

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE MOULOU D MAMMERI, TIZI-OUZOU

FACULTE DES SCIENCES

DEPARTEMENT DE MATHEMATIQUES

MEMOIRE DE MASTER

FILIERE : MATHÉMATIQUES

SPECIALITE : PROBABILITES ET STATISTIQUES

Présenté par :

M^{lle} **HAMIDANI Imane**

THEME

Le Bayésien Empirique, Application Aux Séries Chronologiques

Devant le jury d'examen composé de :

Fellag Hocine

Professeur

U.M.M.T.O ; Président

Belkacem Cherifa

MCB

U.M.M.T.O ; Rapporteur

Atil Lynda

MCA

U.M.M.T.O ; Examinatrice

Soutenu le : 25/09/2024

Remerciements

Je remercie le Tout-Puissant de m'avoir guidé vers les portes du savoir, en illuminant mon chemin de Sa lumière. Sa force m'a insufflé le courage et la persévérance nécessaires pour mener à bien ce travail.

Je tiens à exprimer ma profonde gratitude et mes plus sincères remerciements à M^{me} **Belkacem Cherifa** pour avoir accepté de m'encadrer, mais aussi sa gentillesse, et ses précieux conseils non seulement tout au long de mon travail mais aussi durant mes 3 dernières années d'études.

Je remercie vivement Pr **Fellag Hocine**, pour l'honneur qu'il me fait en acceptant d'être président du jury. Toute ma reconnaissance aux membres du jury pour avoir accepté d'examiner et d'enrichir mon travail par leurs propositions et remarques, je suis honoré de la présence de Mme **Atil Lynda**. Qu'elle trouve ici l'expression de mon profond respect.

Je remercie également toutes les personnes du département de mathématiques, particulièrement tous les enseignants qui nous ont donné des connaissances tout au long de mon cursus, ainsi que toutes celles et ceux qui ont de près ou de loin contribué à la réalisation de ce modeste travail.

En fin, un grand merci à ma chère famille, particulièrement mes parents pour leur protection filiale et leur encouragement, et à tous mes amies et camarades.

Table des matières

Table des matières	2
INTRODUCTION GÉNÉRALE	5
1 GÉNÉRALITÉS SUR LES SÉRIES CHRONOLOGIQUES	8
1.1 Définition et domaines d'applications des séries chronologiques	8
1.1.1 Série Chronologique	8
1.1.2 Objectif de l'analyse des séries chronologiques	12
1.1.3 Les domaines d'applications d'une série chronologique	13
1.2 Composantes d'une série chronologique	13
1.3 Modélisation d'une série chronologique	15
1.3.1 Modèle additif	15
1.3.2 Modèle multiplicatif	16
1.3.3 Modèle mixte	18
1.4 Les Opérateurs dans les séries chronologiques	18
1.4.1 Opérateur retard B (Backward)	18
1.4.2 Opérateur de différence	19
1.5 Stationnarité	19
1.5.1 Stationnarité stricte (forte)	19
1.5.2 Stationnarité faible	20
1.6 Processus stochastique	20
1.6.1 Processus stochastique	20
1.6.2 Réalisation d'un processus stochastique	21
1.6.3 La fonction d'autocovariance	21
1.6.4 La fonction d'autocorrelation (acf)	22
1.7 Bruit Blanc	22
1.8 Modélisation ARIMA (AutoRegressive Integrated Moving Average)	23
1.8.1 Processus autorégressif AR(p)	23
1.8.2 Processus à moyenne mobile MA(q)	24
1.8.3 Causalité et Invérsibilité	24
1.8.4 Processus ARMA(p,q)	26
1.8.5 Le processus intégré I(d)	27
1.8.6 Le processus ARIMA(p,d,q)	27

2	STATISTIQUE BAYÉSIENNE ET BAYÉSIENNE EMPIRIQUE	28
2.1	Inférence Bayésienne	28
2.1.1	Théorème de Bayes	28
2.1.2	Modèle Statistique Bayésienne	29
2.1.3	Description d'un modèle Bayésien	30
2.1.4	Les bases de la décision de l'approche Bayésienne	30
2.1.5	Fonctions coût usuelles	35
2.1.6	Modélisation de l'information a priori	39
2.2	La Statistique Bayésienne Hiérarchique	44
2.3	La Statistique Bayésienne Empirique	46
2.3.1	Le Concept Du Bayésien Empirique	47
2.3.2	Estimation Bayésienne empirique pour la loi de Poisson	48
2.3.3	Estimation Bayésienne empirique pour la loi Normale	63
2.3.4	L'importance des méthodes Bayésiennes empiriques en statistiques	69
2.4	Principales différences Entre les deux méthodes Bayésiennes et Bayésiennes empirique	69
3	ESTIMATION BAYÉSIENNE EMPIRIQUE DANS UN PROCESSUS AUTORÉ- GRESSIF D'ORDRE UN (AR(1))	71
3.1	Introduction	71
3.2	Présentation du modèle	71
3.3	Estimation Bayésienne Empirique	72
3.3.1	La fonction de Vraisemblance	72
3.3.2	La loi a priori	73
3.3.3	Loi a posteriori	73
3.3.4	Estimateur Bayésien	75
3.3.5	Estimateur Bayésien empirique	76
3.4	Propriétés asymptotiques	78
3.5	Estimation du coefficient d'un AR(1) avec diverses approches	79
3.5.1	Présentation du modèle	79
3.5.2	Estimateur de maximum de vraisemblance	79
3.5.3	Estimateur Bayésien	80
3.5.4	L'estimateur Bayésien Empirique	82
3.6	Application numérique	84
3.6.1	Calcul du MSE pour chaque estimation	85
	CONCLUSION GÉNÉRALE	88
A	PROGRAMMATION	89
A.1	Programme 1	89
A.2	Programme 2	89
A.3	Programme 3	90
A.4	Programme 4	92

TABLE DES MATIÈRES

A.5 Programme 5	93
A.6 Programme 6	94
Bibliographie	96

INTRODUCTION GÉNÉRALE

L'analyse statistique est devenue un pilier essentiel dans divers domaines scientifiques, offrant des outils puissants pour interpréter des phénomènes incertains et complexes. Que ce soit en économie, en biologie, en ingénierie ou en finance, les méthodes statistiques permettent de traiter des données et d'en extraire des informations cruciales pour la prise de décisions. Parmi ces approches, l'inférence bayésienne qui tire son nom du révérend Thomas Bayes (1702 – 1761), un mathématicien anglais du 18^{ème} siècle. Ses travaux posthumes publiés en 1763 ont introduit le théorème de Bayes, un principe mathématique fondamental qui permet de mettre à jour les probabilités à mesure que de nouvelles informations deviennent disponibles. Le théorème de Bayes permet de combiner des informations a priori sur un phénomène avec des données observées pour en déduire une probabilité a posteriori, ce qui en fait un outil précieux pour l'inférence statistique.

Cependant, l'approche bayésienne n'a véritablement pris son essor qu'au milieu du 20^{ème} siècle, grâce aux travaux de Harold Jeffreys et Leonard Jimmie Savage, qui ont défendu et développé la théorie bayésienne dans un cadre statistique moderne. Ils ont montré que l'inférence bayésienne pouvait être utilisée non seulement pour estimer des paramètres inconnus, mais aussi pour prendre des décisions dans des situations d'incertitude.

Aujourd'hui, l'inférence bayésienne est largement utilisée dans de nombreux domaines, notamment en économie, en épidémiologie, en intelligence artificielle, et dans l'analyse des données, grâce à sa capacité à intégrer des connaissances préalables et à mettre à jour ces connaissances à mesure que de nouvelles données sont recueillies.

Cependant, dans de nombreux cas pratiques, déterminer une distribution a priori pertinente peut s'avérer difficile, car on ne peut pas connaître la loi a priori avec certitude. L'approche bayésienne pour modéliser l'incertitude consiste à mettre les mesures a priori sur les paramètres de la loi a priori originale. Ce type de modélisation bayésienne est appelé la méthode bayésienne hiérarchique (Berger (1985) et Robert (1992, 2001)). Par contre, la méthode d'analyse bayésienne empirique, qui est un mélange de l'inférence bayésienne et fréquentiste, est une approche pour rectifier l'incertitude concernant la loi a priori. Cette approche a vu le jour au milieu du 20^{ème} siècle, lorsqu'elle a été initialement proposée par Herbert Robbins (1951, 1955, 1964, 1980), qui a d'abord proposé une approche non paramétrique pour estimer la loi marginale (1955), laquelle requiert un échantillon suffisamment grand pour garantir une approximation fiable. En réponse à cette exigence, plusieurs chercheurs ont développé des méthodes alternatives. Par exemple, Morris (1983) a introduit la méthodologie bayésienne empirique paramétrique, qui a trouvé de nombreuses applications

dans divers domaines comme l'apprentissage automatique, la biostatistique et le traitement du signal.

L'idée centrale de L'approche Bayésien Empirique est que les données peuvent elles-mêmes être utilisées pour guider la sélection des paramètres a priori. Cela permet d'éviter le problème du choix arbitraire de la distribution a priori, un défi majeur dans l'approche bayésienne classique. Robbins a montré que cette approche pouvait être appliquée dans de nombreux contextes pratiques, notamment dans les modèles de Poisson et les problèmes de décision statistique.

Au cours des décennies suivantes, le Bayésien Empirique a été largement utilisé dans des applications telles que l'estimation des paramètres dans les modèles hiérarchiques et la prévision de séries chronologiques. La popularité de cette méthode s'explique par sa flexibilité et son efficacité dans des situations où des informations a priori fiables sont difficiles à obtenir. En combinant l'approche fréquentiste et bayésienne, le Bayésien Empirique offre une méthode pragmatique qui utilise le meilleur des deux mondes.

Aujourd'hui, le Bayésien Empirique est une méthode couramment utilisée dans l'analyse des données, notamment dans des domaines tels que la génétique, la bioinformatique, la finance et l'économie. Ses applications dans l'analyse des séries chronologiques permettent de modéliser et de prévoir des phénomènes incertains, tout en tenant compte des tendances cachées et de l'incertitude inhérente aux données.

L'un des champs d'application les plus pertinents pour cette méthode est celui des séries chronologiques, des données collectées et ordonnées au fil du temps. L'étude des séries chronologiques remonte au début du 20ème siècle, avec les travaux pionniers de chercheurs en statistique et en économie. L'une des premières contributions significatives dans ce domaine est celle de l'économiste britannique Sir Francis Ysidro Edgeworth en 1887, qui a commencé à explorer l'analyse des données temporelles. Cependant, c'est dans les années 1920 que la théorie des séries chronologiques a pris son véritable essor, grâce aux travaux d'un autre statisticien britannique, Sir George Udny Yule. Yule a introduit le modèle autorégressif (AR) en 1927, l'un des modèles les plus fondamentaux utilisés pour modéliser les séries chronologiques. Les séries chronologiques sont omniprésentes dans des domaines aussi variés que la finance, la météorologie, l'économie, ou encore la santé publique. Elles permettent d'analyser des tendances passées et d'effectuer des prévisions, ce qui est d'une importance capitale pour anticiper des événements futurs.

En appliquant les techniques du Bayésien Empirique à l'analyse de séries chronologiques, il est possible de mieux capter les tendances cachées et de produire des prévisions plus précises tout en tenant compte de l'incertitude inhérente à ce type de données.

Le présent mémoire s'attache à explorer les fondements théoriques de l'approche Bayésienne Empirique et à démontrer son utilité dans l'analyse des séries chronologiques. Il se décompose en plusieurs chapitres structurés comme suit : Le premier chapitre propose une vue d'ensemble sur les séries chronologiques, en commençant par une définition rigoureuse du concept et une présentation des principales méthodes de modélisation, telles que les modèles additifs, multiplicatifs et mixtes. Nous aborderons également les opérateurs retard et différence, ainsi que le concept de stationnarité, qui est fondamental dans l'analyse des séries temporelles.

Le second chapitre sera consacré à l'étude approfondie de la statistique bayésienne et du bayésien empirique. Nous commencerons par une présentation de l'inférence bayésienne, en introduisant les fondements théoriques avec le théorème de Bayes et son application aux modèles statistiques bayésiens. Cette section abordera également les bases de la prise de décision dans un cadre bayésien, avec une analyse des fonctions de coût usuelles et de la modélisation de l'information a priori.

Ensuite, nous explorerons le concept du bayésien hiérarchique, avant de nous concentrer sur la méthode du bayésien empirique. Plusieurs lois statistiques, comme la loi de Poisson et la loi normale, seront utilisées pour illustrer cette approche, en mettant en avant l'importance des méthodes bayésiennes empiriques dans les statistiques modernes.

Enfin, une comparaison des différences principales entre la méthode bayésienne empirique et la méthode bayésienne traditionnelle sera présentée, permettant de mieux comprendre les avantages et les limites de chacune.

