

UNIVERSITÉ MOHAMMED SEDDIK BENYAHIA  
JIJEL  
FACULTÉ DE SCIENCES EXACTES ET D'INFORMATIQUE

N° d'ordre : ...



**MEMOIRE DE MASTER**

Présentée pour l'obtention du diplôme de :

**MASTER**

En **Mathématique**

**Option** : Probabilités et Statistique

Par :

**Guedda Nadjah**

Thème

**Tests statistiques entre approches classique et  
baysésienne**

Soutenue le : 14/09/2022 , devant le jury composé de :

|                      |                     |            |
|----------------------|---------------------|------------|
| Mme. Yakoubi Fatima  | Grade à institution | Président  |
| Mme. Ghouil Djoweyda | Grade à institution | Encadrante |
| Mme. Roula Amel      | Grade à institution | Examineur  |

# Remerciements

*Au terme de la rédaction de cette thèse, il est de mon devoir d'exprimer en quelques lignes la gratitude que je dois à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce travail. Je remercie Dieu tout-puissant qui m'a donné tant de courage, santé, volonté, patience et abnégation pour mener à bien cette thèse.*

*La première personne à qui je tiens à exprimer ma plus profonde gratitude est ma directrice " Mme. Ghouil Djoweyda ", enseignante au département de mathématiques pour avoir accepté mon encadrement, pour son aide, ses précieuses recommandations et sa disponibilité tout au long ma période de travail.*

*J'adresse également mes sincères remerciements aux membres du jury, " Roula Amel " et " Mme. Yakoubi Fatima " pour l'intérêt porté à cette étude et pour accepter le jugement de ce travail.*

*Un remerciement tout particulier à " Mr. Ghareda Mebrouk ", le défunt, que Dieu lui fasse miséricorde et lui pardonne et le récompense pour sa gentillesse envers nous, ancien enseignant au département de mathématiques qui a servi de père à tous les étudiants, et qui n'a ménagé aucun effort pour que nous apprenions aussi pour Mme " Laoudj Farida " pour son aide précieux, ses conseils et son cœur qui était avec nous malgré les distances, qu'Allah vous récompense mille bien.*

*Je voudrais également remercier tous mes professeurs pour leur participation à mon enseignement et pour m'avoir aidé en particulier. Depuis la première fois que je suis venu de mon état, je ne me suis pas senti aliéné, surtout " Zahra Djeridi ", " Cheraitia Hassen ", " Abdi Zeyneb ", "Madi Meriem " que Dieu vous récompense mille et le fasse dans la balance de vos bonnes actions.*

*Je voudrais remercier mes parents dans cette lettre, sans leur soutien et leur amour je ne serais pas là où je suis.*

# *Dédicace*

*A celui qui a avalé la coupe vide pour me donner une goutte d'amour, à celui qui a récolté les épines de mon chemin pour m'ouvrir la voie de la connaissance, ma mère "Noura Dida" et mon père " Ibrahim ", je dédie à vous mon diplôme et la récolte de ce que j'ai semé pendant de nombreuses années pour le savoir et me voici aujourd'hui à un seuil qui était hier un rêve et aujourd'hui une réalité*

*A mon frère " Abd El-Fattah ", qui a été mon soutien tout au long de mon exil et de mes études, je demande à Dieu que tu t'élèves dans les Daryans, et je n'oublie pas mes frères et mes sœurs le confort de mes yeux " Manal ", "Rehana", "Hadj Al-Eid" et "Ahmed Chouaib" Je vous souhaite un parcours scientifique et pratique réussi.*

*Mes grand-pères et mes grand-mères, "Mabrouka", "Arabi", "Houria" et "Mouhammad" que Dieu vous bénisse. Pour mes tantes et mes oncles*

*À tous ceux qui m'ont aidé et se sont tenus à mes côtés au moment de ma détresse, mes amis "Farid", "Abou-Baker", "Fares", "Bader Al-Dinne", "Iman", "Itiraf", "Haïfa", "Samiha", "Amira", "Jihan", "Fella", "Darine", "Amina". Vous avez mes salutations et mes meilleurs vœux de succès dans votre vie.*

*Et à tous ceux qui ont voulu ma chute, je vous remercie infiniment pour la volonté que vous avez plantée en moi.*

# Table des matières

|   |           |
|---|-----------|
| <b>Liste des tableaux</b>   | <b>iv</b> |
| <b>Résumé</b>   | <b>v</b>  |
| <b>Abstract</b>   | <b>vi</b> |
| <b>Introduction</b>   | <b>vi</b> |
| <b>1 Les tests statistique par l'approche classique</b>                       | <b>1</b>  |
| 1.1 Introduction . . . . .  | 1         |
| 1.2 Bases des tests statistiques . . . . .                                    | 1         |
| 1.2.1 Test de conformité d'une proportion . . . . .                           | 7         |
| 1.2.2 Test de conformité d'une variance . . . . .                             | 8         |
| 1.3 Les tests d'homogénéité : Comparaison de deux valeurs observées . . . . . | 11        |
| 1.3.1 Homogénéité des moyennes . . . . .                                      | 12        |
| 1.3.2 Homogénéité des proportions . . . . .                                   | 14        |
| 1.3.3 Homogénéité des variances . . . . .                                     | 15        |
| 1.4 Analyse de la variance : ANOVA . . . . .                                  | 16        |
| 1.4.1 ANOVA à un facteur contrôlé . . . . .                                   | 17        |
| <b>2 Les tests statistique par l'approche Bayésienne</b>                      | <b>20</b> |
| 2.1 Introduction . . . . .  | 20        |
| 2.2 Principe de la statistique Bayésienne . . . . .                           | 21        |
| 2.3 La loi a posteriori . . . . .   | 22        |
| 2.4 Choix de la loi a priori . . . . .  | 24        |

|          |  |           |
|----------|--|-----------|
| 2.5      | Une introduction à la théorie de la décision . . . . .                               | 25        |
| 2.5.1    | Fonction de perte et risque . . . . .  | 25        |
| 2.5.2    | Exemples . . . . .   | 29        |
| 2.6      | Tests et régions de confiance . . . . .  | 33        |
| 2.6.1    | Régions de confiance . . . . .   | 33        |
| 2.6.2    | Tests . . . . .  | 34        |
| 2.6.2.1  | Approche par la fonction de perte de type 0 – 1 . . . . .                            | 34        |
| 2.7      | Évaluation d’hypothèses informatives à l’aide du facteur de Bayes . . . . .          | 36        |
| 2.7.1    | Le facteur de Bayes . . . . .  | 36        |
| 2.7.2    | Propriétés asymptotiques des facteurs de Bayes . . . . .                             | 37        |
| 2.7.3    | Calcul du facteur de Bayes . . . . .   | 39        |
| <b>3</b> | <b>Exemples et Discussion</b>  | <b>42</b> |
| 3.1      | Exemples . . . . .   | 42        |
| 3.1.1    | Exemple de test d’hypothèse dans l’approche classique . . . . .                      | 42        |
| 3.1.2    | Exemples des test d’hypothèse dans l’approche bayésienne . . . . .                   | 45        |
| 3.2      | Discussion . . . . .   | 47        |
| 3.3      | A/B testing : statistiques bayésiennes vs fréquentistes, un mauvais combat . . . . . | 47        |
| 3.3.1    | Statistiques fréquentistes et bayésiennes . . . . .                                  | 47        |
| 3.3.2    | L’intérêt de l’approche bayésienne et ses limites . . . . .                          | 48        |
| 3.3.3    | Puissance de la méthode fréquentiste . . . . .                                       | 49        |
| 3.3.4    | Quelle approche est meilleure de l’autre ? . . . . .                                 | 49        |
|          | <b>Conclusion</b>  | <b>ix</b> |
|          | <b>Bibliographie</b>   | <b>x</b>  |

# Liste des tableaux

|     |  |    |
|-----|--|----|
| 1.1 | Tablea d'ANOVA [8] . . . . .   | 18 |
| 2.1 | Loi a priori conjugée . . . . .  | 25 |
| 3.1 | Le taux d'hémoglobine chez certains patients et ceux en bonne santé. . . . . | 43 |

# Résumé

Le test d'hypothèse est un problème de décision, il peut être paramétrique ou non paramétrique comme il peut être appliqué sur deux hypothèses ou plus. Il consiste à tester la validité d'une hypothèse sur une mesure de population inconnue ou sur la distribution de la population elle-même sur la base de données d'échantillon aléatoires ; et donc de choisir une des hypothèses parmi les autres.

Ce travail aborde les démarches pour effectuer un test statistique paramétrique par les deux approches : fréquentiste et Bayésienne.

# Introduction Générale

Les statisticiens utilisent les termes «test d'hypothèse» et «test de signification» pour décrire deux processus permettant d'explorer si les paramètres d'un modèle prennent des valeurs spécifiques ou se situent dans certaines plages. À son niveau le plus élémentaire, cela peut signifier tester un seul paramètre.

Le test d'hypothèse signifie tester la validité d'une hypothèse sur une mesure de population inconnue ou sur la distribution de la population elle-même sur la base de données d'échantillon aléatoires.

Les hypothèses d'un point de vue statistique sont des relations qui sont prédites entre deux ou plusieurs variables, car elles représentent des solutions à des problèmes posés par une recherche, qui sont dérivées d'un ensemble de fondements théoriques et des exigences de son étude, et dont la validité est déduite en laissant tomber sur le test, où les hypothèses statistiques sont divisées en deux types de base : l'hypothèse nulle et l'hypothèse alternative.

L'hypothèse nulle est l'hypothèse qu'il n'y a pas de relation entre deux variables statistiques ou qu'il n'y a pas de différences statistiquement significatives entre deux variables A et B. L'hypothèse alternative est celle opposée à l'hypothèse nulle et elle confirme l'existence d'une relation entre deux variables. variables ou l'existence de différences entre elles. Ce dernier peut être dirigé ou non dirigé.

La situation de test d'hypothèse est similaire à la situation devant un tribunal, où l'accusé est initialement présumé innocent (l'hypothèse nulle) et ici deux types d'erreurs sont pos-

sibles : juger un innocent coupable (type I) ou juger un coupable personne innocente (Type II) .

Notez que deux hypothèses sont incluses dans la méthode *Niemann-Pearson* de test d'hypothèse : l'hypothèse nulle et l'hypothèse alternative. Dans le test de signification, seule l'hypothèse nulle est testée ; L'objectif est de "rejeter" l'hypothèse nulle si la valeur statistique testée est suffisamment différente de ce à quoi on pourrait s'attendre sous l'hypothèse nulle, ou de "ne pas la rejeter" si la valeur est pas trop extrême. Aucune hypothèse alternative n'est clairement énoncée. Le terme « *p – valeur* » est utilisé pour décrire la probabilité que nous détectons une valeur statistique testable aussi extrême que celle observée dans la réalité, ou même plus extrême si l'hypothèse nulle est vraie.

Les notions d'hypothèse nulle et de test de signification sont développées pour un large éventail de problèmes, certains tests étant développés et souvent nommés d'après l'un de leurs développeurs originaux (par exemple, le test de *Wald*, le test de *Mann-Whitney*), ou nommés d'après la distribution de la statistique respective testée (par exemple test *T*, test carré de *Kay*).

Les tests d'hypothèse bayésiens sont, au moins superficiellement, plus simples. Selon le théorème de Bayes, nous avons des probabilités dimensionnelles que chaque hypothèse est vraie ; Ensuite, nous pouvons l'utiliser pour choisir l'une des hypothèses. En pratique, les choses sont parfois plus compliquées.

Les tests d'hypothèses statistiques jouent un rôle important dans les statistiques globales et dans l'inférence statistique. Par exemple, *Lyman* (1992) déclare dans une revue de l'article de base de *Niemann et Pearson* (1933) : Il joue un rôle central à la fois dans la théorie et dans la pratique des statistiques, et on peut s'attendre à ce qu'il le fasse dans un avenir prévisible.

Le but de cette recherche est de comparer les tests paramétriques classiques et bayésiens, et d'essayer de trouver lequel est le meilleur.

Ans ce mémoire est organisé comme suit : le premier chapitre est un aperçu des différents types de tests paramétriques dans l'approche fréquentiste, à savoir : les tests de conformité à une valeur théorique, les tests d'homogénéité entre deux valeurs expérimentales, le test ANOVA, le unidirectionnel et le bidirectionnel.

Le deuxième aborde l'approche Bayésienne des tests paramétriques. Enfin dans le troisième chapitre nous avons donné des exemples sur chaque approche avec une discussion.

# Les tests statistique par l'approche classique

## 1.1 Introduction

Un test paramétrique est un test pour lequel on fait une hypothèse paramétrique sur la loi des données sous  $H_0$  (loi normale, loi de Poisson...); Les hypothèses du test concernent alors les paramètres de cette loi.

Dans ce chapitre nous énonçons (ou rappelons) un certain nombre de généralités autour des tests d'hypothèse, l'objectif étant d'être capable de bien formuler un test paramétrique classique : tests de normalité des distributions parentes, tests d'homogénéité des variances. Loi de Fischer Snedecor, analyse de variance.

## 1.2 Bases des tests statistiques

### 1. Hypothèses [4]

On oppose deux hypothèses complémentaires :  $H_0$  et  $H_1$ ,

- l'hypothèse  $H_0$  formule ce que l'on souhaite rejeter/réfuter,
- l'hypothèse  $H_1$  formule ce que l'on souhaite montrer.

Par exemple, si on veut montrer l'hypothèse "lot non conforme",  $H_0$  et  $H_1$  s'opposent sous la forme :

$$H_0 : \text{"lot conforme"} \quad \text{contre} \quad H_1 : \text{"lot non conforme"}.$$

**2. La statistique du test** : c'est une fonction qui résume l'information sur l'échantillon (ou valeur) qu'on veut tester. On la choisit de façon à pouvoir calculer sa loi sous  $H_0$ .

**3. La région critique** : c'est la région de rejet de l'hypothèse nulle  $H_0$  : on rejette  $H_0$  si la

valeur observée de la statistique calculée à partir des données, appartient à la région de rejet.

**4. Test bilatéral** : c'est lorsque la région critique est partagée en deux parties, et dans ce cas

$$\text{le test prend la forme suivante : } \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

La région critique est alors de la forme :  $] - \infty, a] \cup [b, +\infty[$ .

**5. Test unilatéral** : c'est lorsque la région critique est présentée par une seule partie. Ce type

$$\text{de tests prend la forme suivante : } \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \leq \theta_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Et la région critique prend une des deux formes suivantes :  $] - \infty, a]$  ou  $[a, +\infty[$ .

## 2. Le risque d'erreur

Un test statistique doit aboutir à choisir une des deux hypothèses  $H_0$  et  $H_1$ . Il ya quatre solutions dont seulement deux sont justes :

1.  $H_0$  est vraie et on a choisi  $H_0$ .
2.  $H_0$  est fausse et on a rejeté  $H_0$ .
3.  $H_0$  est vraie et on a rejeté  $H_0$ .
4.  $H_0$  est fausse et on a choisi  $H_0$ .

On peut résumer ces différent cas de décisions dans le tableau suivant :

|                   |       | Hypothèse vraie | Hypothèse vraie |
|-------------------|-------|-----------------|-----------------|
|                   |       | $H_0$           | $H_1$           |
| Hypothèse retenue | $H_0$ | $1 - \alpha$    | $\beta$         |
| Hypothèse retenue | $H_1$ | $\alpha$        | $1 - \beta$     |

On appelle risque de première espèce et on note  $\alpha$ , la probabilité de rejeter à tort l'hypothèse nulle  $H_0$  alors qu'elle est vraie.

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

Le risque de première espèce  $\alpha$  est aussi appelé " seuil de signification du test".

La quantité  $1 - \alpha$  est appelée niveau de confiance du test.

