . .	77
3.3	Test O'Brien et Fleming avec 5 groupe d'observation . . . . .	78
3.4	Les valeurs de $\mathbf{P}(Z < z)$ en fonction de $z$ . . . . .	81
3.5	les valeurs de $\mathbf{P}(Z < z)$ en fonction de $z$ (suite) . . . . .	82
3.6	<b>Test de Pocock</b> : Les constantes $C_P(K, \alpha)$ . . . . .	83
3.7	<b>Test de Pocock</b> : Les constantes $R_P(K, \alpha, \beta)$ . . . . .	84
3.8	<b>Test d'O'Brien et Fleming</b> : Les constantes $C_B(K, \alpha)$ . . . . .	85
3.9	<b>Test d'O'Brien et Fleming</b> : Les constantes $R_B(K, \alpha, \beta)$ . . . . .	86

# Introduction Générale

On appelle théorie des tests la seconde branche de la statistique mathématique, celle qui permet de confronter deux hypothèses retenues a priori. On retient un modèle statistique où la variable aléatoire  $X$  suit une loi de probabilité  $P_\theta$ , qui dépend d'un paramètre  $\theta$  inconnu. On dispose cependant ici d'informations supplémentaires qui font penser a priori que la valeur de ce paramètre est égale à une valeur fixée  $\theta_0$  et on cherche à valider (à tester) cette hypothèse, au vu d'un échantillon de la loi de  $X$ .

Réaliser un test statistique consiste à mettre en oeuvre une procédure permettant de confronter une hypothèse avec la réalité, ou plus exactement, avec ce que l'on perçoit de la réalité à travers les observations à disposition et de prendre une décision à la suite de cette confrontation.

La littérature statistique abonde de types de tests statistiques. Les tests permettent de trancher entre deux hypothèses (ou questions) antagonistes sur la base d'un échantillon obtenu en réalisant  $n$  expériences indépendantes. La décision prise dépend alors de la taille de l'échantillon considéré.[17]

Dans certaines situations pratiques la taille doit être très grande pour prendre une bonne décision quand à l'acceptation ou le rejet de l'hypothèse (exp, en médecine). Les tests séquentiels ont alors fait leur apparition.

Les tests séquentiels et plus précisément les tests séquentiels du rapport des probabilités, ont été introduits par le mathématicien Wald pendant la seconde guerre mondiale mais son livre ne fût, pour des raisons évidentes, publié qu'en 1947.

Jusqu'à la deuxième guerre mondiale les méthodes statistiques séquentielles n'ont pas débordé le cadre d'applications très simples prévues pour un usage bien déterminé. Il a fallu

attendre l'année 1943 pour voir s'élaborer, avec les travaux de A. Wald aux États-Unis et ceux de G. A. Barnard en Grande-Bretagne, le début d'une construction théorique connue maintenant sous le nom d'Analyse séquentielle. Ces travaux ont été réalisés dans le cadre de groupes de perfectionnement des méthodes industrielles en temps de guerre. la découverte la plus importante est due à Wald, il s'agit du test séquentiel du rapport des probabilités TSRP. Ce fut le point de départ d'une construction théorique fort séduisante qui se développa très rapidement.[9]

Aujourd'hui encore cet ensemble conserve une place importante dans l'analyse séquentielle et c'est certainement le domaine où le développement théorique est le plus complet.

De par son essence l'analyse séquentielle est très différente de l'analyse classique. Alors que l'analyse classique fixe à l'avance la taille de l'échantillon d'expérimentation pour donner un verdict au moyen d'un résultat interprété avec un certain seuil d'erreur, la démarche de l'analyse séquentielle est pratiquement inverse, car la taille de l'échantillon n'est pas fixée. En revanche, les risques de première espèce et de seconde espèce sont, quant à eux, fixés.[9]

La technique des tests séquentiels est peu enseignée dans le supérieur. Cependant, elle est très utilisée en fiabilité et plus généralement pour tout problème de certification, comme par exemple celle de médicaments nouveaux. En effet cette technique permet une économie de mesures et par là même de temps, de l'ordre de 50% en moyenne.

Sur la base des travaux effectués par Wald, de nombreux mathématiciens ont développé la théorie des tests séquentiels et appliquée en médecine par divers auteurs.

En 1975, Armitage a appliqué l'analyse séquentielle dans les essais cliniques, cette dernière est une étude sur des êtres humains utiliser pour expliquer l'efficacité de traitement expérimental est analyse les données accumulent avec des résultats après chaque observation et pour chaque participants sont remportés, et la décision est soit continue l'étude en nouvelle participation, et soit arrêter l'étude avec le résultat qu'il y a ou non un différent statistiquement entre les traitements.

Aux États-Unis, l'administration des aliments et des médicaments supervise la réglementation et l'approbation des médicaments, des produits biologiques et des dispositifs médi-

caux. Ses processus d'examen et d'approbation influent sur la conception et la conduite de la plupart des nouveaux essais cliniques. Le test préclinique d'une intervention expérimentale est effectué avant que les investigateurs commencent un essai clinique. Ces études sont réalisées en laboratoire et dans des études avec des animaux pour fournir des preuves préliminaires que l'intervention expérimentale sera sûre et efficace pour les humains. L'administration des aliments et des médicaments exige des tests précliniques avant de pouvoir commencer les essais cliniques. Les informations de sécurité provenant des tests précliniques sont utilisées pour appuyer une demande à l'administration des aliments et des médicaments de commencer à tester l'intervention expérimentale dans des études avec des humains.

Le COVID-19 C'est une maladie infectieuse émergente de type zoonose virale causée par la souche de coronavirus SARS-CoV-2. Les symptômes les plus fréquents sont la fièvre, la toux, la fatigue et la gêne respiratoire. Dans les formes les plus graves, l'apparition d'un syndrome de détresse respiratoire aiguë peut entraîner la mort, notamment chez les personnes plus fragiles du fait de leur âge ou en cas de comorbidités. Un nouveau coronavirus (COVID-19) a été identifié en 2019 à Wuhan, en Chine.[14]

Notre travail sera réparti en trois chapitres organisés comme suite : le premier chapitre sera consacré à des définitions sur les tests et ses caractéristiques dont nous nous appuyons pour formuler un test et tests de comparaison à une valeur de référence et un exemple récapitulatif.

Puis le deuxième chapitre présentera les tests séquentiels et plus précisément les tests séquentiels du rapport des probabilités, et ses caractéristiques et une comparaison entre les tests classiques et les tests séquentiels.

Le troisième chapitre nous nous intéressons à l'applications des tests séquentiels du rapport des probabilités dans les essais cliniques, Nous appliquerons le test de Pocock et le test d'O'Brien et Fleming sur un exemple en réel (covid-19).

# Chapitre 1

## Test d'hypothèse

### Introduction

Le test c'est un mécanisme qui permet de trancher entre deux hypothèses antagonistes (une hypothèse nulle et une hypothèse dite alternative), à la vue des résultats d'un échantillon, en quantifiant le risque associé à la décision prise, Un test d'hypothèse il s'agit d'une étape de la statistique très importante qui nous permettant de comparer différents paramètres (variances, moyennes...) afin de valider ou de réfuter une hypothèse de statistique et ce en fonction d'un échantillon tiré d'une population à partir de calculs basé sur une observation c'est à dire des données de l'échantillon on va pouvoir conclure sur l'hypothèse de départ et elle acceptable ou réfutable.

Dans ce chapitre nous rappelons des généralités autour des tests d'hypothèse, L'objectif étant d'être capable de bien formuler un test.

### 1.1 Le test d'hypothèse

#### 1.1.1 Définitions

#### 1. Test statistique

Un test statistique est une procédure de décision entre deux hypothèses concernant un ou plusieurs échantillons.[1]



### 2. Hypothèse nulle

L'hypothèse nulle notée  $H_0$  c'est l'hypothèse que l'on cherche à vérifier qui correspond à une absence de différence ou d'effet. Elle porte sur la loi de probabilité (ou d'autre façon le paramètre  $\theta$  de cette loi) ayant donné naissance à l'échantillon disponible.[1]

$$H_0 : \theta = \theta_0$$

#### Exemple 1.1.1.

$H_0$  : deux populations d'étudiants (de même niveau) ayant suivi des méthodes pédagogiques différentes ont les mêmes notes moyennes aux examens.

### 3. Hypothèse alternative

L'hypothèse alternative qui s'appelle aussi contre hypothèse ou l'hypothèse complémentaire notée  $H_1$  que l'on souhaite prouver en rejetant l'hypothèse nulle  $H_0$ , elle traduit une différence ou un effet statistiquement significatif.[1]

#### D'après l'exemple 1.1.1

$H_1$  : deux populations d'étudiants ayant suivi des méthodes pédagogiques différentes ont des notes moyennes significativement différentes aux examens.

#### Remarque 1.1.1.

Accepter  $H_0$  ne signifie pas que cette hypothèse soit vraie, mais seulement que les observations disponibles ne sont pas incompatibles avec cette hypothèse et que l'on n'a pas de raisons suffisantes pour préférer  $H_1$ . De même, rejeter  $H_0$  ne signifie pas que  $H_1$  soit vraie.

### 4. Hypothèse simple

Une hypothèse est qualifiée de simple si la loi de la v. a . $X$  est totalement spécifiée quand cette hypothèse est réalisée. Dans le cas contraire elle est dite multiple. Nous allons examiner le cas où le paramètre  $\theta$  ne peut prendre que deux valeurs  $\theta_0$  et  $\theta_1$  [2]. Ce qui

correspond au choix entre les deux hypothèse simples suivantes :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

### 5. Hypothèse multiple

Nous allons d'abord considérer le cas d'une hypothèse simple contre une hypothèse multiple de l'une des formes suivantes [2] :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

#### 1.1.2 Test unilatéral et test bilatéral

Un test statistique se base sur deux hypothèse en compétition : l'hypothèse nulle  $H_0$  et l'hypothèse alternative  $H_1$ . Le type d'hypothèse alternative  $H_1$  définit si un test est unilatéral ou bilatéral.

##### • Test bilatéral :

Un test bilatéral s'applique quand nous cherchons une différence entre deux paramètres, ou entre un paramètre et une valeur donnée sans se préoccuper du signe ou du sens de la différence.[3]

Dans ce cas, la zone de rejet de l'hypothèse principale se fait de part et d'autre de la distribution de référence .

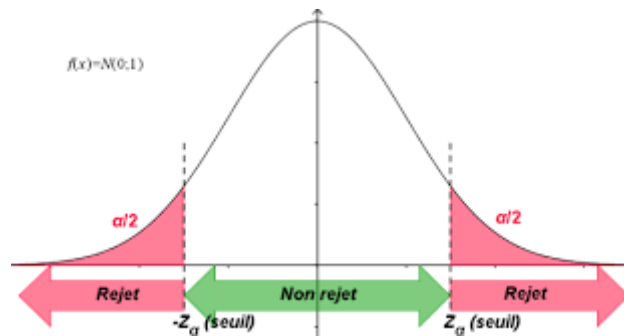


FIGURE 1.1 – Test bilatéral

• **Test unilatéral :**

Un test unilatéral s'applique quand nous cherchons à savoir si un paramètre est supérieur (ou inférieur) à un autre ou à une valeur donnée.[3]

-La zone de rejet de l'hypothèse principale est située d'un seul côté de la distribution de probabilité de référence.

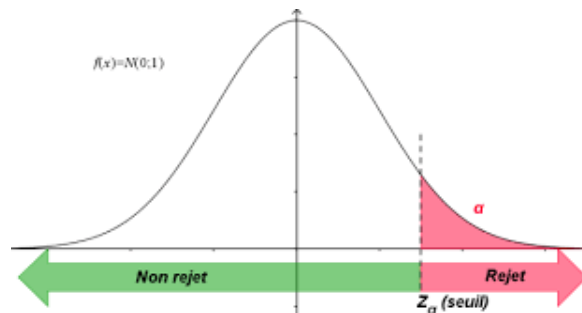


FIGURE 1.2 – Un test unilatéral à droite

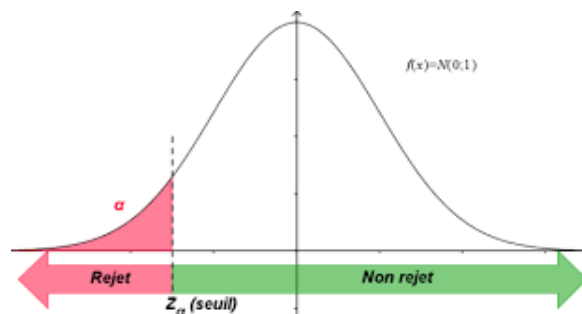


FIGURE 1.3 – Un test unilatéral à gauche

### 1.1.3 La région critique, La région d'acceptation

**La région critique**  $W$  d'un test est l'ensemble des valeurs observées pour lesquelles l'hypothèse nulle  $H_0$  est rejetée. Les valeurs limites de cette région constituent les valeurs critique.[3]

#### Remarque 1.1.2.

*Dans la plupart des situation que vous rencontrerez dans la suite, la région critique  $W$  peut être reliée au risque d'erreur de première espèce  $\alpha$  par :*

$$P_{H_0}[W] = \alpha$$

**La région d'acceptation** de  $H_0$  est le complément de la région critique notée  $\bar{W}$ . Elle est formée par l'ensemble des valeurs observées pour lesquelles l'hypothèse nulle  $H_0$  est acceptée. [3]

#### \* Les risques :

Aucun test d'hypothèse n'est fiable à 100% . Le test étant basé sur des probabilités, il existe toujours un risque de tirer une mauvaise conclusion. Lorsque vous effectuez un test d'hypothèse, vous pouvez rencontrer deux types d'erreurs : l'erreur de 1ère espèce (Type I) et l'erreur de 2ème espèce (Type II). Les risques de ces deux erreurs sont inversement proportionnels et sont déterminés par le seuil de signification et la puissance du test. Par conséquent, vous devez déterminer celle qui présente les conséquences les plus graves dans votre cas avant de définir le risque que vous acceptez pour chaque erreur.

#### -Le risque de première espèce

On appelle le risque de première espèce  $\alpha$  la probabilité sous  $H_0$  de la région de rejet. En d'autres termes, il s'agit de la probabilité que l'on accepte de décider  $H_1$  si la vérité est  $H_0$  [4]. Elle s'écrit :

$$\alpha = P(\text{rejeter } H_0 | H_0 \text{ vraie}) = P(\text{accepter } H_1 | H_0 \text{ vraie}) = P_{H_0}(H_1 | H_0)$$

## Chapitre 1. Test d'hypothèse

---

### Remarque 1.1.3.

Dans la plupart des situations que vous rencontrerez dans la suite, la région d'acceptation  $\bar{W}$  peut être reliée au risque d'erreur de première espèce  $\alpha$  par :

$$P_{H_0}[\bar{W}] = 1 - \alpha$$

-La probabilité complémentaire  $1 - \alpha$  est appelée *niveau de confiance*.

### -Le risque de deuxième espèce

On appelle le risque de deuxième espèce  $\beta$  la probabilité de ne pas rejeter  $H_0$  alors que la vérité est  $H_1$ . Il s'agit d'un risque qui n'est pas fixé a priori par le test, et il est souvent difficile à estimer. On prend ce risque lorsqu'on accepte l'hypothèse  $H_0$  [4]. Elle s'écrit :

$$\beta = P(\text{accepter } H_0 | H_1 \text{ vraie}) = P(\text{rejeter } H_1 | H_1 \text{ vraie}) = P_{H_0}(H_0 | H_1)$$

### Remarque 1.1.4.

Le risque  $\beta$  peut être alors évalué dans certains cas simples mais en général, les calculs sont difficiles.

-La probabilité complémentaire  $1 - \beta$  d'opter pour  $H_1$  en ayant raison et s'appelle **la puissance du test**.

Il y a donc quatre cas possibles qui sont détaillés dans le tableau ci-dessous :

	$H_0$ vraie	$H_1$ vraie
$H_0$ décidée	$1 - \alpha$	$\beta$
$H_1$ décidée	$\alpha$	$1 - \beta$

TABLE 1.1 – Le tableau représente les risques de première et deuxième espèce

### Lien entre les risques :

La situation idéale serait que ces deux erreurs soient nulles mais ce n'est pas possible.[3]

## Chapitre 1. Test d'hypothèse

---

-Pire encore, toutes autres choses étant fixées, ces deux erreurs sont antagonistes.

-Si une des deux erreurs diminue, l'autre augmente.

### 1.1.4 La statistique de test : variable de décision

La conception d'un test passe par l'identification d'une statistique  $T = h(X_1, \dots, X_n)$  de l'échantillon, appelée statistique de test, jouant le rôle de variable de décision .

#### Définition 1.1.1.

*La statistique de test est une variable aléatoire définie indépendamment des données observées. La valeur que prend cette variable aléatoire pour les données observées sera appelée statistique observée et notée  $Z_{obs}$  dans la suite.[4]*

- Elle doit apporter le maximum d'information sur le problème posé.
- Sa loi doit être différente sous  $H_0$  et sous  $H_1$ .
- Il faut que sa loi soit entièrement connue au moins sous  $H_0$ .

#### Remarque 1.1.5.

- *Concevoir une statistique de test n'est pas une question simple. La plupart des tests classiques reposent sur des statistiques dont l'identification a demandé à leurs auteurs beaucoup d'efforts et d'imagination.*
- *Une statistique de test n'a aucune raison d'être unique, et le choix entre plusieurs statistiques candidates est une question difficile, qui déterminera entre autre la puissance du test.*
- *La statistique de test choisie et sa distribution sous  $H_0$  dépendent généralement de la distribution de la population et/ou de la taille de l'échantillon. Il s'agit de préciser ces hypothèses et de les vérifier ultérieurement .*

#### • Le niveau de significativité

Depuis les travaux de Neyman et Pearson, l'erreur de première espèce est limitée à un niveau dit niveau de significativité.

Le fait d'imposer faible conduit à une règle de décision plus stricte.

## Chapitre 1. Test d'hypothèse

---

En effet, dans ce cas, la décision consiste à abandonner  $H_0$  dans des cas rarissimes, et à conserver, plus souvent, à tort  $H_0$ .

### Définition 1.1.2.

Plus grande valeur possible du risque de première espèce (parmi toutes les valeurs possibles sous  $H_0$ ) [6].

$$\alpha = \sup_{\theta \in \Theta_0} P_{\theta}(W)$$

Pour une hypothèse nulle simple :

Risque de première espèce = niveau de significativité .