Enfin, Le troisième chapitre sera axé sur une application pratique du Bayésien Empirique aux séries chronologiques, notamment à travers l'estimation et la prévision des modèles AR(1). À partir de données réelles, nous illustrerons comment cette approche permet de formuler des prévisions plus fiables, en tenant compte de l'incertitude sur les paramètres du modèle.

L'objectif final de ce travail est de démontrer que l'approche Bayésienne Empirique offre une méthodologie flexible et efficace pour l'analyse des séries chronologiques. Elle constitue une alternative intéressante aux méthodes traditionnelles, en permettant une meilleure intégration des données disponibles dans la formulation des hypothèses, ce qui en fait un outil précieux pour les chercheurs et les praticiens confrontés à des données temporelles incertaines.

Chapitre 1

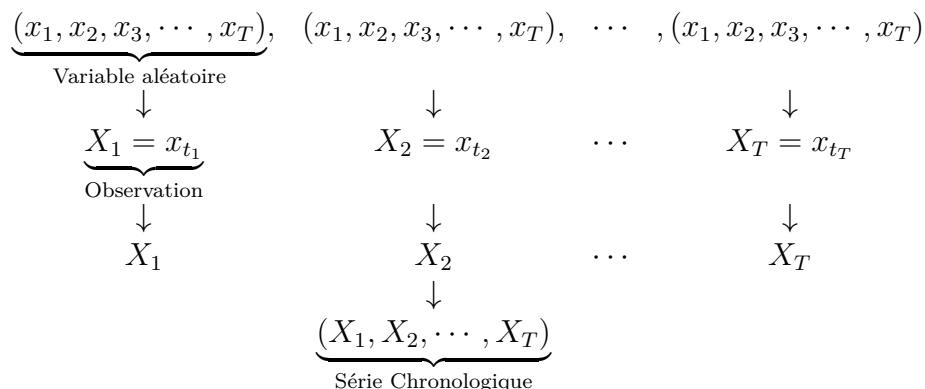
GÉNÉRALITÉS SUR LES SÉRIES CHRONOLOGIQUES

La théorie des séries chronologiques, également connue sous le nom de séries temporelles, constitue une approche analytique visant à étudier l'évolution d'un phénomène au fil du temps, dans le but de décrire, expliquer et anticiper ses variations futures. Cette méthodologie repose sur l'analyse de données observées à des moments spécifiques, formant ainsi une séquence de valeurs numériques indexées temporellement.

1.1 Définition et domaines d'applications des séries chronologiques

1.1.1 Série Chronologique

Définition 1. Une série chronologique ou série temporelle est un ensemble d'observations $(x_t)_{t \in \mathbb{T}}$, chaque observation étant enregistrée à un moment spécifique t . Une série chronologique peut être considérée comme une collection de variables aléatoires $\{X_t : t \in \mathbb{T}\}$, en d'autres termes, chaque observation individuelle dans une série chronologique est une valeur spécifique prise par une variable aléatoire à un instant donné.



1.1. DÉFINITION ET DOMAINES D'APPLICATIONS DES SÉRIES CHRONOLOGIQUES

- **Une série temporelle discrète** est celle pour laquelle l'ensemble \mathbb{T} des moments auxquels les observations sont effectuées est un ensemble discret, dans ce cas $\mathbb{T} \subset \mathbb{Z}$.
Exemple : Évolution d'une population au fil des années.
- **Les séries temporelles continues** sont obtenues lorsque les observations sont enregistrées de manière continue sur un certain intervalle de temps $\mathbb{T} \subset \mathbb{R}$.
Exemple : mesuré la température d'une région tout au long d'une journée.

Exemples de séries chronologiques

Exemple 1. `ts <- AirPassengers`
`plot(ts, xlab='year', ylab='Number of Passengers', type="o")`

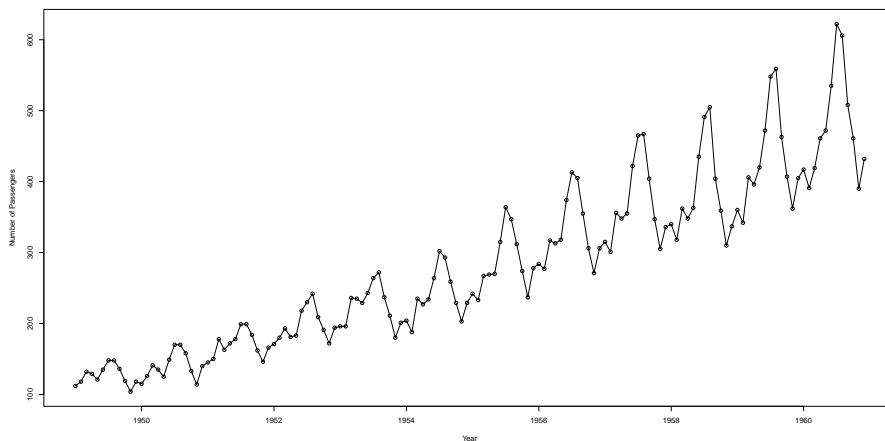


FIGURE 1.1 – Le nombre mensuel de passagers aériens internationaux, en milliers, de janvier 1949 à décembre 1960.

La Figure 1.1 présentant l'évolution du nombre de passagers offre une vision significative des voyages aériens sur une période de 10 ans. La croissance du nombre de passagers est clairement observée, indiquant l'essor continu du secteur du transport aérien.

Des variations saisonnières émergent, illustrant des pics de voyages à certaines périodes de l'année, probablement liés aux vacances ou aux événements saisonniers. Ces variations soulignent l'impact des facteurs saisonniers sur les tendances du secteur aérien.

Les premières années (1949, 1950,...) ont été témoins d'une augmentation constante du nombre de passagers, suggérant un intérêt croissant pour les voyages aériens à l'échelle mondiale. Cette période peut être associée à des avancées technologiques majeures qui ont rendu les voyages plus accessibles.

En arrivant à l'année 1960 on remarque que le transport aérien a connu une croissance remarquable pendant cette décennie, avec une expansion rapide du nombre de passagers.

1.1. DÉFINITION ET DOMAINES D'APPLICATIONS DES SÉRIES CHRONOLOGIQUES

Ce phénomène peut être attribué à divers facteurs, y compris l'essor économique, la démocratisation des voyages aériens (disponibilité des vols pour un public plus large) et les changements culturels favorisant la mobilité.

L'analyse plus approfondie de ces données pourrait permettre de mettre en lumière des tendances à plus long terme, des modèles de croissance spécifiques ou des événements marquants ayant influencé le secteur aérien. Ces informations sont essentielles pour les compagnies aériennes et autres acteurs du secteur afin d'anticiper les demandes futures, de planifier les capacités et de comprendre l'évolution dynamique du paysage des voyages aériens.

Exemple 2. *la série chronologique suivante présente les observations mensuelle du Concentration du CO₂ dans l'air pendant 7ans de Octobre 1977 à Mai 1984. Les données sont pris de [?].*

```
co2_data <- read.table("D:/2CO2.txt", header = FALSE)
colnames(co2_data) <- c("CO2", "YearMonth", "Year", "Month")
co2_data$Date <- as.Date(with(co2_data, paste(Year, Month, "01", sep = "-")),
  "%Y-%m-%d")
plot(co2_data$Date, co2_data$CO2, type = "o", xlab = "Temp",
  ylab = "Concentration du CO2",
  main = "La Concentration du CO2 en fonction du temp ")
```

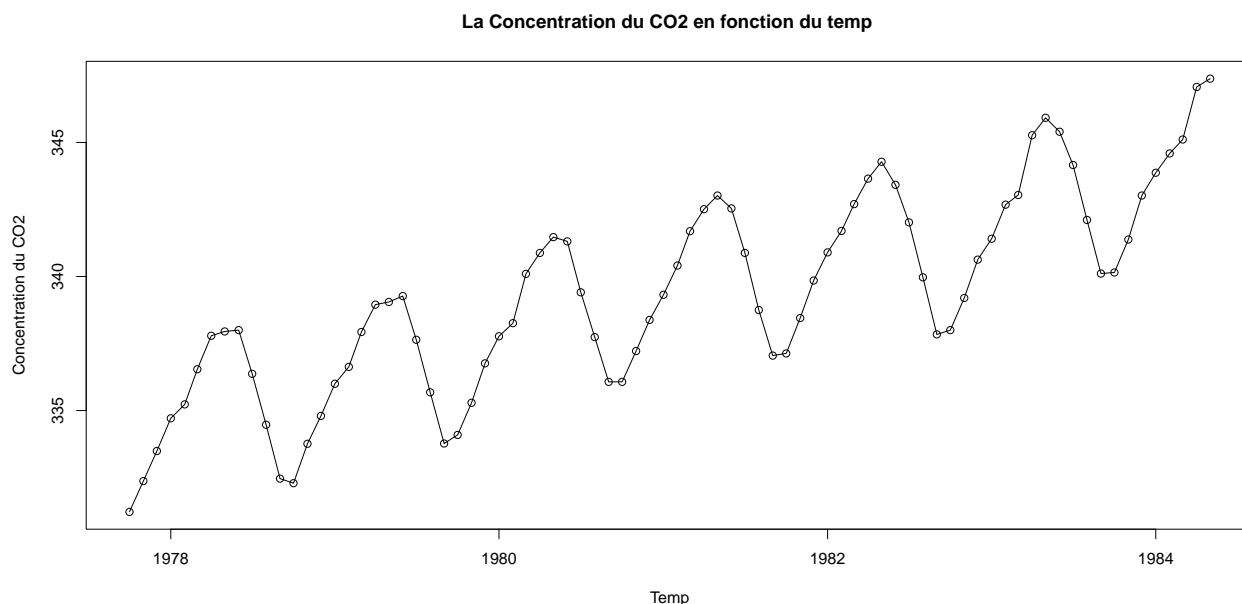


FIGURE 1.2 – Concentration du CO₂ dans l'air de Octobre 1977 à Mai 1984

Exemple 3. Courant à travers une résistance

Si une tension sinusoidale $v(t) = a \cos(vt + \theta)$ est appliquée à une résistance de valeur r et que le courant est enregistré en continu, nous obtenons une série temporelle continue

$$X(t) = r^{-1}a * \cos(vt + \theta)$$

avec :

a est l'amplitude de la tension sinusoidale, v c'est la fréquence de la tension sinusoidale et θ représente la phase de la tension sinusoidale.

pour $r = 1$, $v = 2$, $a = 1$ et $\theta = \pi/3$, on obtient le graph suivant :

```
# Générer les valeurs de t
t <- seq(0, 10, length.out = 100)

# Générer les valeurs de x(t)
x <- cos(2 * t + pi/3)

# Tracer le graphique
plot(t, x, type = "o", lwd = 2, xlab = "t", ylab = "x(t)")
```

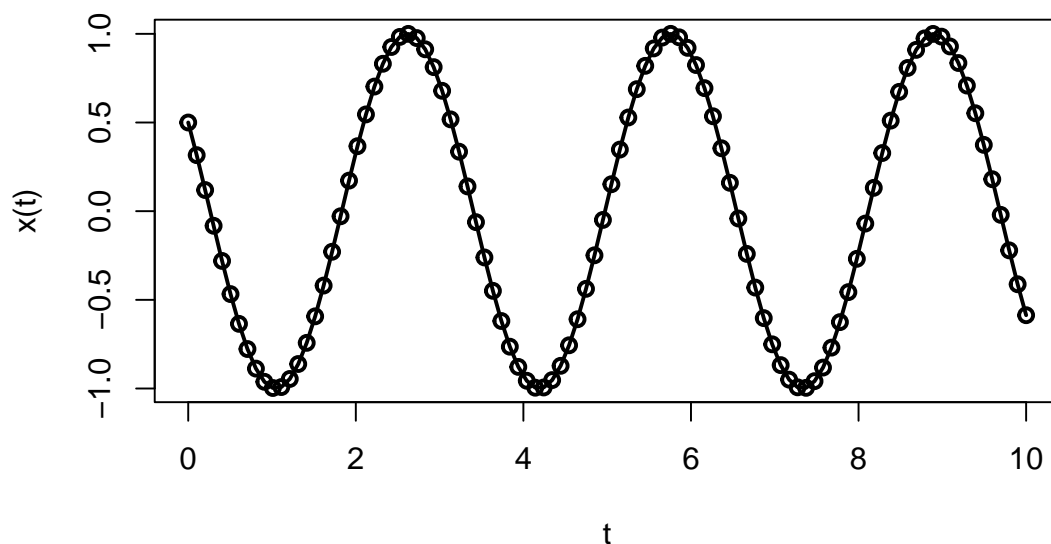


FIGURE 1.3 – 100 observations of the series $X(t) = \cos(2t + \pi/3)$.

1.1.2 Objectif de l'analyse des séries chronologiques

L'analyse des séries chronologiques revêt une importance cruciale dans divers domaines en raison de sa capacité à examiner, à détailler et à expliquer l'évolution des phénomènes à travers le temps.