On appelle risque de deuxième espèce et on note  $\beta$ , la probabilité d'accepter l'hypothèse nulle  $H_0$  alors qu'elle est fausse.

$$\beta = P(\text{accepter } H_0 | H_0 \text{ fausse})$$

$\beta$  se détermine par un calcul de probabilité si  $H_1$  est précisément définie (car  $\alpha$  étant fixé). La quantité  $1 - \beta$ ; qui est la probabilité de rejeter  $H_0$  alors qu'elle est fausse; est appelée "la puissance du test".

### 3. Le mécanisme général d'un test statistique

La méthodologie des tests consiste à l'aide des résultats expérimentaux à une question concernant les paramètres de la loi de probabilité des variables aléatoires.

La réalisation d'un test statistique passe par les étapes suivantes :

1. Position de La question biologique : On formule la problématique à l'aide d'une question simple de sorte qu'elle doit avoir que deux réponses possibles : oui ou non.
2. Formulation des hypothèses : une hypothèse nulle  $H_0$  qui est toujours de non-effet ("il n'y a pas de différence entre ...", "il n'y a pas de relation entre ..."). Et une hypothèse alternative  $H_1$ , établie selon nos connaissances du domaine sous étude.
  - Si on ne connaît rien,  $H_1$  est bilatérale : "il y a une relation entre...".
  - Si on a des connaissances plus détaillées, on peut parfois les utiliser dans le test,  $H_1$  devient unilatérale : "il y a une relation positive entre...".
3. Choix du test : Il faut noter qu'il est nécessaire de définir tout d'abord le type de la variable étudiée, discrète ou continue, puis en fonction définir le nombre d'échantillons. Ensuite choisir parmi les différents tests disponibles, le test adéquat pour répondre à la problématique.
4. Calcul de la statistique du test : Après avoir défini le seuil de signification du test  $\alpha$  (en général on prend  $\alpha = 5\%$  ou  $\alpha = 1\%$ ), on calcule la valeur observée à partir de l'échantillon.
5. Prendre la décision : soit l'acceptation de l'hypothèse nulle  $H_0$ , soit le rejet de cette hypothèse.
6. Faire une interprétation des résultats

## 4. Les tests de conformité : Comparaison à une valeur théorique

Il s'agit de vérifier si les différences constatés entre la distribution théorique et la distribution expérimentale sont liées à la constitution de l'échantillon, via un paramètre donné. Ce paramètre soit toujours une des caractéristiques de la variable étudiée : sa moyenne, sa proportion ou sa variance.

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

On étudie une variable quantitative  $X$  et on cherche à savoir si les observations (un échantillon de données de taille  $n$ , de moyenne observée  $\bar{X}$  et de variance  $\sigma_e^2$ ) provenant d'une population de moyenne  $m$  et de variance  $\sigma^2$  concordent avec une loi théorique de moyenne  $m_0$ .

#### Hypothèses à tester

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m \neq m_0 \end{cases}$$

#### Statistique du test et règle de décision

##### 1. variance $\sigma^2$ connue :

La statistique du test dans ce cas est :  $Z = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$

Sous  $H_0$  :  $z_{obs} = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}}$ .

- Si  $|z_{obs}| > z_\alpha$  : on rejette  $H_0$  au seuil  $\alpha$ .
- Si  $|z_{obs}| \leq z_\alpha$  : on accepte  $H_0$

##### 2. variance $\sigma^2$ inconnue et $n \geq 30$ :

$Z = f(S) \sim \mathcal{N}(0, 1)$

Sous  $H_0$  :  $z = \frac{\bar{X} - m_0}{\frac{S}{\sqrt{n}}}$  où :  $S^2 = \frac{n}{n-1}\sigma^2$  est la variance estimée.

- Si  $|z_{obs}| > z_\alpha$  : on est dans la région critique et on doit rejeter  $H_0$  au seuil  $\alpha$ .

- Si  $|z_{obs}| \leq z_\alpha$  : on accepte  $H_0$

**3. variance  $\sigma^2$  inconnue et  $n < 30$  :**

$$T = \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}} \sim \mathcal{T}(n-1)ddl.$$

$$\text{Sous } H_0 : t_{obs} = \frac{\bar{X} - m_0}{\frac{S}{\sqrt{n}}}.$$

- Si  $|t_{obs}| > t_{\alpha, n-1}$  : on rejette  $H_0$  au seuil  $\alpha$ .
- Si  $|t_{obs}| \leq t_{\alpha, n-1}$  : on accepte  $H_0$

Le tableau suivant résume les différents cas d'un test de conformité d'une moyenne.

| $H_0 : m = m_0 \quad H_1 : m \neq m_0$   |  |  |   |
|--|--|--|---|
| $n \geq 30$  |  | $n < 30$ et X normal   |   |
| $\sigma$ connu   | $\sigma$ inconnu   | $\sigma$ connu   | $\sigma$ inconnu  |
| Statistique<br>$Z = f(\sigma) \sim \mathcal{N}(0, 1)$<br>$z = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}}$ | Statistique<br>$Z = f(S) \sim \mathcal{N}(0, 1)$<br>$z = \frac{\bar{X} - m_0}{\frac{S}{\sqrt{n}}}$ | Statistique<br>$Z = f(\sigma) \sim \mathcal{N}(0, 1)$<br>$z = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}}$ | Statistique<br>$T = f(S) \text{ suit loi de Student}$<br>$t = \frac{\bar{X} - m_0}{\frac{S}{\sqrt{n}}}$ |

**Test statistique [4]**

Un test statistique est une procédure qui vise à apporter une réponse à la question :

Est-ce que les données nous permettent de rejeter  $H_0$ , donc d'accepter  $H_1$ , avec un faible risque de se tromper ?

**Types de test statistique sur un paramètre : [4]**

Lorsque le test statistique porte sur un paramètre inconnu  $\theta$ , on dit que le test est :

- bilatéral si  $H_1$  est de la forme  $H_1 : \theta \neq \dots$
- unilatéral à gauche (sens de  $<$ ) si  $H_1$  est de la forme  $H_1 : \theta < \dots$
- unilatéral à droite (sens de  $>$ ) si  $H_1$  est de la forme  $H_1 : \theta > \dots$

**p-valeur [4]**

La p-valeur est le plus petit réel  $\alpha \in ]0; 1[$  calculé à partir des données tel que l'on puisse se permettre de rejeter  $H_0$  au risque  $100 \alpha\%$ . Autrement écrit, la p-valeur est une estimation ponctuelle de la probabilité critique de se tromper en rejetant  $H_0$  alors que  $H_0$  est vraie.

Les logiciels actuels travaillent principalement avec cette p-valeur. [4]

### **p-valeur : définition mathématique [4]**

La définition mathématique d'une p-valeur repose sur la notion de statistique de test. On appelle statistique de test un estimateur (var fonction d'un ou plusieurs n-échantillons) tel que sa loi (ou, éventuellement, sa loi approchée) soit connue si  $H_0$  est vraie, et qu'elle diffère selon que  $H_0$  ou  $H_1$  soit vraie. Dès lors, la p-valeur est définie par la probabilité qu'une réalisation quelconque de cette statistique de test indique un désaccord avec  $H_0$  au moins aussi élevé que la réalisation de cette statistique de test correspondante aux données, ceci si  $H_0$  était vraie.

### **Degré de significativité [4]**

La p-valeur nous donne un degré de significativité du rejet de  $H_0$ .

Le rejet de  $H_0$  est dit :

- significatif si p-valeur  $\in ]0,01 ; 0,05]$ , symbolisé par  $\star$ ,
- très significatif si p-valeur  $\in ]0,001 ; 0,01]$ , symbolisé par  $\star\star$ ,
- hautement significatif si p-valeur  $< 0,001$ , symbolisé par  $\star\star\star$ .

Il y a non rejet de  $H_0$  si p-valeur  $> 0,05$ .

### **En cas de non-rejet de $H_0$**

S'il y a non-rejet de  $H_0$ , sauf convention, on ne peut rien conclure du tout (avec le risque considéré). Éventuellement, on peut dire que  $H_0$  est plausible (elle "semble pouvoir être admise"). En revanche, peut-être qu'un risque de départ plus élevé ou la disposition de plus de données peuvent conduire à un rejet de  $H_0$ . [4]

## **Erreurs-types**

La conclusion retenue (rejet ou non de l'hypothèse  $H_0$ ) est établie avec une certaine probabilité d'erreur.

Lorsque le test conduit à rejeter l'hypothèse nulle, l'erreur éventuelle, dans le cas où cette hypothèse serait en réalité vraie, est appelée "Erreur de type 1" ou "Erreur alpha".

Lorsqu'au contraire, le test nous indique qu'il ne faut pas rejeter l'hypothèse nulle, l'erreur éventuelle, au cas où cette hypothèse serait en réalité fausse, est appelée "Erreur de type 2" ou "Erreur Bêta".

Ces indicateurs sont interdépendants : quand l'erreur alpha est réduite, l'erreur bêta augmente. Cela signifie que le choix du seuil alpha pour le test à effectuer doit se faire en fonction du coût économique de l'une ou l'autre mauvaise décision.

## Notion de risque [4]

Le risque (de première espèce) est le pourcentage de chances de rejeter  $H_0$ , donc d'accepter  $H_1$ , alors que  $H_0$  est vraie. On veut que ce risque soit aussi faible que possible.

Il s'écrit sous la forme :  $100 \alpha\%$ , avec  $\alpha \in ]0; 1[$  (par exemple, 5%, soit  $\alpha = 0.05$ ).

Le réel  $\alpha$  est alors la probabilité de rejeter  $H_0$  alors que  $H_0$  est vraie.

Le rejet de  $H_0$  est dit "significatif" si elle est rejetée au risque 5%.

**Exemple 1.2.1.** *Avant de lancer un nouveau packaging, une entreprise effectue un test pour vérifier qu'il plaît plus à ses clients que l'ancien.*

*Si l'hypothèse est vérifiée alors qu'elle est fautive, l'entreprise va remplacer l'ancien packaging qui plaît plus par un nouveau moins attirant. Elle va y perdre de l'argent et des clients.*

*En revanche, si le test lui indique que le nouveau packaging est moins attirant alors qu'il l'est plus, elle va perdre une opportunité en ne le lançant pas.*

*La comparaison des coûts de ces deux erreurs permet de fixer les seuils de manière optimale. Notons que les indicateurs alpha et bêta permettent de formaliser un niveau de sécurité pour le résultat obtenu ( $1-\alpha$ ) et un paramètre indiquant la puissance du test ( $1-\beta$ ).*

### 1.2.1 Test de conformité d'une proportion

On souhaite tester la conformité et la représentativité d'un échantillon de taille  $n$ , on étudie une variable  $X$ , qui représente le nombre  $k$  de succès parmi  $n$  tirages. On note  $\hat{p} = \frac{k}{n}$  et on veut vérifier si les observations concordent avec une loi théorique de probabilité de succès  $p_0$ . La loi théorique de  $X$  est une loi binomiale de probabilité  $p$  pour  $n$  tirages. [8]

Hypothèses à tester est :

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

Afin de pouvoir utiliser une loi de référence standard, la taille de l'échantillon doit être suffisamment grande ( $n \geq 30$ ), dans ce cas le théorème de la limite centrale garantit la convergence de la loi binomiale vers la loi normale. [8]

Statistique du test est :

$$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

qui suit la loi normale centrée réduite  $\mathcal{N}(0, 1)$  et la valeur seuil est  $z_\alpha$ .

Règle de décision :

- Si  $|z_{obs}| > z_\alpha$  : on est dans la région critique et on doit rejeter  $H_0$  au seuil  $\alpha$ .
- Si  $|z_{obs}| \leq z_\alpha$  : on accepte  $H_0$  .

### 1.2.2 Test de conformité d'une variance

On souhaite tester la conformité de la variance d'une population normale de variance  $\sigma^2$  à partir d'un échantillon de taille  $n$  et de variance  $\sigma_\varepsilon^2$ . On définit une nouvelle variable aléatoire notée  $\mathcal{X}^2$  appelée variable de "Khi deux" et définie par :

$$\mathcal{X}^2 = (n - 1) \frac{S^2}{\sigma^2}$$

Cette variable aléatoire est une variable continue définie comme la somme des carrés de  $n$  variables aléatoire  $X_i$  telles que :  $X_i \sim \mathcal{N}(0, 1)$ ,  $\forall i = \overline{1, n}$  son espérance mathématique est  $n$  et sa variance est  $2n$ . [8]

Hypothèses à tester est :

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

où la valeur seuil est  $\mathcal{X}_{\alpha, n-1}^2$ .

La statistique du test sous  $H_0$  est :

$$\mathcal{X}^2 = (n - 1) \frac{S^2}{\sigma_0^2}$$

où  $S^2$  désigne l'estimateur de  $\sigma^2$  à partir de l'échantillon.

Règle de décision :

- Pour un test bilatéral  $H_1 : \sigma^2 \neq \sigma_0^2$ , la région d'acceptation de  $H_0$  est donnée par l'intervalle :

$$[\mathcal{X}_{1-\frac{\alpha}{2}, n-1}^2, \mathcal{X}_{\frac{\alpha}{2}, n-1}^2]$$

- Pour un test unilatéral gauche  $H_1 : \sigma^2 < \sigma_0^2$ , la région d'acceptation de  $H_0$  est un intervalle de la forme :

$$[\chi_{1-\alpha, n-1}^2, +\infty[$$

- Pour un test unilatéral droit  $H_1 : \sigma^2 > \sigma_0^2$ , la région d'acceptation de  $H_0$  est un intervalle de la forme :

$$]-\infty, \chi_{\alpha, n-1}^2]$$

**Exemple 1.2.2.** [6]

On a mesuré, après une course de 400 mètres, le pouls (en battements par minute) de 7 étudiants suivants un cours d'éducation physique : Supposons que l'accroissement du pouls est une variable

|   |    |    |    |     |     |    |    |
|---|----|----|----|-----|-----|----|----|
| X | 83 | 96 | 99 | 110 | 130 | 95 | 74 |
|---|----|----|----|-----|-----|----|----|

aléatoire de loi normale  $\mathcal{N}(\mu, \sigma^2)$ , alors à un risque  $\alpha = 5\%$ , peut-on considérer que :

1. Le nombre des pouls est inférieur à 100 battements en moyenne.
2. La variation des pouls est différente de 300.

**Solution.** [6]

Afin de répondre aux questions de l'exercice on aura besoin des quantités suivantes :

La moyenne :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{7}(83 + 96 + 99 + 110 + 130 + 95 + 74) = 98.1429$$

La variance :

$$\begin{aligned} \sigma_c^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \\ &= \frac{1}{6}(83 - 98.1429)^2 + (96 - 98.1429)^2 + \dots + (95 - 98.1429)^2 + (74 - 98.1429)^2 \\ &= 330.4762 \end{aligned}$$

1. Le test à réaliser dans ce cas est le test de conformité d'une moyenne :

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

plus précisément :

$$\begin{cases} H_0 : \mu = 100 \\ H_1 : \mu < 100 \end{cases}$$

• La statistique du test est :

$$T_{obs} = \frac{\bar{X} - \mu_0}{\sqrt{\hat{\sigma}_c^2/n}} = \frac{98.1429 - 100}{\sqrt{330.4762/n}} = -0.2703$$

• La valeur critique du test est  $t_\alpha = t_{(n-1, 1-\alpha)} = t_{(7-1, 1-0.05)} = 1.943$  (de la table de la loi de Student)

On remarque que  $T_{obs} \in ] -t_\alpha, t_\alpha[$ , alors on ne rejette pas  $H_0$ , c'est-à-dire le nombre des pouls est égale à 100 battements en moyenne, avec un risque 5% de se tromper.