### Remarque 1.1.6.

En pratique, seul le risque  $\alpha$  peut être parfaitement contrôlé. Par conséquent on le fixe à des valeurs classiques, qui n'ont aucune propriété particulière si ce n'est d'être classiques : 5%, 1% et 0.1% [7].

### 1.1.5 P-valeur (ou la probabilité critique)

#### Définition 1.1.3.

*P-valeur* (en anglais *P-value*) associée à un test donné. est La probabilité pour que la statistique de test dépasse sous l'hypothèse  $H_0$  la valeur seuil.

En d'autres termes, c'est la probabilité d'observer quelque chose d'au moins aussi surprenant que ce que l'on observe [4].

$$P = p(|Z| > z_{obs})$$

Le cas du test unilatéral se déduit facilement en supprimant les signes de valeur absolue. Plus cette probabilité est proche à 0, plus forte est la contradiction entre  $H_0$  et le résultat observé avec l'échantillon.

### 1.1.6 La décision

La décision d'un test consiste à choisir entre  $H_0$  et  $H_1$ .

-Il existe deux stratégies pour prendre une décision en ce qui concerne un test d'hypothèse : la première stratégie fixe à priori la valeur du seuil de signification  $\alpha$  et la seconde établit la valeur de la probabilité critique  $Z_{obs}$  à posteriori[5][3].

#### Règle de décision 1 :

Sous l'hypothèse «  $H_0$  est vraie » et pour un seuil de signification fixé.

- Si la valeur de la statistique  $Z_{obs}$  calculée appartient à la région critique alors l'hypothèse nulle est rejetée au risque d'erreur  $\alpha$  et l'hypothèse  $H_1$  est acceptée.
- Si la valeur de la statistique  $z_{obs}$  n'appartient pas à la région critique alors l'hypothèse  $H_0$  ne peut être rejetée.

#### Règle de décision 2 :

On peut comparer la p-valeur à  $\alpha$  plutôt que  $Z_{obs}$  et la région de rejet.

- La **p-valeur**  $> \alpha$ , il n'est pas exceptionnel sous  $H_0$  d'observer la valeur effectivement observée. Par conséquent,  $H_0$  est acceptée.
- Si la **p-valeur**  $< \alpha$ , la valeur observée est jugée exceptionnelle sous  $H_0$ . On décide alors de rejeter  $H_0$  et de valider  $H_1$ .

### 1.1.7 Puissance du test

#### Définition 1.1.4.

*Puissance du test notée  $(1 - \beta)$  est la Probabilité de rejeter  $H_0$  alors qu'elle est fausse.[6]*

$$P(H_1/H_1) = P(W/\theta \in \Theta_1) = 1 - \beta$$

Puissance + Risque de deuxième espèce = 1
---

#### Remarques importantes :

En général on connaît la distribution sous  $H_0$ , mais pas sous  $H_1$  :



- Formulation du test non symétrique par rapport à  $H_0$  et  $H_1$ .
- Si la statistique  $z_{obs}$  est dans la zone de rejet, on rejette  $H_0$  (et donc on accepte  $H_1$ ).
- Dans le cas contraire, on ne peut accepter définitivement l'hypothèse  $H_0$ . On ne peut qu'affirmer que les données ne permettent pas de la rejeter .

## 1.2 Mise en oeuvre pratique

### 1.2.1 La démarche d'un test

Après avoir clairement défini la question posée et le modèle statistique sous-jacent, une démarche de test suit généralement les étapes suivantes :

1. Choix de  $H_0$  et  $H_1$ .
2. Choix  $\alpha$ .
3. Détermination de la variable de décision (statistique de test) et de sa loi sous  $H_0$  (en fonction de  $H_0$  et  $H_1$ , des hypothèses sur la distribution de  $X$  et de la taille de l'échantillon prélevé).
4. Détermination d'une région critique en fonction de  $\alpha$ .
5. Décision (voir si les observations sont ou pas dans la région critique).
6. Calcul de la p-valeur si nécessaire.
7. Calcul éventuel de la puissance du test si possible.

#### Exemple 1.2.1. (essais thérapeutiques)

*Pour apaiser un certain type de maux de tête, on a l'habitude de traiter les malades avec un médicament A. Une étude statistique a montré que la durée de disparition de la douleur chez les malades traités avec A était une variable aléatoire de loi normale  $N(\mu_0, \sigma_0^2)$ , avec  $\mu_0 = 30$  mn et  $\sigma_0 = 5$  mn. Un laboratoire pharmaceutique a conçu un nouveau médicament B et désire tester son efficacité. Pour cela, le nouveau médicament a été administré à  $n$  malades cobayes, et on a mesuré la durée de disparition de la douleur pour chacun d'entre eux :  $x_1, \dots, x_n$ . Une étude de statistique descriptive sur ces données a amené les pharmacologues à considérer que cette durée était une variable aléatoire de loi normale  $N(\mu, \sigma^2)$ . [8]*

## Chapitre 1. Test d'hypothèse

---

### Remarque 1.2.1.

En toute rigueur, on ne devrait pas modéliser une durée (positive) par une variable aléatoire qui, comme pour la loi normale, peut prendre des valeurs négatives. En pratique, on peut le faire quand, pour les lois considérées, la probabilité que la variable soit négative est négligeable.

La conséquence du nouveau médicament se traduit sur la valeur de la durée moyenne de disparition de la douleur :

- “ $\mu = \mu_0$ ” : le médicament B a en moyenne le même conséquence que le médicament A.
- “ $\mu < \mu_0$ ” : le médicament B a en moyenne plus efficace que le médicament A.
- “ $\mu > \mu_0$ ” : le médicament B a en moyenne moins efficace que le médicament A.

Nous reviendrons ultérieurement sur l'interprétation de la valeur de l'écart-type  $\sigma$  en termes d'efficacité du médicament.

Pour savoir s'il faut commercialiser B, il faut trancher entre ces 3 hypothèses. L'important est de ne pas se tromper si on décide de changer de médicament : il est préférable de conserver un médicament moins performant que le nouveau que d'adopter un médicament moins performant que l'ancien. Il faut donc que l'hypothèse “ $\mu < \mu_0$ ” corresponde au rejet de  $H_0$ .

Par conséquent, ces hypothèses pouvaient se formaliser comme suit :

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0. \end{cases}$$

au vu de  $n$  réalisations indépendantes  $x_1, \dots, x_n$  de la loi  $N(\mu, \sigma^2)$ .

### Première idée(Olivier Gaudoin)

Une première idée est de conclure que  $\mu < \mu_0$  si et seulement si  $\bar{x}_n < \mu_0$  : la durée moyenne de disparition de la douleur sur les malades traités avec B est plus petite que ce qu'elle est sur les malades traités avec A. Cela revient à proposer comme région critique du test :

## Chapitre 1. Test d'hypothèse

---

$$W = \{(x_1, \dots, x_n); \bar{x}_n < \mu_0\}$$

Si  $\bar{x}_n$  est beaucoup plus petit que  $\mu_0$ , il est en effet très probable que  $B$  soit plus efficace que  $A$ . Mais si  $\bar{x}_n$  est proche de  $\mu_0$  tout en étant plus petit, on risque de se tromper si on affirme que  $\mu < \mu_0$ . La probabilité de cette erreur, qui n'est autre que le risque de première espèce  $\alpha$ , est très facile à calculer :

$$\begin{aligned}\alpha &= \sup_{H_0} \beta(\mu) = \sup_{\mu \geq \mu_0} P(\bar{X}_n < \mu_0; \mu) \\ &= \sup_{\mu \geq \mu_0} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}; \mu\right) = \sup_{\mu \geq \mu_0} \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)\end{aligned}$$

où  $\Phi$  est la fonction de répartition de la loi normale centrée-réduite. En effet, comme on l'a déjà vu, si  $X_1, \dots, X_n$  sont indépendantes et de même loi  $N(\mu; \sigma^2)$ , alors  $\bar{X}_n$  est de loi  $N\left(\mu, \frac{\sigma^2}{n}\right)$  et  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$  est de loi  $N(0, 1)$ .

$\Phi(m)$  est une fonction croissante de  $m$ , donc  $\beta(\mu) = \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$  est une fonction décroissante de  $\mu$ . Par conséquent,  $\alpha = \sup_{\mu \geq \mu_0} \beta(\mu) = \beta(\mu_0) = \Phi(0) = 1/2$ .

Il y a donc une chance sur deux de se tromper si on décide que  $B$  est plus efficace que  $A$  quand  $\bar{x}_n < \mu_0$ . C'est évidemment beaucoup trop.

### Deuxième idée(Olivier Gaudoin)

On voit qu'il faut en fait rejeter  $H_0$  quand  $\bar{x}_n$  est significativement plus petit que  $\mu_0$ . Cela revient à prendre une région critique de la forme :

$$W = \{(x_1, \dots, x_n); \bar{x}_n < l_\alpha\}, \text{ où } l_\alpha < \mu_0$$

La borne  $l_\alpha$  dépend du seuil  $\alpha$  que l'on s'est fixé. Moins on veut risquer de rejeter à tort  $H_0$ , plus  $\alpha$  sera petit, et plus  $l_\alpha$  sera petit. Le sens de l'expression significativement plus petit est lié à la valeur de  $\alpha$ .

Un calcul analogue au précédent montre que :

$$\alpha = \sup_{H_0} \beta(\mu) = \sup_{\mu \geq \mu_0} P(\bar{X}_n < l_\alpha; \mu) = \sup_{\mu \geq \mu_0} \Phi\left(\frac{l_\alpha - \mu}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{l_\alpha - \mu_0}{\sigma/\sqrt{n}}\right)$$

## Chapitre 1. Test d'hypothèse

---

On obtient donc

$$\frac{l_\alpha - \mu_0}{\sigma/\sqrt{n}} = \Phi^{-1}(\alpha), \quad \text{d'où } l_\alpha = \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha) = \mu_0 - \frac{\sigma}{\sqrt{n}}m_{2\alpha},$$

avec les notations habituelles pour les quantiles de la loi normale.

En conclusion, on a :

**Proposition 1.2.1.** *Un test de seuil  $\alpha$  de  $H_0 : \mu \geq \mu_0$  contre  $H_1 : \mu < \mu_0$  est déterminé par la région critique :*

$$W = \{(x_1, \dots, x_n); \bar{x}_n < \mu_0 - \frac{\sigma}{\sqrt{n}}m_{2\alpha}\}$$

### Troisième idée(Olivier Gaudoin)

La région critique proposée ci-dessus pose un problème déjà rencontré à propos des intervalles de confiance : ce test est inutilisable si on ne connaît pas la vraie valeur de  $\sigma$ , ce qui est toujours le cas en pratique. Pour pallier cet inconvénient, on utilise la même procédure que pour les intervalles de confiance : on remplace  $\sigma$  par son estimateur  $S'_n$ , ce qui nécessite de remplacer la loi normale par la loi de Student.

Rappelons en effet que si  $X_1, \dots, X_n$  sont indépendantes et de même loi  $N(\mu, \sigma^2)$ , alors  $\frac{\bar{X}_n - \mu}{S'_n/\sqrt{n}}$  est de loi  $St(n-1)$ . Alors, à partir d'une région critique de la forme

$$W = \{(x_1, \dots, x_n); \bar{x}_n < l_\alpha\}, \quad \text{on obtient :}$$

$$\begin{aligned} \alpha &= \sup_{H_0} \beta(\mu) = \sup_{\mu \geq \mu_0} P(\bar{X}_n < l_\alpha; \mu) = \sup_{\mu \geq \mu_0} P\left(\frac{\bar{X}_n - \mu}{S'_n/\sqrt{n}} < \frac{l_\alpha - \mu}{S'_n/\sqrt{n}}; \mu\right) \\ &= \sup_{\mu \geq \mu_0} F_{St(n-1)}\left(\frac{l_\alpha - \mu}{S'_n/\sqrt{n}}\right) = F_{St(n-1)}\left(\frac{l_\alpha - \mu_0}{S'_n/\sqrt{n}}\right) \end{aligned}$$

D'où  $\frac{l_\alpha - \mu_0}{S'_n/\sqrt{n}} = F_{St(n-1)}^{-1}(\alpha) = -t_{n-1, 2\alpha}$ , avec les notations habituelles pour les quantiles

de la loi de Student. Finalement,  $l_\alpha = \mu_0 - \frac{S'_n}{\sqrt{n}}t_{n-1, 2\alpha}$ .

En conclusion, on a :

## Chapitre 1. Test d'hypothèse

---

**Proposition 1.2.2.** *Un test de seuil  $\alpha$  de  $H_0 : \mu \geq \mu_0$  contre  $H_1 : \mu < \mu_0$  est déterminé par la région critique :*

$$W = \{(x_1, \dots, x_n); \bar{x}_n < \mu_0 - \frac{s'_n}{\sqrt{n}} t_{n-1, 2\alpha}\}$$

**Remarque 1.2.2.** *La région critique peut aussi s'écrire :*

$$W = \{(x_1, \dots, x_n); \frac{\bar{x}_n - \mu_0}{s'_n/\sqrt{n}} < -t_{n-1, 2\alpha}\}$$

*Cette forme met en évidence l'utilisation de la variable aléatoire  $\frac{\bar{X}_n - \mu_0}{S'_n/\sqrt{n}}$  qui n'est autre que la fonction pivotale déjà vue à l'occasion des intervalles de confiance. C'est cette forme que l'on conservera dans la suite.*

### Application

Avec le médicament A, la durée moyenne de disparition de la douleur était 30 mn. On a administré le médicament B à 12 malades et relevé les durées de disparition de la douleur suivants :

25 28 20 32 17 24 41 28 25 30 27 24

La moyenne empirique de ces données est  $\bar{x}_n = 26.75$  et l'écart-type estimé est  $s'_n = 6.08$ . On décide de ne commercialiser B que si on est sûr à 95% qu'il est plus efficace que A. Cela revient donc à faire un test de  $H_0 : \mu \geq 30$  contre  $H_1 : \mu < 30$  au seuil  $\alpha = 5\%$ . On voit qu'il s'agit finalement de déterminer si 26.75 est suffisamment inférieur à 30 pour que l'on puisse conclure que le médicament B réduit vraiment la durée de disparition de la douleur.

D'après ce qui précède, on rejettera  $H_0$  Si  $\frac{\bar{x}_n - \mu_0}{s'_n/\sqrt{n}} < -t_{n-1, 2\alpha}$ .

$$\text{Or } \frac{\bar{x}_n - \mu_0}{s'_n/\sqrt{n}} = \frac{26.75 - 30}{6.08/\sqrt{12}} = -1.853 \text{ et } t_{n-1, 2\alpha} = t_{11, 0.1} = 1.796.$$

$-1.853 < -1.796$ , donc les observations sont dans la région critique. On rejette donc  $H_0$ , ce qui signifie que l'on conclut que B est plus efficace que A, avec moins de 5% de chances

## Chapitre 1. Test d'hypothèse

---

de se tromper. Par conséquent, on peut lancer la commercialisation du médicament B.

Si on avait pris  $\alpha = 1\%$ , on aurait eu  $t_{11,0.02} = 2.718$ . Comme  $-1.853 > -2.718$ , on n'aurait pas rejeté  $H_0$ , donc on n'aurait pas adopté le médicament B.

Ce phénomène est normal : se fixer un seuil  $\alpha$  petit revient à éviter au maximum d'adopter à tort le médicament B. Or un bon moyen de ne pas prendre ce risque, c'est de conserver le médicament A. Le test de seuil  $\alpha = 0$  consiste à conserver le médicament A quelles que soient les observations : la probabilité de rejeter à tort  $H_0$  est nulle quand on ne rejette jamais  $H_0$ .

Il est donc fondamental de bien savoir évaluer les risques et de choisir  $\alpha$  en connaissance de cause.

Cet exemple avec  $\alpha = 1\%$  permet également de comprendre la nuance entre "ne pas rejeter  $H_0$ " et "accepter  $H_0$ " : on va conclure que rien ne prouve que B est plus efficace que A, mais on ne va évidemment pas conclure que A est plus efficace que B.

La remarque précédente met en évidence l'existence d'un seuil critique  $\alpha_c$  tel que pour tout seuil  $\alpha$  supérieur à  $\alpha_c$ , on rejettera  $H_0$ , et pour tout seuil  $\alpha$  inférieur à  $\alpha_c$ , on ne rejettera pas  $H_0$ . Cette valeur  $\alpha_c$  est la p-valeur.

$\alpha_c$  vérifie  $\frac{\bar{x}_n - \mu_0}{s'_n/\sqrt{n}} < -t_{n-1,2\alpha_c}$ . Sur l'exemple, la table de la loi de Student permet de constater que

$-t_{11,0.05} = -2.201 < -1.853 < -1.796 = -t_{11,0.1}$ . On en déduit que  $5\% < 2\alpha_c < 10\%$ , d'où  $2.5\% < \alpha_c < 5\%$ .

Pour calculer exactement la p-valeur, on écrit :

$$-t_{n-1,2\alpha_c} = F_{St(n-1)}^{-1}\left(\frac{2\alpha_c}{2}\right) = F_{St(n-1)}^{-1}(\alpha_c) \implies \alpha_c = F_{St(n-1)}\left(\frac{\bar{x}_n - \mu_0}{s'_n/\sqrt{n}}\right)$$

et on obtient ici  $\alpha_c = 0.04547$ . La réponse à un test de seuil fixé est binaire : on rejette ou on ne rejette pas  $H_0$ . Fournir une p-valeur est une réponse plus riche puisqu'elle permet de connaître le résultat du test pour n'importe quel choix du seuil. C'est pourquoi le traitement des tests d'hypothèses par les logiciels de statistique consiste à fournir des p-valeurs.

En R, la commande permettant d'effectuer un test sur la moyenne d'une loi normale est

## Chapitre 1. Test d'hypothèse

---

t.test. L'option alternative permet de préciser lequel du test bilatéral et des deux tests unilatéraux on choisit. Sur l'exemple, on obtient :

```
> medic < -c(25, 28, 20, 32, 17, 24, 41, 28, 25, 30, 27, 24)
> t.test(medic, alternative="less", mu = 30)
```

### One Sample $t$ - test

```
data : medic
t = -1.8526, df = 11, p - value = 0.04547
alternative hypothesis : true mean is less than 30
95 percent confidence interval :
```

```
 -Inf 29.90056
```

```
sample estimates :
```

```
mean of x
```

```
26.75
```

La p-valeur est ici  $\alpha_c = 4.5\%$ . Cela signifie que, pour tout seuil supérieur à  $4.5\%$  (c'est le cas de  $5\%$ ), on rejettera  $H_0$ , donc on conclura que B est plus efficace que A, et pour tout seuil inférieur à  $4.5\%$  (c'est le cas de  $1\%$ ), on ne rejettera pas  $H_0$ , donc on conclura que B n'est pas plus efficace que A.