Cette approche permet non seulement de comprendre les tendances et les motifs qui se manifestent dans les données historiques, mais aussi de prendre des décisions éclairées sur leur base. Par exemple, dans la gestion des stocks, l'analyse des séries chronologiques aide à prévoir la demande future en fonction des données passées, optimisant ainsi les niveaux de stock pour éviter les surplus ou les ruptures. De même, dans le domaine industriel comme le contrôle des processus chimiques, elle est utilisée pour détecter les variations, prévenir les défaillances et améliorer la qualité du produit final.

Cependant, l'objectif principal de cette analyse reste la prévision. Cela implique la capacité à anticiper les valeurs futures de la série chronologique, représentées par X_{T+h} pour $h = 1, 2, 3, \dots$, en se basant sur les observations antérieures X_1, X_2, \dots, X_T .

Cette prévision est essentielle pour planifier efficacement les ressources, ajuster les stratégies opérationnelles et répondre aux défis futurs avec une meilleure préparation.

1.1.3 Les domaines d'applications d'une série chronologique

Les séries chronologiques ont une large gamme d'applications dans divers domaines. Voici quelques-uns des domaines d'application principaux :

1. **Économie et Finance** : L'analyse des séries chronologiques est fréquemment utilisée pour modéliser les tendances économiques, prédire les rendements financiers, et étudier les cycles économiques.
2. **Météorologie** : Les données relatives aux températures maximales et minimales quotidiennes, aux précipitations annuelles, aux indices de sécheresse collectés à intervalles réguliers, et aux vitesses du vent enregistrées à des moments spécifiques de chaque heure donnent lieu à des ensembles de données temporelles. Ces ensembles génèrent des séries chronologiques, offrant ainsi la possibilité d'analyser et d'identifier des tendances, des modèles saisonniers, des événements extrêmes, et d'autres caractéristiques climatiques au fil du temps.
3. **Biologie et Médecine** : Dans le domaine de la biologie, les séries chronologiques peuvent être utilisées pour suivre l'évolution des populations d'espèces, analyser des séquences génétiques, etc. En médecine, elles peuvent être utilisées pour surveiller les paramètres vitaux des patients au fil du temps.
4. **Marketing et Ventes** : Les entreprises utilisent les séries chronologiques pour suivre les ventes au fil du temps, prévoir la demande de produits, et ajuster leurs stratégies de marketing.
5. **Recherche Opérationnelle et Gestion de la Chaîne d'Approvisionnement** : Les séries chronologiques sont utilisées pour la planification des stocks, la gestion de la chaîne d'approvisionnement, et la prévision de la demande.
6. **Production d'Énergie** : Dans les domaines de la production d'énergie, les séries chronologiques sont utilisées pour surveiller la consommation d'énergie, gérer les réseaux électriques, et planifier la production d'électricité.
7. **Éducation et Sciences Sociales** : Les séries chronologiques sont utilisées dans l'éducation pour suivre les performances des élèves au fil du temps. En sciences sociales, elles peuvent être utilisées pour analyser des tendances démographiques, économiques ou sociales.

Ces applications mettent en lumière la polyvalence des séries chronologiques.

1.2 Composantes d'une série chronologique

Les composantes d'une série chronologique se réfèrent aux différentes parties qui contribuent à la variation observée dans les données au fil du temps. Une décomposition classique d'une série chronologique suppose généralement trois composantes principales :

1.2. COMPOSANTES D'UNE SÉRIE CHRONOLOGIQUE

- **Tendance** : La tendance ou trend notée (m_t) capte l'orientation à long terme de la série et indique si la série a une croissance, une décroissance ou reste relativement stable.
- **Composante saisonnière** : La composante saisonnière noté (s_t) capte un comportement qui se répète dans des période fixe.
Sa période est égale à 12 pour les séries mensuelles, et à 4 pour les séries trimestrielles, 7 pour les séries Hebdomadaire...
Si p désigne la période du mouvement saisonnier alors :

$$s_t = s_{t+p} = s_{t+2p} = \dots$$

- **La composante résiduelle** : (ou bruit ou résidu) noté (e_t) regroupe tout ce qui n'a pas été pris en compte par la tendance et la saisonnalité, il correspond à des fluctuations irrégulières dues à des facteurs perturbateurs non permanents , en général de faible intensité mais de nature aléatoire et de moyenne nulle.

Exemple 4. Dans la Figure 1.4 La tendance peut être décomposée en trois phases comme indiqué par les flèches :

- Le cours baisse du début des observations l'observation n°7 jusqu'à l'observation n°15.
- De l'observation n°16 à l'observation n°41 le cours diminue légèrement.
- Il augmente rapidement à partir de l'observation n°42.

la série n'est soumis à aucun effet saisonnier visible. La composante accidentelle (résiduelle) est visualisée par les petites variations de cours d'un jour à l'autre.

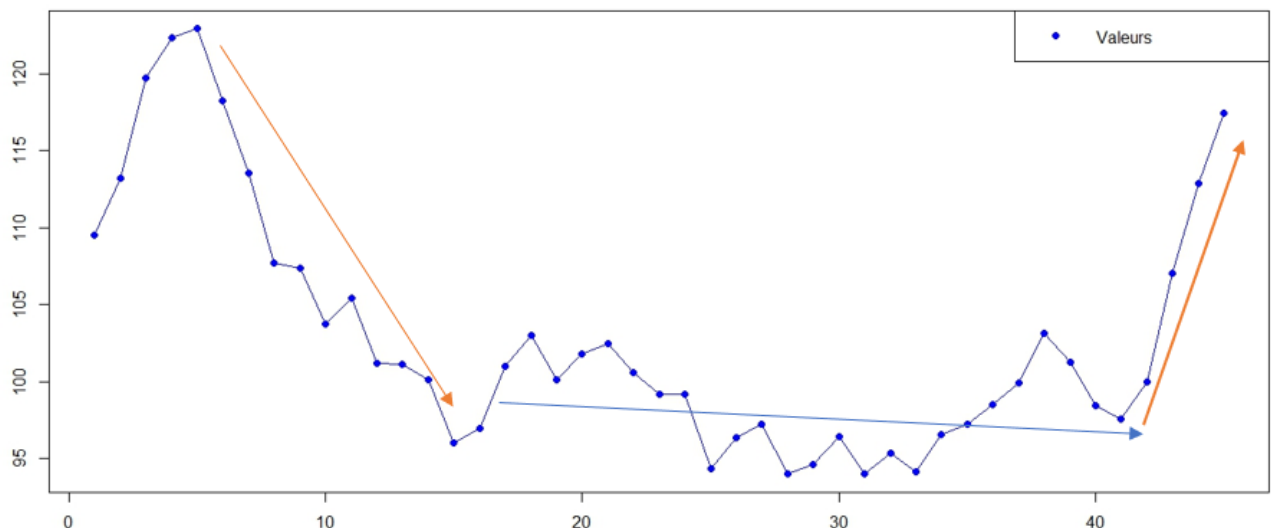


FIGURE 1.4 – cours journalier du titre Alcatel (du 4 janvier 1999 au 5 mars 1999(45 jours))

1.3 Modélisation d'une série chronologique

La décomposition d'une série chronologique peut s'effectuer selon trois types de modèles pour tout $t = 1, \dots, n$

1.3.1 Modèle additif

On adapte ce modèle lorsque la série chronologique $\{X_t\}$ peut être écrite comme somme des trois composantes

$$X_t = m_t + s_t + e_t \quad t \in \{1, 2, \dots, n\} \quad \text{où } E(e_t) = 0 \quad (1.1)$$

Où (m_t) , (s_t) et (e_t) sont définis précédemment.

REMARQUE. Pour savoir si le modèle additif est adapté, on trace les lignes polygonales passant par *les pics positifs* d'une part et *négatifs* d'autre part.

Si celles-ci sont proches de droites et si la largeur de la bande délimitée par celles-ci est essentiellement constante, on choisit d'utiliser le modèle additif, ce qui est bien le cas dans la Figure 1.2.

Exemple 5. On utilise dans cet exemple les données de [?] (Exemple 2), le programme A.1 donne :

Le graphe 1.5 montre bien un modèle additif.

Graphe des données avec lignes spécifiques

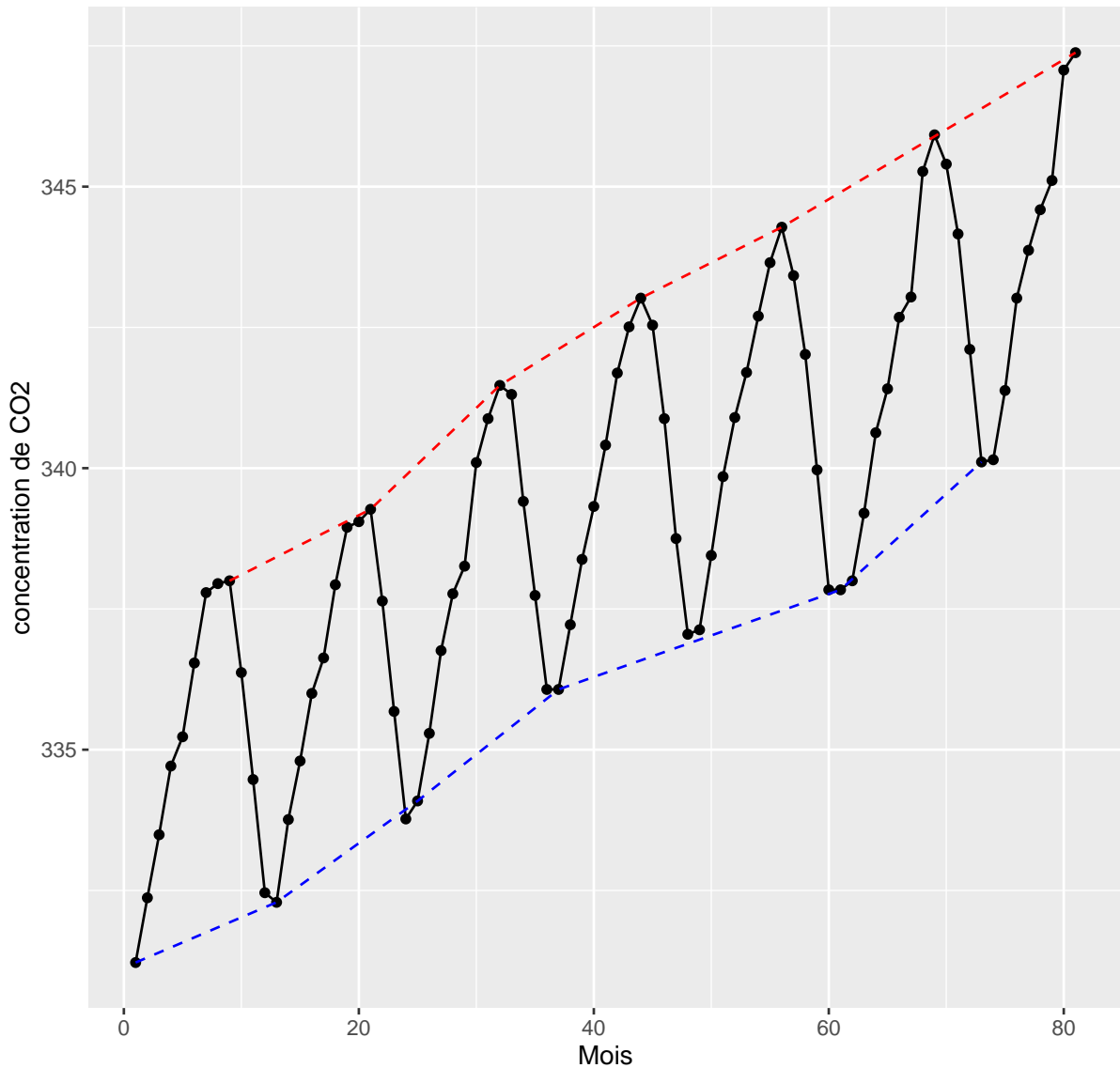


FIGURE 1.5 – Concentration du CO2 dans l'air de Octobre 1977 à Mai 1984

1.3.2 Modèle multiplicatif

On parle de modèle multiplicatif lorsque la série chronologique $\{X_t\}$ se décompose sous la forme :

$$X_t = m_t * s_t * e_t \quad t \in \{1, 2, \dots, n\} \quad \text{où } E(e_t) = 1 \quad (1.2)$$

REMARQUE. Pour savoir si le modèle multiplicatif est adapté, on trace les lignes polygonales passant par les pics positifs d'une part et négatifs d'autre part. Si celles-ci sont proches de droites et si la largeur de la bande délimitée par celles-ci est clairement croissante ou décroissante, on choisit d'utiliser le modèle multiplicatif, ce qui est bien le cas

dans la Figure 1.6.