2. Le test à réaliser dans ce cas est le test de conformité d'une variance :

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

plus précisément :

$$\begin{cases} H_0 : \sigma^2 = 300 \\ H_1 : \sigma^2 \neq 300 \end{cases}$$

• La statistique du test est :

$$\mathcal{X}^2 = (n-1) \frac{\sigma_c^2}{\sigma^2} = \frac{6 \times 330.4762}{300} = 6.6095$$

• Les valeurs critiques du test sont :

$$\mathbb{P}(\mathcal{X}_{(n-1)}^2 > a_\alpha) = 1 - \frac{\alpha}{2} \implies \mathbb{P}(\mathcal{X}_{(6)}^2 > a_\alpha) = 0.975 \implies a_\alpha = 1.237$$

$$\mathbb{P}(\mathcal{X}_{(n-1)}^2 > b_\alpha) = \frac{\alpha}{2} \implies \mathbb{P}(\mathcal{X}_{(6)}^2 > b_\alpha) = 0.025 \implies b_\alpha = 14.449$$

D'où l'intervalle du test de la variance  $\sigma^2$  est  $[1.237; 14.449]$ .

On remarque que  $\mathcal{X}^2 \in ]a_\alpha, b_\alpha[$ , alors on ne rejette pas  $H_0$ , c'est-à-dire la variation des pouls est égale à 300, avec un risque 5% de se tromper.

## 1.3 Les tests d'homogénéité : Comparaison de deux valeurs observées

Ces tests permettent de comparer des résultats obtenus pour une variable, sur deux groupes d'observations, en vue de déterminer si ces résultats sont significativement différents d'un groupe à l'autre. Il peut s'agir, par exemple, d'un test de deux packagings ou de deux messages publicitaires, en vue d'évaluer la version la plus appréciée par les personnes interrogées.

Les tests paramétriques de comparaison les plus fréquents sont les tests de différence entre deux moyennes ou entre deux pourcentages.

Le premier s'applique sur des variables numériques. Il peut porter sur des échantillons indépendants ou appariés. A titre d'exemple, si on fait goûter une boisson à un groupe de femmes et à un groupe d'hommes pour voir s'il y a une différence d'appréciation selon le sexe, on réalise là un test sur des échantillons indépendants. En revanche, si on fait goûter deux boissons différentes à un même groupe d'individus, pour voir s'il y a une préférence significative pour l'une des deux, il s'agit d'une mesure sur des échantillons appariés.

Dans le premier cas, le test compare la moyenne pour le 1er et pour le 2ème groupe puis cherche à évaluer si cette différence est significativement différente de 0. Si tel est le cas, on peut considérer que les hommes n'apprécient pas la boisson de la même manière que les femmes. Pour savoir quel groupe l'apprécie le plus, il n'est pas forcément besoin de choisir que le test se fasse de manière unilatérale puisqu'il suffit de jeter un coup d'oeil sur les moyennes.

Dans le deuxième cas, le test consiste à calculer les différences entre les 2 notes données par chaque individu aux produits testés. Ensuite le test calcule la moyenne de ces différences puis essaie de voir si cette moyenne est significativement différente de 0. Si tel est le cas, on peut conclure que les produits sont notés de manière différente. Là aussi, l'appréciation du meilleur produit peut se faire par l'examen de la moyenne de chacun des deux ou alors, en demandant au départ un test unilatéral.

Le test de comparaison de deux pourcentages est également extrêmement utile pour évaluer la différence entre deux échantillons pour une modalité de réponse donnée (ou un regroupement de modalités). Ainsi, une enseigne de distribution peut comparer la proportion de clients satisfaits dans deux de ses magasins pour savoir si cette différence est significative.

Ce type de tests s'intéresse à comparer directement deux valeurs expérimentales, au lieu de comparer une valeur observée à une valeur d référence.

### 1.3.1 Homogénéité des moyennes

#### Cas des grands échantillon [8]

On étudie deux variables  $X_1$  et  $X_2$  sur deux échantillons de tailles  $n_1$  ( $n_1 \geq 30$ ) et  $n_2$  ( $n_2 \geq 30$ ), de moyennes observées  $\bar{x}_1, \bar{x}_2$ , et de variances observées  $\sigma_{\varepsilon_1}^2$  et  $\sigma_{\varepsilon_2}^2$  issues respectivement de deux populations de moyennes  $m_1$  et  $m_2$  et de variances  $\sigma_1^2, \sigma_2^2$ . On cherche à vérifier si ces observations proviennent de la même loi théorique. Hypothèses à tester Les hypothèses à tester sont donc :

$$\begin{cases} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{cases}$$

Statistique du test On a

- $X_1 \sim \mathcal{N}(\bar{x}_1, \bar{x}_1)$  et  $\bar{x}_1 \sim \mathcal{N}(m_1, \frac{\sigma_1^2}{n_1})$
- $X_2 \sim \mathcal{N}(\bar{x}_2, \bar{x}_2)$  et  $\bar{x}_2 \sim \mathcal{N}(m_2, \frac{\sigma_2^2}{n_2})$

La statistique du test est :

$$\begin{aligned} Z &= \frac{(\bar{x}_1 - \bar{x}_2) - \mathbb{E}[\bar{x}_1 - \bar{x}_2]}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \end{aligned}$$

On distingue deux cas :

**Les variances  $\sigma_1^2$  et  $\sigma_2^2$  sont connues :**

La statistique du test sous  $H_0$  est :

$$z_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$$

Règle de décision :

- Si  $|z_{obs}| > z_\alpha$  : on rejette  $H_0$  et on conclut que les deux populations ont des moyennes différentes au seuil  $\alpha$ .
- Si  $|z_{obs}| \leq z_\alpha$  : on accepte  $H_0$  et on constate que les moyennes des deux populations sont égales.

**Les variances  $\sigma_1^2$  et  $\sigma_2^2$  sont inconnues :**

Dans ce cas on les remplace par leurs estimations respectivement :

$$S_1^2 = \sigma_{\varepsilon 1}^2 \frac{n_1}{n_1 - 1} \text{ et } S_2^2 = \sigma_{\varepsilon 2}^2 \frac{n_2}{n_2 - 1}$$

la statistique du test est :

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{T}_\alpha(n_1 + n_2 - 2) \text{ ddl}$$

La règle de décision reste la même, seulement la valeur seuil est celle de la loi de Student  $t(n_1 + n_2 - 2, \alpha)$

### Cas des petits échantillons [8]

On sait que la distribution d'échantillonnage des moyennes d'échantillons de taille  $n$ , issus d'une population normale de moyenne  $m$  et d'écart-type  $\sigma$ , suit la loi normale  $\mathcal{N}(m, \frac{\sigma}{n})$ , ce résultat est vrai quelque soit  $n$ . Par suite, le rapport  $\frac{\bar{x}-m}{\frac{\sigma}{n}}$  est une variable centrée réduite de Gauss. Si la population n'est pas normale, mais la taille de l'échantillon est suffisante ( $n \geq 30$ ), le théorème de la limite centrale assure que la distribution des moyennes est approximativement gaussienne; Le rapport  $\frac{\bar{x}-m}{\frac{\sigma}{n}}$  est lui aussi approximativement une variable centrée réduite de Gauss.

Cependant, en pratique il est rare que l'on connaisse la valeur de  $\sigma$ , on ne connaît qu'une estimation  $S$  valeur calculée de l'estimateur  $S^2$ .

Que peut-on dire alors de la variable  $\frac{\bar{x}-m}{\frac{\sigma}{n}}$  ?

Sous réserve que le caractère étudié soit distribué dans la population selon une loi normale, on peut démontrer que ce rapport suit une loi de Student à  $n - 1$  ddl, et que cette loi qui converge rapidement vers la loi de Gauss lorsque  $n$  augmente peut être remplacée par elle dès que  $n \geq 30$ . Pour le cas des petits échantillons et sous réserve que les échantillons proviennent des populations normales et de mêmes variances  $\sigma_1^2 = \sigma_2^2$ , la statistique du test est alors :

$$t_{obs} = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{2}{n_2}}} \sim \mathcal{T}_\alpha(n_1 + n_2 - 2) \text{ ddl}$$

Où :  $\hat{\sigma}^2 = \frac{n_1\sigma_{\varepsilon 1}^2 + n_2\sigma_{\varepsilon 2}^2}{n_1 + n_2 - 2}$  et la règle de décision reste la même.

### 1.3.2 Homogénéité des proportions

La démarche est la même que pour le cas précédent : on souhaite comparer deux populations par rapport à la proportions d'individus pour lesquels la variable prend une certaine modalité "A". On tire de ces deux populations deux échantillons indépendants de taille respectives  $n_1$  et  $n_2$  sur lesquels on détermine les proportions d'individus de type "A". elles valent respectivement  $\hat{p}_1 = \frac{k_1}{n_1}$  et  $\hat{p}_2 = \frac{k_2}{n_2}$ . On cherche à savoir si ces observations proviennent de la même loi théorique, une loi binomiale de probabilité de succès  $p$ . Ici aussi, on utilise la convergence de la loi binomiale vers la loi normale.[8]

On cherche à tester :

$$\begin{cases} H_0 : p_1 = p_2 = p \\ H_1 : p_1 \neq p_2 \end{cases},$$

telle que  $p$  est estimée par  $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{k_1 + k_2}{n_1 + n_2}$

Étant donné que  $n_1 \geq 30$  et  $n_2 \geq 30$ .

Statistique du test

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}(0, 1)$$

Règle de décision

On accepte  $H_0$  si  $|z_{obs}| \leq z_\alpha$  (ou  $z_\alpha$  est le fractile de la loi normale centrée réduite), alors la différence entre  $p_1$  et  $p_2$  n'est pas significative.

Sinon on la rejette et on constate que la différence entre  $p_1$  et  $p_2$  est significative au seuil  $\alpha$ .

**Exemple 1.3.1.** [?]

*Un biologiste a mis 80 souris malades sous traitement, et 60 souris malades d'un autre type sous le même traitement. 1 mois plus tard, il reste en vie respectivement 50 souris sur 80 et 40 souris sur 60.*

*La réaction des souris au traitement peut-elle être considérée comme identique ?*

**Solution.**

Il s'agit de comparer deux échantillons indépendants via la proportion.

$$X_{i_1} = \begin{cases} 1 & \text{si la souris de type 1 survie après 1 mois} \\ 0 & \text{sinon} \end{cases} ; X_{i_1} \sim B(p_1) \longrightarrow X_1 \sim B(n_1, p_1)$$

$$X_{i_2} = \begin{cases} 1 & \text{si la souris de type 1 survie après 1 mois} \\ 0 & \text{sinon} \end{cases} ; X_{i_2} \sim B(p_2) \longrightarrow X_2 \sim B(n_2, p_2)$$

On mène le test suivant :

$$\begin{cases} H_0 : p_1 = p_2 = p \\ H_1 : p_1 \neq p_2 \end{cases}$$

$p_1$  et  $p_2$  sont les proportions des souris qui sont restés en vie dans les populations.

Sous  $H_0$  et puisque  $n_1 > 30$  et  $n_2 > 30$  :

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

telle que :

$$\hat{p}_1 = \frac{k_1}{n_1} = \frac{50}{80} = 0.625$$

$$\hat{p}_2 = \frac{k_2}{n_2} = \frac{40}{60} = 0.67$$

$$\hat{p} = \frac{k_1 + k_2}{n_1 + n_2} = \frac{90}{140} = 0.64$$

alors ;

$$\begin{aligned} |z_{obs}| &= \left| \frac{0.625 - 0.67}{\sqrt{(0.64)(0.36)\left(\frac{1}{80} + \frac{1}{60}\right)}} \right| \\ &= 0.55 \end{aligned}$$

$|z_{obs}| < z_\alpha = 1.96$  donc on accepte  $H_0$  .

**Conclusion du test :**

La réaction des souris au traitement dans les deux populations est identique au seuil 5%

### 1.3.3 Homogénéité des variances

D'une Façon analogue aux deux précédente, on peut s'interroger sur l'égalité des variances  $\sigma_1^2$  et  $\sigma_2^2$  de deux populations indépendantes.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Pour cela on tire de ces deux populations deux échantillons indépendants de tailles respectives  $n_1$  et  $n_2$ , de moyennes  $\bar{x}_1$  et  $\bar{x}_2$  et de variances  $\sigma_{\varepsilon_1}^2$  et  $\sigma_{\varepsilon_2}^2$ . Il n'est pas nécessaire que  $n_1$  et  $n_2$  soient grands, mais il est impératif que les deux populations soient normales. [8]

On calcule la quantité

$$F_{obs} = \frac{S_1^2}{S_2^2} \sim F_{\alpha}(n_1 - 1, n_2 - 1)$$

telle que :  $S_1^2 \geq S_2^2$ ,  $S_1^2 = \frac{n_1}{n_1-1}\sigma_{\varepsilon_1}^2$  et  $S_2^2 = \frac{n_2}{n_2-1}\sigma_{\varepsilon_2}^2$

### Règle de décision

Si  $F_{obs} > F_{\frac{\alpha}{2}, n_1-1, n_2-1}$ , on rejette  $H_0$  au risque  $\alpha$  et on constate que les variances sont différentes et les deux populations sont hétérogènes.

## 1.4 Analyse de la variance : ANOVA

Contrairement à ce que pourrait laisser penser son nom, l'analyse de la variance n'est pas une méthode qui permet d'étudier les différences de variances entre populations, mais une méthode pour étudier les différences de moyenne entre populations (par exemple, trois populations ont-elles la même moyenne ? ou autrement dit, les différences de moyenne entre les trois populations sont-elles significatives ?). Cette méthode, néanmoins, doit son nom au fait qu'elle utilise des mesures de variance afin de déterminer le caractère significatif, ou non, des différences de moyenne mesurées sur les populations.[8]

Il s'agit d'une généralisation à  $k$  populations du classique test de comparaison de moyennes de deux échantillons : le célèbre test de student  $\mathcal{T}$ .

### Idée générale

L'idée de l'analyse de la variance repose sur un modèle qu'on se donne a priori des données. On suppose ainsi, par exemple, qu'une variable mesurée  $\mathbf{Y}$  vérifie une relation linéaire avec un ensemble de  $p$  variables explicatives dénotées  $\mathbf{X}_i$ . La relation est du type suivant :

$$Y = \mu + \sum_{i=1}^p \alpha_i X_i + \varepsilon$$

avec ;

- $\mu$  un paramètre commun à toutes les observations, c'est-à-dire une ordonnée à l'origine (dont on pourra tester éventuellement la nullité plus tard).
- $\varepsilon$  représente la variabilité aléatoire du modèle, non contrôlable.

On s'attache ensuite à l'étude de la contribution de ces différents termes à la variance de  $\mathbf{Y}$ , grace à une décomposition dite de **l'analyse de la variance**. [8]

### Conditions d'application de l'ANOVA

- Le caractère étudié suit la loi normale.
- Les variances des populations sont toutes égales (**Homoscédasticité**).
- Les échantillons sont prélevés aléatoirement et indépendamment dans les populations.

**Remarque 1.4.1.** *Il est important de comprendre que l'ANOVA n'est pas un test pour classer des moyennes, mais plutôt c'est un test permettant de comparer les moyennes de différents groupes et dire, si parmi l'ensemble, au moins une d'entre elles diffère des autres, mais on ne sait ni laquelle ni combien d'entre elles.*[8]

#### 1.4.1 ANOVA à un facteur contrôlé

On parle d'**ANOVA** à un facteur lorsque les groupes analysés se distinguent par un seul facteur qualitatif (par exemple, Biologiste et statisticien). Elle a pour but de comparer les moyennes de  $n$  populations à partir d'échantillons aléatoires et indépendants prélevés dans chacune d'elles.

**Hypothèses à tester :**  $\begin{cases} H_0 : \text{toutes les moyennes sont identiques.} \\ H_1 : \text{au moins une des moyennes est différente des autres.} \end{cases}$

**Procédure des calculs** Il s'agit en pratique de décomposer la variabilité selon (au moins) deux critères :

- Variabilité non expliquée, ou résiduelle, entre un terme estimé et la vraie valeur mesurée, qu'on appellera **SCR**, on parle aussi de variance intra-classe.
- Variabilité expliquée par le modèle, c'est-à-dire la différence entre l'estimation de moyenne d'une classe et la moyenne totale des observations, qu'on appellera **SCF**, pour la variance due au facteur **A**, c'est la variance inter-classe.