De manière générale, plus la p-valeur est petite, moins on prend de risque en rejetant  $H_0$ .

## Conclusion

Dans ce chapitre nous avons rappelé quelques définition de base de test d'hypothèse pour pouvoir les utiliser dans les chapitres prochains.

# Chapitre 2

## Test séquentiel

Ce chapitre a pour but, de présenter le test séquentiel du rapport des probabilités introduit par Wald en 1943 et d'en donner les principales propriétés.

Nous considérerons  $(X_1, \dots, X_n)$  un échantillon de variables aléatoires réelles indépendantes définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et identiquement distribuées dont on notera  $x_1, \dots, x_n$  les réalisations.

La loi de  $X_1$  dépend d'un paramètre  $\theta \in \Theta \subset \mathbb{R}$  et est donnée par  $f(x, \theta)$  qui désigne la densité de  $X_1$  au point  $x$  pour le paramètre  $\theta$  si  $X_1$  est absolument continue ou la probabilité que  $X_1$  prenne la valeur  $x$  pour le paramètre  $\theta$  si  $X_1$  est une variable aléatoire discrète.

On veut alors tester l'hypothèse  $H_0 : \theta = \theta_0$  contre l'hypothèse  $H_1 : \theta = \theta_1$ , où  $f(x, \theta)$  est la densité de  $X_1$  (ou probabilité que  $X_1$  soit égale à  $x$ ) lorsque l'hypothèse  $H_0$  est vraie. De la même manière,  $f(x, \theta_1)$  désigne la densité de  $X_1$  (ou probabilité que  $X_1$  soit égale à  $x$ ) lorsque  $H_1$  est vraie.

Enfin, pour tout évènement  $E \in \mathcal{A}$ , on notera  $\mathbb{P}_\theta(E) = \mathbb{P}(E|\theta \text{ est le vrai paramètre})$ .



### 2.1 Expériences séquentielles

L'expérimentation séquentielle est une branche de la statistique qui, en plus de son importance pratique, présente un grand intérêt théorique.

#### 2.1.1 Définition du test séquentiel du rapport des probabilités

Le test séquentiel du rapport des probabilités permet de choisir entre deux hypothèses simples. Supposons nous cherchions à tester l'hypothèse  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta = \theta_1$

Lorsque l'échantillon a une taille fixe (test non séquentiel), Neyman et Pearson ont fourni la solution optimale. Ils ont montré que pour une taille donnée  $n$  on pouvait obtenir le test plus puissant en considérant le rapport de vraisemblance  $l_m$  :

$$\begin{aligned} l_m &= \frac{\text{Probabilité des résultats observés si } \theta_1 \text{ est vrai}}{\text{Probabilité des résultats observés si } \theta_0 \text{ est vrai}} \\ &= \prod_{i=1}^m \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)}, \end{aligned}$$

Le test conduira à choisir ou non  $\theta_0$  selon que  $l_m$  est inférieur ou supérieur à une certaine constante.[10]

#### Définition 2.1.1.

*Le test séquentiel du rapport des probabilités (ou TSRP)  $S(A, B)$  de Wald [10] est analogue au précédent et possède une propriété d'optimalité de même nature. Il est défini comme suit :*

On se donne deux constantes positives  $A$  et  $B$  généralement telles que  $B < 1 < A$  et à chaque étape  $m$  de l'expérience, on calcule le rapport des probabilités  $l_m$  et on adopte la règle de décision suivante :

- Si  $B < l_m < A$  on continue l'expérience en prenant une nouvelle observation.
- Si  $l_m \geq A$  on rejette  $H_0$  ( on accepte  $H_1$  ).
- Si  $l_m \leq B$  on accepte  $H_0$ .

De manière générale, il est pratique pour les calculs de prendre le logarithme du rapport des vraisemblances  $l_m$ . On a :

## Chapitre 2. Test Séquentiel

---

$$\log l_m = \sum_{i=1}^m \log \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)},$$

Avec :

$$Z_i = \log \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)}. \quad (2.1)$$

Il est alors possible de décrire la règle de décision du test séquentiel à l'étape  $m$  de la manière suivante :

- $\log B < \sum_{i=1}^m Z_i < \log A$ , alors on continue l'échantillonnage.
- $\sum_{i=1}^m Z_i \geq \log A$ , on rejette  $H_0$ .
- $\sum_{i=1}^m Z_i \leq \log B$ , on accepte  $H_0$ .

### Remarque 2.1.1.

- Si  $f(X_i, \theta_0) = f(X_i, \theta_1) = 0$ , alors on pose  $l_m = 1$ .
- Si  $f(X_i, \theta_1) > 0$  et  $f(X_i, \theta_0) = 0$  alors  $l_m \geq A$  et on termine par conséquent le test à l'étape  $m$  par le rejet de  $H_0$ . [10]

### Les régions du test séquentiel du rapport des probabilités :

De façon plus formelle, on peut définir trois régions à chaque étape  $m$  de la procédure séquentielle :

- $D_m^1$  la région de rejet.
- $D_m^0$  la région d'acceptation.
- Et  $D_m$  la région de continuation.

L'ensemble des échantillons (d'observations) de  $D_m$  conduisent à continuer la procédure séquentielle. [10]

D'où  $(x_1, \dots, x_m) \in D_m$  si :

$$\forall n \leq m \quad B < l_n < A$$

De même  $D_m^1$  est l'ensemble des échantillons (d'observations) conduisant à décider  $H_1$ .

Alors  $(x_1, \dots, x_m) \in D_m^1$  si :

$$\forall \quad n < m \quad B < l_n < A \text{ et } l_m \geq A$$

## Chapitre 2. Test Séquentiel

---

enfin  $D_m^0$  est l'ensemble des échantillons (d'observations) conduisant à décider  $H_0$ . Alors  $(x_1, \dots, x_m) \in D_m^0$  si :

$$\forall n < m \quad B < l_n < A \text{ et } l_m \leq B$$

Soit  $\theta \in \Theta$  fixé. Pour tout test séquentiel du rapport des probabilités et tout  $n \geq 1$ , on peut écrire :

$$\begin{aligned} & \mathbb{P}_\theta(\text{Continuer la procédure à l'étape } n) + \mathbb{P}_\theta(\text{Décider } H_0 \text{ avant ou à l'étape } n) \\ & + \mathbb{P}_\theta(\text{Décider } H_1 \text{ avant ou à l'étape } n) = 1 \end{aligned}$$

qui en termes de  $D_m^1$ ,  $D_m^0$  et  $D_m$ , s'écrit :

$$\mathbb{P}_\theta((X_1, \dots, X_n) \in D_m) + \sum_{m=1}^n \mathbb{P}_\theta((X_1, \dots, X_m) \in D_m^1) + \sum_{m=1}^n \mathbb{P}_\theta((X_1, \dots, X_m) \in D_m^0) = 1 \quad (2.2)$$

Par continuité inférieure de la probabilité, on obtient que :

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P}_\theta((X_1, \dots, X_n) \in D_m) &= \mathbb{P}_\theta\left(\bigcap_{m \geq 1} \{(X_1, \dots, X_m) \in D_m\}\right) \\ &= \mathbb{P}_\theta(\text{le test ne s'arrête pas}) \end{aligned}$$

et par  $\sigma$ -additivité on en déduit que

$$\begin{aligned} \sum_{m \geq 1} \mathbb{P}_\theta((X_1, \dots, X_m) \in D_m^0) &= \mathbb{P}_\theta\left(\bigcup_{m \geq 1} \{(X_1, \dots, X_m) \in D_m^0\}\right) \\ &= \mathbb{P}_\theta(\text{décider } H_0) \end{aligned}$$

et

$$\begin{aligned} \sum_{m \geq 1} \mathbb{P}_\theta((X_1, \dots, X_m) \in D_m^1) &= \mathbb{P}_\theta\left(\bigcup_{m \geq 1} \{(X_1, \dots, X_m) \in D_m^1\}\right) \\ &= \mathbb{P}_\theta(\text{décider } H_1) \end{aligned}$$

On peut alors passer à la limite quand  $n \rightarrow +\infty$  dans (2.2) et on obtient la relation fondamentale :

$$\mathbb{P}_\theta(\text{le test ne s'arrête pas}) + \mathbb{P}_\theta(\text{Décider } H_1) + \mathbb{P}_\theta(\text{Décider } H_0) = 1. \quad (2.3)$$

## Chapitre 2. Test Séquentiel

---

Nous allons démontrer que, pour tout  $\theta \in \Theta$ , le test séquentiel du rapport des probabilités s'arrête en un nombre fini d'observations avec une probabilité égale à 1. On pourra alors en déduire que

$$\mathbb{P}_\theta(\text{Décider } H_1) + \mathbb{P}_\theta(\text{Décider } H_0) = 1. \quad (2.4)$$

### Remarque 2.1.2.

*Tous les tests séquentiels ne s'arrêtent pas en un nombre fini d'observations avec une probabilité égale à 1. Il s'agit d'une particularité du TSRP.[10]*

### 2.1.2 Détermination des constantes A et B du TSRP

Afin de pouvoir définir correctement le test séquentiel du rapport des probabilités  $S(A,B)$ , il faut choisir les constantes A et B.

On se fixe un couple  $(\alpha, \beta)$  avec  $0 < \alpha < 1$  et  $0 < \beta < 1$  et  $\alpha + \beta < 1$  et les constantes A et B sont alors, en théorie, déterminées de façon à ce que le test ait pour erreurs de première et deuxième espèces respectivement.

Cependant, en pratique, il est difficile de calculer exactement les constantes A et B. C'est pourquoi nous allons d'abord nous intéresser aux relations qui existent entre A, B,  $\alpha$  et  $\beta$  et ensuite nous chercherons à obtenir une approximation de ces constantes.[10]

#### Relations entre A, B, $\alpha$ et $\beta$

##### Théorème 2.1.1. (Wald 1943)

*Soit un test séquentiel du rapport des probabilité  $S(A,B)$  d'erreurs de première et deuxième espèces respectivement  $\alpha$  et  $\beta$ . Alors :*

$$A \leq \frac{1 - \beta}{\alpha} \text{ et } B \geq \frac{\beta}{1 - \alpha}. \quad (2.5)$$

##### Démonstration 2.1.1.

*Lorsqu'un échantillon d'observations  $(x_1, \dots, x_m)$  conduit à décider  $H_0$ , on dit qu'il est de type 0. De même lorsqu'un échantillon conduit à décider  $H_1$ , on dit qu'il est de type 1.*

*Nous noterons alors  $D_j$ ,  $j = 0, 1$ , l'ensemble des échantillons de type  $j$ .*

*Par conséquent  $(x_1, \dots, x_m) \in D_0$  si*

$$B < l_n(x_1, \dots, x_n) < A \quad \forall n < m \text{ et } l_m(x_1, \dots, x_m) \leq B.$$

## Chapitre 2. Test Séquentiel

---

De même  $(x_1, \dots, x_m) \in D_1$  si

$$B < l_n(x_1, \dots, x_n) < A \quad \forall n < m \text{ et } l_m(x_1, \dots, x_m) \geq A.$$

On a donc

$$D_1 = \bigcup_{m \geq 1} D_m^1 \text{ et } D_0 = \bigcup_{m \geq 1} D_m^0.$$

où  $D_m^1$  et  $D_m^0$  ont été définis comme précédents.

- Considérons  $(x_1, \dots, x_m)$  un échantillon de  $D_1$ , alors

$$\begin{aligned} l_m &\geq A \\ \Leftrightarrow \prod_{i=1}^m \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} &\geq A \\ \Leftrightarrow \prod_{i=1}^m f(x_i, \theta_1) &\geq A \prod_{i=1}^m f(x_i, \theta_0). \end{aligned} \quad (2.6)$$

Si on intègre (2.6) sur  $D_{1,m}$  on obtient :

$$\int_{D_{1,m}} \prod_{i=1}^m f(x_i, \theta_1) dx_1 \dots dx_m \geq A \int_{D_{1,m}} \prod_{i=1}^m f(x_i, \theta_0) dx_1 \dots dx_m$$

et en sommant sur  $m$  on obtient que

$$\sum_{m \geq 1} \int_{D_{1,m}} \prod_{i=1}^m f(x_i, \theta_1) dx_1 \dots dx_m \geq A \sum_{m \geq 1} \int_{D_{1,m}} \prod_{i=1}^m f(x_i, \theta_0) dx_1 \dots dx_m \quad (2.7)$$

D'autre part,

$$\begin{aligned} 1 - \beta &= \mathbb{P}(\text{Accepter } H_1 | H_1 \text{ vraie}) \\ &= \mathbb{P}_{\theta_1}(\exists m \text{ tel que } (X_1, \dots, X_m) \in D_{1,m}) \\ &= \sum_{m \geq 1} \mathbb{P}_{\theta_1}((X_1, \dots, X_m) \in D_{1,m}) \text{ par } \sigma\text{-additivité} \\ &= \sum_{m \geq 1} \int_{D_{1,m}} \prod_{i=1}^m f(x_i, \theta_1) dx_1 \dots dx_m. \end{aligned} \quad (2.8)$$

De la même manière, on peut montrer que

$$\alpha = \mathbb{P}(\text{Accepter } H_1 | H_0 \text{ vraie})$$

## Chapitre 2. Test Séquentiel

---

$$= \sum_{m \geq 1} \int_{D_{1,m}} \prod_{i=1}^m f(x_i, \theta_0) dx_1 \dots dx_m. \quad (2.9)$$

Enfin, en utilisant (2.8) et (2.9), on peut réécrire (2.7) comme suit :

$$1 - \beta \geq A\alpha.$$

$$D'où  $A \leq \frac{1 - \beta}{\alpha}$ .$$

- Pour démontrer l'inégalité  $B \geq \frac{\beta}{1 - \alpha}$ , on procède de la même manière en considérant les échantillons de type 0.

Ces inégalités s'avèrent très importantes car elles vont permettre d'éviter le calcul des constantes du test et fournir une approximation suffisante en pratique.

### Remarque 2.1.3.

- ▷ Lorsque l'on écrit  $\mathbb{P}(\text{Accepter } H_1 | H_1 \text{ vraie}) = 1 - \beta$  on utilise le fait que le test séquentiel s'arrête en un nombre ni d'observations avec une probabilité égale à 1. Plus précisément, on utilise la relation (2.4) pour  $\theta = \theta_1$ .
- ▷ De même en utilisant la relation (2.4) pour  $\theta = \theta_0$ , on obtient

$$\mathbb{P}(\text{Accepter } H_0 | H_0 \text{ vraie}) = 1 - \alpha.$$

### Approximation de A et B

Les inégalités (2.5) du théorème 2.1.1, suggèrent d'approximer A et B respectivement par

$$A^* = \frac{1 - \beta}{\alpha}$$

et

$$B^* = \frac{\beta}{1 - \alpha}$$

## Chapitre 2. Test Séquentiel

---

### Remarque 2.1.4.

- Les inégalités (2.5) impliquent  $B^* < A^*$  car  $B < A$ . En multipliant cette inégalité par  $\alpha(1 - \alpha) > 0$ , on obtient après simplification  $0 < 1 - \beta - \alpha$ .

On en déduit finalement que  $A^* > 1$  et  $B^* < 1$ . [10]

- Les risques de première et deuxième espèces pouvant être modifiés par l'approximation, notons  $\alpha^*$  et  $\beta^*$  les erreurs de première et deuxième espèces du test  $S(A^*, B^*)$  et analysons les conséquences de cette approximation. [10]

### Théorème 2.1.2. (Wald 1943)

Les risques du test  $S(A^*, B^*)$  vérifient :

$$\alpha^* \leq \frac{\alpha}{1 - \beta} \quad \text{et} \quad \beta^* \leq \frac{\beta}{1 - \alpha}$$

mais aussi

$$\alpha^* + \beta^* \leq \alpha + \beta$$

### Démonstration 2.1.2.

En appliquant le théorème 2.1.1 au test  $S(A^*, B^*)$ , on obtient

$$\frac{1}{A^*} \geq \frac{\alpha^*}{1 - \beta^*}$$

$$B^* \geq \frac{\beta^*}{1 - \alpha^*}$$

C'est-à-dire

$$\frac{\alpha^*}{1 - \beta^*} \leq \frac{\alpha}{1 - \beta}, \quad (2.10)$$

$$\frac{\beta^*}{1 - \alpha^*} \leq \frac{\beta}{1 - \alpha}. \quad (2.11)$$

Ce qui implique

$$\alpha^* \leq \frac{\alpha}{1 - \beta},$$

## Chapitre 2. Test Séquentiel

---

$$\beta^* \leq \frac{\beta}{1 - \alpha}.$$

-Si on multiplie (2.10) par  $(1 - \beta)(1 - \beta^*)$  et (2.11) par  $(1 - \alpha)(1 - \alpha^*)$ ; il en découle que

$$(1 - \beta)\alpha^* \leq (1 - \beta^*)\alpha,$$

$$(1 - \alpha)\beta^* \leq (1 - \alpha^*)\beta.$$

En additionnant ces deux inégalités, on obtient après simplification :

$$\alpha^* + \beta^* \leq \alpha + \beta.$$

Dans la pratique, si  $\alpha$  et  $\beta$  sont choisis dans l'intervalle  $[0.01; 0.05]$ , alors les proportions dans lesquelles  $\alpha^*$  peut excéder  $\alpha$  et  $\beta^*$  peut excéder  $\beta$  sont très faibles (car  $\frac{1}{1-\beta}$  et  $\frac{1}{1-\alpha}$  sont proches de 1 ) et peuvent donc être négligées.

### Exemple 2.1.1.

Si  $\alpha = 0.01$  et  $\beta = 0.05$ , alors  $\alpha^* \leq 0.01053$  et  $\beta^* \leq 0.05051$ .

La deuxième inégalité du théorème 2.1.2 montre que nécessairement une des deux inégalités  $\alpha^* \leq \alpha$  et  $\beta^* \leq \beta$  est réalisée. En conséquence, l'approximation de  $A$  par  $A^*$  et de  $B$  par  $B^*$  conduit au plus à l'augmentation d'un des risques et cela dans une proportion très faible.