Exemple 6. *On utilise dans cette exemple les même données que l'Exemple 1, le programme A.2 donne :*

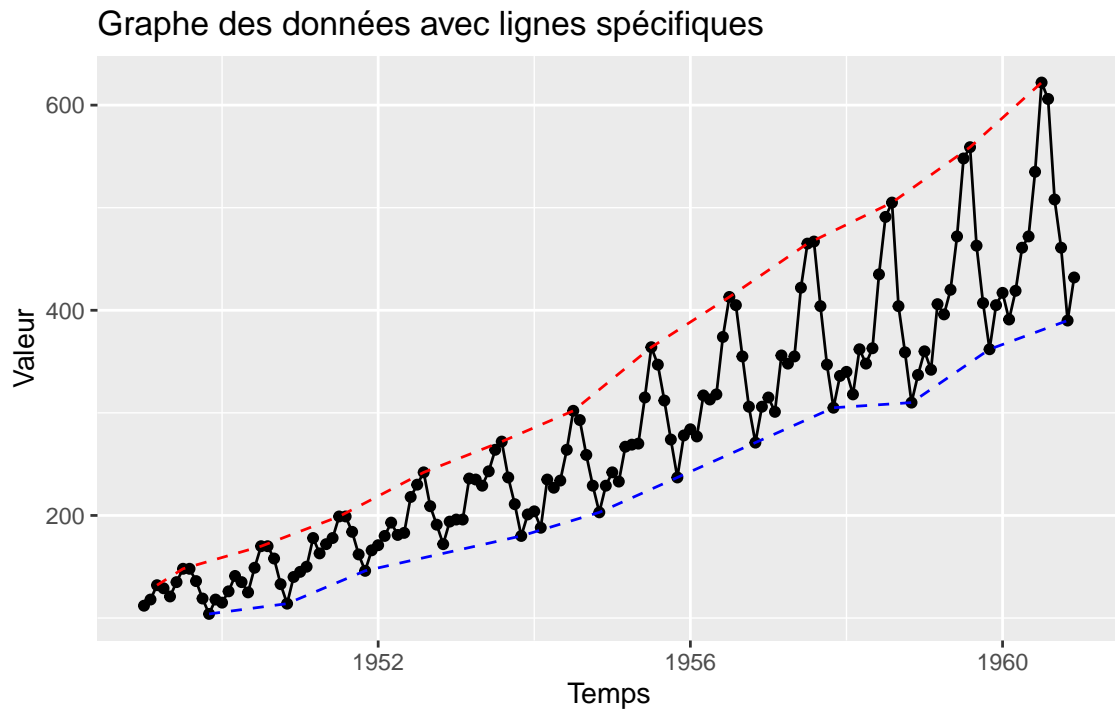


FIGURE 1.6 – Le nombre mensuel de passagers aériens internationaux, en milliers. de janvier 1949 à décembre 1960.

Le graphe 1.6 montre bien un modèle Multiplicatif.

REMARQUE. Une transformation logarithmique du modèle multiplicatif le ramène au modèle additif.

$$\log(X_t) = \log(m_t) + \log(s_t) + \log(e_t)$$

1.3.3 Modèle mixte

Il s'agit de modèles où l'addition et la multiplication sont utilisées :

$$X_t = (m_t + s_t) * e_t \quad t \in \{1, 2, \dots, n\} \quad \text{où } E(e_t) = 1 \quad (1.3)$$

1.4 Les Opérateurs dans les séries chronologiques

1.4.1 Opérateur retard B (Backward)

L'opérateur retard (Lag operator) noté B (backwards) est l'opérateur qui fait passer de X_t à X_{t-1} , tel que :

$$BX_t = X_{t-1}$$

On a donc :

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}$$

et donc $\forall n \in \mathbb{N}$

$$B^n X_t = X_{t-n}$$

Cet opérateur est linéaire, il est inversible, et son inverse est défini par $FX_t = X_{t+1}$. F est appelé opérateur de décalage. L'opérateur retard simplifie grandement l'écriture des équations relatives aux séries. Il permet d'écrire une équation de récurrence comme un polynôme de l'opérateur retard appliqué à une série. On présente ci-dessus quelques propriétés de l'opérateur :

Propriétés

L'opérateur retard a les mêmes propriétés qu'un coefficient de multiplication.

- soit α une constante, l'opérateur retard et la constante sont cumulatifs.

$$B(\alpha X_t) = \alpha BX_t$$

- soit w_t une autre série chronologique, L'opérateur retard est distributif.

$$B(X_t + w_t) = BX_t + Bw_t$$

- L'opérateur retard multiplié par une constante est égale à la constante elle-même.

$$\alpha B = \alpha$$

Exemple 7. Soit λ_1 et λ_2 des constantes, on a :

$$\begin{aligned}(1 - \lambda_1 B)(1 - \lambda_2 B)X_t &= (1 - \lambda_1 B - \lambda_2 B + \lambda_1 \lambda_2 B^2)X_t \\ &= (1 - [\lambda_1 + \lambda_2]B + \lambda_1 \lambda_2 B^2)X_t \\ &= X_t - (\lambda_1 + \lambda_2)x_{t-1} + \lambda_1 \lambda_2 X_{t-2}\end{aligned}$$

1.4.2 Opérateur de différence

L'Opérateur de différence $\nabla = 1 - B$ donne ce qu'on appelle la différence première par :

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1}$$

qui est la série des accroissements, alors que la différence seconde ∇^2 donne la série des « accroissements des accroissements », on a :

$$\nabla^2 X_t = \nabla(\nabla X_t) = (1 - B)^2 X_t = (1 - 2B + B^2)x_t = X_t - 2X_{t-1} + X_{t-2}.$$

et donc $\forall n \in \mathbb{N}$

$$\nabla^n X_t = (1 - B)^n X_t$$

1.5 Stationnarité

1.5.1 Stationnarité stricte (forte)

On dit que la série temporelle $\{X_t\}$ est strictement stationnaire ou fortement stationnaire si et seulement si :

$$(X_1, X_2, \dots, X_t) \text{ et } (X_{1+k}, X_{2+k}, \dots, X_{t+k}) \text{ suivent la même loi.}$$

Ce qui est équivalent à :

$$\forall k, F_{X_1, X_2, \dots, X_t}(x_1, x_2, \dots, x_t) = F_{X_{1+k}, X_{2+k}, \dots, X_{t+k}}(x_1, x_2, \dots, x_t)$$

c'est-à-dire que, quels que soient le nombre de dates et les dates choisis, quand on décale ces dates d'une même quantité (k), la distribution ne change pas.

Cette condition est généralement difficile à vérifier c'est pour cela qu'on utilise une version plus faible de stationnarité, la stationnarité faible ou du second ordre qui est souvent suffisante.

1.5.2 Stationnarité faible

On dit que la série temporelle $\{X_t\}$ est faiblement stationnaire ou Stationnaire de 2nd ordre ou tout simplement stationnaire si :

1. $E(X_t) = \mu \quad \forall t.$
2. $V(X_t) = \sigma^2$ (cste).
3. $COV(X_t, X_{t+k})$ est indépendant du temps et ne dépend que de k.

REMARQUE.

- La stationnarité forte implique la stationnarité faible.
- La stationnarité faible implique la stationnarité forte si la série chronologique est Gaussienne.

1.6 Processus stochastique

Les premières approches dans l'analyse des séries chronologiques est la sélection d'un modèle mathématique adéquat (ou une classe de modèles) pour les données.

Pour permettre une prédiction possible pour des futures observations, il est naturel de supposer que toute observation x_t est une réalisation d'une certaine variable aléatoire X_t .

La série chronologique $\{X_t, t \in \mathbb{T}_0\}$ est donc une réalisation de la famille des variables aléatoires $(X_t)_{t \in \mathbb{T}_0}$.

Cette démarche suppose la modélisation des données comme une réalisation (ou une partie d'une réalisation) d'un processus stochastique $\{X_t, t \in \mathbb{T}\}$ où $\mathbb{T}_0 \subseteq \mathbb{T}$.

Pour clarifier cette idée on a besoin de définir le processus stochastique et ses réalisations.

1.6.1 Processus stochastique

Définition 2. Soit $(\Omega, \mathfrak{F}, P)$ un espace de probabilités, un processus stochastique est une famille $\{X_t, t \in \mathbb{T}\}$ de fonctions $X_t := X(t, \omega)$ définie dans $\mathbb{T} \times \Omega$, avec \mathbb{T} est l'un des ensembles suivants : \mathbb{N} , \mathbb{Z} , $[0, \infty)$ ou \mathbb{R} ;

Tel que pour un t fixé, $X(t, \omega)$ est une variable aléatoire sur $(\Omega, \mathfrak{F}, P)$.

Pour un ω fixé, $X(t, \omega)$ est une fonction de t dite réalisation.

Sous la condition que les observations x_t ont été générées par un processus stochastique $\{X_t, t \in \mathbb{T}\}$ (famille de variables aléatoires), la série temporelle $\{x_t\}$ est une réalisation de ce processus stochastique.

Un processus stochastique sert donc comme un modèle d'une série chronologique observée.

1.6.2 Réalisation d'un processus stochastique

Les fonctions $\{X.(w), \omega \in \Omega\}$ sur \mathbb{T} sont connues sous le nom de réalisations d'un processus $\{X_t, t \in \mathbb{T}\}$.

Exemple 8. (*Sinusoïde avec une phase et une amplitude aléatoires*)

Soient A et Θ des variables aléatoires indépendantes telles que $A \geq 0$ et Θ distribué uniformément sur $[0, 2\pi)$. Un processus stochastique $\{X(t), t \in \mathbb{R}\}$ peut alors être défini en termes de A et Θ pour toute valeur de $v \geq 0$ et $r > 0$ par

$$X_t = r^{-1}A * \cos(vt + \Theta) \quad (1.4)$$

ou plus explicitement,

$$X_t(\omega) = r^{-1}A(\omega)\cos(vt + \Theta(\omega)) \quad (1.5)$$

où ω est un élément de l'espace probabiliste Ω sur lequel A et θ sont définis.

Les réalisations du processus défini par (1.5) sont les fonctions de t obtenues en fixant w , c'est-à-dire des fonctions de la forme

$$X(t) = r^{-1}a * \cos(vt + \theta) \quad (1.6)$$

La série temporelle représentée dans la Figure (1.3) est l'une de ces réalisations.

1.6.3 La fonction d'autocovariance

Définition 3. Si $\{X_t, t \in \mathbb{T}\}$ est un processus à temps continue tel que $Var(X_t) < \infty$ pour tout $t \in \mathbb{T}$, alors la fonction d'autocovariance $\gamma_x(\cdot, \cdot)$ de $\{X_t\}$ est définie par :

$$\gamma_x(t, s) = Cov(X_t, X_s) = E[(X_t - E(X_t))(X_s - E(X_s))] = E(X_t X_s) - E(X_t)E(X_s) \quad t, s \in \mathbb{T} \quad (1.7)$$

Propriétés

On présente ici quelques propriétés de base

1. on peut écrire : $\gamma_x(s, t) = \gamma_{|t-s|}$
2. $\gamma_x(s, t) = \gamma_x(t, s) \quad \forall t, s$
3. Si $\gamma_x(s, t) = 0$, on dit que X_s et X_t n'ont aucune liaison linière.(non corrélées)
4. $\gamma_{|t-s|} = \gamma_{-|t-s|} \quad \forall t, s$
5. pour $s=t$, L'autocovariance se réduit à une variance, car :

$$\gamma_x(t, t) = E[(X_t - E(X_t))^2] = var(X_t)$$

REMARQUES :

REMARQUE. 1. Pour un processus à temps discret $\{X_t, t \in \mathbb{T}\}$, $\mathbb{T} = 0, \pm 1, \pm 2, \pm 3, \dots$, on pose $s = t+k$ pour $k \in \mathbb{T}$, et on écrit :

$$\gamma_k = \gamma_x(k) = \gamma_x(k, 0) = \text{Cov}(X_t, X_{t+k}) = E(X_t X_{t+k}) - E(X_t)E(X_{t+k}) \quad (1.8)$$

REMARQUE. 2. La différence entre la covariance et l'autocovariance est que la covariance mesure la relation entre deux variables distinctes X et Y , tandis que l'autocovariance évalue la relation d'une variable avec elle-même à différents décalages temporels X_t et X_{t+k} .

1.6.4 La fonction d'autocorrelation (acf)

Définition 4. La fonction d'autocorrelation $\rho(h)$ mesure la corrélation entre X_t et X_{t+h} et elle est définie par :

$$\rho(h) = \frac{\text{cov}(X_t, X_{t+h})}{\sqrt{(\text{var}(X_t)\text{var}(X_{t+h}))}} = \frac{\gamma(h)}{\gamma(0)} = \text{Corr}(X_{t+h}, X_t) \quad \forall t, h \in \mathbb{T} \quad (1.9)$$

La fonction d'Autocorrelation satisfait toutes les propriétés de la fonction d'auto covariance. De plus, elle satisfait à la condition $\rho(0) = 1$.