Cette décomposition se fait par l'**équation d'analyse de variance** suivante :

$$\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2$$

*variation Totale                  variation INTRA                  variation INTER*

$$\mathbf{SCT} = \mathbf{SCR} + \mathbf{SCF}$$

tels que :

- $n$  : nombre total d'individus.
- $y_{ij}$  : observation de la variable endogène  $\mathbf{Y}$ .
- $n_i$  : nombre d'individus du niveau  $i$ .
- $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$  : la moyenne des individus du niveau  $i$ .
- $\bar{y} = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} = \frac{1}{n} \sum_{i=1}^I n_i \bar{y}_i$  : La moyenne totale.

Si le facteur  $\mathbf{A}$  a un effet sur la variable endogène  $\mathbf{Y}$ , la variation INTER sera importante par rapport à la variation INTRA.

A partir de cette définition, on va comparer les espérances des variances **SCF** et **SCR** en faisant leur rapport. Il se trouve (comme on peut le voir dans la décomposition mathématique) que les deux termes sont tous les deux une estimation de la variabilité résiduelle si le facteur  $\mathbf{A}$  n'a pas d'effet. De plus, ces deux termes suivent chacun une loi de **khi-deux**, **leur rapport suit donc une loi de F** (voir plus loin pour les degrés de liberté de ces lois).

#### Tableau d'ANOVA

| Source de variation   | Somme des carrés (SC)  | ddl     | Moyenne somme des carrés (MC) | Statistique F de Fisher |
|-----------------------|--|---------|-------------------------------|-------------------------|
| INTER(due au facteur) | $SCF = \sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2$             | $I - 1$ | $MCF = \frac{SCF}{I-1}$       |                         |
| INTRA (résiduelle)    | $SCR = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ | $n - I$ | $MCR = \frac{SCR}{n-I}$       | $F_c = \frac{MCF}{MCR}$ |
| Totale                | $SCT = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$   | $n - 1$ | $MCT = \frac{SCT}{n-1}$       |                         |

TABLE 1.1 – Tablea d'ANOVA [8]

#### La règle de décision :

Au risque  $\alpha$ , l'hypothèse  $H_0$  sera rejetée si  $F_c > F_t = F_{I-1, n-I, \alpha}$

Où  $F_{I-1, n-I, \alpha}$  est la valeur seuil de la loi de Fisher.

#### Exemple 1.4.1. [6]

*Nous souhaitons comparer quatre traitements, notés A, B, C et D. Nous répartissons par tirage au sort les patients, et nous leur affectons l'un des quatre traitements. Nous mesurons sur chaque patient la durée, en jours, séparant de la prochaine crise d'asthme. Les mesures sont reportées dans le tableau ci-dessous :*

| Traitement A           | Traitement B           | Traitement C         | Traitement D           |
|------------------------|------------------------|----------------------|------------------------|
| 36 ; 37 ; 35 ; 38 ; 41 | 42 ; 38 ; 39 ; 42 ; 44 | 26 ; 26 ; 30 38 ; 34 | 42 ; 45 ; 50 ; 56 ; 58 |

Pouvons-nous conclure, à un seuil de risque 1%, que les facteur traitement a une influence sur le critère retenue ? (On donne  $SCT = 1324,55$ .)

**Solution.**[6]

Calculons les moyennes :

$$\bar{X}_A = \frac{1}{n_{A_i}} \sum_{i=1}^5 x_{A_i}$$

$$\bar{X}_A = \frac{1}{5} \sum_{i=1}^5 x_{A_i} = 37.40$$

$$\bar{X}_B = \frac{1}{5} \sum_{i=1}^5 x_{B_i} = 41.00$$

$$\bar{X}_C = \frac{1}{5} \sum_{i=1}^5 x_{C_i} = 30.80$$

$$\bar{X}_D = \frac{1}{5} \sum_{i=1}^5 x_{D_i} = 50.20$$

$$\bar{X} = \frac{1}{5} \sum_{i=1}^{20} x_i = 39.85$$

En exploitant ces dernières quantités pour le calcul des différentes variations on obtient :

| Source de variation | SC               | ddl | CM               | Statistique F   |
|---------------------|------------------|-----|------------------|-----------------|
| INTER-groupes       | $SCF = 961.2667$ | 3   | $MCF = 320.4222$ |                 |
| INTRA-groupes       | $SCR = 363.2833$ | 16  | $MCR = 22.7052$  | $F_c = 14.1123$ |
| Total               | $SCT = 1324.55$  | 19  |                  |                 |

On constate que  $F_c > F_\alpha = 5.29$  cela signifie qu'on doit rejeter  $H_0$  .

C'est-à-dire le facteur traitement a une influence significative sur les durées séparant deux crise d'asthme.

# Les tests statistique par l'approche Bayésienne

## 2.1 Introduction

L'approche bayésienne fournit un cadre naturel pour résoudre les problèmes d'inférence statistiques. Elle diffère des statistiques traditionnelles en ce que les paramètres du modèle sont considérés comme des variables aléatoires. peut être présenté comme une généralisation approche classique. Les paramètres ne sont plus des inconnues fixes, mais des variables aléatoires. Nous croyons que c'est l'objectif principal de statistiques construites à partir d'observations de phénomènes aléatoires Raisonner sur les lois qui produisent des observations pour analyser des phénomènes passés ou prédire des événements futurs. En particulier, nous nous référons à la surface de décision de l'inférence bayésienne. La statistique est une science mathématique dont le but est d'expliquer ce qui s'est passé et de prédire ce qui se passera dans le futur. Nous nous appuyons sur des observations de phénomènes naturels pour suggérer des interprétations, souvent à travers des modèles probabilistes. Pour l'inférence bayésienne, un phénomène observé nécessaire pour pouvoir le remplacer Elle conduit à une complétude qui permet l'analyse et la généralisation.

Ce chapitre présente quelques concepts de base liés à : Statistiques bayésiennes (loi a priori, loi a posteriori, loi prédictive, . . . )

## 2.2 Principe de la statistique Bayésienne

Dans le cadre de la statistique bayésienne l'expression latine *a priori* est beaucoup utilisée. Elle signifie au préalable en français, ou plus précisément en se fondant sur des données antérieures à l'expérience. Étymologiquement cette expression vient de "a priori ration" qui signifie en latin par une raison qui précède, et s'oppose à *a posteriori*. L'expression latine *a posteriori* est également très utilisée dans le cadre bayésien. Elle signifie après coup en français, ou plus précisément en s'appuyant sur l'expérience, sur les faits constatés. Étymologiquement cette expression vient de "a posteriori ration" qui signifie en latin par une raison qui vient après, et s'oppose à *a priori*. Action formalisant les connaissances d'un expert pour permettre de les partager, e.g. de les incorporer à un modèle. La statistique est une science mathématique, dont l'objectif est de décrire ce qui s'est produit et de faire des projections quant à ce qu'il peut advenir dans le futur. Elle s'appuie sur l'observation de phénomènes naturels pour en proposer une interprétation, souvent à travers des modèles probabilistes.[3]

Cette approche consiste à éliciter la loi *a priori* d'après sa loi marginale empirique, et donc à estimer la distribution *a priori* à partir des données. Cela revient donc à se donner des hyperparamètres et à chercher à les estimer de manière fréquentiste (par exemple par maximum de vraisemblance) par  $\hat{\eta}$ , avant d'injecter cet estimateur dans la distribution *a priori* et donc d'obtenir la distribution *a posteriori*  $p(\theta/y, \hat{\eta})$ . Cette approche bayésienne empirique qui combine bayésien et fréquentiste peut sembler aller à l'encontre de la notion d'*a priori*, puisque l'on utilise alors déjà les données pour choisir l'*a priori*. Néanmoins, on peut voir l'approche bayésienne empirique comme une approximation de l'approche bayésienne complète. Son utilisation résulte en une distribution *a posteriori* plus resserrée qu'avec un *a priori* faiblement informatif (diminution de la variance), au prix de l'introduction d'un biais dans l'estimation (on "utilise les données 2 fois" !). Cette approche illustre une fois de plus le compromis existant entre biais et variance, classique dans toute procédure d'estimation. L'approche bayésienne fournit un cadre naturel pour résoudre des problèmes d'inférence statistique. Elle se distingue de la statistique classique parce qu'elle considère le(s) paramètre(s) du modèle comme des variables aléatoires.

Dans ce chapitre, nous rappelons tout d'abord quelques notions de base sur la statistique bayésienne tel que le modèle bayésien et le choix *a priori*.

## Modèle statistique paramétrique bayésien

L'ensemble des observation est noté  $x = (x_1, \dots, x_i, \dots, x_n)$ , autrement dit on dispose d'un échantillon de taille  $n$ . Le cadre statistique dans ce cas est le meme que celui de la statistique inférentielle, les observation  $x_i$  sont donc considérées comme des réalisations de variables aléatoires notées  $x_i$ .

Et soit  $\theta$  l'essemble des paramètres ( $\theta \in \Theta$ )

### Définition 2.2.1.

*On entend Par information a priori sur le paramètre  $\theta$  toute information disponible sur  $\theta$  en dehors de celle apportée par les observations*

*L'information a priori sur  $\theta$  et attachée d'incertitude ( si ce n'était pas le cas, le paramètre  $\theta$  serait connu avec certitude et n'aurait pas à l'estime). Il est naturel de modéliser cette information a priori au travers d'une loi de propabilité, appelée loi a priori. Sa densité est notée  $\pi(\theta)$ .*

*Le modèle statistique paramétrique bayésien consiste en la donnée d'une loi a priori et de la loi des observations. On appelle loi des observations, la loi conditionnelle se sachant  $\theta$ . Sa densité est notée  $f(x/\theta)$ , que la variable soit discrète ou continue. Si  $x$  est discrète  $f(x/\theta)$  représente lhypothèse que , sachant  $\theta$  les variables aléatoires  $x_i$  sont indépendantes , autrement dit on aura :*

$$L(x/\theta) = \prod_{i=1}^n L(x_i/\theta)$$

## 2.3 La loi a posteriori

L'objectif est donc d'utiliser cette information supplémentaire. Sachant que l'information contenue dans les observations  $x$  est contenue dans  $L(x/\theta)$  et l'information a priori sur  $\theta$  dans  $\pi(\theta)$ , on peut utiliser la règle de Bayes pour combiner ces deux types d'informations en définissant la densité a posteriori par :

$$\pi(\theta/x) = \frac{L(x/\theta)\pi(\theta)}{\int L(x/\theta)\pi(\theta)d\theta}$$

qui contiendra donc toutes les informations sur  $\theta$ .

On remarque que l'inversion de cause à effet est ici beaucoup plus naturelle. Elle se fait d'une manière cohérente, car l'état de connaissance a priori sur  $\theta$  traduite par la densité priori

$\pi(\theta)$  est transformé, après les observations  $x$ , en état de connaissance a posteriori par la densité a posteriori  $\pi(\theta/x)$ .

Remarquons que la densité a posteriori peut s'écrire :

$$\pi(\theta/x) \propto L(x/\theta)\pi(\theta)$$

### L'espérance a posteriori

L'espérance a posteriori est définie par :

$$\mu_p = \mathbb{E}[\theta/y] = \mathbb{E}_{\theta/y}(\theta)$$

$$\theta \in (\Theta)$$

À noter que le calcul de cette espérance a posteriori n'est pas toujours facile car il suppose le calcul d'une intégrale *dots* C'est l'estimateur qui minimise l'erreur moyenne quadratique a posteriori (ou coût quadratique). En effet, pour un estimateur quelconque  $\hat{\theta}$ , l'erreur quadratique moyenne a posteriori se décompose comme :

$$\mathbb{E}_{\theta} \left[ (\hat{\theta} - \theta)^2 / y \right] = \hat{\theta}^2 - 2\mathbb{E}_{\theta}[\theta/y] + \mathbb{E}_{\theta}[\theta^2/y]$$

En dérivant l'expression ci-dessus par rapport à  $\hat{\theta}$ , on montre aisément que le minimum est obtenu pour :

$$\hat{\theta} = \mathbb{E}_{\theta}[\theta/y]$$

### Le maximum a posteriori

Le maximum a été beaucoup utilisé, surtout car il est plus facile (ou en tout cas moins difficile) à calculer. En effet, il ne requiert aucun calcul d'intégrale, mais une simple maximisation de  $f(x/\theta)\pi(\theta)$  (car le dénominateur  $f(x)$  ne dépend pas de  $\theta$ ). L'estimateur du mode s'appelle le maximum a posteriori (souvent noté MAP). Le MAP peut être vu comme une régularisation (par la loi a priori) de l'estimateur du maximum de vraisemblance, dont il est proche.

### La médiane a posteriori

La médiane est également un résumé possible de la distribution a posteriori. Comme son nom l'indique, il s'agit de la médiane de  $\pi(\theta/x)$ . Il s'agit de l'estimateur ponctuel optimal au sens de l'erreur absolue (fonction de coût linéaire).

## La loi marginale de X

Sa densité est notée  $m(x)$  et on a :

$$m(X) = \int_{\Theta} L(X/\theta)\pi(\theta)d\theta$$

## Le calcul de la loi a posteriori

### Une situation simple

On dispose d'un vecteur d'observation  $x = (x_1, x_2, \dots, x_n)$  et on considère le modèle bayésien suivant :

$$x_i/\theta \sim \text{Bernoulli}(\theta) \text{ et } \theta \sim \text{Beta}(a, b)$$

On a :

$$\begin{aligned} L(x/\theta) &= \prod_{i=1}^n P(x = x_i/\theta) \\ &= \prod_{i=1}^n L(x_i/\theta) \\ &= \theta^s (1 - \theta)^{n-s} \quad \text{ou } s = \sum_{i=1}^n x_i \end{aligned}$$

et comme  $\theta \sim \text{Beta}(a, b)$ , on a :

$$\pi(\theta) = \frac{1}{\mathcal{B}(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbf{1}_{[0,1]}(\theta)$$

Alors ;

$$\pi(\theta/x) = \frac{L(x/\theta) \pi(\theta)}{\int_{\theta} L(x/\theta) \pi(\theta) d\theta} = \frac{\theta^s (1 - \theta)^{n-s} \frac{1}{\mathcal{B}(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbf{1}_{[0,1]}(\theta)}{m(X)}$$

Dans un contexte bayésien on a  $\pi(\theta/x) \propto f(x/\theta) \pi(\theta)$  en tant que fonction de  $\theta$ , les deux expressions  $\pi(\theta/x)$  et  $f(x/\theta) \pi(\theta)$  sont effectivement proportionnelles, la constante  $a$  qui apparaît dans la définition est égale ici à  $\frac{1}{m(x)}$ , à noter que cette quantité est bien une constante au sens ou elle ne dépend pas de  $\theta$ .

## 2.4 Choix de la loi a priori

L'aspect de l'analyse bayésienne le plus critiqué et le plus délicat est certainement le choix de la loi a priori des paramètres. Deux types d'approches sont généralement considérés :

- Une approche dite subjective, ou informative, qui revient à tenir compte (lorsqu'elles existent) d'informations a priori sur le paramètre (expériences précédentes, avis d'experts, connaissances extérieures au processus d'observation, etc.)
- Une approche dite objective, ou non informative, qui revient à modéliser l'absence d'information a priori.