Cependant cette approximation accroît le nombre d'observations nécessaires au test pour prendre une décision. Mais il est possible de montrer que l'augmentation du nombre d'observations causée par l'approximation est relativement faible comme l'illustre la figure 2.1

En pratique si le coût d'une nouvelle observation n'est pas trop élevé, si l'on désire obtenir un test séquentiel d'erreur de première espèce inférieure à  $\alpha$  et d'erreur de deuxième espèce inférieure à  $\beta$ , on pose  $A = \frac{1 - \beta}{\alpha}$ ,  $B = \frac{\beta}{1 - \alpha}$  et on effectue le test de la définition 2.1.1.[10]



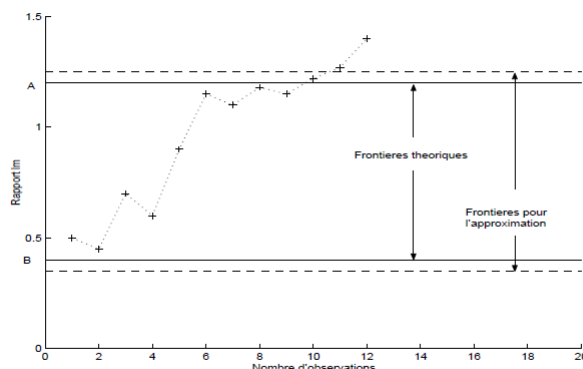


FIGURE 2.1 – Augmentation du nombre d’observations

### 2.1.3 Mise en oeuvre du test séquentiel sur un exemple

On suppose que l’on dispose d’un échantillon de variables aléatoires  $X_1, \dots, X_n$  indépendantes et de même loi  $\mathcal{N}(\theta, \sigma^2)$  avec  $\sigma$  connu. On veut tester au moyen du TSRP, l’hypothèse  $H_0 : \theta = \theta_0$  contre l’hypothèse  $H_1 : \theta = \theta_1$  avec  $\theta_1 > \theta_0$ . [10]

Si l’on souhaite obtenir un test tel que

$$\mathbb{P}(\text{rejeter } H_0 | H_0) \leq \alpha \text{ et } \mathbb{P}(\text{accepter } H_0 | H_1) \leq \beta$$

, où  $\alpha, \beta \in [0, 01; 0, 05]$ , on utilise l’approximation  $A = \frac{1 - \beta}{\alpha}$  et  $B = \frac{\beta}{1 - \alpha}$ .

A l’issue de la  $m$ -ième observation :

- on accepte  $H_0$  Si  $\sum_{i=1}^m Z_i \leq \log B$ ,
- on accepte  $H_1$  Si  $\sum_{i=1}^m Z_i \geq \log A$ ,
- on continue le test si  $\log B < \sum_{i=1}^m Z_i < \log A$ ,

avec  $Z_i$  défini par la relation (2.1).

Or

$$Z_i = \frac{\theta_1 - \theta_0}{\sigma^2} X_i - \frac{1}{2\sigma^2} (\theta_1^2 - \theta_0^2).$$

Donc

## Chapitre 2. Test Séquentiel

---

$$\sum_{i=1}^m Z_i = \frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^m X_i - \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2).$$

Pour résumer :

-On accepte  $H_0$  Si

$$\frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^m X_i - \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2) \leq \log \frac{\beta}{1 - \alpha}$$

$$\Leftrightarrow \sum_{i=1}^m X_i \leq \frac{\sigma^2}{\theta_1 - \theta_0} \left[ \log \frac{\beta}{1 - \alpha} + \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2) \right]$$

$$\Leftrightarrow \sum_{i=1}^m X_i \leq \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{\beta}{1 - \alpha} + n \frac{\theta_1 - \theta_0}{2} \stackrel{\text{déf}}{=} a_m$$

- on accepte  $H_1$  si

$$\frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^m X_i - \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2) \geq \log \frac{\beta}{1 - \alpha}$$

$$\Leftrightarrow \sum_{i=1}^m X_i \geq \frac{\sigma^2}{\theta_1 - \theta_0} \left[ \log \frac{1 - \beta}{\alpha} + \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2) \right]$$

$$\Leftrightarrow \sum_{i=1}^m X_i \geq \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{1 - \beta}{\alpha} + n \frac{\theta_1 - \theta_0}{2} \stackrel{\text{déf}}{=} r_m$$

-on prend une observation supplémentaire si

$$a_m < \sum_{i=1}^m X_i < r_m.$$

## Chapitre 2. Test Séquentiel

---

En pratique on trace les deux droites parallèles :

$$L_0 : y = \frac{\theta_1 + \theta_0}{2}x + \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{\beta}{1 - \alpha},$$

$$L_1 : y = \frac{\theta_1 + \theta_0}{2}x + \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{1 - \beta}{\alpha}.$$

Ensuite on place sur le même graphique les points de coordonnées  $(m, \sum_{i=1}^m x_i)$  tant qu'ils sont dans la bande de plan définie par les deux droites et dès que l'un des points sort de la bande de plan, on décide  $H_0$  si  $L_0$  est franchie et  $H_1$  si  $L_1$  est franchie.

### Application :

Nous avons simulé des données selon plusieurs valeurs du paramètre  $\theta$  et appliqué le test séquentiel décrit ci-dessus. Les figures Fig. 2.2 et Fig. 2.3 représentent les résultats obtenus respectivement pour  $\theta = 1$  et  $\theta = 2$  lorsque  $\theta_1 = 2$ ,  $\theta_0 = 1$ ,  $\sigma = 1$ ,  $\alpha = 0,01$  et  $\beta = 0,05$ . [10]

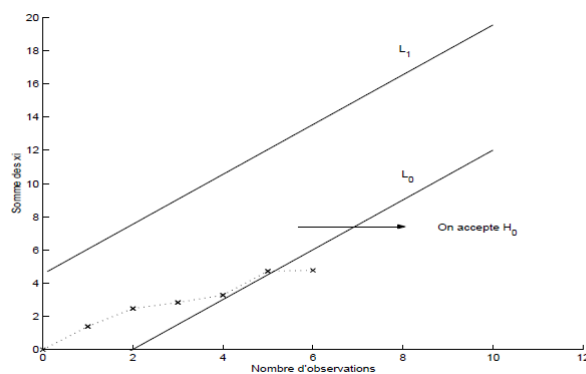


FIGURE 2.2 – Représentation du rejet de  $H_0(\theta = 1)$

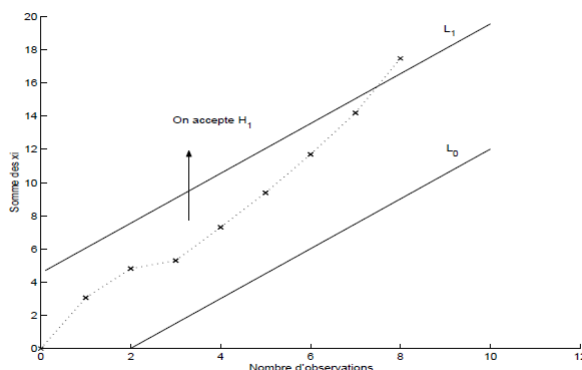


FIGURE 2.3 – Représentation de l'acceptation de  $H_0(\theta = 2)$

## 2.2 Propriétés importantes du TSRP

### 2.2.1 La courbe d'efficacité

Dans la mesure où le test séquentiel du rapport des probabilités pour tester une hypothèse  $H_0 : \theta = \theta_0$  contre une hypothèse  $H_1 : \theta = \theta_1$  sera appliqué à des valeurs du paramètre  $\theta \neq \theta_0$  et  $\theta \neq \theta_1$ , il est intéressant de calculer la fonction d'efficacité car elle est définie pour toutes les valeurs possibles de  $\theta$ . [9]

#### Définition 2.2.1.

La fonction d'efficacité du test séquentiel du rapport des probabilités est la fonction :

$$\theta \mapsto P(\theta) = \mathbb{P}_\theta(\text{Accepter } H_0)$$

#### Théorème 2.2.1. (Wald 1943)

On peut donner, au moins en première approximation, la relation suivante pour définir la fonction d'efficacité

$$P(\theta) = P(\theta/\alpha, \beta, \theta_0, \theta_1) \simeq \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

$P(\theta/\alpha, \beta, \theta_0, \theta_1)$  est la probabilité de choisir  $\theta_0$ , et  $h(\theta)$  est la solution non nulle, si elle existe, zéro sinon de

$$\sum_x \left[ \frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta) = 1$$

## Chapitre 2. Test Séquentiel

---

si  $f(x, \theta)$  est discrète et de

$$\int_{-\infty}^{+\infty} \left[ \frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta) dx = 1$$

si  $f(x, \theta)$  est continue.

### Démonstration 2.2.1.

On prend

$$A = \frac{1 - \beta}{\alpha} \quad \text{et} \quad B = \frac{\beta}{1 - \alpha}$$

donc la fonction d'efficacité est

$$P(\theta) \simeq \frac{\left[ \left( \frac{1 - \beta}{\alpha} \right)^h - 1 \right]}{\left[ \left( \frac{1 - \beta}{\alpha} \right)^h - \left( \frac{\beta}{1 - \alpha} \right)^h \right]},$$

où  $h(\theta)$  est la solution de

$$\theta \left( \frac{\theta_1}{\theta_0} \right)^h + (1 - \theta) \left( \frac{1 - \theta_1}{1 - \theta_0} \right)^h = 1,$$

Wald suggère de résoudre cette équation par rapport à  $\theta$  pour des valeurs données de  $h$ , plutôt que le contraire.

Nous nous contenterons d'indiquer les étapes de la démonstration.

Supposons que l'on puisse trouver  $h(\theta)$  tel que

$$\int_{-\infty}^{+\infty} \left[ \frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta) dx = 1,$$

alors

$$f^*(x, \theta) = \left[ \frac{f(x, \theta_1)}{f(x, \theta_0)} \right]^{h(\theta)} f(x, \theta)$$

peut s'interpréter comme une fonction de distribution. Le TSRP permettant de tester  $f^*(x, \theta)$  contre  $f(x, \theta)$ , pour  $\theta$  donné, est de la forme

$$B_\theta^* < \frac{f^*(x, \theta)}{f(x, \theta)} < A_\theta^* \tag{2.12}$$

## Chapitre 2. Test Séquentiel

---

Sachant que

$$A_{\theta}^* = A^{h(\theta)}$$

et

$$B_{\theta}^* = B^{h(\theta)},$$

On a le TSRP  $B < l_n < A$  s'arrête et décide pour  $\theta_0$  ou  $\theta_1$  et la probabilité pour que ce TSRP décide en faveur de  $\theta_0$  est  $P(\theta)$ .

Au point où le TSRP (2.12) s'arrête et décide pour  $f(x, \theta)$  ou  $f^*(x, \theta)$ .

Supposons maintenant que  $f(x, \theta)$  soit la distribution vraie de  $x$ , la probabilité pour que ce TSRP décide en faveur de  $f(x, \theta)$  lorsque cette fonction est la fonction de distribution vraie de  $x$ .

Si maintenant nous définissons pour le TSRP (2.12)

$$Pr\{f(x, \theta) \text{ est vraie et le TSRP (2.12) décide en faveur de } f(x, \theta)\} = 1 - \alpha^*$$

$$Pr\{f(x, \theta) \text{ est vraie et le TSRP (2.12) décide en faveur de } f^*(x, \theta)\} = \alpha^*$$

$$Pr\{f^*(x, \theta) \text{ est vraie et le TSRP (2.12) décide en faveur de } f^*(x, \theta)\} = 1 - \beta^*$$

$$Pr\{f^*(x, \theta) \text{ est vraie et le TSRP (2.12) décide en faveur de } f(x, \theta)\} = \beta^*$$

le raisonnement ci-dessus permet alors d'écrire

$$P(\theta) = 1 - \alpha^*$$

Appliquons ensuite le théorème 2.1.1 au TSRP (2.12) :

$$A_{\theta}^* \simeq \frac{1 - \beta^*}{\alpha^*}$$

$$B_{\theta}^* \simeq \frac{\beta^*}{1 - \alpha^*}$$

Grâce aux équations  $A_{\theta}^* = A^{h(\theta)}$  et  $B_{\theta}^* = \frac{\beta^*}{1 - \alpha^*}$  le théorème est démontré.

### Remarque 2.2.1.

Pour certains points, il est facile de calculer la valeur de la fonction d'efficacité.

Par définition de  $P(\theta)$ , on a  $P(\theta_1) = \beta$  et d'après la remarque suivant le théorème 2.1.1,

$P(\text{Accepter } H_0 | H_0 \text{ vraie}) = 1 - \alpha$ . Donc  $P(\theta_0 = 1 - \alpha)$ . [10]

## Chapitre 2. Test Séquentiel

---

### Exemple 2.2.1.

On se place à nouveau dans l'exemple d'un échantillon de loi  $N(\theta, \sigma^2)$  avec  $\sigma$  connu.

Si on utilise les approximations A et de B, on a la formule d'approximation de la fonction d'efficacité :

$$P(\theta) \simeq \frac{\left[ \left( \frac{1-\beta}{\alpha} \right)^{h(\theta)} - 1 \right]}{\left[ \left( \frac{1-\beta}{\alpha} \right)^{h(\theta)} - \left( \frac{\beta}{1-\alpha} \right)^{h(\theta)} \right]}$$

Pour l'exemple des lois normales, on obtient

$$\forall \theta, h(\theta) = \frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0}$$

Par conséquent pour tout  $\theta$ ,

$$P(\theta) = \frac{e^{\frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0} \log \frac{1-\beta}{\alpha}} - 1}{e^{\frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0} \log \frac{1-\beta}{\alpha}} - e^{\frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0} \log \frac{\beta}{1-\alpha}}}$$

La figure suivante représente l'approximation la fonction d'efficacité pour un échantillon de loi normale lorsque  $\alpha = 0.01$ ,  $\beta = 0.05$  et  $\theta_1 = 2$ ,  $\theta_0 = 1$ . [10]

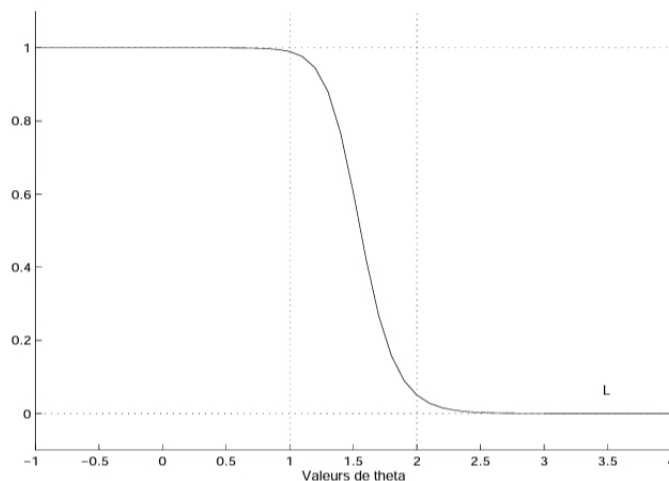


FIGURE 2.4 – Approximation de la fonction d'efficacité

### 2.2.2 La fonction de l'effectif moyen de l'échantillon (ASN)

On peut obtenir des informations supplémentaires sur le test séquentiel du rapport des probabilités, plus particulièrement sur le nombre d'observations nécessaires pour prendre une décision, en considérant la fonction de l'effectif moyen de l'échantillon (Average Sample Number en anglais).

Notons  $N$  le nombre d'observations nécessaires au test pour prendre une décision. Ce nombre dépend du résultat des observations : il s'agit donc d'une variable aléatoire discrète dont nous noterons  $n$  la réalisation.[9]

#### Définition 2.2.2.

On appelle la fonction de l'effectif moyen de l'échantillon (ASN) la fonction

$$\theta \mapsto E_{\theta}(N)$$

Nous allons exprimer cette fonction d'une manière différente afin de pouvoir en donner une approximation dont l'expression sera plus facile à calculer.[10]

Notons  $S_n := Z_1 + \dots + Z_n$  avec  $Z_i$  défini par la relation :

$$Z_i = \log \frac{f(X_i, \theta_1)}{f(X_i, \theta_0)}$$

#### Théorème 2.2.2. (Wald 1943)

Étant donné le TSRP avec des probabilités d'erreur  $\alpha$  et  $\beta$ , une approximation de la taille moyenne de l'échantillon est donnée par :

$$E(n|\theta) \simeq \{P(\theta) \log B + (1 - P(\theta)) \log A\} / E(z|\theta)$$

#### Démonstration 2.2.2.

On doit remarquer que cette approximation n'est plus valable si  $E(z|\theta) = 0$  ; dans ce cas Wald a établi une formule particulière. Ce théorème comporte deux approximations, tout d'abord l'erreur résultant du franchissement des frontières est négligée, ensuite on calcule  $P(\theta)$  à l'aide de la formule approchée précédente. Wald étant en mesure de calculer des limites pour cette fonction ASN cela constituait une justification, au moins partielle, des approximations effectuées.



## Chapitre 2. Test Séquentiel

---

D'autre part, si la distribution de  $z_i$  est une fonction de  $z$  et de  $\theta$ , on peut constater facilement que, à partir d'une valeur initial  $Z$ , et pour  $\theta$  donné, la taille de l'échantillon satisfait en probabilité à l'équation

$$E(n|Z, \theta) = 1 + \int_{\log A}^{\log B} E(n|y, \theta) \varphi(y - Z, \theta) dy$$

On trouverait pour la fonction d'efficacité une équation intégrale analogue ( sans constante dans la second membre).

### Exemple 2.2.2 (10).

Reprenons l'exemple d'un échantillon de loi normale  $N(\theta, \sigma^2)$  avec  $\sigma = 1$ .