1.7 Bruit Blanc

Définition 5. Un bruit blanc $\{\varepsilon_t\}$ est une suite de v.a non corrélées (mais pas nécessairement indépendantes) de moyenne nulle et de variance constante σ_ε^2 . C'est donc une série faiblement stationnaire. On note

$$\varepsilon_t \sim BB(0, \sigma_\varepsilon^2) \quad (1.10)$$

Bruit blanc gaussien

Un bruit blanc gaussien $\{\varepsilon_t\}$ est une suite de v.a i.i.d. qui suit une loi normale de paramètres $(0, \sigma_\varepsilon^2)$, on note :

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (1.11)$$

REMARQUE. Un bruit blanc gaussien est une série strictement stationnaire. En effet

$$\begin{aligned} & P(\varepsilon_{t_1} \leq x_1, \varepsilon_{t_2} \leq x_2, \dots, \varepsilon_{t_n} \leq x_n) \\ &= P(\varepsilon_{t_1} \leq x_1)P(\varepsilon_{t_2} \leq x_2) \dots P(\varepsilon_{t_n} \leq x_n) && \text{(par indépendance)} \\ &= P(\varepsilon_{t_1-k} \leq x_1)P(\varepsilon_{t_2-k} \leq x_2) \dots P(\varepsilon_{t_n-k} \leq x_n) && \text{(identiquement distribuer)} \\ &= P(\varepsilon_{t_1-k} \leq x_1, \varepsilon_{t_2-k} \leq x_2, \dots, \varepsilon_{t_n-k} \leq x_n) && \text{(par indépendance)} \end{aligned}$$

Théorème de Wold (1938)

Si X_t est un processus stationnaire du second ordre alors, on peut écrire :

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i u_{t-i} \quad (1.12)$$

où

1. les coefficients ψ_i sont des constantes avec $\psi_0 = 1$
2. u est un processus bruit blanc vérifiant

$$\forall t, E(u_t) = 0, \quad (1.13)$$

$$E(u_t u_s) = \begin{cases} \text{une constante } \sigma_u^2, & \text{si } t = s \\ 0, & \text{si } t \neq s. \end{cases} \quad (1.14)$$

3. μ une constante.

1.8 Modélisation ARIMA (AutoRegressive Integrated Moving Average)

1.8.1 Processus autorégressif AR(p)

Le processus autorégressif permet de modéliser les relations linéaires entre les observations successives, où la valeur actuelle dépend des p valeurs précédentes.

Définition 6. $\{X_t, t \in \mathbb{T}\}$ est un processus autorégressif d'ordre P (AR (P)) si :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (1.15)$$

$$\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$$

où X_t est la valeur de la série temporelle à l'instant t et c est une constante qui est nulle en général, $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ sont les valeurs passées de la série temporelle, $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients autorégressifs, tel que :

$$\forall j < p, \quad \phi_j \in \mathbb{R}$$

$$\phi_p \neq 0$$

REMARQUE. On peut aussi écrire :

$$\phi(B)X_t = \varepsilon_t \sim AR(p)$$

Où ϕ est un polynôme d'ordre p .

Autre écritures :

$$X_t = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p + \varepsilon_t \quad (1.16)$$

Par exemple :

$$AR(2) : X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \quad (1.17)$$

1.8.2 Processus à moyenne mobile MA(q)

Le processus moyenne mobile est essentielle pour saisir les erreurs de prédiction et les fluctuations temporaires dans une série chronologique.

Définition 7. $\{X_t, t \in \mathbb{T}\}$ est un processus moyenne mobile d'ordre q (MA(q)) si :

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} \quad (1.18)$$

$$\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$$

où X_t est la valeur de la série temporelle à l'instant t , μ sa moyenne et $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients de la moyenne mobile, avec $\theta_q \neq 0$.

REMARQUE. On peut aussi écrire :

$$X_t = \Theta(B)\varepsilon_t \sim MA(q)$$

Où Θ est un polynôme d'ordre q.

Propriétés

Un MA(q) est toujours stationnaire quelles que soient les valeurs des θ .

Par exemple :

$$MA(2) : X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} \quad (1.19)$$

1.8.3 Causalité et Invérsibilité

Causalité

Un processus est dit causal s'il existe une suite a_k réelle telle que $\sum_{k=0}^{\infty} a_k < \infty$ et que

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \quad \text{où} \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

Théorème 1. On dit que $\{X_t, t \in \mathbb{T}\}$ est causal si et seulement si les racines de ϕ sont à l'extérieure du cercle unité .

$$|z| > 1$$

Exemple 9. Soit le processus AR(2) suivant :

$$X_t = 0.9X_{t-1} - 0.7X_{t-2} + \varepsilon_t$$

montrons que X_t est causal.

$$X_t = 0.9X_{t-1} - 0.7X_{t-2} + \varepsilon_t$$

$$X_t = 0.9BX_t - 0.7B^2X_t + \varepsilon_t$$

$$X_t - 0.9BX_t + 0.7B^2X_t = \varepsilon_t$$

$$X_t(1 - 0.9B + 0.7B^2) = \varepsilon_t$$

1.8. MODÉLISATION ARIMA (AUTOREGRESSIVE INTEGRATED MOVING
AVERAGE)

on remplace B par z :

$$\begin{aligned} X_t(1 - 0.9z + 0.7z^2) &= \varepsilon_t \\ \Phi(z) &= 1 - 0.9z + 0.7z^2 \\ \implies \Delta &= (0.9)^2 - 4(0.7) = -1,99 \\ z_i &= \frac{1.99 \pm i\sqrt{1.99}}{2(0.7)} \\ |z_1| = |z_2| &= \sqrt{\left(\frac{0.9}{1.4}\right)^2 + \left(\frac{\sqrt{1.99}}{1.4}\right)^2} \\ &= 1.19 > 1 \end{aligned}$$

et puisque $|z| > 1$, donc le processus est bien causal.

Inversibilité

Un processus est dit inversible s'il existe une suite b_k réelle telle que

$$\sum_{k=0}^{\infty} |b_k| < \infty \quad \text{et}$$

$$\varepsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k} \quad \varepsilon_t \text{ est un bruit blanc.}$$

Théorème 2. On dit que $\{X_t, t \in \mathbb{T}\}$ est inversible si et seulement si les racines de Θ sont à l'extérieur du cercle unité .

$$|z| > 1$$

REMARQUE.

- Un AR(p) est toujours inversible.
- Un MA(q) est toujours causal.

Exemple 10. Soit le processus MA(2) suivant :

$$X_t = \varepsilon_t + 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$$

On sait que les MA sont toujours causales, montrons que celui-ci est inversible.

$$\begin{aligned} X_t &= \varepsilon_t + 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} \\ X_t &= \varepsilon_t + 0.6B\varepsilon_t + 0.3B^2\varepsilon_t \\ X_t &= \varepsilon_t(1 + 0.6B + 0.3B^2) \end{aligned}$$

on remplace B par z

$$\begin{aligned} X_t &= \varepsilon_t(1 + 0.6z + 0.3z^2) \\ \Theta(z) &= 1 + 0.6z + 0.3z^2 \end{aligned}$$

les solutions du polynôme caractéristique $\Theta(z)$ sont donné par

$$z_1 = 3.08 \quad z_2 = -1.08$$

On voit que $|z_i| > 1$, $i=\{1, 2\}$, donc le processus est bien inversible.

1.8.4 Processus ARMA(p,q)

Pour être considéré comme un modèle ARMA(p, q), le processus doit être stationnaire. La stationnarité implique que les propriétés statistiques du processus restent constantes au fil du temps.

L'utilisation du modèle ARMA(p, q) offre une flexibilité dans la modélisation de diverses séries chronologiques, permettant de capturer à la fois les relations linéaires avec les observations passées et les fluctuations temporaires.

Définition 8. $\{X_t, t \in \mathbb{T}\}$ obéit à un modèle ARMA(p, q) s'il est stationnaire et vérifie :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (1.20)$$

Ou bien :

$$\phi(B)X_t = \Theta(B)\varepsilon_t$$

Avec

$$\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$$

REMARQUE. La condition de causalité et d'inversibilité d'un ARMA(p,q) est la réunion des condition des models AR et MA.

À savoir Θ et ϕ ont leur racines à l'extérieure du cercle unité

$$|z| > 1$$

Exemple 11. Soit le processus ARMA(1,1) suivant :

$$X_t = 0.8X_{t-1} + \varepsilon_t + 0.7\varepsilon_{t-1}$$

Vérifiant de la causalité et l'invisibilité.

$$\begin{aligned} X_t &= 0.8X_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-1} \\ X_t - 0.8X_{t-1} &= \varepsilon_t - 0.7\varepsilon_{t-1} \\ X_t(1 - 0.8B) &= \varepsilon_t(1 - 0.7B) \\ \Phi(z) &= 1 - 0.8z \quad , \quad \Theta(z) = 1 - 0.7z \\ z &= \frac{1}{0.8} > 1 \quad , \quad z = \frac{1}{0.7} > 1 \end{aligned}$$

Le processus est donc causal et inversible.

1.8.5 Le processus intégré I(d)

La composante intégrée est une partie essentielle de la modélisation ARIMA et est souvent utilisée pour stationnariser des séries chronologiques non stationnaires rendant ainsi possible l'application des composantes autoregressive et à moyenne mobile.

Définition 9. $\{X_t, t \in \mathbb{T}\}$ est un processus intégrée d'ordre d (I(d)), si :

$$\nabla^d X_t = \varepsilon_t \tag{1.21}$$

Où $\nabla^d = (1 - B)^d$ avec B est l'opérateur retard.

1.8.6 Le processus ARIMA(p,d,q)

Nous avons déjà discuté de l'importance de la classe des modèles ARMA pour représenter des séries stationnaires. Une généralisation de cette classe, qui incorpore une large gamme de séries non stationnaires, est fournie par les processus ARIMA, c'est-à-dire des processus qui, après différenciation un nombre fini de fois, se réduisent à des processus ARMA.

Définition 10. Le processus $\{X_t, t \in \mathbb{T}\}$ est un processus ARIMA(p, d, q) si le processus $Y_t = \nabla^d X_t$ est un processus ARMA(p, q) causal.

Le modèle ARIMA revient à appliquer un modèle ARMA sur le processus différencié :

$$Y_t = \text{ARMA}(p,q) \iff X_t = \text{ARIMA}(p,d,q)$$

L'équation d'un modèle ARIMA(p,d,q) est donc donnée par :

$$\phi(B)Y_t = \theta(B)\varepsilon_t \iff \phi(B)\nabla^d X_t = \theta(B)\varepsilon_t$$

Où ϕ et θ sont deux polynômes de degrés respectifs p et q .

Le "I" de ARIMA signifie "integrated", comme réciproque de la différenciation.

Évidemment, le degré d n'est généralement pas connu. Pour le déterminer, on peut agir par tâtonnements ou avoir recours à des tests de stationnarité : puisqu'un processus ARMA(p,q) est stationnaire, on cherche le d tel qu'on puisse accepter l'hypothèse de stationnarité pour le processus $Y_t = \nabla^d X_t$. De façon générale, on cherche à un principe de parcimonie et à une valeur satisfaisant la meilleure efficacité en d .

REMARQUE.

1. Un ARMA(p, q) est un ARIMA(p, 0, q).
2. Le processus ARMA(p, q) est stationnaire si et seulement si $d = 0$

Il existe bien d'autres types de séries chronologiques, comme les ARCH, les GARCH...

Chapitre 2

STATISTIQUE BAYÉSIENNE ET BAYÉSIENNE EMPIRIQUE

La statistique Bayésienne est une approche puissante qui repose sur le théorème de Bayes pour modéliser l'incertitude et estimer des résultats à partir des données observées. Contrairement à l'approche fréquentiste, elle permet d'incorporer des informations a priori, puis de les ajuster à la lumière des nouvelles observations, offrant ainsi un cadre flexible et évolutif pour la prise de décision.

Dans ce chapitre, nous examinerons en détail les concepts clés de la statistique Bayésienne, ainsi que l'approche Bayésienne empirique, qui combine les principes Bayésiens avec l'analyse fréquentiste pour améliorer l'estimation des paramètres lorsque l'information a priori est incertaine ou difficile à définir.