Le tableau suivant présente quelques lois a priori conjuguée pour quelques familles exponentielles usuelles :

| $f(x/\theta)$                                    | $\pi(\theta)$                               | $\pi(\theta/x)$  |
|--|---|--|
| Normale<br>$N(\theta, \sigma^2)$                 | Normale<br>$N(\mu, \tau^2)$                 | Normale<br>$N(\rho(\sigma^2\mu + \tau^2x), \rho\sigma^2\tau^2)$<br>avec $\rho = \frac{1}{\sigma^2 + \tau^2}$ |
| Poisson<br>$P(\theta)$                           | Gamma<br>$G(\alpha, \beta)$                 | Gamma<br>$G(\alpha + x, \beta + 1)$  |
| Gamma<br>$G(\nu, \theta)$                        | Gamma<br>$G(\alpha, \beta)$                 | Gamma<br>$G(\alpha + \nu, \beta + x)$  |
| Binomiale<br>$B(n, \theta)$                      | Beta<br>$Beta(\alpha, \beta)$               | Beta<br>$Beta(\alpha + x, \beta + n - x)$  |
| Normale<br>$N(\mu, \frac{1}{\theta})$            | Gamma<br>$G_a(\alpha, \beta)$               | Gamma<br>$G(\alpha + 0.5, \beta + (\mu - x)^2/2)$  |
| Binomiale négative<br>$Neg(m, \theta)$           | Beta<br>$Be(\alpha, \beta)$                 | Beta<br>$Beta(\alpha + m, \beta + x)$  |
| Multinomiale<br>$M_k(\theta_1, \dots, \theta_k)$ | Dirichlet<br>$D(\alpha_1, \dots, \alpha_k)$ | Dirichlet<br>$D(\alpha_1 + x_1, \dots, \alpha_k + x_k)$  |

TABLE 2.1 – Loi a priori conjuguée

## 2.5 Une introduction à la théorie de la décision

### 2.5.1 Fonction de perte et risque

Pour le modèle  $X \in (\chi, \beta, \{P_\theta, \theta \in \Theta\})$ , on définit  $\mathcal{D}$  l'ensemble des décisions possibles. C'est-à-dire l'ensemble des fonctions de  $\Theta$  dans  $g(\Theta)$  où  $g$  dépend du contexte :

— si le but est d'estimer  $\theta$  alors  $\mathcal{D} = \Theta$

— pour un test,  $\mathcal{D} = \{0, 1\}$

La fonction de perte est une fonction mesurable de  $(\theta \times \mathcal{D})$  à valeurs réelles positives :  $L : \Theta \times \mathcal{D} \rightarrow \mathbb{R}_+$ . Elle est définie selon le problème étudié et constitue l'armature du problème statistique. Dans ce qui suit, nous allons donner quelques fonctions de perte et risques les plus usuelles.

**Définition 2.5.1. La perte quadratique**

Une fonction de perte quadratique est une fonction  $l : (\Theta \times \mathcal{D}) \rightarrow \mathbb{R}_+$  donnée par

$$l(\theta, \delta) = (\theta - \delta)^2$$

**Définition 2.5.2. La perte absolue**

une fonction de perte absolue est définie comme suit :

$$l(\delta, \theta) = |\theta - \delta|$$

**Définition 2.5.3. Risque fréquentiste**

Pour  $(\theta, \delta) \in \Theta \times \mathcal{D}$ , le risque fréquentiste est défini par :

$$\begin{aligned} R(\theta, \delta) &= E_\theta[L(\theta, \delta(X))] \\ &= \int_{\mathcal{X}} L(\theta, \delta(X)) f(X|\theta) d\mu(X) \end{aligned}$$

C'est une fonction de  $\theta$  et ne définit donc par un ordre total sur  $\mathcal{D}$  et ne permet donc pas de comparer toutes décisions et estimateurs. Il n'existe donc pas de meilleur estimateur dans un sens absolu. Ainsi, l'approche fréquentiste restreint l'espace d'estimation en préférant la classe des estimateurs sans biais dans laquelle il existe des estimateurs de risque uniformément minimal ; l'école bayésienne ne perd pas en généralité en définissant un risque *a posteriori*. L'idée est d'intégrer sur l'espace des paramètres pour pallier cette difficulté.

**Définition 2.5.4. (Risque a posteriori)**

Une fois données la loi a priori sur le paramètre et la fonction de perte, le risque a posteriori est défini par :

$$\begin{aligned}\rho(\pi, \delta|X) &= E^\pi(L(\theta, \delta(X))|X) \\ &= \int_{\Theta} L(\theta, \delta(X)) d\Pi(\theta|X)\end{aligned}$$

Ainsi, le problème change selon les données ; ceci est dû à la non existence d'un ordre total sur les estimateurs.

**Définition 2.5.5.** (*Risque intégré*)

Pour une fonction de perte donnée, le risque intégré est défini par :

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta) d\Pi(\theta)$$

**La considération de la fonction de coût**

Pour le problème de test d'hypothèse

$$H_0 : \theta \in \Theta_0 \text{ contre } H_1 : \theta \in \Theta_0^c$$

nous étudions maintenant les formes raisonnables pour la fonction de coût  $L(\theta, \Phi)$  pour évaluer la valeur de l'estimateur  $\Phi(x)$  de  $I_{\Theta_0}(\theta)$ , (voir Juinn Tzon Hwang, George Casella, Christian Robert, Martin T, Wells et Roger H. Farrell 1992). Depuis que notre paramètre intéressé eu plus que deux valeurs, la fonction de coût est donnée par :

$$L(\theta, \Phi) = \begin{cases} L(1, \Phi(x)) & \text{si } \theta \in \Theta_0 \\ L(0, \Phi(x)) & \text{si non} \end{cases}$$

Le coût a posteriori attendu pour les fonctions de coût précédentes sachant  $X=x$  est :

$$\begin{aligned}E(L(\theta, \Phi(x))/X = x) &= \int_{\Theta} (L(\theta, \Phi(x)))\pi(\theta/x)d\theta \\ &= L(1, \Phi(x))P(\theta \in \Theta_0/x) + L(0, \Phi(x))P(\theta \in \Theta_0^c/x)\end{aligned}$$

où

$$\pi(\theta/x) = \frac{f(x/\theta)\pi(\theta)}{\int_{\Theta} f(x/\theta)\pi(\theta)d\theta}$$

Et la distribution a posteriori est :

$$P(\theta \in \Theta_0/x) \int_{\Theta_0} \pi(\theta/x)d\theta$$

La borne stratégique bayésienne dit la vérité, et le bon estimateur de  $I_{\Theta_0}(\theta)$  est la bonne évaluation bayésienne de la probabilité de son apparition, donc  $L(\theta, \Phi(x))$  est propre si :

$$\min_{\Phi(x)} E[L(\theta, \Phi(x))/X = x] = E[L(\theta, p(\theta \in \Theta_0/x))/X = x]$$

Comme exemple, il y a un risque équivalent entre les tests randomisés et l'estimateur de  $I_{\Theta_0}(\theta)$ . Ce fait explique en partie pourquoi  $L(\theta, \Phi(x)) = |\Phi(x) - I_{\Theta_0}(\theta)|$  ce qui conduit à la solution de Bayes 0-1. Cette équivalence est simple à voir si on écrit le risque de la règle de décision  $\Phi(x)$  comme suit :

$$\begin{aligned} R(\theta, \Phi(x)) &= \int_{-\infty}^{+\infty} |I_{\Theta_0} - \Phi(x) f(x/\theta)| dx \\ &= I_{\Theta_0^c}(\theta) \int_{-\infty}^{+\infty} (1 - \Phi(x)) f(x/\theta) dx \end{aligned}$$

Il est également possible dans les conditions de régularité d'établir une réciproque, c'est que le coût  $L_1$  est le seul coût en vertu duquel il y a une correspondance entre l'estimateur de  $I_{\Theta_0}(\theta)$  et les tests randomisés. Le fait que le coût  $L_1$  ce qui est si étroitement lié au coût 0-1 de Neymann Pearson conduit aux estimateurs qu'ils ne seront pas lisses.

**Définition 2.5.6.** (*Estimateur bayésien*)

*Un estimateur bayésien est un estimateur vérifiant :*

$$r(\pi, \delta^\pi) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta) < \infty$$

Pour obtenir la valeur de l'infimum du risque intégré il faut donc en théorie minimiser une intégrale double  $\delta$ . L'introduction du risque intégré se justifie par le théorème suivant. Il suffira de minimiser une grandeur qui ne dépend plus que des données, ceci permet donc d'arriver à des estimateurs satisfaisants.

**Théorème 2.5.1.** (*Méthode de calcul*)

*Si  $\exists \delta \in \mathcal{D}, r(\pi, \delta) < \infty$  et  $\forall X \in \mathcal{X} \delta^\pi(X) = \text{Argmin}_\delta \rho(\pi, \delta|X)$ , alors ;  $\delta^\pi(X)$  est un estimateur bayésien.*

*Démonstration.*

$$\begin{aligned}
r(\pi, \delta) &= \int_{\Theta} R(\theta, \delta) d\Pi(\theta) \\
&= \int_{\Theta} \int_X L(\theta, \delta(X)) f(X|\theta) d\mu(X) d\Pi(\theta) \\
&= \int_X \int_{\Theta} L(\theta, \delta(X)) \frac{f(X|\theta) d\Pi(\theta)}{m_{\pi}(X)} m_{\pi}(X) d\mu(X) \quad \text{Fubini} \\
&= \int_X \int_{\Theta} L(\theta, \delta(X)) d\Pi(\theta|X) m_{\pi}(X) d\mu(X) \\
&= \int_X \rho(\pi, \theta|X) m_{\pi}(X) d\mu(X)
\end{aligned}$$

Ainsi, pour  $\delta \in \mathcal{D}$ ,  $\rho(\pi, \delta^{\pi}|X) \leq \rho(\pi, \theta|X) \Rightarrow r(\pi, \delta) \leq r(\pi, \theta)$ . Ce qui permet de conclure.  $\square$

## 2.5.2 Exemples

**Exemple 2.5.1.** *Sous la perte quadratique*

$$\mathcal{D} = \Theta \subset \mathbb{R}^{d^3} \text{ et } L(\theta, \delta) = \|\theta - \delta\|^2$$

Comme la norme au carré est une fonction convexe deux fois dérivable sur  $\Theta$ , pour trouver  $\delta^{\pi}$  estimateur bayésien, il suffit de déterminer les points critiques du risque a posteriori.

$$\begin{aligned}
\rho(\pi, \theta|X) &= E^{\pi}(\|\theta - \delta\|^2 |X) \\
\frac{\partial \rho(\pi, \theta|X)}{\partial \delta} &= -2 \int_{\Theta} (\pi - \delta(X)) d\Pi(\theta, X)
\end{aligned}$$

Donc

$$\frac{\partial \rho(\pi, \theta|X)}{\partial \delta} = 0 \iff \delta(X) = E^{\pi}(\theta|X)$$

D'après l'inégalité de Jensen,  $\delta \rightarrow \rho(\pi, \delta|X)$  est aussi convexe. L'estimateur bayésien vaut donc  $\delta^{\pi}(X) = E^{\pi}(\theta|X)$   $\mu$ -pp.

**Exemple 2.5.2.** *Perte  $L^1$*

$$\mathcal{D} = \Theta \subset \mathbb{R} \text{ et } L(\theta, \delta) = \sum_{i=1}^d |\theta_i - \delta_i|$$

Dans le cas simple où  $d = 1$  :

$$\begin{aligned}
\rho(\pi, \theta|X) &= \int_{\Theta} |\theta - \delta| d\Pi(\theta, X) \\
&= \int_{-\infty}^{\delta} (\delta - \theta)\pi(\theta|X) d\theta + \int_{\delta}^{+\infty} (\theta - \delta)\pi(\theta|X) d\theta
\end{aligned}$$

Comme  $\delta \rightarrow \rho(\pi, \delta|X)$  est convexe et dérivable  $\mu$ -presque partout, il suffit là encore de déterminer les points critiques :

$$\frac{\partial \rho(\pi, \theta|X)}{\partial \delta} = \int_{-\infty}^{\delta} \pi(\theta|X) d\theta - \int_{\delta}^{+\infty} \pi(\theta|X) d\theta$$

Donc

$$\frac{\partial \rho(\pi, \theta|X)}{\partial \delta} = 0 \iff \delta(X) = P^\pi(\theta \leq \delta|X) = P^\pi(\theta \geq \delta|X)$$

C'est-à-dire  $\delta^\pi(X)$  est la médiane de  $\pi(\theta|X)$ .

**Exemple 2.5.3.** Sous la perte 0 – 1

Cette fonction de perte est utilisée dans le contexte des tests. Un test est la donnée d'une partition de  $\Theta$  en  $\Theta_0$  et  $\Theta_1$ .  $\theta \in \Theta_i$  correspond à l'hypothèse  $H_i$ ,  $H_0$  est appelée l'hypothèse nulle. Le principe du test (décisions)  $\delta$  est défini comme suit :

$$\delta = \begin{cases} 0 & \text{si } \theta \in \Theta_1 \\ 1 & \text{si } \theta \in \Theta_0 \end{cases}$$

La fonction de perte correspondant au test est définie par :

$$L(\theta, \delta) = \mathbf{1}_{\theta \in \Theta_1} \times \mathbf{1}_{\delta=0} + 0 \quad \text{si } \delta = \times \mathbf{1}_{\delta=1}$$

Le risque a posteriori est alors le suivant :

$$\rho(\pi, \delta|X) = \mathbf{1}_{\delta=0} P^\pi(\Theta_1|X) + \mathbf{1}_{\delta=1} P^\pi(\Theta_0|X)$$

Ainsi :

$$\delta^\pi(X) = 1 \iff P^\pi(\Theta_0|X) \leq P^\pi(\Theta_1|X)$$

C'est-à-dire que l'estimation permet d'accepter  $H_0$  si c'est l'hypothèse la plus probable a posteriori, ce qui est une réponse naturelle.

Une variante du test 0 – 1 est le test de Neymann Person qui permet de distinguer risques de première et de deuxième espèce :

$$L(\theta, \delta) = \begin{cases} 0 & \text{si } \delta = \mathbf{1}_{\theta \in \Theta_1} \\ a_0 & \text{si } \delta = \mathbf{1}_{\theta \in \Theta_0} \\ a_1 & \text{si } \delta = 0, \theta \in \Theta_1 \end{cases}$$

Le risque a posteriori est donné par :

$$\rho(\pi, \delta|X) = a_0\delta P^\pi(\Theta_1|X) + a_1(1 - \delta) P^\pi(\Theta_0|X)$$

Ainsi :

$$\delta^\pi(X) = 1 \iff a_0P^\pi(\Theta_0|X) \leq a_1P^\pi(\Theta_1|X)$$

Ceci permet de modéliser des raisonnements de la forme « j'ai plus ou moins tort » en jouant sur le rapport  $\frac{a_1}{a_0}$  et accorder plus ou moins d'importance relative à  $\Theta_0$  selon des arguments a priori.

### Remarque 2.5.1.

Le calcul d'une loi a posteriori mène à une loi. Ainsi, le résultat de l'inférence est beaucoup plus informatif que dans le cas fréquentiste : on a accès beaucoup plus facilement à des intervalles de confiances pour une estimation de  $\theta$  (en prenant par exemple le maximum de la loi a posteriori).

En pratique, on calculera la loi a posteriori empirique. Dans le cas continu pour  $X$ , on considère une réalisation  $x_1, \dots, x_n$  de  $X$ , on aura alors : La loi de  $X$  est continue et la loi a priori  $\pi$  sur  $\theta$  est discrète. La loi a posteriori est une loi discrète, comme la loi a priori et elle est définie par les probabilités suivantes :

$$\pi(\theta|X = (x_1, \dots, x_n)) = \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{\int_{\theta \in \Theta} f(X|\theta)\pi(\theta)d\theta}$$

Ainsi  $\Pi[. / \mathbf{X}] = \text{Gamma}(a + n, b + n\bar{X}_n)$ . La moyenne a posteriori vaut  $\frac{a+n}{b+n\bar{X}_n}$  (la moyenne d'une  $\text{Gamma}(a, b)$  est  $a/b$ ).

• On procède de même. On part de  $\Pi = \text{Beta}(a, b)$ . Dans le modèle de Bernoulli  $\mathcal{B}(p)^{\otimes n}$ , la loi a posteriori  $\Pi[. / \mathbf{X}]$  a pour densité

$$\begin{aligned} \pi(p|\mathbf{X}) &\propto \pi(p) \prod_{i=1}^n p_p(X_i) \\ &\propto p^{a-1}(1-p)^{b-1} \mathbf{1}_{0 < p < 1} \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} \\ &\propto p^{a+n\bar{X}_n-1}(1-p)^{b+n-n\bar{X}_n-1} \mathbf{1}_{0 < p < 1} \end{aligned}$$

Ainsi  $\Pi[. / \mathbf{X}] = \text{Beta}(a + n\bar{X}_n, b + n - n\bar{X}_n)$ . La moyenne a posteriori vaut  $\frac{a+n\bar{X}_n}{a+b+n}$  (la moyenne d'une  $\text{Beta}(a, b)$  est  $a/(a+b)$ ).