Déterminons la fonction ASN pour  $\theta = \theta_1$  et  $\theta = \theta_0$ . On obtient dans le cas particulier des lois normales :

$$E_{\theta_1}(Z_1) = \frac{1}{2}(\theta_1 - \theta_0)^2 \neq 0$$

et

$$E_{\theta_0}(Z_1) = -\frac{1}{2}(\theta_1 - \theta_0)^2 \neq 0$$

En prenant  $A = \frac{1-\beta}{\alpha}$  et  $B = \frac{\beta}{1-\alpha}$ , grâce au théorème précédent, les approximations de la fonction ASN en  $\theta_0$  et  $\theta_1$  sont données par :

$$- \text{ Sous } H_0 : \quad E_{\theta_0}(N) \sim \frac{\alpha \log A + (1-\alpha) \log B}{-\frac{1}{2}(\theta_1 - \theta_0)^2} \text{ car } P(\theta_0) = 1 - \alpha,$$

$$- \text{ Sous } H_1 : \quad E_{\theta_1}(N) \sim \frac{(1-\beta) \log A + \beta \log B}{\frac{1}{2}(\theta_1 - \theta_0)^2} \text{ car } P(\theta_1) = \beta$$

Pour  $\theta$  quelconque on a

$$E_{\theta}(Z_1) = -\frac{1}{2}(\theta_1^2 - \theta_0^2) + (\theta_1 - \theta_0)\theta$$

et

$$E_{\theta}(Z_1) = 0 \Leftrightarrow \theta = \frac{\theta_1 + \theta_0}{2}$$

Donc pour  $\theta \neq \frac{\theta_1 + \theta_0}{2}$ , l'approximation de la fonction ASN est donnée par la formule

$$E_{\theta}(N) \sim \frac{(1 - P(\theta)) \log A + P(\theta) \log B}{-\frac{1}{2}(\theta_1^2 - \theta_0^2) + (\theta_1 - \theta_0)\theta}$$

## Chapitre 2. Test Séquentiel

---

En utilisant l'approximation de Wald de la fonction d'efficacité, on est en mesure de tracer une approximation de la fonction ASN (figure 2.5) pour  $\alpha = 0.01, \beta = 0.05, \theta_0 = 1$  et  $\theta_1 = 2$ .

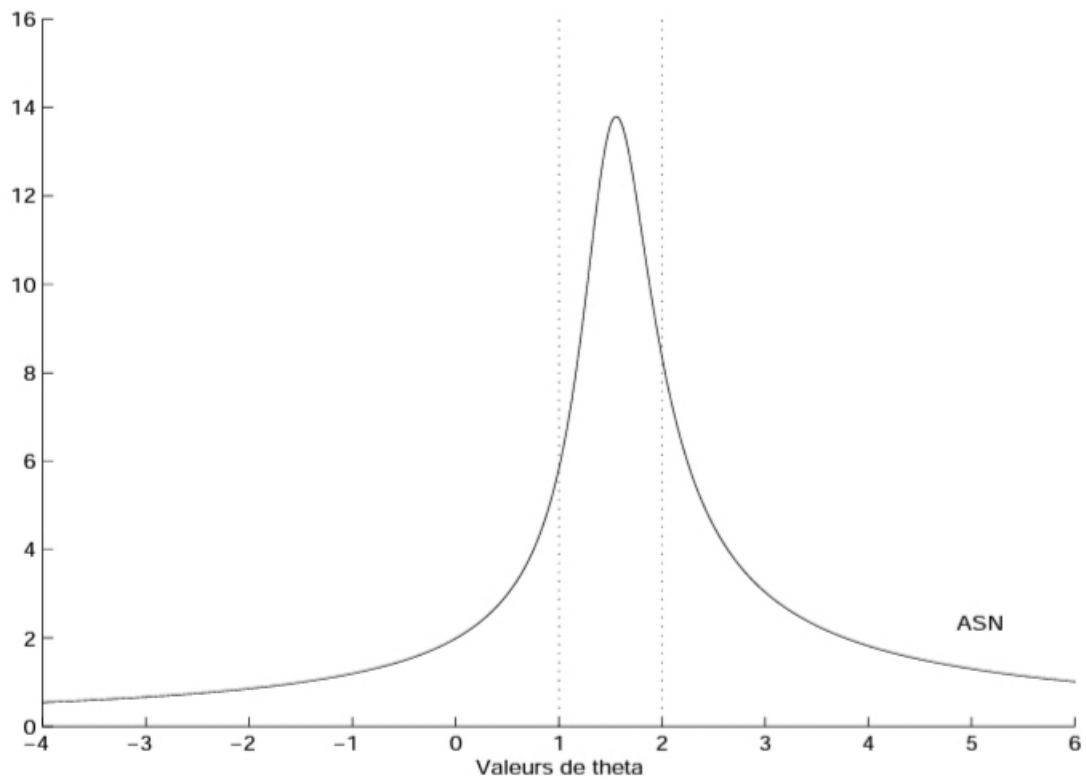


FIGURE 2.5 – Approximation de La fonction de l'effectif moyen de l'échantillon (ASN)

## 2.3 Comparaison entre tests classiques et tests séquentiels

### 2.3.1 Avantages des tests séquentiels

Ces tests sont explicitement intéressants par ce qu'ils permettent, pour des raisons éthiques, d'exposer le moins de sujets possible aux traitements les moins efficaces et ils sont plus économes concernant le temps et le coût que les tests classiques.[11]

### 2.3.2 Inconvénients des tests séquentiels

Contrairement au tests classiques la durée de l'étude de tests séquentiels est aléatoire, le nombre de sujets est aléatoire en plus si la procédure s'arrête rapidement, les estimations ponctuelles peuvent être très imprécises.[11]

### 2.3.3 Approche décisionnelle entre le test classique et le test séquentiel

Test classique c'est de Choisir l'une des 2 décisions suivantes :

Hypothèse nulle  $H_0$ .

Hypothèse alternative  $H_1$ .

Contrairement au test séquentiel qui choisit l'une entre ces 3 décisions :

Hypothèse nulle  $H_0$ .

L'hypothèse alternative  $H_1$ .

Information insuffisante, besoin de plus de données.

#### **Exemple 2.3.1.**

*Soit un nouveau lot de pièces d'une société qui a toujours donné satisfaction jusque là.*

*Allons-nous accepter ce lot de nouveau, ou bien le rejeter ?[12]*

#### **Test classique**

*Tirer du lot un échantillon de taille  $n$  fixée.*

*En notant  $p$  la proportion de pièces défectueuses et faire le test suivant avec un risque  $\alpha$*

## Chapitre 2. Test Séquentiel

---

(erreur grave) fixé :  $H_0 : p = p_0$  et  $H_1 : p = p_1$  ( $p_1 > p_0$ ).

### Test séquentiel

*Tirer du lot une pièce*

*Regarder si la pièce tirée est bonne ou défectueuse*

*Décider :*

- soit d'accepter le lot (accepter  $H_0$ ).
- soit de refuser le lot (rejeter  $H_0$ ).
- soit de tirer une nouvelle pièce ( $n \leftarrow n + 1$ ) c-à-d on reprend le test avec  $(n + 1)$  observations.

### 2.3.4 Gain de la méthode séquentielle

Nous allons étudier un exemple afin de comparer le test séquentiel du rapport des probabilités à un test classique et constater le gain en nombre d'observations de la méthode séquentielle.

Supposons que l'on dispose d'un échantillon  $X_1, \dots, X_n$  ( $n$  grand) de variables aléatoires indépendantes et de loi normale  $N(\theta, 1)$ . On cherche à tester  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta = \theta_1$  avec  $\theta_0 < \theta_1$ .

On désire pour cela utiliser un test de niveau  $\alpha$  et de puissance  $1 - \beta$ .

On utilisera tout d'abord un test classique et ensuite on mettra en oeuvre un test séquentiel. Enfin on comparera les deux méthodes au niveau du nombre d'observations nécessaires pour prendre une décision. [10]

– Pour un **test classique** :

Soient donc  $\alpha$  et  $\beta$  fixés, on va déterminer  $n(\alpha, \beta)$  le nombre d'observations nécessaires pour obtenir une puissance de test égale à  $1 - \beta$ .

Par application du lemme de Neyman-Pearson, on obtient un test  $\psi$  de niveau  $\alpha$  et de puissance maximale. Le test  $\psi$  s'écrit

$$\psi = \mathbb{I}_{\{\bar{X}_n > k\}} \quad \text{où} \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

## Chapitre 2. Test Séquentiel

---

Puisque le test doit être de niveau  $\alpha$  et de puissance  $1 - \beta$ , il doit vérifier les deux équations :

$$\mathbb{P}_{\theta_0}(\bar{X}_n \leq K) = 1 - \alpha$$

$$\mathbb{P}_{\theta_1}(\bar{X}_n \leq K) = \beta$$

qui vont nous permettre de déterminer  $k$  et  $n(\alpha, \beta)$ .

D'une part on a

$$\mathbb{P}_{\theta_0}(\bar{X}_n \leq K) = 1 - \alpha$$

$$\Leftrightarrow \mathbb{P}_{\theta_0}(\sqrt{n}(\bar{X}_n - \theta_0)) \leq \sqrt{n}(K - \theta_0) = 1 - \alpha$$

$$\Leftrightarrow \psi(\sqrt{n}(K - \theta_0)) = 1 - \alpha,$$

où  $\psi$  désigne la fonction de répartition d'une loi normale centrée réduite car  $\sqrt{n}(\bar{X}_n - \theta_0)$  suit une loi  $N(0, 1)$ .

Et d'une part on a

$$\mathbb{P}_{\theta_1}(\bar{X}_n \leq K) = \beta$$

$$\Leftrightarrow \mathbb{P}_{\theta_1}(\sqrt{n}(\bar{X}_n - \theta_1)) \leq \sqrt{n}(K - \theta_1) = \beta$$

$$\Leftrightarrow \psi(\sqrt{n}(K - \theta_1)) = \beta$$

En notant  $\mu_1$  et  $\mu_0$  les réels tels que  $\psi(\mu_0) = 1 - \alpha$  et  $\psi(\mu_1) = \beta$ , on obtient alors

$$\begin{cases} \mu_0 = \sqrt{n}(K - \theta_0) \\ \mu_1 = \sqrt{n}(K - \theta_1). \end{cases}$$

Finalement, en résolvant ce système d'équations on obtient

$$\begin{cases} K = \theta_0 + \sqrt{n}\mu_0 \\ \sqrt{n} = \frac{\mu_1 - \mu_0}{\theta_1 - \theta_0} \end{cases}$$

On en déduit donc que

$$n(\alpha, \beta) = \frac{(\mu_1 - \mu_0)^2}{(\theta_1 - \theta_0)^2}$$

## Chapitre 2. Test Séquentiel

---

– Pour le **test séquentiel** :

On utilise l'approximation de A par  $\frac{1-\beta}{\alpha}$  et l'approximation de B par  $\frac{\beta}{1-\alpha}$ .

D'après le théorème (2.2.4) on a

$$E_{\theta}(N) \sim \frac{(1-P(\theta)) \log A + P(\theta) \log B}{E_{\theta}(Z_1)}$$

avec  $A = \frac{1-\beta}{\alpha}$  et  $B = \frac{\beta}{1-\alpha}$ .

Donc

$$E_{\theta_0}(N) \sim \frac{\alpha \log A + (1-\alpha) \log B}{E_{\theta}(Z_1)},$$

et

$$E_{\theta_1}(N) \sim \frac{(1-\beta) \log A + \beta \log B}{E_{\theta}(Z_1)}.$$

Nous avons vu dans l'exemple précédent que

$$E_{\theta_1}(Z_1) = \frac{1}{2}(\theta_1 - \theta_0)^2 \quad \text{et} \quad E_{\theta_0}(Z_1) = -\frac{1}{2}(\theta_1 - \theta_0)^2.$$

On a finalement

$$\frac{E_{\theta_1}(N)}{n(\alpha, \beta)} = \frac{2}{(\mu_1 - \mu_0)^2} [(1-\beta) \log A + \beta \log B]$$

et

$$\frac{E_{\theta_0}(N)}{n(\alpha, \beta)} = -\frac{2}{(\mu_1 - \mu_0)^2} [(1-\alpha) \log B + \alpha \log A]$$

On peut remarquer que ces expressions ne dépendent pas de  $\theta_0$  et  $\theta_1$ .

Pour le test utilisant les approximations les pourcentages moyens d'économie sont

- sous  $H_0$  :  $100[1 - \frac{E_{\theta_0}(N)}{n(\alpha, \beta)}]$ ,

- sous  $H_1$  :  $100[1 - \frac{E_{\theta_1}(N)}{n(\alpha, \beta)}]$ .

Dans le test séquentiel du rapport des probabilités, on utilise les approximations de A et B, ce qui implique comme nous l'avons vu précédemment une augmentation

## Chapitre 2. Test Séquentiel

---

du nombre d'observations. Donc l'économie obtenue avec les valeurs théoriques ne peut être que supérieur à l'économie obtenue par le test séquentiel utilisant les approximations de A et B.

### Application numérique :

$\alpha = 0.01$  et  $\beta = 0.05$  d'où  $\psi(\mu_0) = 0.99$  et  $\psi(\mu_1) = 0.05$ .

En utilisant la table de la loi  $N(0, 1)$  on trouve  $\mu_0 = 2.33$  et  $\mu_1 = -1.64$ .

Le pourcentage d'économie sous  $H_1$  est de l'ordre de 47% tandis que le pourcentage d'économie sous  $H_0$  est d'environ 63%. [10]

### Preuve que le TSRP s'arrête p.s. et moments de N

Le test séquentiel du rapport des probabilités est défini par récurrence sur le nombre d'observations. On pourrait, a priori, imaginer que le rapport des vraisemblances  $l_m$  reste constamment dans l'intervalle  $]A, B[$  et que le test ne s'arrête pas. Or il s'avère que presque sûrement  $l_m(X_1, \dots, X_m)$  sort de cet intervalle. C'est ce que nous allons démontrer. [10]

Le test s'arrête à l'étape  $m$  si  $m$  est le plus petit entier tel que

$$Z_1 + \dots + Z_m \geq A \text{ ou } Z_1 + \dots + Z_m \leq B$$

Posons  $c := |\log B| + |\log A|$  et notons les  $k$  premières séquences

$$\gamma_1 = Z_1 + \dots + Z_r$$

$$\vdots \quad \vdots$$

$$\gamma_k = Z_{r(k-1)+1} + \dots + Z_{kr} \text{ où } r \text{ est un entier fixé.}$$

## Chapitre 2. Test Séquentiel

---

**Lemme 2.3.1.** (Wald 1943)

S'il existe un entier  $k$  tel que  $\gamma_k \geq c$ , alors le test  $S(A,B)$  s'arrête avant ou à la  $k$ -ième séquence.

**Démonstration 2.3.1.**

Si le test ne s'est pas arrêté à la  $(k-1)$ -ième séquence on a :

$$\log B < \gamma_1 + \dots + \gamma_{k-1} < \log A$$

- Si  $\gamma_k \geq c$  alors

$$\begin{aligned} \sum_{i=1}^k \gamma_i &\geq \log B + |\log B| + |\log A| \\ &\geq |\log A| \text{ car } \log B + |\log B| \geq 0, \\ &\geq \log A \text{ donc le test s'arrête à la } k\text{-ième séquence.} \end{aligned}$$

- Si  $\gamma_k \leq -c$  alors

$$\begin{aligned} \sum_{i=1}^k \gamma_i &\leq -\log A - |\log B| - |\log A| \\ &\leq -|\log B| \text{ car } \log A - |\log A| \leq 0, \\ &\leq -\log B \text{ donc le test s'arrête à la } k\text{-ième séquence.} \end{aligned}$$

**Remarque 2.3.1.**

Donc s'il existe  $k$  tel que  $|\gamma_k| \geq c$ , alors  $N \leq kr$  où  $N$  est la variable aléatoire représentant le nombre d'observations au moment de la décision.

**Théorème 2.3.1.** (Wald 1943)

Soit  $\theta \in \Theta$  fixé. Si  $\text{Var}_\theta(Z_1) < +\infty$  alors le test séquentiel du rapport des probabilités s'arrête en un nombre fini d'observations avec une probabilité égale à 1 pour le paramètre  $\theta$ .

**Démonstration 2.3.2.**

Si le test ne s'arrête pas, on a alors  $|\gamma_k| < c \forall k \geq 1$  d'après le lemme précédent.

Par conséquent,

$$\mathbb{P}_\theta(\text{Le test ne s'arrête pas}) \leq \mathbb{P}_\theta(\forall K, |\gamma_k|^2 < c^2).$$

On peut alors écrire



$$\begin{aligned}\mathbb{P}_\theta(\forall K, |\gamma_k|^2 < c^2) &= \lim_{j \rightarrow +\infty} \mathbb{P}_\theta(\forall K = 1, \dots, j, |\gamma_k|^2 < c^2) \text{ par continuité inférieure,} \\ &= \lim_{j \rightarrow +\infty} \mathbb{P}_\theta(|\gamma_1|^2 < c^2)^j \text{ car les } \gamma_i \text{ sont i.i.d.}\end{aligned}$$

Montrons que  $\mathbb{P}_\theta(|\gamma_1|^2 < c^2) < 1$ , on aura ainsi  $\mathbb{P}_\theta(\forall K, |\gamma_k|^2 < c^2) = 0$ .

Si  $\mathbb{P}_\theta(|\gamma_1|^2 < c^2) = 1$ , c'est à dire si  $\mathbb{P}_\theta(|\gamma_1|^2 \geq c^2) = 0$ , alors :

$$\begin{aligned}E_\theta(\gamma_1^2) &= E_\theta(\gamma_1^2 \mathbb{I}_{\{\gamma_1^2 < c^2\}} + \gamma_1^2 \mathbb{I}_{\{\gamma_1^2 \geq c^2\}}) \\ &= E_\theta(\gamma_1^2 \mathbb{I}_{\{\gamma_1^2 < c^2\}}) + E_\theta(\gamma_1^2 \mathbb{I}_{\{\gamma_1^2 \geq c^2\}}) < 0 + c^2.\end{aligned}$$

Or on a supposé  $E_\theta(Z_i^2) > 0$ , donc il est possible de trouver  $r$  tel que  $E_\theta(\gamma_1^2) \geq c^2 = 0$ . Ce qui fournit une contradiction.

Pour ce choix de  $r$ , on a  $\mathbb{P}_\theta(|\gamma_1|^2 < c^2) < 1$  et on peut alors conclure que

$$\mathbb{P}_\theta(\text{Le test s'arrête}) = 1.$$

### 2.3.5 Contrôle par échantillonnage

En 1996, Dodge et König ont proposé la forme la plus élémentaire de procédure séquentielle, c'est le plan double échantillonnage. Il s'agit simplement d'un plan où, dans une première étape, on prend un échantillon de taille fixe ; si le résultat recherché apparaît avec une évidence suffisante on s'arrête à cette première étape. Sinon on tire un deuxième échantillon de taille fixe et l'expérience se termine là.

### 2.3.6 Plan d'échantillonnage

#### Définition 2.3.1.

*Le plan d'échantillonnage c'est un plan consiste à choisir la taille d'un telle analyse des lots, chaque lot contient un grand nombre d'élément donné, chacun élément pouvant être qualifié de bon ou défectueux, qui permet en fournir des règles pour l'acceptation ou le rejeter d'un lot.[9]*

## Chapitre 2. Test Séquentiel

---

**Définition 2.3.2.** (*Plan d'échantillonnage simple*)

Pour chaque lot on analyse un nombre donné  $N$  d'éléments. Le lot est accepté lorsque nous ne dépasserons pas l'élément défectueux  $c$ . Par contre le lot est rejeté, si l'on rencontre ou plus l'élément défectueux  $c$ .