2.1 Inférence Bayésienne

L'inférence Bayésienne est basée sur ce théorème

2.1.1 Théorème de Bayes

Soit $(\mathcal{X}, \mathcal{A}, \mathcal{P}_\theta)$ un espace probabilisé et A et B deux événements aléatoires tels que $P[A] \neq 0$. La probabilité de B conditionnellement à la réalisation de A est donnée par la relation suivante :

$$P[B|A] = \frac{P[B, A]}{P[A]} \quad (2.1)$$

$P[B, A]$ est la probabilité que les deux événements A et B aient lieu simultanément. Puisque $P[B, A] = P[A, B]$, la théorie de Bayes (2.1) peut s'écrire aussi :

$$P[A|B] = \frac{P[A, B]}{P[B]}$$

Donc

$$P[B, A] = P[B|A] \times P[A]$$

On remplace $P[A, B]$ par $P[B, A]$, en déduisant la relation entre les deux probabilités conditionnelles $P[A|B]$ et $P[B|A]$:

$$P[A|B] = \frac{P[A] \times P[B|A]}{P[B]}$$

Ce théorème (ou formule de Bayes) est l'un des plus célèbres de la statistique. Bayes (1763) donne en une version continue de ces résultats, à savoir, pour deux variables aléatoires X et Y , de distributions conditionnelles $f(X|Y)$ et marginale $g(Y)$, la distribution conditionnelle de Y sachant X est :

$$g(Y|X) = \frac{f(X|Y)g(Y)}{\int_{\Theta} f(X)g(Y)dY}$$

2.1.2 Modèle Statistique Bayésienne

On appelle modèle statistique Bayésien la donnée d'un modèle statistique paramétré $(\mathcal{X}, \mathcal{A}, \mathcal{P}_{\theta})$ avec $\theta \in \Theta$, plus la loi a priori.

Définissant les espaces intervenant dans un modèle Bayésien :

- **Espace des observations**

Noté \mathcal{X} , il représente l'ensemble des résultats observé (observations) suite à une étude d'un phénomène $(X_1, X_2, X_3, \dots X_n)$.

- **Espace des actions**

Noté \mathcal{A} , il représente l'ensemble des actions ou décisions à prendre après l'obtention de l'information.

- \mathcal{P}_{θ} présente la loi (fonction de répartition) des X_i en fonction de θ et $f(x|\theta)$ sa densité.

- **Espace des états de la nature**

Noté Θ , il représente l'espace des paramètres inconnus θ .

- **Espace des règles de décisions**

Noté \mathcal{D} , il représente l'ensemble des règles de décisions qu'on définit comme une application de \mathcal{X} dans \mathcal{A} notée par $\delta(x)$.

$$\begin{aligned} \delta : \mathcal{X} &\rightarrow \mathcal{A} \\ x &\rightarrow \delta(x) = a \end{aligned}$$

Avec $a \in \mathcal{A}$ est une action ou une décision

REMARQUE.

- La règle de décision est un estimateur.
- L'action est l'estimation
- L'estimation est la valeur de l'estimateur au point d'observation x .

$$\delta(x) = a$$

2.1.3 Description d'un modèle Bayésien

Définissons les lois de probabilités intervenant dans l'analyse Bayésienne :

- **La vraisemblance**

C'est la loi des observations ou encore la loi conditionnelle de x sachant θ , notée par $f(x|\theta)$, elle est donnée par :

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

- **Loi a priori**

La loi a priori notée par $\pi(\theta)$ désigne la distribution du paramètre inconnu θ elle porte l'information sur ce dernier. Autrement dit elle exprime nos croyances initiales sur les différentes hypothèses avant de tenir compte des données observées x .

En d'autres termes, c'est ce que nous croyons être vrai avant de voir les preuves.

- **Loi a posteriori**

Dans le contexte du raisonnement Bayésien, la loi a priori est combinée avec la vraisemblance des données pour calculer la loi a posteriori, qui représente nos croyances mises à jour après avoir pris en compte les données observées.

La loi a posteriori est la loi conditionnelle de θ sachant x , notée par :

$$\pi(\theta|x) = \frac{\pi(\theta)f(x|\theta)}{f(x)} \quad (1)$$

$$= \frac{\pi(\theta)f(x|\theta)}{\int_{\theta \in \Theta} \pi(\theta)f(x|\theta) d\theta} \quad (2.2)$$

$$= \frac{\pi(\theta) \prod_{i=1}^n f(x_i|\theta)}{\int_{\theta \in \Theta} \pi(\theta) \prod_{i=1}^n f(x_i|\theta) d\theta} \quad (2.3)$$

- $f(x) = \int_{\theta \in \Theta} \pi(\theta)f(x|\theta) d\theta$ est la densité marginale de x .

2.1.4 Les bases de la décision de l'approche Bayésienne

La fonction coût

Définition 11. La fonction coût (Loss function), ou bien la fonction de perte, dénotée par $L(\theta, \delta(x))$ évalue le coût d'une décision a quand le paramètre vaut θ

$$L : \Theta \times \mathcal{A} \rightarrow \mathcal{R}^+ \\ (\theta, a) \rightarrow L(\theta, a) = L(\theta, \delta(x))$$

REMARQUE.

- Un coût négatif correspond à un gain.
- Dans la statistiques Bayésienne on ne considère pas le coût négatif, par conséquent une décision est considérée bonne si elle conduit à un coût nul. $L(\theta, a) = 0$

Puisque θ est inconnu, on ne peut pas résoudre l'équation $L(\theta, a) = 0$, donc on est amené à utiliser la fonction de risque fréquentiste

La fonction de risque fréquentiste

Définition 12. La fonction de **risque fréquentiste** $\mathcal{R}(\theta, \delta)$ de δ associée à $L(\theta, \delta(x))$ appelé aussi l'espérance mathématique du coût d'une règle de décision ou le coût moyen est définie comme :

$$\mathcal{R}(\theta, \delta) = E_f[L(\theta, \delta)] = \int_{\mathcal{X}} L(\theta, \delta(x))f(x|\theta) dx$$

REMARQUE. Le risque fréquentiste nous permet de comparer entre les décisions.

Critères de comparaison

On dira que la stratégie δ_1 est préférable à la stratégie δ_2 et on note :

$$\delta_1 < \delta_2 \quad \text{si} \quad \mathcal{R}(\theta, \delta_1) \leq \mathcal{R}(\theta, \delta_2) \quad \forall \theta \in \Theta$$

et :

$$\exists \theta_0 \in \Theta \quad \text{tel que} \quad \mathcal{R}(\theta_0, \delta_1) < \mathcal{R}(\theta_0, \delta_2)$$

REMARQUE. Ce critère ne nous permet pas de comparer deux règles de décision, tel que :

$$\mathcal{R}(\theta_1, \delta_1) < \mathcal{R}(\theta_1, \delta_2) \quad \text{et} \quad \mathcal{R}(\theta_2, \delta_1) > \mathcal{R}(\theta_2, \delta_2)$$

Pour cela on utilisera le critère minimax et le Risque Bayésien.

Critères Minimax

On appelle **risque minimax** associé à la fonction coût L la valeur $\bar{\mathcal{R}}(\delta)$, tel que :

$$\bar{\mathcal{R}}(\delta) = \inf_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta)$$

La stratégie δ_m est dite **estimateur minimax** si :

$$\sup_{\theta} \mathcal{R}(\theta, \delta_m) = \inf_{\delta \in \mathcal{D}} \sup_{\theta} \mathcal{R}(\theta, \delta) = \bar{\mathcal{R}}$$

Avec \mathcal{D} est l'ensemble des stratégies ou règles de décision.

On dit qu'une stratégie δ_0 n'est *pas minimax* ssi :

$$\exists \delta^* \in \mathcal{D} \quad \text{tel que} : \quad \sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta^*) < \sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta_0)$$

Le Risque a posteriori

Ou le **coût a posteriori** noté $\rho(\pi, \delta)$ définit comme étant la moyenne du coût par rapport à la loi à posteriori.

$$\rho(\pi, \delta) = E_{\pi(\cdot|X)}[L(\theta, \delta)] = \int_{\Theta} L(\theta, \delta)\pi(\theta|X) d\theta$$

Le Risque Bayésien (Intégré)

Il s'agit de la moyenne du risque fréquentiste suivant la loi a priori, on note $r(\pi, \delta)$ et on a :

$$\begin{aligned} r(\pi, \delta) &= E_{\pi(\cdot)}[\mathcal{R}(\theta, \delta)] \\ &= \int_{\Theta} \mathcal{R}(\theta, \delta) \pi(\theta) d\theta \end{aligned}$$

Remarquons que le risque intégré satisfait :

$$r(\pi, \delta) = \int_{\Theta} \rho(\theta, \delta) f(x) d\theta$$

En effet :

$$\begin{aligned} r(\pi, \delta) &= \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta) dx \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta) \pi(\theta) dx d\theta \end{aligned}$$

On à :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)} \implies f(x|\theta)\pi(\theta) = \pi(\theta|x)f(x)$$

$$r(\pi, \delta) = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) \pi(\theta|x) f(x) dx d\theta$$

En effet, comme $L(\theta, \delta) \geq 0$, d'après le théorème de fubini, on obtient :

$$\begin{aligned} r(\pi, \delta) &= \int_{\mathcal{X}} \int_{\Theta} L(\theta, \delta(x)) \pi(\theta|x) d\theta f(x) dx \\ &= \int_{\mathcal{X}} \rho(\pi, \delta(x)) f(x) dx = E_x[\rho(\pi, \delta(x))] \end{aligned}$$

D'où, le risque de Bayes $r(\pi, \delta)$ est aussi la moyenne du risque a posteriori $\rho(\pi, \delta(x))$ suivant la loi marginale $f(x)$.

Estimateur de Bayes

En appelle, estimateur de Bayes associer à la fonction coût L et à une distribution a priori π , tous décision δ^B qui minimise le risque $r(\pi, \delta)$

$$\delta^B = \arg \min_{\delta \in \mathcal{D}} r(\pi, \delta)$$

Théorème 3. *Le risque de Bayes $r(\pi, \delta^B)$ est plus petit que le risque minimax.*

$$r(\pi, \delta^B) < \inf_{\delta} \sup_{\theta} \mathcal{R}(\theta, \delta)$$

Démonstration.

$$\begin{aligned} r(\pi, \delta^B) &< r(\pi, \delta) && \forall \delta \in \mathcal{D} \\ r(\pi, \delta^B) &\leq \int_{\Theta} \mathcal{R}(\theta, \delta) \pi(\theta) d\theta && \forall \delta \in \mathcal{D} \\ r(\pi, \delta^B) &\leq E_{\pi(\cdot)}[\mathcal{R}(\theta, \delta)] && \forall \delta \in \mathcal{D} \\ &< \sup_{\theta} \mathcal{R}(\theta, \delta) && \forall \delta \in \mathcal{D} \\ r(\pi, \delta^B) &< \inf_{\delta} \sup_{\theta} \mathcal{R}(\theta, \delta) \end{aligned}$$

□

Le risque maximin

Définition 13. On définit le risque maximin par

$$\underline{\mathcal{R}} = \sup_{\pi} r(\pi, \delta^B) = \sup_{\pi} \inf_{\delta} r(\pi, \delta)$$

Admissibilité

Définition 14. Un estimateur δ_0 est dit **inadmissible** s'il existe un estimateur δ_1 qui domine δ_0 , c'est à dire :

$$\forall \theta \in \Theta, \quad \mathcal{R}(\theta, \delta_0) \geq \mathcal{R}(\theta, \delta_1)$$

et, pour au moins une valeur θ_0 du paramètre,

$$\mathcal{R}(\theta_0, \delta_0) > \mathcal{R}(\theta_0, \delta_1)$$

Sinon, δ_0 est dit admissible.

Les résultats suivants fournissent les conditions suffisantes d'admissibilité.

Proposition 1. *S'il existe un unique estimateur mini-max, cet estimateur est admissible.*

Démonstration. Si δ^* est le seul estimateur minimax, pour tout estimateur $\tilde{\delta} \neq \delta^*$,

$$\sup_{\theta} \mathcal{R}(\theta, \tilde{\delta}) > \sup_{\theta} \mathcal{R}(\theta, \delta^*)$$

Donc, $\tilde{\delta}$ ne peut pas dominer δ^* .

□

Proposition 2. *Si l'estimateur de Bayes associé à la loi π est unique, alors il est admissible.*

Démonstration. Supposons que δ^B est un estimateur Bayésien inadmissible :

$$\exists \delta_0 \in \mathcal{D}, \forall \theta \in \Theta, \mathcal{R}(\theta, \delta) \geq \mathcal{R}(\theta, \delta_0) \quad \text{et} \quad \exists \theta_0 \in \Theta, \mathcal{R}(\theta_0, \delta) > \mathcal{R}(\theta_0, \delta_0)$$

En intégrant la première inégalité :

$$\int_{\Theta} \mathcal{R}(\theta, \delta_0) d\theta \leq \int_{\Theta} \mathcal{R}(\theta, \delta^B) \pi(\theta) d\theta = r(\pi),$$

donc δ_0 est aussi un estimateur Bayésien associé à L et π et $\delta_0 \neq \delta^B$ d'après la seconde inégalité, donc la proposition se déduit par contraposée.

Cette proposition s'applique notamment dans le cas d'un risque fini et d'une fonction coût convexe.

En outre, l'unicité et d'estimateur Bayésien implique la finitude du risque :

$$r(\pi) = \int_{\Theta} \mathcal{R}(\theta, \delta^B) \pi(\theta) d\theta < \infty$$

Sinon tout estimateur minimise le risque. □

Proposition 3. *Si la distribution a priori π est strictement positive sur Θ , de risque de Bayes fini et la fonction de risque $\mathcal{R}(\theta, \delta)$ est une fonction continue de Θ pour toute δ , alors l'estimateur de Bayes δ^B est admissible.*

Proposition 4. *Si l'Estimateur δ_0 est admissible est de risque constant, alors δ_0 est l'unique estimateur minimax.*

Exemple 12.