• Soit  $\Pi = \mathcal{N}(\mu, \sigma^2)$ . Il suffit de montrer que, si on prend  $\Pi$  comme loi a priori sur  $\theta$  dans

le modèle  $\{\mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R}\}$ , la loi a posteriori  $\Pi[\cdot / \mathbf{X}]$  est de la forme  $\mathcal{N}(\mu', \sigma'^2)$ . Ici  $\mathbf{X} = (X_1, \dots, X_n)$ , et  $\mathbf{X}/\theta \sim \mathcal{N}(\theta, 1)^{\otimes n}$ . On a

$$d\Pi(\theta) = \pi(\theta)d\theta, \quad \pi(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(\theta-\mu)^2}$$

$$dP_\theta(x) = p_\theta(x)dx, \quad p_\theta(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2}$$

La densité a posteriori s'écrit

$$\begin{aligned} \pi(\theta/\mathbf{X}) &\propto \pi(\theta) \prod_{i=1}^n p_\theta(X_i) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\theta-\mu)^2 - \frac{1}{2}\sum_{i=1}^n (X_i - \theta)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\theta^2 - 2\mu\theta) - \frac{1}{2}\sum_{i=1}^n (n\theta^2 - 2n\bar{X}_n\theta)\right\} \\ &\propto \exp\left\{-\frac{\sigma^{-2} + n}{2}\left(\theta^2 - 2\theta\frac{\mu\sigma^{-2} + n\bar{X}_n}{\sigma^{-2} + n}\right)\right\} \\ &\propto \exp\left\{-\frac{\sigma^{-2} + n}{2}\left(\theta - \frac{\mu\sigma^{-2} + n\bar{X}_n}{\sigma^{-2} + n}\right)\right\} \end{aligned}$$

Ainsi

$$\Pi[\cdot / \mathbf{X}] = \mathcal{N}\left(\frac{\mu\sigma^{-2} + n\bar{X}_n}{\sigma^{-2} + n}, \frac{1}{\sigma^{-2} + n}\right)$$

Donc la famille considérée est conjuguée. La moyenne a posteriori vaut  $\mathbb{E}[\theta/\mathbf{X}] = \frac{\mu\sigma^{-2} + n\bar{X}_n}{\sigma^{-2} + n}$ . On peut noter que cette moyenne a posteriori est une moyenne pondérée entre la moyenne  $\mu$  de la loi a priori et la moyenne empirique  $\bar{X}_n$  des observations :

$$\mathbb{E}[\theta/\mathbf{X}] = \frac{\sigma^{-2}}{\sigma^{-2} + n}\mu + \frac{n}{\sigma^{-2} + n}\bar{X}_n$$

En particulier, plus le nombre  $n$  d'observations est grand et moins la loi a priori affecte la loi a posteriori.

**Exemple 2.5.4.** Soit  $x_1, \dots, x_n$  un  $n$ -échantillon. On considère le modèle suivant :  $x_i \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a, b)$

$$f(x/\theta) = \prod_{i=1}^n P(X = x_i/\theta) = \theta^s(1-\theta)^{n-s}, \quad s = \sum_{i=1}^n x_i$$

$$\pi(\theta) = \frac{1}{\text{Beta}(a, b)} \theta^{a-1} (1 - \theta)^{b-1}, \quad 0 < \theta < 1$$

$$\pi(x/\theta) = \frac{f(x/\theta)\pi(\theta)}{\int_{\Theta} f(x/\theta)\pi(\theta)d\theta}$$

$$\begin{aligned} \int_{\Theta} f(x/\theta)\pi(\theta)d\theta &= \int_{\Theta} \frac{1}{\text{Beta}(a, b)} \theta^{s+a-1} (1 - \theta)^{n+b-s-1} d\theta \\ &= \frac{1}{\text{Beta}(a, b)} \int_{\Theta} \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta \\ &= \frac{\text{Beta}(\alpha, \beta)}{\text{Beta}(a, b)} \end{aligned}$$

Ou ;  $\alpha = s + a$        $\beta = n + b - s$

$$f(x/\theta)\pi(\theta) = \frac{1}{\text{Beta}(a, b)} \theta^{s+a-1} (1 - \theta)^{n+b-s-1}$$

Par conséquence :

$$\theta/x \sim \text{Beta}\left(a + \sum_{i=1}^n x_i, b +, - \sum_{i=1}^n x_i\right)$$

Les informations préalables sur les paramètres  $\theta$  s'entendent comme impliquant toute information En plus des informations fournies par l'observation, peuvent également être utilisés. L'information a priori sur  $\theta$  préalable s'accompagne d'incertitudes (sinon Dans ce cas, les paramètres  $\theta$  sont bien connus et n'ont pas à être estimés). Il est naturel de modéliser cette information a priori par des lois distributionnelles. e distribution a priori, notée  $\pi(\theta)$ .

Le choix des lois a priori est une étape fondamentale en statistique bayésienne et constitue une différence notable avec la statistique fréquentiste. Les différents choix possibles peuvent être motivés par différents points de vue :

- Choix basé sur des expériences du passé ou sur une intuition du statisticien.
- Choix basé sur la faisabilité des calculs.
- Choix basé sur la volonté de n'apporter aucune information nouvelle pouvant biaiser l'estimation.

## 2.6 Tests et régions de confiance

### 2.6.1 Régions de confiance

**Définition 2.6.1.** *Région  $\alpha$ -crédible*

Une région  $C$  de  $\Theta$  est dite  $\alpha$ -crédible si et seulement si  $P^\pi(\theta \in C|X) > 1 - \alpha$ .

Notons que le paradigme bayésien permet une nouvelle fois de s'affranchir d'un inconvénient de l'approche fréquentiste. En effet, au sens fréquentiste, une région de confiance  $C$  est définie par  $\forall \theta, P_\theta(\theta \in C) \geq 1 - \alpha$  et correspond à l'interprétation suivante. En refaisant l'expérience un grand nombre de fois, la probabilité que  $\theta$  soit dans  $C$  est plus grand que  $1 - \alpha$ . Une région de confiance n'a donc de sens que pour un très grand nombre d'expériences tandis que la définition bayésienne exprime que la probabilité que  $\theta$  soit dans  $C$  au vue des celles déjà réalisées est plus grande que  $1 - \alpha$ . Il n'y a donc pas besoin ici d'avoir recours à un nombre infini d'expériences pour définir une région  $\alpha$ -crédible, seule compte l'expérience effectivement réalisée.

Il y a une infinité de régions  $\alpha$ -crédibles, il est donc logique de s'intéresser à la région qui a le volume minimal. Le volume étant défini par  $\text{vol}(C) = \int_C d\nu(\theta)$ , si  $\pi(\theta|X)$  est absolument continue par rapport à une mesure de référence  $\nu$ . [10]

**Définition 2.6.2.** Région HPD (highest posteriori density)

$C_\alpha^\pi$  est une région HPD si et seulement si  $C_\alpha^\pi = \{\theta, \pi(\theta|X) \geq h_\alpha\}$  où  $h_\alpha$  est défini par

$$h_\alpha = \sup\{h, P^\pi(\{\theta, \pi(\theta|X) \geq h\}|X) \geq 1 - \alpha\}$$

$C_\alpha^\pi$  est parmi les régions qui ont une probabilité supérieure à  $1 - \alpha$  de contenir  $\theta$  (et qui sont donc  $\alpha$ -crédibles) et sur lesquelles la densité a posteriori ne descend pas sous un certain niveau (restant au dessus de la valeur la plus élevée possible). [10]

**Théorème 2.6.1.** Régions HPD

$C_\alpha^\pi$  est parmi les régions  $\alpha$ -crédibles celle de volume minimal si et seulement si elle est HPD. [10]

## 2.6.2 Tests

Comme nous l'avons vu précédemment, un test se formalise de la manière suivante : on définit une hypothèse nulle  $H_0 : \theta \in \Theta_0$  et une hypothèse alternative  $H_1 : \theta \in \Theta_1$  de telle sorte que  $\Theta_0 \cap \Theta_1 = \emptyset$  et  $\Theta_0 \cup \Theta_1 \subset \Theta$ .

Pour le modèle paramétrique  $X \in (\chi, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ ,  $\Theta_0 \cup \Theta_1 \subset \Theta$ . Le cas propre correspond au cas où la loi a priori est telle que si  $\theta$  suit la loi  $\pi$  alors  $\pi(\Theta_0) + \pi(\Theta_1) = 1 = \pi(\Theta)$ . Cette dernière condition peut s'exprimer aussi sous la forme suivante :  $\pi(\Theta \setminus (\Theta_0 \cup \Theta_1)) = 0$ . [10]

### 2.6.2.1 Approche par la fonction de perte de type 0 – 1

Dans cette partie [10], l'espace des décisions est  $\mathcal{D} = \{0, 1\}$ , la décision  $i$  correspondant à accepter  $H_i$ ,  $i = 0, 1$ . La fonction de perte est alors la suivante :

$$L(\theta, \delta) = \begin{cases} 0 & \text{si } \delta = \mathbb{1}_{\theta=\theta_1} \\ 1 & \text{si } \delta = \mathbb{1}_{\theta=\theta_0} \end{cases}$$

Comme cela a été montré précédemment, l'estimateur bayésien associé à cette fonction de perte est :

$$\delta^\pi(X) = \begin{cases} 1 & \text{si } P^\pi(\Theta_1|X) \geq P^\pi(\Theta_0|X) \\ 0 & \text{sinon} \end{cases}$$

**Exemple 2.6.1.** *Loi normale [10]*

$X \sim \mathcal{N}(\theta, \sigma^2)$  avec  $\sigma^2$  connu,  $\theta \sim \mathcal{N}(\mu, \tau^2)$  et pour :

$$L(\theta, \delta) = \begin{cases} 0 & \text{si } \delta = \mathbb{1}_{\theta=\theta_1} \\ a_0 & \text{si } \delta = \mathbb{1}_{\theta=\theta_0} \\ a_1 & \text{si } \delta = 0, \theta \in \Theta_1 \end{cases}$$

Alors :

$$\delta^\pi(X) = 1 \iff a_0 P^\pi(\Theta_0) < a_1 P^\pi(\Theta_1|X)$$

Ainsi,

$$\delta^\pi(X) = 1 \iff P^\pi(\Theta_1|X) > \frac{a_0}{a_0 + a_1}$$

Pour tester l'hypothèse  $H_0 : \theta < 0$  contre  $H_1 : \theta \leq 0$ , on calcule

$$\begin{aligned} \pi(\theta|X) &\propto e^{-\frac{(X-\theta)^2}{2\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\tau^2}} \\ &\propto e^{-\frac{(\theta-\mu(X))^2}{2V^2}} = \mathcal{N}(\mu(X), V) \end{aligned}$$

où

$$\mu(X) = \frac{X\tau^2}{\sigma^2 + \tau^2} + \frac{\mu\sigma^2}{\sigma^2 + \tau^2} \quad \text{et} \quad V = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}$$

Notons une dualité entre la fonction de perte  $\mu$  et la loi a priori  $(a_0, a_1)$ . On peut prendre à l'une pour donner à l'autre comme l'illustre l'exemple précédent. La détermination d'une fonction de risque ne peut se dissocier de celle de la loi a priori puisqu'elle correspond à une

façon de sanctionner l'erreur. Ceci s'explique par le fait que le statisticien a une idée a priori sur la répartition de l'erreur et donc sur la loi a priori du paramètre. Prendre  $\mu \gg 0$  équivaut à  $a_0 \ll a_1$ .

Ce dernier point se retrouve dans le cas des fonctions de perte quadratiques de la forme  $L(\theta, \delta) = \omega(\theta)(\theta - \delta)^2$  où seul le produit  $\omega \cdot \pi$  compte dans l'estimateur bayésien associé

$$\delta^\pi(X) = \frac{\int_{\Theta} \theta \omega(\theta) f(X|\theta) d\Pi(\theta)}{\int_{\Theta} \omega(\theta) f(X|\theta) d\Pi(\theta)}$$

Afin de rendre absolue la fonction de perte et qu'elle ne dépende plus de la loi a priori, nous introduisons dans la partie suivante le facteur de Bayes.

## 2.7 Évaluation d'hypothèses informatives à l'aide du facteur de Bayes

Le facteur de Bayes est un critère bayésien de sélection de modèle, comme il est un outi pour comparer la crédibilité de deux hypothèses. On va voir le facteur de Bayes pour tester les paramètres et les modèles avec quelques propriétés asymptotique du facteur de Bayes.

### 2.7.1 Le facteur de Bayes

#### Cas pour tester les paramètres

On considère le test d'hypothèse suivant :

$$H_0 : \theta \in \Theta_0 \text{ contre } H_1 : \theta \in \Theta_1$$

La loi a priori est donnée par le mélange de lois

$$\pi(\theta) = \pi_0 g_0(\theta) I\{\theta \in \Theta_0\} + \pi_1 g_1(\theta) I\{\theta \in \Theta_1\}$$

Où  $\pi_i(\theta) = P(\theta = H_i)$  et  $g_i(\theta)$  la fonction de densité a priori de  $\theta$  sous  $\Theta_i$ .

La fonction de vraisemblance sous  $H_i$  s'écrit :

$$P(X/H_i) = \int_{\theta \in \Theta_i} L(X/H_i, \theta) \pi(\theta) d\theta$$

Le facteur de Bayes de l'hypothèse  $H_0$  en faveur de l'hypothèse  $H_1$  s'exprime par :

$$B_{01} = \frac{P(X/H_0)}{P(X/H_1)}$$

Le facteur de Bayes est un critère bayésien de sélection de modèle, comme il est un outil pour comparer la crédibilité de deux hypothèses.

## Interprétation du facteur de Bayes :

Dans la pratique on compare  $B_{01}$  avec 1

- Si  $B_{01} > 1$  on accepte l'hypothèse  $H_0$ .
- Si  $B_{01} < 1$  on accepte l'hypothèse  $H_1$

**Exemple 2.7.1.** Pour  $x \sim \mathcal{B}(n, p)$  et  $p \sim U_{[0,1]}$ , on veut tester

$$H_0 : p > \frac{1}{2} \text{ contre } H_1 : p \leq \frac{1}{2}$$

$$\begin{aligned} P(H_0/x) &= P(p > \frac{1}{2}/x) \\ &= \int_{\frac{1}{2}}^1 \pi(p/x) dp \\ &= \frac{\int_{\frac{1}{2}}^1 p^x (1-p)^{n-x} dp}{\int_0^1 p^x (1-p)^{n-x} dp} \end{aligned}$$

et

$$\begin{aligned} P(H_0/x) &= P(p \leq \frac{1}{2}/x) \\ &= \int_0^{\frac{1}{2}} \pi(p/x) dp \\ &= \frac{\int_0^{\frac{1}{2}} p^x (1-p)^{n-x} dp}{\int_0^1 p^x (1-p)^{n-x} dp} \end{aligned}$$

pour  $x = 10$  et  $n = 15$

Donc  $P(p \leq \frac{1}{2}/x) = 15 \times 10^{-4}$  et  $P(p > \frac{1}{2}/x) = 0.0283$

Le facteur de Bayes associé est alors :

$$\begin{aligned} B_{01}(x) &= \frac{P(p > \frac{1}{2}/x)}{P(p \leq \frac{1}{2}/x)} \frac{\pi(\frac{1}{2}, 1]}{\pi(0, \frac{1}{2})} \\ &= \frac{\int_{\frac{1}{2}}^1 p^x (1-p)^{n-x} dp}{\int_0^{\frac{1}{2}} p^x (1-p)^{n-x} dp} \\ &= 18.87 \text{ pour } x = 10, \text{ et } n = 15 \end{aligned}$$

$B_{01}(x) = 18.87 > 1$ , donc on accepte  $H_0$

## 2.7.2 Propriétés asymptotiques des facteurs de Bayes

Cette section [10] cherche en outre à établir un lien entre le facteur de Bayes et la vraisemblance pénalisée. Le contexte est désormais usuel :  $X_n = (X_1, \dots, X_n)$

et  $X^n \sim^{iid} f(X; \theta)$ ,  $\theta \sim \pi$ , le test s'écrit toujours  $H_0 : \theta \in \Theta_0$  et  $H_1 : \theta \in \Theta_1$  dans le cadre un peu plus restreint  $\Theta \subset \mathbb{R}^d$ .