Pour le plan d'échantillonnage simple on a le point représentatif du résultat sur la droite  $r + s = N$ , elle se trouve plus haut, puisque on observe  $N$  éléments dans chaque lot. Les points de rejet c'est les points dans cette droite qui vérifient  $r \geq c$ , et les points d'accepter c'est les points qui satisfont  $r \leq c - 1$ .

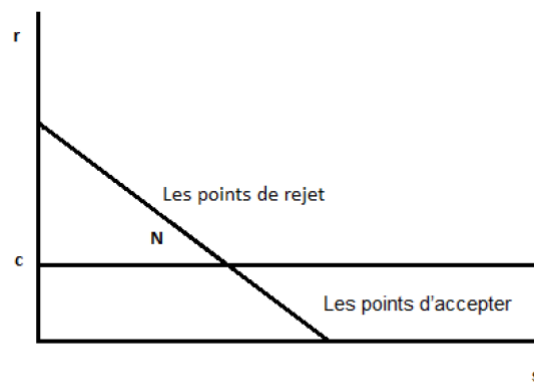


FIGURE 2.6 – Plan d'échantillonnage simple

### 2.3.7 Contrôle tronqué

Le plan tronqué contient les résultats dans deux axes, l'axe vertical représente le nombre des éléments défectueux  $r$ , et l'axe horizontal représente le nombre des éléments acceptables  $s$ , telle que  $r + s = N$ .

**Définition 2.3.3.** (*Plan tronqué*)

C'est un plan d'échantillonnage qui contrôle les éléments de  $r \geq c$  et arrête l'observation et rejeter le lot. Et l'élément de  $s \geq N - c + 1$  et arrête l'observation et accepte le lot.[9]

**Exemple 2.3.2.**

Pour  $N = 20$  et  $c = 3$  on le plan tronqué suivant :

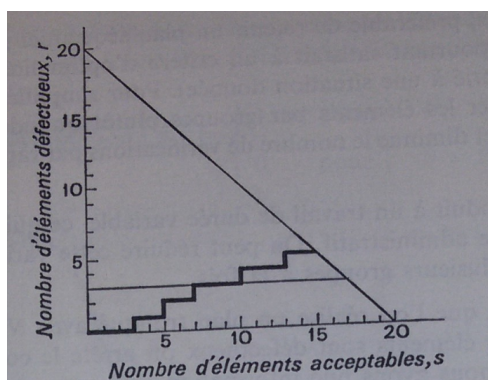


FIGURE 2.7 – Plan tronqué

Le graphique est défini par deux frontières, la première équation  $r = c$  c'est l'axe horizontal et la deuxième équation  $s = N - c + 1$  c'est l'axe vertical, et on arrête le contrôle si que l'une des frontières est remplie. Cette plan est séquentiel car pour chaque élément observé on arrête pour si l'un ou l'autre des conditions est satisfait et on continuerons l'observation, et en plus la taille d'échantillon n'est pas fixe elle est entre  $c$  et  $N$  avec  $c$  le première élément est défectueux.

Le plan séquentiel est plus compliqué puisque plan de taille fixe, et l'observation est plus lente. Les frontières de décision de graphique (2.7) sont modifiées et améliorer les propriétés de la taille moyenne de l'échantillon sans modifications à mener des plans plus en plus compliqué, en pratique il est difficile pour définir les condition pour remplir d'échantillonnage.

Pour simplifier les plans séquentiels et diminuer le nombre de vérification par rapport aux frontières de décision on mettre les éléments par groupe.

### 2.3.8 Les propriétés du plan tronqués

1. Dans le plan tronqué on définit la fonction d'efficacité c'est la probabilité de lot qui accepter comme suit :

$$\mathbf{P}(\text{acceptation}|\theta) = \mathbf{P}(r \leq c - 1|\theta) = \sum_{r=0}^{c-1} \binom{N}{r} \theta^r (1 - \theta)^{N-r}$$

## Chapitre 2. Test Séquentiel

---

avec  $\theta$  le proportion de défectueux dans le lot, et  $c$  l'élément de défectueux et  $r$  les nombre des éléments défectueux. La courbe d'efficacité est représentée sur la figure suivante :

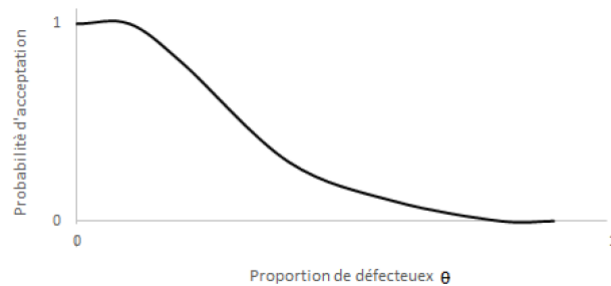


FIGURE 2.8 – Courbe d'efficacité

En théorie nécessaire d'inspecter la totalité de chaque lot pour tracer cette courbe au moins  $\theta \neq 0$  ou  $\theta \neq 1$  pour la façon suivant :

$$\mathbf{P}(\text{acceptation}|\theta) = \begin{cases} 1 & \text{si } \theta < \theta_0 \\ 0 & \text{si } \theta > \theta_1 \end{cases}$$

Et  $[\theta_0, \theta_1]$  c'est intervalle de région d'indifférence et on ne nécessite pas la décision qui on a utilisé, pour simplifier on a la figure :

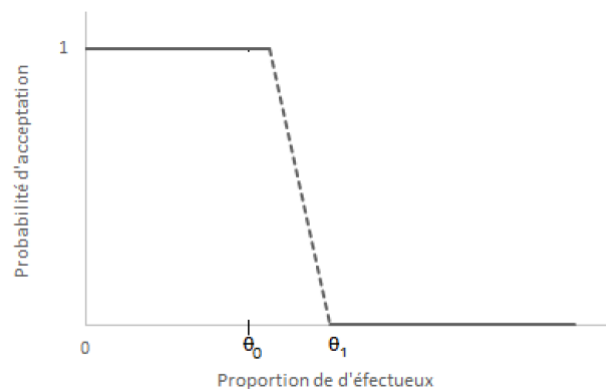


FIGURE 2.9 – Courbe d'efficacité idéale pour  $[\theta_0, \theta_1]$

Même dans le cas de moins désagréable. On n'éviter pas le contrôle complet de

## Chapitre 2. Test Séquentiel

---

chaque lot, mais on a la modification suivant :

$$\mathbf{P}(\text{acceptation}|\theta) = \begin{cases} 1 - \alpha & \text{si } \theta = \theta_0 \\ \beta & \text{si } \theta = \theta_1 \end{cases} \quad (2.13)$$

Avec la taille  $\alpha$  et  $\beta$  la puissance des tests d'hypothèse de Neymann-Pearson, et les paramètres  $N$  et  $c$  du plan tronqué choisir pour satisfait la relation (2.13)

2. Pour la distribution de la taille de l'échantillon. Un lot rejeté avec la taille de l'échantillon entre  $c$  et  $N$  on a :

Pour  $n = c$  :  $\mathbf{P}(n) = \theta^c$

Pour  $n = c + r$  :  $\mathbf{P}(n) = \binom{c+r-1}{c-1} \theta^c (1-\theta)^r$ , pour  $r = 0, 1, \dots, N-c$ .

Et pour un lot accepter avec la taille de l'échantillon entre  $N - c + 1$  et  $N$ , pour  $n = N - c + 1 + r$  et  $r = 0, 1, \dots, c - 1$  :  $\mathbf{P}(n) = \binom{N-c+r}{r} \theta^r (1-\theta)^{N-c+1}$

Dans le plan tronqué d'effectif moyen dépende de  $\theta$  présenter un maximum aux environs de  $\theta = \frac{c}{N}$ .

L'ordre de l'effectif moyen pour  $\theta$  à des valeurs faibles est  $\frac{N-c+1}{1-\theta}$  et  $\frac{c}{N}$  c'est l'ordre de  $\theta$  à des valeurs élevées sur la figure (2.9), on a l'axe horizontal indique la taille de l'échantillon fixe.

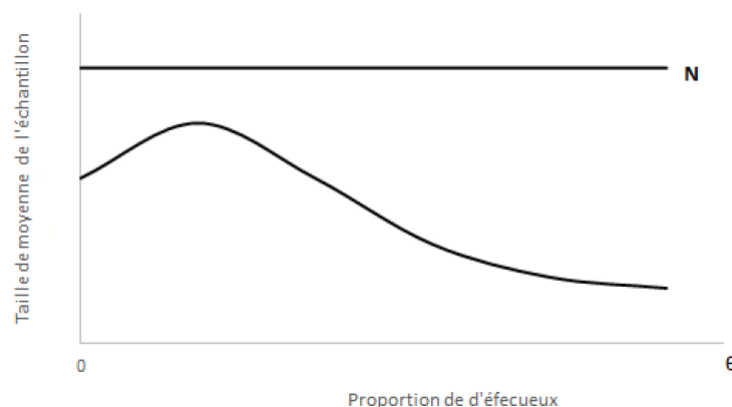


FIGURE 2.10 – La fonction de l'effectif de moyenne de l'échantillon dans le cas de plan tronqué

## Chapitre 2. Test Séquentiel

---

3. Plan de contrôle par échantillonnage ce sont des frontières pour l'acceptation ou rejeter un lot par rapport l'élément de défectueux  $c$ , tandis que la taille de l'échantillon est faible pour tous les éléments inférieure à  $c$  et la taille de l'échantillon est élevées pour les éléments supérieure à  $c$ .
4. Il y a deux type des frontières pour les valeurs faibles et élevées de  $\theta$  qui remplir les condition (2.11). Wald définir cette frontières comme des droites parallèles et une probabilité non nul de la taille de l'échantillon grande comme suivant :  
Les frontières fermées ont la taille de l'échantillon à une limite supérieure dans la figure suivante :

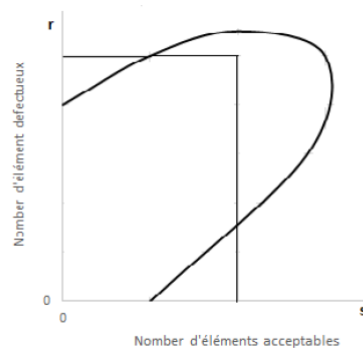


FIGURE 2.11 – Frontières fermées

## Chapitre 2. Test Séquentiel

---

Les frontières ouvertes ont la taille de l'échantillon infinie dans la figure suivant :

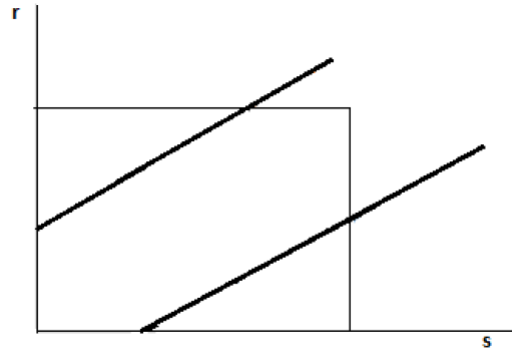


FIGURE 2.12 – Frontières ouvertes

# Quelques tests séquentiels

## 2.4 Tests bilatéraux

### 2.4.1 Tests bilatéraux pour comparer deux traitements avec une réponse normale de la variance connue :

Ici nous nous limitons à tester la différence de réponse moyenne de deux traitements lorsque les observations sont normalement réparties avec variance connue et commune. Indiquant la différence de moyennes par  $\theta$ , l'hypothèse nul  $H_0 : \theta = 0$  indique que les réponses suivent la même distribution sous les deux traitements. L'hypothèse alternative,  $H_A : \theta \neq 0$ , contient deux cas,  $\theta < 0$  et  $\theta > 0$ , ce qui correspond à un traitement supérieur à l'autre et vice versa.

Dans cette comparaison, la statistique de test normalisée,  $Z$ , est distribuée symétriquement environ 0 sous  $H_0$ , et un test d'échantillonnage fixe rejette  $H_0$  si  $|Z| > c$  pour une constante  $c$ . Le signe de  $Z$  détermine le traitement à privilégier lorsque  $H_0$  est rejeté. La probabilité d'erreur de type I d'un test est définie comme la probabilité de rejetant à tort l'hypothèse nulle,

$$P_{\theta=0}\{|Z| > c\}.$$

La puissance d'un test est la probabilité de rejeter l'hypothèse nulle lorsqu'elle le fait ne tient pas,  $P_{\theta}\{|Z| > c\}$ , pour des valeurs de  $\theta \neq 0$ . La puissance dépend de  $\theta$ , augmentant lorsque  $\theta$  s'éloigne de 0, mais il est commode d'indiquer une puissance requise à un valeur ou valeurs spécifiques de  $\theta$ , par exemple,

$$P_{\theta=\delta}\{|Z| > c\} = P_{\theta=-\delta}\{|Z| > c\} = 1 - \beta, \quad (2.14)$$



## Chapitre 2. Test Séquentiel

---

où  $\delta$  représente une différence de traitement que les enquêteurs espèrent détecter avec une forte probabilité. La petite probabilité  $\beta$  est appelée Type II probabilité d'erreur à  $\theta = \pm\delta$  [12].

**Remarque 2.4.1.** (*Chr. Jennison. 2000*)

Concrètement, il n'est pas souhaitable de rejeter  $H_0$  dans faveur de  $\theta < 0$  alors qu'en fait  $\theta > 0$  car cela impliquerait une recommandation de l'inférieur des deux traitements. Dans cet esprit, la puissance requise ci-dessus peut être remplacé par

$$P_{\theta=\delta}\{Z > c\} = P_{\theta=-\delta}\{Z < -c\} = 1 - \beta. \quad (2.15)$$

Heureusement, pour les valeurs de  $c$  et  $\delta$  que nous allons considérer, les probabilités  $P_{\theta=\delta}\{Z < -c\}$  et  $P_{\theta=-\delta}\{Z > c\}$  sont extrêmement petits et les deux les exigences (3.1) et (3.2) sont égales à toutes fins pratiques.

### 2.4.2 Un test d'échantillonnage fixe

Soit  $X_{Ai}$  et  $X_{Bi}$ ,  $i = 1, 2, \dots$ , indiquent les réponses des sujets deux traitements, A et B. Supposons les réponses des sujets recevant un traitement A sont normalement distribués avec la variance  $\sigma^2$  et la moyenne  $\mu_A$ , que nous écrivons  $X_{Ai} \sim N(\mu_A, \sigma^2)$ ,  $i = 1, 2, \dots$ . De même, supposons  $X_{Bi} \sim N(\mu_B, \sigma^2)$ ,  $i = 1, 2, \dots$  et toutes les observations sont indépendantes. Considérez le problème de tester l'hypothèse nulle de pas de différence de traitement  $H_0 : \mu_A = \mu_B$  contre la alternative bilatérale  $\mu_A \neq \mu_B$  avec probabilité d'erreur de type I  $\alpha$  et puissance  $1 - \beta$  à  $\mu_A - \mu_B = \pm\delta$ .

Si  $n$  sujets sont affectés à chaque traitement, la statistique standardisée

$$\begin{aligned} Z &= \frac{1}{\sqrt{(2n\sigma^2)}} \left( \sum_{i=1}^n X_{Ai} - \sum_{i=1}^n X_{Bi} \right) \\ &\sim N((\mu_A - \mu_B)\sqrt{\{n/(2\sigma^2)\}}, 1). \end{aligned}$$

Ainsi  $Z \sim N(0, 1)$  sous  $H_0$  et le test symétrique bilatéral avec erreur de type I la probabilité  $\alpha$  rejette  $H_0$  si  $|Z| > F^{-1}(1 - \alpha/2)$ , où  $F$  désigne la fonction de répartition.

Pour satisfaire la puissance requise, nous avons aussi besoin de

$$P\{|Z| > F^{-1}(1 - \alpha/2)\} = 1 - \beta$$

## Chapitre 2. Test Séquentiel

---

lorsque  $Z \sim N(\pm\delta\sqrt{\{n/(2\sigma^2)\}}, 1)$ . Nous pouvons ignorer les très faibles probabilités que  $Z < -F^{-1}(1 - \alpha/2)$  lorsque  $\mu_A - \mu_B = \delta$  ou  $Z > F^{-1}(1 - \alpha/2)$  lorsque  $\mu_A - \mu_B = -\delta$ ; nous avons donc besoin

$$E(Z) = F^{-1}(1 - \alpha/2) + F^{-1}(1 - \beta)$$

à l'alternative positive,  $\mu_A - \mu_B = \delta$ , et en assimilant cette valeur attendue à  $\delta\sqrt{\{n/(2\sigma^2)\}}$  nous trouvons la taille d'échantillon nécessaire pour être

$$n_f(\alpha, \beta, \delta, \sigma^2) = \{F^{-1}(1 - \alpha/2) + F^{-1}(1 - \beta)\}^2 2\sigma^2 / \delta^2 \quad (2.16)$$

sujets par bras de traitement. En pratique,  $n_f$  doit bien entendu être arrondi à un entier taille de l'échantillon [12].

### 2.4.3 Tests séquentiels de groupe

Dans le test séquentiel de groupe, par opposition à un test entièrement séquentiel, les données accumulées sont analysées à intervalles plutôt qu'après chaque nouvelle observation. La surveillance continue des données peut être un lourd fardeau pratique, et l'introduction de tests séquentiels de groupe a conduit à une utilisation beaucoup plus large de méthodes. Leur impact a été particulièrement évident dans les essais cliniques, ils offrent également de nombreuses possibilités d'arrêt précoce et peut obtenir la plupart des avantages des tests entièrement séquentiels en termes de faibles tailles d'échantillon attendues et longueurs moyennes d'études plus courtes.

Dans la comparaison de base de deux traitements, un maximum nombre de groupes,  $K$ , et une taille de groupe,  $m$ , sont choisis, les sujets sont attribués aux traitements selon un schéma de randomisation contraint qui assure  $m$  sujets reçoivent chaque traitement dans chaque groupe et les données accumulées sont analysés après chaque groupe de réponses de  $2m$ .