Comme exemple, si on considère la fonction de perte usuelle :

$$L(\delta(x), \theta) = (\delta(x) - \theta)^2$$

alors

$$\begin{aligned} r(\pi, \delta) &= \iint L(\delta(x), \theta) f(x|\theta) g(\theta) d\theta dx \\ &= \iint (\delta(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\ &= \iint (\delta(x) - \delta^B(x) + \delta^B(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\ &= \iint (\delta(x) - \delta^B(x))^2 f(x|\theta) g(\theta) d\theta dx \\ &\quad + \iint (\delta^B(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\ &\quad + 2 \iint (\delta(x) - \delta^B(x)) (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta dx \\ &= r(\pi, \delta^B) + \iint (\delta(x) - \delta^B(x))^2 f(x|\theta) g(\theta) d\theta dx \\ &\quad + 2 \iint (\delta(x) - \delta^B(x)) (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta dx \end{aligned}$$

Donc, pour minimiser $r(\pi, \delta)$ par rapport à δ , il suffit que :

$$\frac{\partial}{\partial \delta} (2 \iint (\delta(x) - \delta^B(x))(\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta dx) = 0,$$

ou encore, il suffit que

$$\int (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta = 0.$$

Ceci équivaut à

$$\begin{aligned} \delta^B(x) &= \frac{\int \theta f(x|\theta) g(\theta) d\theta}{\int f(x|\theta) g(\theta) d\theta} \\ &= \frac{\int \theta f(x|\theta) g(\theta) d\theta}{f(x)} \\ &= \int \theta \pi(\theta|x) d\theta \\ &= E_{\theta|x}(\theta). \end{aligned}$$

L'estimateur Bayésien ou la moyenne a posteriori est alors noté $E_{\theta|x}(\theta)$.

Voici une proposition très pratique pour les calculs des risques intégrés, qui est une déduction immédiate de la démonstration de la proposition 4.

Proposition 5. *Le risque intégré peut s'écrire comme suit*

$$r(\pi, \delta) = r(\pi, \delta^B) + \int (\delta(x) - \delta^B(x))^2 f(x) dx, \quad (2.4)$$

où

$$\delta^B(x) = \frac{\int \theta f(x|\theta) g(\theta) d\theta}{\int f(x|\theta) g(\theta) d\theta}. \quad (2.5)$$

2.1.5 Fonctions coût usuelles

L'objectif de l'inférence est de fournir une décision au statisticiens.

Les différentes décisions sont comparées au moyen d'un critère d'évaluation donné par la fonction coût.

L'inférence consiste à choisir une règle de décision $\delta \in \mathcal{D}$ concernant $\theta \in \Theta$ sur la base d'observation $x \in \mathcal{X}$

Définition 15. On appelle fonction coût, toute fonction "L" de $\Theta \times \mathcal{A}$ dans \mathbb{R}^+ , C'est-à-dire :

$$\begin{aligned} L : \Theta \times \mathcal{A} &\longrightarrow \mathbb{R}^+ \\ \theta, \delta(x) &\longrightarrow L(\theta, \delta(x)) \end{aligned}$$

$L(\theta, \delta(x))$ évalue le coût associé à la décision $a = \delta(x)$ quand le paramètre vaut θ . elle permet de quantifier la perte encourue par une mauvaise décision.

Nous allons définir 3 fonction coût usuelles. Elles sont appréciées car elles nous permettent d'obtenir aisément l'estimateur Bayésien.

La fonction coût quadratique

Définition 16. La fonction coût quadratique est la fonction définie par

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2$$

Proposition 6. L'estimateur de Bayes δ^B associé à la loi à priori π et au coût quadratique est la moyenne a posteriori θ .

$$\delta^B = E(\theta|x) \tag{2.6}$$

Cette estimateur est aussi appelé minimum mean square error (MMSE).

Démonstration. Par définition, l'estimateur de Bayes δ^B minimise le coût à posteriori

$$\begin{aligned} \rho(\pi, \delta) &= E_{\pi(\cdot|x)}(L(\theta, \delta)) \\ \rho(\pi, \delta) &= E_{\pi(\cdot|x)}(\theta_\delta)^2 = E_{\pi(\cdot|x)}(\theta^2 + \delta^2 - 2\delta\theta) \\ &= E_{\pi(\cdot|x)}(\theta^2) + \delta^2 - 2\delta E_{\pi(\cdot|x)}(\theta) \\ \frac{d\rho(\pi, \delta)}{d\delta} &= 2\delta - 2E_{\pi(\cdot|x)}(\theta) = 0 \\ &\Rightarrow \delta^B = E_{\pi(\cdot|x)}(\theta) \Rightarrow \delta^B \text{ est l'estimateur de Bayes} \\ \frac{d^2\rho(\pi, \delta)}{d\delta^2} &= 2 > 0 \end{aligned}$$

□

Lemme 1. Soit $x|\mu \sim \mathcal{N}(\mu, \sigma^2)$ et $\mu \sim \mathcal{N}(M, A)$. Alors l'estimateur Bayésien, $\delta^B(x)$, est :

$$\delta^B(x) = M + \frac{A}{A + \sigma^2}(x - M) \tag{2.7}$$

Démonstration.

La loi a posteriori de μ sachant x , $\pi(\mu|x)$, est proportionnelle à :

$$\begin{aligned}
 \pi(\mu|x) &\propto f(x|\mu)\pi(\mu) \\
 &\propto \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \exp\left(-\frac{1}{2A}(\mu-M)^2\right) \\
 &\propto \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 - \frac{1}{2A}(\mu-M)^2\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \frac{\mu^2}{A} - 2\frac{x\mu}{\sigma^2} - 2\frac{M\mu}{A}\right)\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(\frac{\mu^2}{\alpha^2} - 2\mu\left(-\frac{x}{\sigma^2} + \frac{M}{A}\right)\right)\right) \\
 &\propto \exp\left(-\frac{1}{2}\frac{1}{\alpha^2}\left(\mu^2 - 2\mu\alpha^2\left(\frac{x}{\sigma^2} + \frac{M}{A}\right)\right)\right) \\
 &\propto \exp\left(-\frac{1}{2\alpha^2}\left(\mu - \alpha^2\left(\frac{x}{\sigma^2} + \frac{M}{A}\right)\right)^2\right)
 \end{aligned}$$

avec $\frac{1}{\alpha^2} = \frac{1}{\sigma^2} + \frac{1}{A}$, et on a d'après la proposition(2.1.5)

$$\delta^B(x) = \mathbb{E}_{\mu|x}(\mu).$$

Soit alors

$$\begin{aligned}
 \delta^B(x) &= \alpha^2 \left(\frac{x}{\sigma^2} + \frac{M}{A}\right) \\
 \delta^B(x) &= \frac{\alpha^2 x}{\sigma^2} + \frac{\alpha^2 M}{A} \\
 \delta^B(x) &= \frac{A}{A + \sigma^2}x + \frac{\sigma^2}{A + \sigma^2}M \\
 \delta^B(x) &= M + \frac{A}{A + \sigma^2}(x - M)
 \end{aligned}$$

Déduisons aussi que cette loi a posteriori de μ sachant x est :

$$\pi(\mu|x) \sim \mathcal{N}\left(M + \frac{A}{A + \sigma^2}(x - M), \frac{\sigma^2 A}{A + \sigma^2}\right). \quad (2.8)$$

De même, la distribution de la fonction marginale f de x est déduite de la façon

suivante :

$$\begin{aligned}
 f_X(x) &= \int f(x|\mu)\pi(\mu)d\mu \\
 &\propto \int e^{-\frac{1}{2\sigma^2}(x-\mu)^2} e^{-\frac{1}{2A}(\mu-M)^2} d\mu \\
 &\propto \int e^{-\frac{1}{2\sigma^2}(x^2-2x\mu+\mu^2)} e^{-\frac{1}{2A}(\mu^2-2M\mu+M^2)} d\mu \\
 &\propto \int e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^2}\right)} e^{\mu^2\left(\frac{1}{\sigma^2}+\frac{1}{A}\right)-2\mu\left(\frac{x}{\sigma^2}+\frac{M}{A}\right)} d\mu \\
 &\propto e^{-\frac{1}{2\sigma^2}x^2} e^{\frac{1}{2}\alpha^2\left(\frac{1}{\alpha^2}\left(\frac{x}{\sigma^2}+\frac{M}{A}\right)\right)^2} \int e^{-\frac{1}{2}\alpha^2\left(\mu-\frac{1}{\alpha^2}\left(\frac{x}{\sigma^2}+\frac{M}{A}\right)\right)^2} d\mu \\
 &\propto e^{-\frac{1}{2\sigma^2}x^2} e^{-\frac{1}{2}\alpha^2\left(\frac{1}{\alpha^2}\left(\frac{x}{\sigma^2}+\frac{M}{A}\right)\right)^2} \\
 &\propto \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{1}{2\alpha^2}\left(\frac{x}{\sigma^2} + \frac{M}{A}\right)^2\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(x^2\left(\frac{1}{\sigma^2} - \frac{1}{\alpha^2\sigma^2}\right) - \frac{2xM}{\alpha^2\sigma^2 A}\right)\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(x^2 a - \frac{2xM}{\alpha^2\sigma^2 A}\right)\right) \quad \text{où } a = \frac{1}{\sigma^2 + A} \\
 &\propto \exp\left(-\frac{1}{2}a\left(x - \frac{2xM}{a\alpha^2\sigma^2 A}\right)^2\right)
 \end{aligned}$$

Nous remarquons que $a\alpha^2\sigma^2 A = 1$.

On obtient alors :

$$f_X(x) \propto \exp\left(-\frac{1}{2}a(x - M)^2\right). \quad (2.9)$$

Par conséquent,

$$x \sim \mathcal{N}(M, \sigma^2 + A). \quad (2.10)$$

□

La fonction coût absolue

Définition 17. La fonction coût absolue est la fonction définie par : $L(\theta, \delta) = |\theta - \delta|$

$$L(\theta, \delta) = \begin{cases} \theta - \delta & \text{si } \theta > \delta \\ \delta - \theta & \text{si } \theta \leq \delta \end{cases}$$

Proposition 7. L'estimateur de Bayes à la loi a priori π et à la fonction coût absolue est la médiane de la loi a posteriori, tel que :

$$\int_{-\infty}^{\delta^B} \pi(\theta|x) d\theta = \frac{1}{2}$$

La fonction Coût 0-1

Définition 18. Soit la fonction L défini comme suite

$$L(\theta, \delta) = \begin{cases} 0 & \text{si } |\theta - \delta| < \varepsilon \\ 1 & \text{si } |\theta - \delta| \geq \varepsilon \end{cases}$$

où $\varepsilon > 0$ est un nombre petit est fini.

Proposition 8. *L'estimateur de Bayes associé à la loi a priori π et a la fonction coût 0-1 est le mode de la loi distribution a posteriori $\pi(\theta|x)$. On l'appelle également le maximum a posteriori MAP.*

REMARQUE.

1. Notons que l'estimateur MAP maximise

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)},$$

ceci revient à maximiser $f(x|\theta).\pi(\theta)$

2. Le MAP est utilisé lorsque MMSE est difficile à mettre en oeuvre pour des raisons de calcul d'intégrale compliqué.

2.1.6 Modélisation de l'information a priori**Lois a priori conjuguées**

Définition 19. Une famille \mathbb{F} de distribution sur Θ est dite conjuguée pour la loi $f(X|\theta)$, si pour tout $\pi(\cdot) \in \mathbb{F}$, la distribution a posteriori $\pi(\cdot|X) \in \mathbb{F}$ aussi.

REMARQUE. : Une loi conjuguée peut être déterminée en considérant la forme de la vraisemblance $f(x|\theta)$ et en prenant une loi a priori de la même forme que cette dernière. Les lois a priori conjuguées obtenues par ce procédé sont dite conjuguées naturelles.

Famille exponentielle

On appelle famille exponentielle, toute famille de loi de probabilité P_θ et de densité de la forme suivante :

$$f(x|\theta) = \exp\left\{\sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta)\right\}.h(x)$$

Où $\eta_i(\cdot), B(\cdot)$ sont des fonctions du paramètre θ et $T_i(\cdot)$ sont des statistiques et h est une fonction de x

Exemple 13. Soit la loi exponentielle de paramètre $\lambda = \frac{1}{\theta}$

$$X \sim \exp\left(\frac{1}{\theta}\right)$$

Ça densité est donné par :

$$\begin{aligned} f(X|\theta) &= \frac{1}{\theta} \exp\left(\frac{-x}{\theta}\right) \\ &= \exp\left\{\frac{-x}{\theta} + \ln\left(\frac{1}{\theta}\right)\right\} \\ &= \exp\left\{\frac{-x}{\theta} - \ln(\theta)\right\} \end{aligned}$$

On a bien : $s = 1, \eta_1(\theta) = \frac{-1}{\theta}, T_1(x) = x, B(\theta) = \ln(\theta), h(x) = 1$ ou $h(x) = 1_{[0,+\infty[}(x)$

Donc la loi exponentielle appartient à la famille exponentielle.