Si le modèle est régulier et si  $\theta_0 \in \Theta_0$  tel que  $\pi_1(\theta_0) > 0$  et de surcroît  $d\Pi_j(\theta) = \pi_j(\theta)d(\theta)$ , alors dans le cas où :

$$d\Pi(\theta) = \alpha\pi_0(\theta)\mathbf{1}_{\Theta_0}(\theta) + (1 - \alpha)\pi_1(\theta)\mathbf{1}_{\Theta_1},$$

on peut définir une sorte de rapport de vraisemblance par :

$$B_{0/1} = \frac{\int_{\Theta_0} e^{\log(\theta_n)} d\Pi_0(\theta)d\theta}{\int_{\Theta_1} e^{\log(\theta_n)} d\Pi_1(\theta)d\theta}$$

où  $\log(\theta_n)$  désigne  $\log(f(X^{nj}\theta))$  la valeur de la log-vraisemblance.

Les propriétés asymptotiques de  $B_{0/1}$  découlent du maximum du rapport de vraisemblance et sont les suivantes :

- si  $\theta_0 \in \Theta_0$ ,  $B_{0/1} \xrightarrow{P_{\theta_0}} \infty$ ,  $n \rightarrow +\infty$ .
- si  $\theta_0 \in \Theta_1$ ,  $B_{0/1} \xrightarrow{P_{\theta_0}} \infty$ ,  $n \rightarrow 0$ .

Ainsi le facteur de Bayes distingue la bonne hypothèse, avec une probabilité qui tend exponentiellement rapidement avec  $n$  vers 1. Le facteur de Bayes est dit consistant.

Remarquons que l'hypothèse  $\pi_1(\theta_0) > 0$  peut paraître illusoire étant donnée la propriété  $\Theta_0 \cap \Theta_1 = \emptyset$ . En réalité, cela peut être le cas si par exemple  $\Theta_0$  est dans l'adhérence de  $\Theta_1$  ; en particulier, si  $\Theta_1$  est de dimension  $k$ , on peut trouver de tels  $\Theta_0$  de dimension  $k - 1$ . Même si cela semble contraignant, il faut se rendre compte que ce cas recouvre un grand nombre de problèmes comme nous le verrons plus loin.

Dans le cadre d'hypothèses « pseudo ponctuelles » de la forme  $\Theta_0 = \{\theta_0\}$  ou  $\Theta = \theta_1; \theta_2$  et  $\Theta_0 = \theta_1 = \{\theta_{1,0}\}$  et pour  $\Theta_1 = \Theta/\Theta_0$ , si  $\pi_1$  est absolument continu par rapport à la mesure de Lebesgue alors  $\pi_1(\Theta_0) = 0$  et  $\pi(\Theta_1) = \pi(\Theta) = 1$ . Comme l'illustre l'exemple suivant, l'hypothèse ainsi formulée est « trop petite ». L'idée est que  $\theta_0$  est autant dans  $\Theta_1$  que dans  $\Theta_0$  au sens où  $\theta_0$  appartient à l'adhérence de  $\Theta_1$ .

**Exemple 2.7.2.** *Hypothèse pseudo-ponctuelle [10]*

*On regarde un modèle que l'on suppose mélangé au sens où il se pourrait que certains éléments*

soient tirés suivant une loi d'un certain paramètre et d'autres suivant la loi avec un autre paramètre. La vraisemblance est alors :

$$p \cdot f(X/\theta_1) + (1 - p)f(X/\theta_2)$$

Ainsi,  $\Theta = (p, \theta_1, \theta_2)$  et on définit :  $H_0 : p=1$  ou  $\theta_1=\theta_2$ .

Cela correspond à n'avoir qu'un paramètre ou deux :  $\Theta_1$  est de dimension 2,  $\Theta_0$  est de dimension 1. Le facteur de Bayes converge exponentiellement vers 0 sous  $H_0$  et de façon polynômiale vers l'infini sous  $H_1$ . Le problème est que l'on teste une hypothèse «trop petite» et qu'il est difficile d'accepter  $H_0$ .

Considérons le cadre particulier d'un choix de modèle dans une famille croissante de modèles :  $\forall j \Theta_j \subset \Theta_{j+1}$  tels que  $\pi_j(\Theta_j) = 1$ . L'inclusion doit se comprendre comme le fait que lorsqu'on est dans un modèle on est aussi dans un autre «plus grand». L'exemple du nombre de variables explicatives dans une regression linéaire est en ce sens parlant. Le but du choix de modèle est d'établir une procédure qui permet de sélectionner le modèle «le plus petit» possible, c'est-à-dire le plus précis. Si  $\theta_0$ , la vraie valeur, est contenue dans  $\Theta_j \setminus \Theta_{j-1}$  alors on veut sélectionner  $\Theta_j$  comme étant le plus petit modèle contenant  $\theta_0$  et ce avec un probabilité tendant vers 1 lorsque  $n \rightarrow +\infty$ .

### 2.7.3 Calcul du facteur de Bayes

la règle de Bayes consiste à choisir l'hypothèse de plus grande probabilité a posteriori. Supposons aussi que nous disposons des probabilités a priori des hypothèses

$$\begin{aligned} \pi_0 &= P(\theta \in \Theta_0) = P(H_0) \\ \pi_1 &= P(\theta \in \Theta_1) = P(H_1) = 1 - \pi_0 \end{aligned}$$

L'odds a priori (the odds prior) est le rapport des probabilités a priori de  $H_0$  relativement à  $H_1$

$$\frac{\pi_0}{\pi_1}$$

Ce rapport égal à 1 signifie que les hypothèses sont les mêmes avant d'observer les données. De même nous pouvons définir l'odds a posteriori (the odds posterior) comme

$$\frac{p_0}{p_1}$$

où

$$\begin{aligned} p_0 &= P(H_0|x) = P(\theta \in \Theta_0|x) \\ p_1 &= P(H_1|x) = P(\theta \in \Theta_1|x); \end{aligned}$$

sont les probabilités a posteriori sous les hypothèses  $H_0$  et  $H_1$  respectivement.

et le facteur de Bayes en faveur de  $H_0$  relativement à  $H_1$  est le rapport des deux odds

$$B_{0/1} = \frac{\text{odds a posteriori}}{\text{odds a priori}} = \frac{p_0 \pi_1}{\pi_1 \pi_0}$$

notons que la densité marginale de  $x$  sous  $\pi(\theta)$  peut être exprimée par

$$\begin{aligned} m_\pi(x) &= \int_{\Theta} f(x|\theta)\pi(\theta) d\theta \\ &= \pi_0 \int_{\Theta_0} f(x|\theta)g_0(\theta) d\theta + (1 - \pi_0) \int_{\Theta_1} f(x|\theta)g_1(\theta) d\theta \\ &= \pi_0 P(X|H_0) + (1 - \pi_0)P(X|H_1) \end{aligned}$$

d'où la distribution a posteriori de  $\theta$  sachant  $X = x$  est donnée selon la formule de Bayes par

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta) \pi(\theta)}{m_\pi(x)} \\ &= \begin{cases} \frac{\pi_0 f(x|\theta) g_0(\theta)}{m_\pi(x)} & \text{si } \theta \in \Theta_0 \\ \frac{(1-\pi_0) f(x|\theta) g_1(\theta)}{m_\pi(x)} & \text{si } \theta \in \Theta_1. \end{cases} \end{aligned}$$

et on en déduit que :

$$\begin{aligned} P(H_0|X) &= P(\theta \in \Theta_0|X) \\ &= \frac{\pi_0 \int_{\Theta_0} f(x|\theta) g_0(\theta) d\theta}{m_\pi(x)} \\ &= \frac{\pi_0 \int_{\Theta_0} f(x|\theta) g_0(\theta) d\theta}{\pi_0 \int_{\Theta_0} f(x|\theta) g_0(\theta) d\theta + (1 - \pi_0) \int_{\Theta_1} f(x|\theta) g_1(\theta) d\theta} \\ &= \frac{\pi_0 P(X|H_0)}{\pi_0 P(X|H_0) + (1 - \pi_0) P(X|H_1)} \end{aligned}$$

en remplaçant ces probabilités dans la formule (??), le facteur de Bayes est défini comme suit

$$\begin{aligned} B_{01} &= \frac{P(X|H_0)}{P(X|H_1)} \\ &= \frac{\int_{\Theta_0} f(x|\theta) g_0(\theta) d\theta}{\int_{\Theta_1} f(x|\theta) g_1(\theta) d\theta} \end{aligned}$$

Ce rapport est une analogie Bayésienne du rapport de vraisemblances des tests classiques, il évalue la modification de la vraisemblance de l'ensemble  $\Theta_0$  par rapport à celle de l'ensemble  $\Theta_1$ .

Les facteurs de Bayes sont très flexibles pour la comparaison des hypothèses multiples et des modèles.

## Exemples et Discussion

### 3.1 Exemples

#### 3.1.1 Exemple de test d'hypothèse dans l'approche classique

L'anémie est une maladie dans laquelle le sang ne contient pas suffisamment de globules rouges sains pour transporter suffisamment d'oxygène vers les tissus de l'organisme.

L'anémie (également connue sous le nom de faible taux d'hémoglobine) peut provoquer une sensation de fatigue et de faiblesse. Il existe plusieurs types d'anémie, et chacun a sa propre cause.

Ses causes comprennent la malnutrition, la carence en fer, l'insuffisance rénale, l'état pathologique (femme enceinte), l'état génétique (anémie falciforme), une augmentation des globules rouges.

En ce qui concerne cette maladie, il existe des différences entre les taux d'hémoglobine chez les femmes et les hommes, où pour les femmes, moins de 10 est considéré comme une anémie, tandis que pour les hommes, il est inférieur à 11.

Dans le but de comparer l'état du sang chez des personnes atteintes de cette maladie et d'autres qui sont en bonne santé; nous avons collecter des données à partir du laboratoire d'analyses médicales de l'hôpital Al-Ruqaiba, Wilaya de Oued Souf, pour étudier cette maladie.

Le tableau suivant : décrit le pourcentage d'hémoglobine chez certains patients et ceux en bonne santé.

| cas normal | cas de maladie | cas normal | cas de maladie |
|------------|----------------|------------|----------------|
| 11,2       | 8              | 13,2       | 5,3            |
| 16,1       | 9              | 16,9       | 8,1            |
| 13,9       | 7,1            | 12,7       | 8,6            |
| 15,2       | 9,3            | 10,4       | 7,3            |
| 12,7       | 6,1            | 17,3       | 7,2            |
| 14,2       | 8,4            | 12,2       | 9,5            |
| 10,1       | 9,1            | 11         | 9,8            |
| 12,1       | 9,2            | 14,8       | 8,8            |
| 10,5       | 7,6            | 11,4       | 7,1            |
| 11,9       | 8,1            | 14         | 7,9            |
| 10,5       | 9,6            | 11,1       | 7              |
| 13,1       | 7,4            | 14,2       | 7,5            |
| 16,2       | 9,7            | 11,6       | 8,3            |
| 12,4       | 9,8            | 14,3       | 8,2            |
| 10,6       | 7,7            | 10         | 5,4            |
| 10,4       | 8,7            | 15,6       | 6,3            |
| 11,8       | 7,8            | 15,1       | 6,2            |
| 10,8       | 9,4            |            |                |

TABLE 3.1 – Le taux d’hémoglobine chez certains patients et ceux en bonne santé.

Nous visons à comparer le taux d’hémoglobine, en utilisant un test d’homogénéité via les moyennes des échantillons.

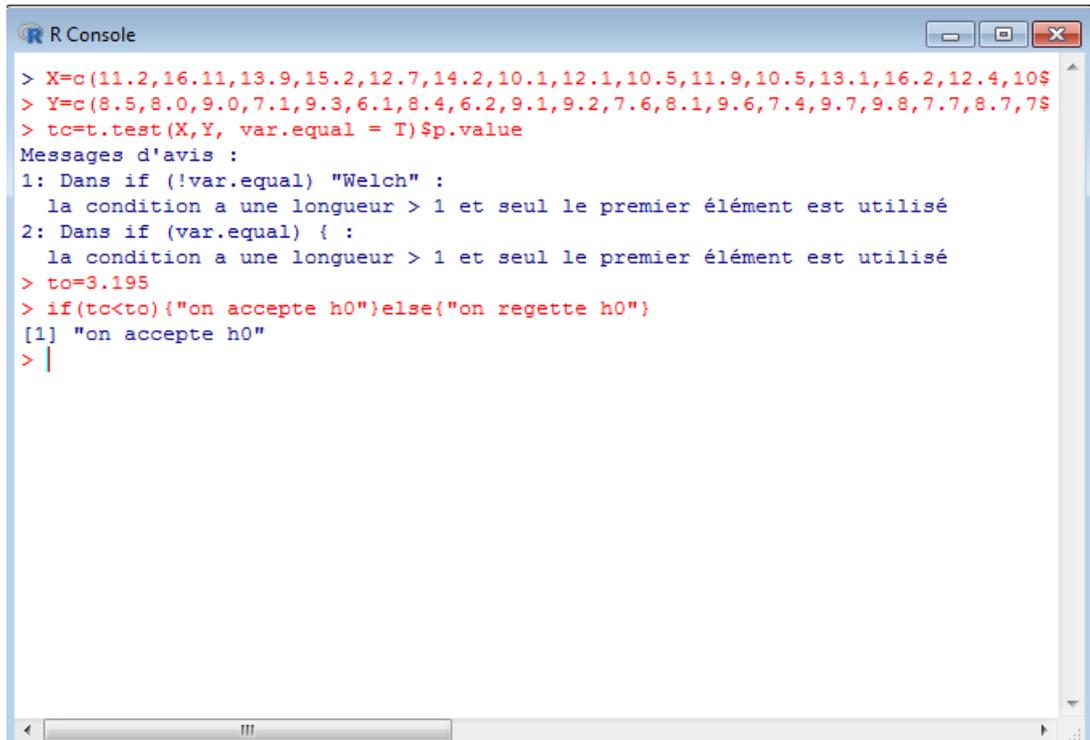
### 1. Premier étape : Construction des hypothèses

Nous avons les hypothèses suivantes :

$$\left\{ \begin{array}{l} H_0 : \text{Le taux d'hémoglobine est identique chez les malades et} \\ \quad \text{les personne en bonne santé : } m_1 = m_2 \\ H_1 : \text{Le taux d'hémoglobine est : } m_1 < m_2 \end{array} \right.$$

## 2. Deuxième étape : Vérification de la normalité

Avant de mener ce test, il faut vérifier la normalité du caractère étudié. Pour cela, nous avons effectué un test de Shapiro-Wilk, en utilisant le langage R.



```
R Console
> X=c(11.2,16.11,13.9,15.2,12.7,14.2,10.1,12.1,10.5,11.9,10.5,13.1,16.2,12.4,10.5)
> Y=c(8.5,8.0,9.0,7.1,9.3,6.1,8.4,6.2,9.1,9.2,7.6,8.1,9.6,7.4,9.7,9.8,7.7,8.7,7.5)
> tc=t.test(X,Y, var.equal = T)$p.value
Messages d'avis :
1: Dans if (!var.equal) "Welch" :
   la condition a une longueur > 1 et seul le premier élément est utilisé
2: Dans if (var.equal) { :
   la condition a une longueur > 1 et seul le premier élément est utilisé
> tc=3.195
> if(tc<to){"on accepte h0"}else{"on rejette h0"}
[1] "on accepte h0"
> |
```

On remarque que les données suivent la loi normal et sont homogènes au seuil  $\alpha = 0.001$ . Donc on peut appliquer le test paramétrique d'homogénéité.