Pour chaque  $k = 1, \dots, K$ , un standard statistique  $Z_k$  est calculée à partir des  $k$  premiers groupes d'observations, et le test se termine par le rejet de  $H_0$  si  $|Z_k|$  dépasse une valeur critique  $c_k$ . Si le test poursuit l'analyse  $K$  ème et  $|Z_k| < c_k$ , il s'arrête à ce point et  $H_0$  est accepté. La séquence de valeurs critiques,  $\{c_1, \dots, c_k\}$ , est choisi pour atteindre une erreur

## Chapitre 2. Test Séquentiel

---

de type I spécifiée et différents types de test séquentiel de groupe donnent lieu à différentes séquences. La taille du groupe,  $m$ , est déterminée séparément par une puissance état.[12]

### 2.5 Test de Pocock

#### Définition 2.5.1.

les observations  $X_{Ai} \sim N(\mu_A, \sigma^2)$  et  $X_{Bi} \sim N(\mu_B, \sigma^2)$ ,  $i = 1, 2, \dots$ , sont disponibles dans les groupes de traitement A et B, respectivement. Le test de Pocock utilise la statistique standardisée après chaque groupe d'observations,

$$Z = \frac{1}{\sqrt{(2mk\sigma^2)}} \left( \sum_{i=1}^{mk} X_{Ai} - \sum_{i=1}^{mk} X_{Bi} \right), \quad K = 1, \dots, K. \quad (2.17)$$

Si, à la  $k$ ème analyse,  $k = 1, \dots, K$ , la valeur absolue de  $Z_k$  est suffisamment grande, l'étude s'arrête avec le rejet de  $H_0$ . Si  $H_0$  n'a pas été rejeté par la finale analyse,  $K$ , il est accepté.[12]

Le test de Pocock conserve la forme générale d'un test de signification appliqué à plusieurs reprises, rejetant  $H_0$  à l'étape  $k$  si  $|Z_k| \geq C_p(K, \alpha)$ ,  $k = 1, \dots, K$ , mais la valeur critique  $C_p(K, \alpha)$  est choisi pour donner une erreur globale de Type I  $\alpha$ , c'est-à-dire,

$$P_{\mu_A - \mu_B = 0} \{ \text{Rejeter } H_0 \text{ à l'analyse } k = 1, k = 2, \dots \text{ ou } k = K \} = \alpha.$$

Officiellement, le test est de la forme suivant :

Après le groupe  $k = 1, \dots, K - 1$

$$\begin{array}{ll} \text{Si } |Z_k| \geq C_p(K, \alpha) & \text{arrêt, rejet } H_0 \\ \text{Sinon} & \text{continuer à grouper } k + 1, \end{array} \quad (2.18)$$

Après le groupe K

$$\begin{array}{ll} \text{Si } |Z_k| \geq C_p(K, \alpha) & \text{arrêt, rejet } H_0 \\ \text{Sinon} & \text{arrêtez, acceptez } H_0. \end{array}$$

## Chapitre 2. Test Séquentiel

---

**Remarque 2.5.1.** (Chr. Jennison. 2000)

La valeur  $K = 1$  dans le tableau (3.6)(voir l'annexe) est appelée le cas non-séquentiel de  $H_0$ . Notez qu'à chaque analyse, le test rejette  $H_0$  si le niveau signification des deux côtés d'un test non-séquentiel de  $H_0$  en utilisant les données disponibles à ce moment est au-dessous

$$\alpha' = 2[1 - F(C_p(K, \alpha))]$$

telle que  $\alpha'$  est le niveau de signification nominale.

**Exemple 2.5.1.**

Si  $K = 5$  et  $\alpha = 0.05$  Le niveau de signification nominale appliqué à chaque analyse est

$$\alpha' = 2(1 - F(2.413)) = 0.0158$$

On remarque que  $\alpha' = 0.0158 < \alpha = 0.05$ , c-à-d un niveau de signification nominal inférieur à la probabilité globale d'erreur de 1<sup>re</sup> espèce est nécessaire pour éviter le problème des "multiples visions".

### 2.5.1 L'exigence de puissance

Le test défini par (3.5) a une probabilité d'erreur de type I  $\alpha$  pour toute taille de groupe,  $m$ . C'est la puissance requise

$$P_{\mu_A - \mu_B = \pm \delta} \{\text{Rejeter } H_0\} = 1 - \beta$$

qui détermine la taille de groupe appropriée. La taille d'échantillon maximale que le test séquentiel de groupe peut avoir besoin dépend de  $K$ ,  $\alpha$  et  $\beta$  et est proportionnel à  $\sigma^2/\delta^2$ . Etant donné que la taille de l'échantillon d'un test d'échantillonnage fixe avec la même erreur de type I  $\alpha$  et la puissance  $1 - \beta$  à  $\mu_A - \mu_B = \pm \delta$  est également proportionnelle à  $\sigma^2/\delta^2$ , il suffit de spécifier le rapport  $R_p(K, \alpha, \beta)$  de la taille maximale de l'échantillon du test séquentiel de groupe à la taille d'échantillon fixe, et ce rapport s'appliquera alors pour tous les  $\delta$  et  $\sigma^2$ . Nous allons voir également que les mêmes ratios peuvent être utilisés dans d'autres problèmes de test bilatéraux et avec d'autres types de données. Les valeurs de

## Chapitre 2. Test Séquentiel

---

$R_p(K, \alpha, \beta)$  dans le tableau (3.7) ont été trouvées par le même type de calcul numérique utilisé pour déterminer le  $C_p(K, \alpha)$ . [12]

**Définition 2.5.2.** (Chr. Jennison. 2000)

La taille de l'échantillon par bras de traitement,  $n_f(\alpha, \beta, \delta, \sigma^2)$ , requise par un échantillon fixe essai avec erreur de type I  $\alpha$  et puissance  $1 - \beta$  à  $\mu_A - \mu_B = \pm\delta$  est donné par (3.3).

Un test séquentiel de groupe Pocock avec  $K$  groupes nécessite un maximum d'observations  $R_p(K, \alpha, \beta)n_f(\alpha, \beta, \delta, \sigma^2)$  par bras de traitement et groupes de

$$m = R_p(K, \alpha, \beta)n_f(\alpha, \beta, \delta, \sigma^2)/K$$

observations sur chaque traitement. Arrondir  $m$  à une valeur entière aura un léger effet sur la puissance réelle du test, et il est habituel d'arrondir  $m$  vers le haut pour obtenir un test conservateur, c'est-à-dire avec un peu de puissance supérieur à  $1 - \beta$ . [12]

## 2.6 Test d'O'Brien et Fleming

En 1979, O'Brien et Fleming ont donné un test dans lequel le niveau de signification nominale nécessaires pour rejeter  $H_0$  à chaque augmentation de l'analyse, Ainsi que une alternative au test de signification répétée avec constante nominale.

**Définition 2.6.1.** (Chr. Jennison. 2000)

Les sujets sont attribués aux traitements en groupes, et les données accumulées sont analysées après terminés des groupes des données deviennent disponibles, et il est le terme de statistique standardisé  $Z_k$  de la forme (3.4), et  $H_0$  est rejeté après le groupe  $k$  si  $|Z_k| \geq c_k$  pour une séquence de valeurs critiques  $c_1, \dots, c_K$ . [12]

Le test se la forme :

Après le groupe  $k = 1, \dots, K - 1$

Si  $|Z_k| \geq C_B(K, \alpha)\sqrt{K/k}$       arrêt, rejet  $H_0$

    Sinon                                      continuer à grouper  $k + 1$ ,                                      (2.19)

Après le groupe  $K$

## Chapitre 2. Test Séquentiel

---

*Si*  $|Z_k| \geq C_B(K, \alpha)$                       *arrêt, rejet*  $H_0$   
*Sinon*    *arrêtez, acceptez*  $H_0$ .

Avec  $c_k = C_B(K, \alpha)\sqrt{(\frac{K}{k})}$ ,  $k = 1, \dots, K$  et  $C_B(K, \alpha)$  des valeurs pour les tests bilatéraux avec  $K$  groupes des observations et probabilité globale de type 1<sup>re</sup> espèce  $\alpha$  est fournie dans le tableau (3.8)(voir l'annexe).

D'autre part au niveau de signification nominale, si le niveau de signification bilatéral de  $H_0$  est inférieur à  $\alpha'_k = 2[1 - F(C_B(K, \alpha)\sqrt{(\frac{K}{k})})]$ , alors  $H_0$  est rejeté à l'analyse  $k$ ,  $k = 1, \dots, K$ . Puisque  $C_B(K, \alpha)\sqrt{(\frac{K}{k})}$  diminue avec l'augmentation de  $k$ , et  $\alpha'_k$  augmente.

### 2.6.1 L'exigence de puissance

Pour calculer la taille de l'échantillon on utilisons les constantes  $R_B(K, \alpha, \beta)$  pour le test de Pocock. Avec les constantes  $R_B(K, \alpha, \beta)$  représenter dans le tableau (3.9)(voir l'annexe), pour déterminer la tailles des groupes de tests bilatéraux avec  $K$  groupes d'observations, et probabilité d'erreur de 1<sup>re</sup> espèce  $\alpha$  et puissance  $1 - \beta$  la taille maximale de l'échantillon sur chaque bras de traitement est

$$R_B(K, \alpha, \beta)n_f(\alpha, \beta, \delta, \sigma^2)$$

Ou  $n_f(\alpha, \beta, \delta, \sigma^2)$  est le nombre d'observation par traitement nécessaire par un test de taille d'échantillon fixe avec la même erreur de 1<sup>re</sup> espèce et la même puissance, donnée par la forme (3.3) et la taille de groupe nécessaire par bras de traitement convenablement arrondi est

$$m = \frac{R_B(K, \alpha, \beta)n_f(\alpha, \beta, \delta, \sigma^2)}{K}$$

#### exemple

Supposons que  $\sigma^2 = 4$ , Donc  $X_{Ai} \sim N(\mu_A, 4)$  et  $X_{Bi} \sim N(\mu_B, 4)$ ,  $i = 1, 2, \dots$  et il est nécessaire de tester  $H_0 : \mu_A = \mu_B$  avec une erreur de type I  $\alpha = 0,05$  et puissance  $1 - \beta = 0,9$  à  $\mu_A - \mu_B = \pm 1$ . Le test d'échantillonnage fixe ce problème nécessite

## *Chapitre 2. Test Séquentiel*

---

$\{F^{-1}(0, 975) + F^{-1}(0, 9)\}^2 \times 2 \times 4 = 84,1$  sujets sur chacun traitement, que nous arrondissons à 85. L'hypothèse nulle est rejetée si :

$$\left| \sum_{i=1}^{85} X_{Ai} - \sum_{i=1}^{85} X_{Bi} \right| \geq 1,96\sqrt{(85 \times 2 \times 4)} = 51,1$$

et accepté autrement.[12]

# Chapitre 3

## Application numérique

### Introduction

Le SARS-COV-2 provoque actuellement une épidémie de Covid-19. Ce virus, de la famille des Coronaviridae, partage beaucoup de similitudes avec le SARS-CoV-1 de 2003, et partiellement avec le MERS-CoV (actuel), bien qu'il soit moins mortel.

Le SARS-CoV 2 est un  $\beta$  coronavirus du groupe 2b.

Le virus est un virus à ARN, qui pénètre dans la cellule via les récepteurs ACE-2.

Le SARS-CoV-2 exprime par ailleurs une exoribonucléase, lui conférant une certaine résistance aux antiviraux. Le taux de mutation est élevé (x8 au moins actuellement), et la recherche explore actuellement les mutations possibles en Iran. 2 sérotypes :

- Sérotype L (70 %) responsable de l'épidémie originale au Wuhan.
- Sérotype S (30 %, mais en augmentation), responsable de la majorité des cas hors Chine.

Le sérotype L était supposé plus agressif, mais des différences de mortalité initiale entre les sérotypes ne sont plus d'actualité. La prévalence actuelle du SARS-CoV-2 ne doit pas faire oublier les autres infections virales, ou les co-infections, qui même si elles sont souvent bénignes, représentent entre 3 et 5 % des cas.



### **3.1 le COVID-19**

#### **Qu'est-ce que le COVID-19 ?**

La maladie à coronavirus 2019 (COVID-19) est une infection aiguë des voies respiratoires apparue fin 2019<sup>1,2</sup>. Les flambées initiales en Chine concernaient 13,8% de cas avec des cours sévères et 6,1% avec des cours critiques<sup>3</sup>. Cette présentation sévère correspond à l'utilisation d'un récepteur viral qui s'exprime majoritairement dans le poumon<sup>2,4</sup>. En provoquant l'apparition précoce de symptômes sévères, on pense que ce même tropisme des récepteurs a déterminé la pathogénicité, mais a également aidé à contrôler le syndrome respiratoire aigu sévère (SRAS) en 2003<sup>5</sup>. Cependant, il y a des rapports de cas de COVID-19 avec des voies respiratoires supérieures légères symptômes, suggérant un potentiel de transmission pré- ou oligosymptomatique<sup>6-8</sup>. Il existe un besoin urgent d'informations sur la réplication virale, l'immunité et l'infectiosité spécifiques au site corporel.[14]

#### **Diagnostic virologique**

Le diagnostic de certitude repose sur l'identification du virus par RT-PCR effectuée sur prélèvements respiratoires. Cependant, devant l'accessibilité à la RT-PCR parfois limitée et la fréquence élevée de faux signes cliniques évocateurs et d'une image scannographique compatible.

L'utilisation de la sérologie est en cours d'évaluation.[13]

Le COVID-19 constitue une famille de virus qui sont connus pour leur capacité à franchir les barrières d'espèces et à en infecter de nouvelles. Ainsi nous avons assisté, depuis le début du XXI<sup>e</sup> siècle, à l'émergence de trois COVID-19 très pathogènes pour l'homme, la dernière ayant résulté en une pandémie mondiale avec des conséquences économiques et sociales qui se sont révélées dévastatrices. Une course contre la montre, à l'échelle mondiale, s'est engagée pour faire face à cette pandémie, en développant des traitements et un vaccin. Nous devons néanmoins aussi penser au futur et nous préparer au risque d'épidémies dues à l'émergence de nouveaux COVID-19 à transmission zoonotique. Cela nécessite, sans doute, une prise de conscience des modalités permettant le maintien de la barrière entre espèces et, surtout, le développement de traitements qui seraient efficaces

### *Chapitre 3. Application numérique*

---

contre toute forme de COVID-19, qu'ils soient très ou peu pathogènes pour l'homme, leur risque de transmission et leur dangerosité ne pouvant être, a priori, anticipés. L'identification de cibles thérapeutiques adéquates nécessite donc d'approfondir les connaissances que nous avons du cycle et de la pathogenèse de ces virus.[15]

#### **3.1.1 Appliquées des donnée réelle dans le COVID-19**

(D'après le Gautret, Philippe, et al. "Hydroxychloroquine and azithromycin as a treatment of COVID-19 : results of an open-label non-randomized clinical trial." International journal of antimicrobial agents (2020) : 105949.)

Le monde fait désormais face à la vague COVID-19, Ainsi, il existe un besoin urgent d'un traitement efficace pour traiter patients symptomatiques. Parmi les médicaments candidats pour traiter le COVID-19, le repositionnement les médicaments anciens à utiliser comme traitement antiviral est une stratégie intéressante car les connaissances sur la sécurité le profil, les effets secondaires, la posologie et les interactions médicamenteuses sont bien connus.

Un article récent a rapporté un effet inhibiteur du chloroquine (un ancien médicament antipaludique) sur la croissance du SRAS-CoV-2 in vitro. Il a été démontré que l'hydroxychloroquine (un analogue de la chloroquine) a un Activité du SRAS-CoV in vitro. Le profil de sécurité clinique de l'hydroxychloroquine est meilleur que celle de la chloroquine.

Donc ils ont mené un essai clinique visant à évaluer l'effet de hydroxychloroquine chez les patients infectés par le SRAS-CoV-2.

Cette étude en cours est coordonnée par l'Institut CHU Méditerranée Infection à Marseille. Les patients hospitalisés avec COVID-19 confirmé ont été inclus dans cette étude s'ils remplissaient deux critères principaux : i) âge > 12 ans ; ii) Portage du SARS-CoV-2 documenté par PCR. Les patients ont été vus au départ pour le recrutement, la collecte initiale des données et le traitement au jour 0, et à nouveau pour un suivi quotidien pendant 14 jours.

Les patients ont été regroupés en trois catégories : infection asymptomatique des voies respiratoires supérieures (URTI) en cas de rhinite, de pharyngite ou de fièvre isolée de bas grade et de myalgie, et infections des voies respiratoires inférieures (LRTI) en cas

### Chapitre 3. Application numérique

---

de symptômes de pneumonie ou bronchite. les résultats détaillés sont disponibles dans le tableau supplémentaire (3.1) Démographie et présentation clinique, Les résultats présentés ici sont donc ceux de 36 patients (20 patients traités par hydroxychloroquine et 16 contrôle des patients).

	Temps entre l'apparition des symptômes et l'inclusion (j)		
	moyen $\pm$ SD	t	p-valeur
N=20	4.1 $\pm$ 2.6	-0.15	0.88
N=16	3.9 $\pm$ 2.8	-0.15	0.88
Total (36)	4 $\pm$ 2.6		

TABLE 3.1 – Caractéristiques de la population étudiée

Soit  $X_{Ai} \sim N(\mu_A, 2.6)$  et  $X_{Bi} \sim N(\mu_B, 2.6)$ ,  $i = 1, 2, \dots$ , indiquent les réponses des sujets deux traitement, A et B telle que A est le Hydroxychloroquine et B est le contrôle et il est nécessaire de tester  $H_0 : \mu_A = \mu_B$  avec une erreur de type I  $\alpha = 0, 05$  et puissance  $1 - \beta = 0, 9$  et d'après le tableau (3.1)  $\sigma^2 = 2.6$  et  $\mu_A - \mu_B = \pm 0.2$ . A partir de (1.3), le test d'échantillonnage fixe ce problème nécessite  $\{F^{-1}(0, 975) + F^{-1}(0, 9)\}^2 \times 2 \times 2.6 / 0.2 = 273.325$  sujets sur chacun traitement, que nous arrondissons à 274. L'hypothèse nulle est rejetée si

$$\left| \sum_{i=1}^{274} X_{Ai} - \sum_{i=1}^{274} X_{Bi} \right| \geq 1.96 \sqrt{(274 * 2 * 2.6)} = 73, 98$$

et accepté autrement. Nous reviendrons sur cet exemple pour illustrer et comparer différents types de test séquentiel de groupe.