## 3. Troisième étape : Calcul de la statistique du test

On applique le test d'homogénéité pour deux moyennes. on a  $\sigma_1$  et  $\sigma_2$  inconnu et égal,  $n_1$  et  $n_2$  supérieure ou égal à 30, la statistique du test est donc

$$\begin{aligned} Z_{obs} &= \frac{(\bar{x}_1 - \bar{x}_2) - \mathbb{E}[\bar{x}_1 - \bar{x}_2]}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \end{aligned}$$

$$\bar{x}_1 = \sum_{i=1}^{37} \frac{1}{n} x_i = 12.74$$

$$\bar{x}_2 = \sum_{i=1}^{37} \frac{1}{n} x_i = 8.05$$

$$\sigma_{\varepsilon_1}^2 = \frac{1}{n} \sum_{i=1}^{37} x_{1_i}^2 - \bar{x}_1^2 = 4.48$$

$$\sigma_{\varepsilon_2}^2 = \frac{1}{n} \sum_{i=1}^{37} x_{2_i}^2 - \bar{x}_2^2 = 1.46$$

Comme les variances sont inconnue, on le remplace par leurs estimateurs :

$$S_1^2 = \sigma_{\varepsilon_1}^2 \frac{n_1}{n_1 - 1} \text{ et } S_2^2 = \sigma_{\varepsilon_2}^2 \frac{n_2}{n_2 - 1}$$

$$S_1^2 = 4.61$$

$$S_2^2 = 1.50$$

D'ou la statistique du test est :

$$Z_{obs} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$$

En utilisant le langage R cette valeur vaut :  $Z_{obs} = 11.54$ .

#### 4. Quatrième étape : Règle de décision

D'après la table statistique de la loi normale, on trouve  $Z_c = 3.195$ .

On remarque que  $Z_{obs} > Z_c$ , donc on rejette l'hypothèse nulle  $H_0$  et on conclut que le taux d'hémoglobine est inférieur chez les malades.

### 3.1.2 Exemples des test d'hypothèse dans l'approche bayésienne

**Exemple 3.1.1.** Soit le test de

$$\begin{cases} H_0 : \theta \in \Theta \\ H_1 : \theta \notin \Theta \end{cases}$$

Alors  $\mathcal{D} = 0, 1$ , ou 1 représente l'acceptation de  $H_0$  et 0 son rejet. Pour la fonction de coût  $0 - 1$ , qui vaut

$$L(\theta, d) = \begin{cases} 1 - d, & \text{si } \theta \in \Theta_0 \\ d, & \text{sinon} \end{cases}$$

le risque associé est

$$R(\theta, \delta) = \mathbb{E}[L(\theta, \delta(\theta))] = \begin{cases} P_\theta(\delta(\theta) = 0), & \text{si } \theta \in \Theta_0 \\ P_\theta(\delta(\theta) = 1), & \text{sinon} \end{cases}$$

Ce qui donne exactement les erreurs de première et deuxième espèce qui sous tendent la Théorie de Neyman-Pearson.

**Exemple 3.1.2.** *Considérons un test pour contrôler le taux du sucre dans le sang d'une personne diabétique deux heures après son petit déjeuner. Il est d'intérêt de savoir si les médicaments qu'elle a pris ont contrôlé ce taux du sucre.*

Assumons que le résultat de ce test  $X$  suit une loi normale avec une moyenne  $\theta$  et une variance égale à 100, c'est-à-dire  $X \sim \mathcal{N}(\theta, 100)$  où  $\theta$  représente la vraie valeur de ce taux qui est inconnue. Dans la population (diabétiques qui suivent ce traitement) le paramètre  $\theta$  est distribué selon une loi normale de moyenne = 100 et variance = 900, i.e.  $\pi(\theta) = \mathcal{N}(100, 900)$ . La distribution marginale de  $X$  est donc une loi normale  $\mathcal{N}(100, 1000)$ .

Ainsi, la loi a posteriori de  $\theta$  est encore normale avec une moyenne

$$\text{moyenne} = \frac{900}{100} x + \frac{100}{1000} 100 = 0.9 x + 10$$

et une variance

$$\text{variance} = \frac{100 \times 900}{1000} = 90$$

Donc

$$\pi(\theta|x) = \mathcal{N}(0.9 x + 10, 90)$$

Supposons qu'on veut tester  $H_0 : \theta \leq 130$  contre  $H_1 : \theta > 130$ .

Si le test du taux du sucre dans le sang de cette personne donne un résultat  $x = 130$ , que peut on conclure ? notons que dans ce cas la loi a posteriori de  $\theta$  devient  $\pi(\theta|x) = \mathcal{N}(127, 90)$ . alors, on obtient :

$$P(\theta \leq 130|X = 130) = \Phi\left(\frac{130 - 127}{\sqrt{90}}\right) = \Phi(0.316) = 0.624$$

$$P(\theta > 130|X = 130) = 0.376$$

Le posterior odds donne dans ce cas :

$$\frac{P_0}{P_1} = \frac{0.624}{0.376} = 1.66$$

et comme

$$\pi_0 = P^\pi(\theta \leq 130) = \Phi\left(\frac{130 - 100}{30}\right) = \Phi(1)$$

le prior odds est donc égale à

$$\frac{\pi_0}{\pi_1} = \frac{\Phi(1)}{1 - \Phi(1)} = \frac{0.8413}{0.1587} = 5.3$$

Ainsi, le facteur de Bayes est

$$B_F = \frac{1.66}{5.3} = 0.313$$

.

## 3.2 Discussion

### 3.3 A/B testing : statistiques bayésiennes vs fréquentistes, un mauvais combat

Une drôle de rumeur semble (vouloir) faire du bruit dans le microcosme de l'A/B testing : une nouvelle méthode statistique viendrait révolutionner la pratique et donner des résultats plus probants, plus rapidement et plus simplement.

En résumé, la méthode dite «fréquentiste», utilisée depuis une dizaine d'années par toutes les solutions d'A/B testing ne serait pas la bonne : l'approche dite «bayésienne» (déjà bien connue des statisticiens depuis le XVIII<sup>ème</sup> siècle) serait plus adaptée à la pratique. Tout le monde se serait-il donc trompé pendant toutes ces années au point de pousser désormais certains à brûler ce qu'ils ont adoré ?

#### 3.3.1 Statistiques fréquentistes et bayésiennes

Deux écoles statistiques se sont affrontées depuis les Lumières : fréquentiste et les bayésienne. La statistique fréquentiste est basée sur la loi des observations et peut être appelée *expérimentale* ou *inductive*, tandis que la statistique bayésienne L'approche Bayésienne traite le paramètre inconnu comme une variable et lui attribue une loi de probabilité, donc elle permet d'introduire l'information a priori disponible sur le paramètre. Elle ne s'intéresse plus à la vraisemblance comme l'approche fréquentiste mais plutôt à la loi a posteriori, qui est une loi de probabilité contrairement à la vraisemblance. Elle peut être appelée *théorique* ou *déductive*. Pour obtenir des

informations a posteriori, les informations contenues dans les données peuvent être combinées avec les résultats a priori d'études antérieures et d'opinions d'experts.

Prenons un exemple simple pour mieux comprendre la différence entre ces deux approches. Lançons une pièce 10 fois. Si nous nous appuyons sur la modélisation fréquentiste, il existe une "vraie" probabilité d'obtenir une queue de valeur  $p$ . Par exemple, si vous obtenez 6 queues sur 10, la probabilité d'obtenir une queue à partir des résultats de cette expérience est égale à  $6/10 = 0,6$ . Selon l'approche bayésienne, on ne s'intéresse pas à cette probabilité, mais à la distribution précédente. En fait, si la pièce est équilibrée, la probabilité d'atterrir face est la même que la probabilité d'atterrir face a priori. Donc  $1/2 = 0,5$ . Cette probabilité a priori provient des résultats d'autres expériences réalisées dans le passé. En fait, après avoir jeté une pièce plusieurs fois, nous constatons que la probabilité calculée par la méthode fréquentiste converge vers 0,5.

### 3.3.2 L'intérêt de l'approche bayésienne et ses limites

Par conséquent, l'approche bayésienne est d'un grand intérêt lorsque des expériences passées complètement similaires peuvent être considérées. Par conséquent, elle est utilisée dans plusieurs domaines tels que la détection de spam. La connaissance préalable du spam nous permet d'attribuer des probabilités correspondant aux fréquences d'occurrences des parties du discours. Cette probabilité dérivée empiriquement rend un mot typique du spam. Le principal avantage de cette méthode est donc la possibilité de s'affranchir du laps de temps fixé et d'obtenir des résultats le plus rapidement possible. Vous n'avez pas à déterminer à l'avance la taille de l'échantillon et la quantité de trafic dont nous avons besoin pour exécuter Nos tests. Les résultats sont affichés pour toutes les expériences et disponibles plus rapidement.

La statistique bayésienne dérive la probabilité d'un événement en considérant les probabilités d'autres événements qui ont déjà été évalués. Dans les tests A/B, les connaissances préalables sont soumises à des variations saisonnières et à des tendances simples qui peuvent fausser vos résultats. Cela signifie que le risque de détecter des faux positifs est beaucoup plus élevé. Le spam n'est pas toujours sérieux. C'est un problème encore plus important avec les tests A/B.

Les méthodes bayésiennes ont également l'inconvénient d'être beaucoup plus difficiles à comprendre. Les statisticiens bayésiennes tentent de calculer des distributions de probabilité, qui sont des concepts plus complexes que de simples mesures de confiance. Pour les tests A/B, cette

distribution de probabilité est basée sur les conversions gagnantes ou perdantes. Si nous simplifions à l'extrême cette distribution et la réduisons à un simple intervalle tel que  $[-0.5\%, +2\%]$  sales, le responsable marketing n'aura pas suffisamment d'informations pour lire les résultats (c'est-à-dire,  $-0,5\%$ , ou  $+2\%$  ?). D'autant plus que la distribution est évidemment basée sur l'intervalle  $[-\infty, +\infty]$ . Le "seuil" pour l'intervalle  $[-0.5\%, +2\%]$  est arbitraire et basé sur le seuil au-dessus duquel les poids statistiques sont considérés comme négligeables.

### 3.3.3 Puissance de la méthode fréquentiste

Les tests A/B, qui dominent les entreprises et les soins de santé, ont été dominés par des méthodes fréquentistes depuis le début. Cette méthode est basée uniquement sur des données de test sous la stricte condition d'égalité des variantes (d'où sa réputation de méthode "data-driven"). Les lacunes de l'approche fréquentiste sont bien connues et détaillées dans une série de trois forums sur la signification statistique.

1. «Attention à la vraie valeur statistique de vos tests A/B»
2. «Comprendre les dessous d'un test A/B»
3. «A/B testing : votre trafic est-il suffisant ?»

En particulier, le volume de trafic requis ne permet pas tous les types de tests en toutes circonstances. De plus, la fiabilité des résultats n'est réelle qu'à la fin du test. Les résultats intermédiaires n'ont pas de sens, nous devons donc savoir comment résister à la tentation de "sélectionner plusieurs fois" lors des tests.

### 3.3.4 Quelle approche est meilleure de l'autre ?

L'une des analyses les plus approfondies opposant les approches fréquentiste et bayésienne a été menée par le statisticien Valen Johnson et résumée dans un article de 2013 publié dans les Actes de l'Académie nationale des sciences. Le but de son analyse fréquentiste était d'examiner les données recueillies pour identifier des effets significatifs qui ne pouvaient être expliqués que par des hypothèses empiriques. Son analyse bayésienne a fait face à deux hypothèses et a pesé la probabilité que l'une soit vraie par rapport à l'autre sur la base des données disponibles au moment de l'expérience et des informations précédemment connues sur le sujet. Sa conclusion est que le niveau de signification statistique de 95% de l'approche bayésienne communément acceptée est insuffisant pour conclure que le test est significatif. Cela ne fait que renforcer

l'approche souvent adoptée par les éditeurs de solutions d'A/B testing. La loi de Bayes devrait-elle donc être disqualifiée? Non, car elle possède un bien très précieux, si les circonstances le permettent. Il s'avère que le monde des tests A/B adopte logiquement une approche fréquentiste car la fiabilité accrue et la complexité réduite de la lecture des résultats l'emportent largement sur les inconvénients ci-dessus.

De manière générale, évaluer entre méthodes fréquentistes et bayésiennes devient un débat d'experts très éloigné des intérêts des équipes marketing. Absolument aucune méthode n'est meilleure qu'une autre. La clé est de bien comprendre la logique sous-jacente ou de demander conseil à quelqu'un de plus compétent.

# Conclusion Générale

Dans ce mémoire nous avons abordé un aspect très important de l'inférence statistique qui est les tests paramétriques. Nous avons donné les démarches pour effectuer un test d'un point de vue fréquentiste, ainsi que d'un point de vue Bayésien.

Nous avons constaté qu'il faut pas privilégier une approche ou l'autre ; chacune a ses avantages et ses inconvénients ; et selon les conditions disponibles et le domaine d'application, une approche peut être mieux que l'autre.

# Bibliographie

- [1] **Béland S., Raïche G. et Magis D.** (2015). "*L'utilisation du facteur de Bayes pour identifier les étudiants qui répondent au hasard*". *Revue des sciences de l'éducation*, 41(3), 385–407. <https://doi.org/10.7202/1035310ar>
- [2] **Berger, James O, Lawrence D. Brown and Robert L. Wolpert.** "*A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing*", *The Annals of Statistics* (1994) : 1787-1807.
- [3] **Boris Hejblum**, *Introduction à l'analyse bayésienne et à ses méthodes numériques pour la biostatistique*, Polycopié du Cours Master 2 Biostatistique, Université de Bordeaux, 2021
- [4] **Christophe Chesneau.** "*Introduction aux tests statistiques avec R*". Licence. France. 2016. cel- 01387707v2
- [5] **Christian P Robert.** "*Des spécificités de l'approche bayésienne et de ses justifications en statistique inférentielle*". 2013. hal-00870124
- [6] **CHERFAOUI Mouloud**, "*Statistiques Appliquées à l'Expérimentation En Sciences Biologique*", Polycopié du Cours, Université de Biskra, 2016
- [7] **Eric Johannesson.** "*Classical versus bayesian statistics*". *Philosophy of Science*, 87(2) :302–318, 2020.
- [8] **Ghouil Djowya**, Cours Biostatistique Master 2 probabilités et statistiques, Université de Jijel
- [9] **Jean-Michel Marin and Christian P Robert.** "*Les bases de la statistique bayésienne*". *Techniques de l'ingénieur*, pages 1–25, 2009.
- [10] **Judith Rousseau.** "*Statistique bayésienne notes de cours*". ENSAE ParisTech troisième année, 2009

## Bibliographie

- [11] **Lydia Larbi**. *"Sur la décision statistique dans le contexte Bayésien"*. Mémoire de master, UMMTO, 2011.
- [12] **Oumoul-Hair** et **Illa Amadou**. *"Approche Bayésienne des tests sur les modèles auto-régressifs vectoriels"*. Mémoire de master, UMMTO, 2013.
- [13] **Yann Traonmilin** et **Adrien Richou**, *Introduction aux Statistiques Bayésiennes*, Cours de l'UE, 2018
- [14] **Yasmina Bouzeriba** et **Ahlam Boumendje**, *"L'utilité de l'approche bayésienne dans l'estimation de la fonction de survie"*, Mémoire de master, Université de jijel, 2018.
- [15] **Zeyneb Abdi**, *"Les tests bayesiens dans les essais cliniques"*, Magister en mathématique, Université Mentouri Constantine.