#### calculer $Z_k$

#### Dans le cas de test Pocock

De Tableau (3.7),  $R_p(5, 0.05, 0.1) = 1.207$ . Par conséquent, un test de Pocock avec cinq groupes de les observations ont une taille d'échantillon maximale par bras de traitement

$$R_p(5, 0.05, 0.1)n_f(0.05, 0.1, 0.2, 2.6) = 1.207 * 273.325 = 329.9$$

### Chapitre 3. Application numérique

---

et la taille de groupe nécessaire est  $m = 329.9/5 = 65.98$  par traitement, que nous arrondir à 66. Le test s'arrête avec le rejet de l'hypothèse nulle à l'analyse  $k$ ,  $k = 1, \dots, 5$  si

$$\left| \left( \sum_{i=1}^{66k} X_{Ai} - \sum_{i=1}^{66k} X_{Bi} \right) \right| \geq 2.413 \sqrt{(66K * 2 * 2.6)} = 44.7 \sqrt{K}$$

Si l'hypothèse nulle n'a pas été rejetée par l'analyse 5, elle est acceptée à ce point.

Pour simplifier la limite du test de Pocock par cet exemple qui a donné la figure (3.1) et comme suivant :

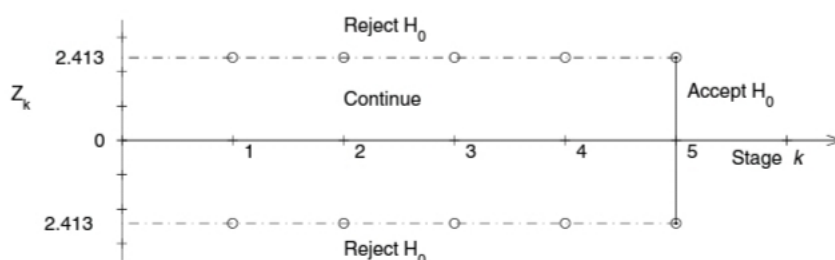


FIGURE 3.1 – Test de Pocock avec 5 groupe d'observation

- Le diagramme représente les valeurs critiques appliquées aux statistiques standardisées  $Z_K$ , sont des constants sur les 5 analyses pour le test de Pocock.

#### Dans le cas de test O'Brien et Fleming

De Tableau (3.9),  $R_B(5, 0.05, 0.1) = 1.026$ . Par conséquent, un test de Pocock avec cinq groupes de les observations ont une taille d'échantillon maximale par bras de traitement

$$R_B(5, 0.05, 0.1) n_f(0.05, 0.1, 0.2, 2.6) = 1.026 * 273.325 = 280.4$$

et la taille de groupe nécessaire est  $m = 280.4/5 = 56.08$  par traitement, que nous arrondir à 57. Le test s'arrête avec le rejet de l'hypothèse nulle à l'analyse  $k$ ,  $k = 1, \dots, 5$  si

$$\left| \left( \sum_{i=1}^{57k} X_{Ai} - \sum_{i=1}^{57k} X_{Bi} \right) \right| \geq 2.040 \sqrt{5/k} \sqrt{(57K * 2 * 2.6)} = 78.5$$

### Chapitre 3. Application numérique

---

Si l'hypothèse nulle n'a pas été rejetée par l'analyse 5, elle est acceptée à ce point.

Pour simplifier la limite du test de d'O'Brien et Fleming par cet exemple qui a donné la figure (3.2) et le tableau (3.2) suivant :

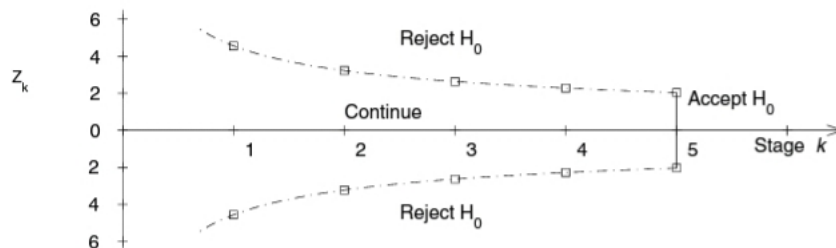


FIGURE 3.2 – Test O'Brien et Fleming avec 5 groupe d'observation

- Le diagramme représente les valeurs critiques appliquées aux statistiques standardisées  $Z_K$ , sont des constants sur les 5 analyses pour le test O'Brien et Fleming.

calculer  $D_k$

Dans le cas de test Pocock

K	$D_k$
1	44.70
2	63.22
3	77.42
4	89.4
5	99.95

TABLE 3.2 – Test de Pocock avec 5 groupe d'observation

- Le tableau représente, le test de Pocock est formulé en terme de différence non normalisées des sommes d'échantillon sur les traitements A et B,

$$D_K = \sum_{i=1}^{66K} X_{Ai} - \sum_{i=1}^{66K} X_{Bi}, K = 1, \dots, 5,$$

- Le test de Pocock utilise le nombre total maximum de sujets sur les deux bras de traitement nécessaire est 660. Ceci est un peu plus grand que le 548 nécessaire pour un test d'échantillon fixe.

#### Dans le cas de test O'Brien et Fleming

K	$D_k$
1	78.5
2	78.5
3	78.5
4	78.5
5	78.5

TABLE 3.3 – Test O'Brien et Fleming avec 5 groupe d'observation

- Le tableau représente, le test d'O'Brien et Fleming est formulé en terme de différence non normalisées des sommes d'échantillon sur les traitements A et B,

$$D_K = \sum_{i=1}^{57K} X_{Ai} - \sum_{i=1}^{57K} X_{Bi}, K = 1, \dots, 5,$$

- Le test de O'Brien et Fleming utilise le nombre total maximum de sujets sur les deux bras de traitement nécessaire est 570. Ceci est un peu plus grand que le 548 nécessaire pour un test d'échantillon fixe.

#### 3.1.2 Propriétés des tests Pocock et O'Brien et Fleming :

##### Comparaison des tests séquentiels de groupe :

A partir de la comparaison des tests séquentiels de groupe peuvent être considérées un certain nombre de propriétés.

### *Chapitre 3. Application numérique*

---

- 1 Pour la comparaison qualitative entre le teste de Pocock et le teste O'Brien et Fleming sont représentés sur les figures (3.1) et (3.2) et les tableaux (3.2) et (3.3) : le test de Pocock a premièrement des limites plus étroites, ce qui permet d'un arrêt très pressé. Tandis que le test O'Brien et Fleming a des limites plus étroites à la fin des analyses et une plus petite taille maximale de l'échantillon.
- 2 Pour la comparaison quantitative, un certain nombre de tests sont disponibles. Les probabilités d'erreur sont une propriété déterminante de tout test, et il est important de comparer correctement des tests "appariés". Cependant, les probabilités d'erreur de 1<sup>re</sup> espèce et de 2<sup>e</sup> espèce sont précisées, et la puissance du test est énormément déterminée sur la plage des valeurs des paramètres possibles.
- 3 La taille de l'échantillon fixe de 548 sujets présente un point de référence par rapport à la comparaison des tests séquentiels, la taille d'échantillon à  $\mu_A - \mu_B = 0$  est plus attendue que cette valeur de référence, puisque pour accepter  $H_0$ , les tests séquentiels du groupe doivent continuer à leur taille maximale d'échantillon c'est pour le test de Pocock est 660, et le test de O'Brien et Fleming est 570. Et la différence de traitement est grande il est important pour l'éthique d'arrêter et puisque la décision précoce évite la randomisation.

### **Conclusion**

L'utilisation de test séquentiel est de diminuer la taille de l'échantillon. La distribution de la taille de l'échantillon dépend de la valeur de  $\mu_A - \mu_B$ , et il faut considérer la taille de l'échantillon sur une plage des valeurs de  $\mu_A - \mu_B$  lors de réaliser des comparaisons. Les tests Pocock ont des tailles d'échantillon maximales assez élevées et leurs tailles d'échantillon attendues sont plus élevées que le test d'échantillon fixe lorsque  $|\mu_A - \mu_B|$  est petit.

## **Annexe :Les tableaux statistiques**



## Chapitre 3. Application numérique

---

### Théorème de la moyenne

**Théorème 3.1.1.** (*E.Hairer.2000*)

Si  $f : [a, b] \rightarrow \mathbb{R}$  est une fonction continue, alors il existe un  $\xi \in [a, b]$  tel que

$$\int_a^b f dx = f(\xi)(b - a)$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5039	0.5079	0.5119	0.5159	0.5199	0.5239	0.5279	0.5318	0.5358
0.10	0.5398	0.5437	0.5477	0.5517	0.5556	0.5596	0.5635	0.5674	0.5714	0.5753
0.20	0.5792	0.5831	0.5870	0.5909	0.5948	0.5987	0.6025	0.6064	0.6102	0.6140
0.30	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6405	0.6443	0.6480	0.6517
0.40	0.6554	0.6590	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6843	0.6879
0.50	0.6914	0.6949	0.6984	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.60	0.7257	0.7290	0.7323	0.7356	0.7389	0.7421	0.7453	0.7485	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7703	0.7733	0.7763	0.7793	0.7823	0.7852
0.80	0.7881	0.7910	0.7938	0.7967	0.7995	0.8023	0.8051	0.8078	0.8105	0.8132
0.90	0.8159	0.8185	0.8212	0.8238	0.8263	0.8289	0.8314	0.8339	0.8364	0.8389
1.00	0.8413	0.8437	0.8461	0.8484	0.8508	0.8531	0.8554	0.8576	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8707	0.8728	0.8749	0.8769	0.8789	0.8810	0.8829
1.20	0.8849	0.8868	0.8887	0.8906	0.8925	0.8943	0.8961	0.8979	0.8997	0.9014
1.30	0.9031	0.9049	0.9065	0.9082	0.9098	0.9114	0.9130	0.9146	0.9162	0.9177
1.40	0.9192	0.9207	0.9221	0.9236	0.9250	0.9264	0.9278	0.9292	0.9305	0.9318
1.50	0.9331	0.9344	0.9357	0.9369	0.9382	0.9394	0.9406	0.9417	0.9429	0.9440

TABLE 3.4 – Les valeurs de  $\mathbf{P}(Z < z)$  en fonction de  $z$

### Chapitre 3. Application numérique

---

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.60	0.9452	0.9463	0.9473	0.9484	0.9494	0.9505	0.9515	0.9525	0.9535	0.9544
1.70	0.9554	0.9563	0.9572	0.9581	0.9590	0.9599	0.9607	0.9616	0.9624	0.9632
1.80	0.9640	0.9648	0.9656	0.9663	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.90	0.9712	0.9719	0.9725	0.9731	0.9738	0.9744	0.9750	0.9755	0.9761	0.9767
2.00	0.9772	0.9777	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9816
2.10	0.9821	0.9825	0.9829	0.9834	0.9838	0.9842	0.9846	0.9849	0.9853	0.9857
2.20	0.9860	0.9864	0.9867	0.9871	0.9874	0.9877	0.9880	0.9883	0.9886	0.9889
2.30	0.9892	0.9895	0.9898	0.9900	0.9903	0.9906	0.9908	0.9911	0.9913	0.9915
2.40	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.50	0.9937	0.9939	0.9941	0.9942	0.9944	0.9946	0.9947	0.9949	0.9950	0.9952
2.60	0.9953	0.9954	0.9956	0.9957	0.9958	0.9959	0.9960	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9971	0.9972	0.9973
2.80	0.9974	0.9975	0.9975	0.9976	0.9977	0.9978	0.9978	0.9979	0.9980	0.9980
2.90	0.9081	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.00	0.9986	0.9986	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9989	0.9989
3.10	0.9990	0.9990	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992
3.20	0.9993	0.9993	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994
3.30	0.9995	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996
3.40	0.9996	0.9996	0.9996	0.9996	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
3.50	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.60	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.70	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.80	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.90	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

TABLE 3.5 – les valeurs de  $\mathbf{P}(Z < z)$  en fonction de  $z$  (suite)

K	$C_P(K, \alpha)$		
	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$
1	2.576	1.960	1.645
2	2.772	2.178	1.875
3	2.873	2.289	1.992
4	2.939	2.361	2.067
5	2.986	2.413	2.122
6	3.023	2.453	2.164
7	3.053	2.485	2.197
8	3.078	2.512	2.225
9	3.099	2.535	2.249
10	3.117	2.555	2.270
11	3.133	2.572	2.288
12	3.147	2.588	2.304
15	3.182	2.626	2.344
20	3.225	2.672	2.392

TABLE 3.6 – **Test de Pocock** : Les constantes  $C_P(K, \alpha)$

$R_P(K, \alpha, \beta)$						
K	1 - $\beta=0.8$			1 - $\beta=0.9$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	1.000	1.000	1.000	1.000	1.000	1.00
2	1.092	1.110	1.121	1.084	1.100	1.110
3	1.137	1.166	1.184	1.125	1.151	1.166
4	1.166	1.202	1.224	1.152	1.183	1.202
5	1.187	1.229	1.254	1.170	1.207	1.228
6	1.203	1.249	1.277	1.185	1.225	1.249
7	1.216	1.265	1.296	1.197	1.239	1.266
8	1.226	1.279	1.311	1.206	1.252	1.280
9	1.236	1.291	1.325	1.215	1.262	1.292
10	1.243	1.301	1.337	1.222	1.271	1.302
11	1.250	1.310	1.348	1.228	1.279	1.312
12	1.257	1.318	1.357	1.234	1.287	1.320
15	1.272	1.338	1.381	1.248	1.305	1.341
20	1.291	1.363	1.411	1.264	1.327	1.367

TABLE 3.7 – **Test de Pocock** : Les constantes  $R_P(K, \alpha, \beta)$

K	$C_B(K, \alpha)$		
	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$
1	2.576	1.960	1.645
2	2.580	1.977	1.678
3	2.595	2.004	1.710
4	2.609	2.024	2.733
5	2.621	2.040	2.751
6	2.631	2.053	2.765
7	2.640	2.063	2.776
8	2.648	2.072	2.786
9	2.654	2.080	2.794
10	2.660	2.087	2.801
11	2.665	2.092	2.807
12	2.670	2.098	2.813
15	2.681	2.110	2.826
20	2.695	2.126	2.842

TABLE 3.8 – **Test d’O’Brien et Fleming** : Les constantes  $C_B(K, \alpha)$

$R_B(K, \alpha, \beta)$						
K	1 - $\beta=0.8$			1 - $\beta=0.9$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	1.000	1.000	1.000	1.000	1.000	1.00
2	1.001	1.008	1.016	1.001	1.007	1.014
3	1.007	1.017	1.027	1.006	1.016	1.025
4	1.011	1.024	1.035	1.010	1.022	1.032
5	1.015	1.028	1.040	1.014	1.026	1.037
6	1.017	1.032	1.044	1.016	1.030	1.041
7	1.019	1.035	1.047	1.018	1.032	1.044
8	1.021	1.037	1.049	1.020	1.034	1.046
9	1.022	1.038	1.051	1.021	1.036	1.048
10	1.024	1.040	1.53	1.022	1.037	1.049
11	1.025	1.041	1.054	1.023	1.039	1.051
12	1.026	1.042	1.055	1.024	1.040	1.052
15	1.028	1.045	1.058	1.026	1.042	1.054
20	1.030	1.047	1.061	1.029	1.045	1.057

TABLE 3.9 – **Test d’O’Brien et Fleming** : Les constantes  $R_B(K, \alpha, \beta)$

# Conclusion

Dans ce travail on a traité le test séquentiel du rapport des probabilités et nous l'avons appliqué sur un exemple réel(COVID-19). L'utilisation du test séquentiel du rapport des probabilités permettait souvent une économie moyenne d'au moins 50% sur le nombre d'observations par rapport à une procédure classique. Donc On peut alors être tenté d'abandonner les tests classiques au profit des tests séquentiels.

# Bibliographie

- [1] Fabrice, Heitz. "STATISTIQUES. Télécom Physique Strasbourg. Cours I : Test d'hypothèses." Module 2101. Octobre (2014).
  
- [2] Lecoutre, Jean-Pierre. "Statistique et probabilités." Dunod, (2002).
  
- [3] Bertrand, Frédéric and Maumy-Bertrand, Myriam, "Initiation à la statistique avec R : cours, exemples, exercices et problèmes corrigés" (2010).
  
- [4] Etienne, Birmelé. "Tests Statistiques." M1IMSV.
  
- [5] Jean-Jacques, Ruch. "STATISTIQUE , TESTS D'HYPOTHESES.", Préparation à l'Agrégation Bordeaux 1.(2012-2013).
  
- [6] Pierre, Neuvial. Démarche Statistique 1 , Introduction aux tests. Evry, M1 SGO, automne 2014. <http://stat.genopole.cnrs.fr/members/pneuvial/demstat>.
  
- [7] Savy, Nicolas. "Probabilités et statistiques pour modeliser et decider : tests, validation, régression, plans d'expérience." Ellipses, (2006).
  
- [8] Gaudoin, Olivier, and Maryse BÉGUIN. "Principes et méthodes statistiques." Ensimag-2ème Année, INP Grenoble (2009).



### *Chapitre 3. Application numérique*

---

- [9] Wetherill, G. Barrie. Méthodes séquentielles en statistique : " Sequential methods in statistics", par G. Barrie Wetherill,... Traduit par P. Pierre Gabe,... Dunod, (1969).
  
- [10] Illig, A. "Probabilités de confiance après décision lors d'un test séquentiel avec un temps de bon fonctionnement de loi exponentielle." Mémoire de DEA sous la direction du professeur B. Garel, Université Toulouse III (2001).
  
- [11] HÉDELIN, Guy. "Principe des tests séquentiels."
  
- [12] CHAPMAN and HALL/CRC, GROUP SEQUENTIAL METHODS with APPLICATIONS to CLINICAL TRIALS, Boca Raton London New York Washington, D.C.
  
- [13] DE BONNES PRATIQUES, DE PRISE EN, et al. "ROYAUME DU MAROC." (2020).
  
- [14] Ehmman, Katrin Zwirgmaier, et al. "Virological assessment of hospitalized cases of coronavirus disease 2019." (2020).
  
- [15] Juckel, Dylan, Jean Dubuisson, and Sandrine Belouzard. "Les coronavirus, ennemis incertains." médecine/sciences 36.6-7 (2020) : 633-641.
  
- [16] Gautret, Philippe, et al. "Hydroxychloroquine and azithromycin as a treatment of COVID-19 : results of an open-label non-randomized clinical trial." International journal of antimicrobial agents (2020) : 105949.
  
- [17] Khaled Dib, Tests D'hypothèses dans un Modèle de Censure, Mémoire Option Probabilités et Statistiques.