Proposition 9. Soit $f(x|\theta)$ une densité appartenant à une famille exponentielle, alors une famille de loi a priori conjuguée pour $f(x|\theta)$ est donnée par :

$$\pi(\theta|\mu, \lambda) = K(\mu, \lambda) \exp\{\theta \cdot \mu - \lambda B(\theta)\}.$$

Où $K(\mu, \lambda)$ est une constante de normalisation.

La loi a posteriori est de la forme :

$$\pi(\theta|x) \propto \exp\{\theta(\mu + x) - (\lambda + 1)B(\theta)\}$$

L'avantage de cette famille est avant tout de simplifier les calculs qui s'avèrent parfois long ou très complexes.

Exemples de lois conjuguées

Le tableau ci-dessous donne des exemples de lois a priori conjuguées usuelles.

Exemple 14. Soit X une variable aléatoire d'une loi de Pareto de paramètres $\theta > 0$ et $a > 0$, de densité de probabilités :

$$f(x|\theta) = \frac{\theta a^\theta}{x^{\theta+1}} 1_{[a,+\infty)}(x) \tag{2.11}$$

On suppose que θ est connu, On suppose que a est connu,

$$\begin{aligned} f(x|\theta, a) &= \theta e^{\theta \ln(a)} e^{-(\theta+1) \ln(x)} \\ &= \theta e^{\theta \ln(a) - \theta \ln(x)} e^{-\ln(x)} \\ &= \theta e^{\theta \ln\left(\frac{a}{x}\right)} e^{-\ln(x)} \propto \theta e^{\theta \ln\left(\frac{a}{x}\right)} \\ &= \theta^{2-1} e^{-\theta \ln\left(\frac{x}{a}\right)} \end{aligned}$$

Cette loi, par rapport à θ est de la forme d'une loi gamma(2, $\ln\left(\frac{x}{a}\right)$), on peut donc prendre une loi a posteriori de type gamma.

TABLE 2.1 – Exemples de lois a priori conjuguées

$f(X \theta)$	$\pi(\theta)$	$\pi(\theta X)$	$E(\theta X)$
$N(\theta, \delta^2)$	$N(\mu, \tau^2)$	$N\left(\frac{x}{\delta^2} + \frac{\mu}{\tau^2}, \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\right)$	$\frac{x}{\delta^2} + \frac{\mu}{\tau^2}$
$P(\theta)$	$g(a, b)$	$g(a + x, b + 1)$	$\frac{a+x}{b+1}$
$B(n, \theta)$	$Beta(a, b)$	$Beta(a+n, b+x)$	$\frac{a+n}{a+n+b+x}$
$g(a, \theta)$	$g(\alpha, \beta)$	$g(a + \alpha, \beta + x)$	$\frac{a+\alpha}{\beta+x}$

Lois non informatives

Définition 20. Une loi non informative est une loi qui porte une information sur le paramètre à estimer dont le poids dans l'inférence est réduit.

REMARQUE. Certains auteurs considèrent la loi non informative comme une loi qui n'apporte aucune information sur le paramètre à estimer, Ou encore, on la considère comme une loi qui ne donne pas davantage de poids à telle ou telle valeur de paramètre. On choisit ces lois lorsque l'on dispose de peu d'informations sur θ et que l'on souhaite que les données parlent d'elle-même.

Loi a priori impropre

Soit $\pi(\cdot)$ une application de Θ dans $[0, +\infty]$, telle que :

$$\int_{\Theta} \pi(\theta) d\theta = +\infty$$

On l'appelle loi a priori impropre.

Estimation de Bayes généralisée

On se donne une loi à priori impropre, caractérisée par la fonction $\pi(\theta)$
 On suppose que l'intégrale $\int_{\Theta} f(x|\theta) \cdot \pi(\theta) d\theta$ est convergente.
 On considère la densité de la loi à posteriori définie par :

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

REMARQUE.

- Ceci est une définition et non une application du théorème de Bayes, puisque le théorème de Bayes peut seulement être appliqué lorsque les distributions ne sont pas impropres.

- Rappelons que pour une distribution a priori propre, l'estimateur Bayésien minimise la moyenne à posteriori de la fonction perte.
- Lorsque la distribution a priori est impropre, un estimateur qui minimise la moyenne à posteriori de la fonction perte est appelé estimateur Bayésien généralisé.
- L'inférence Bayésienne étant basées sur $\pi(\theta|x)$, l'usage des lois a priori impropres est justifié si la loi posteriori est "propre".

Exemple 15. *Considérons le modél Bayésien suivant :*

$$\begin{cases} X|\theta & \rightsquigarrow \exp(\theta) & \theta > 0 \\ \theta & \rightsquigarrow U[0, +\infty] \end{cases}$$

◇ *Vérifions si la loi a priori est propre ou non :*

La loi a priori : $\pi(\theta) = 1_{[0, +\infty]} = 1$

$$\int_0^{+\infty} \pi(\theta) d\theta = \int_0^{+\infty} 1 d\theta = \theta \Big|_0^{+\infty} = +\infty$$

$\pi(\theta)$ est impropre.

la loi a posteriori est donc :

$$L(x, \theta) = \prod_{i=1}^n f(x|\theta) = \theta^n e^{-\theta \sum x_i}$$

$$\pi(\theta|x) = \frac{\theta^n e^{-\theta \sum x_i} \times 1}{\int_0^{+\infty} \theta^n e^{-\theta \sum x_i} d\theta} = \frac{N}{D}$$

$$D = \int_0^{+\infty} \theta^{n+1-1} e^{-\theta \sum x_i} d\theta \cdot \frac{(\sum x_i)^{n+1}}{\Gamma(n+1)} \cdot \frac{\Gamma(n+1)}{(\sum x_i)^{n+1}}$$

$$\theta|x \sim \text{gamma}(n+1, n\bar{X})$$

Sous la fonction coût quadratique, l'estimateur $\delta_\pi(x) = \frac{n+1}{n\bar{X}}$ est appelé estimateur Bayésien généralisé.

Loi a priori uniforme.

La loi a priori non informative la plus simple et la plus communément utilisée est la loi uniforme. En effet, ce choix repose sur l'équiprobabilité des valeurs possibles du paramètre θ sur son domaine de définition, et donc elle n'apporte aucune information supplémentaire sur θ .

Ainsi, supposons que Θ soit un ensemble de taille k , alors $\pi(\theta) = \frac{1}{k}$

La loi de Jeffreys.

En 1961, Jeffreys propose une méthode de construction de loi a priori non informatives, cette méthode utilise l'information de Fisher $I(\theta)$ qui représente une mesure de la quantité d'informations sur θ contenu dans l'observation.

Définition 21. Soit θ étant un paramètre réel, on appelle loi à priori non informative de Jeffreys la loi a priori de la forme :

$$\pi_J(\theta) = c\sqrt{I_X(\theta)} \quad 1_{\Theta}(\theta)$$

c est une constante de normalisation, et $I_X(\theta)$ l'information de Fisher apportée par X sur θ définie par :

si le domaine de X dépend de θ :

$$I_X(\theta) = E \left[\left(\frac{d}{d\theta} \ln f(x|\theta) \right)^2 \right]$$

si le domaine de X ne dépend pas de θ :

$$I_X(\theta) = -E \left[\frac{d^2}{d\theta^2} \ln f(x|\theta) \right]$$

REMARQUE. L'a priori de Jeffreys offre une méthode automatisée pour obtenir une loi a priori non informative pour n'importe quel modèle paramétrique.

REMARQUE. Plus $I_X(\theta)$ est grand, plus l'observation apporte de l'information. c'est pour cela que l'on favorise les valeurs de θ pour lesquels $I_X(\theta)$ est grande .

Exemple 16. Soit $x \sim B(n, p)$,

$$f(x|p) = C_n^x p^x (1-p)^{n-x}$$

$$\frac{\partial^2 f(x|p)}{\partial p^2} = \frac{x}{p^2} - \frac{n-x}{(1-p)^2}$$

$$I(p) = n \left[\frac{1}{p} + \frac{1}{1-p} \right] = \frac{n}{p(1-p)}$$

Donc la loi de Jeffreys pour ce modèle est

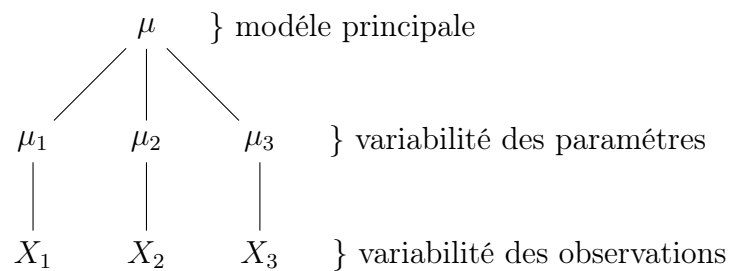
$$\pi(p) \propto [p(1-p)]^{-1/2}$$

et est alors propre, car il s'agit de la distribution $Be(1/2, 1/2)$.

2.2 La Statistique Bayésienne Hiérarchique

Le modèle Bayésien hiérarchique est une approche statistique puissante qui permet de modéliser des données avec une structure complexe et hiérarchique.

Dans un modèle Bayésien hiérarchique, les paramètres sont eux-mêmes modélisés comme des variables aléatoires, et des distributions a priori sont spécifiées pour ces paramètres. Ces distributions a priori expriment les connaissances préalables ou les croyances sur les paramètres avant de voir les données observées. En combinant ces distributions a priori avec les données observées, le modèle peut estimer les paramètres postérieurs, qui capturent à la fois l'information des données et les connaissances a priori.



Les modèles Bayésiens hiérarchiques sont donc des outils essentiels pour l'analyse statistique lorsque les données présentent une structure hiérarchique ou lorsque des informations a priori sont disponibles sur les paramètres du modèle.

Soit X un vecteur de variables aléatoires observées

$$X = (X_1, X_2, \dots, X_n),$$

et θ un vecteur de paramètres inconnus

$$\theta = (\theta_1, \theta_2, \dots, \theta_n).$$

Le modèle Bayésien hiérarchique suppose que le paramètre θ suivant une structure hiérarchique définie par des niveaux de l'hiérarchie (n niveaux).

À chaque niveau de la hiérarchie, les paramètres θ sont modélisés comme des variables aléatoires conditionnelles aux paramètres des niveaux supérieurs. En notation mathématique, un modèle Bayésien hiérarchique de niveau n peut être représenté par :

$$x|\theta \sim f(x|\theta)$$

Avec $f(x|\theta)$ est la densité conditionnelle de x sachant θ .

Ont modélise une hiérarchie de niveaux n comme suit

$$\begin{aligned}
 \theta|\theta_1 &\sim \pi_1(\theta|\theta_1) \\
 \theta_1|\theta_2 &\sim \pi_2(\theta_1|\theta_2) \\
 &\vdots \\
 \theta_{n-1}|\theta_n &\sim \pi_n(\theta_{n-1}|\theta_n) \\
 \theta_n &\sim \pi_{n+1}(\theta_n)
 \end{aligned}$$

avec

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1)\pi_2(\theta_1|\theta_2) \dots \pi_{n+1}(\theta_n) d\theta_1 \dots d\theta_n$$

Les paramètres θ_i sont appelés hyperparamètres de niveau i ($1 \leq i \leq n$) et $\pi_{n+1}(\theta_n)$ une loi marginale.

Le modèle Bayésien hiérarchique utilise le théorème de Bayes (2.1.3) pour estimer la distribution a posteriori des paramètres, en tenant compte des données observées x :

$$\pi(\theta|x, \lambda) = \frac{f(x|\theta)\pi(\theta|\lambda)}{m(x|\lambda)}$$

Où :

- λ est un vecteur des hyperparametres.
- $\pi(\theta|x, \lambda)$ est la distribution a posteriori des paramètres.
- $\pi(\theta|\lambda)$ est une densité a priori de θ .
- $m(x|\lambda)$ est la probabilité marginale des données X , tel que :

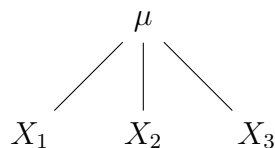
$$m(x|\lambda) = \int_{\Theta} f(x|\theta)\pi(\theta|\lambda) d\theta$$

En utilisant des méthodes d'inférence Bayésienne telles que MCMC (Markov Chain Monte Carlo) ou des algorithmes de "variational inference", on peut obtenir des échantillons de la distribution a posteriori pour estimer les paramètres et quantifier l'incertitude dans les prédictions.

Exemple 17. *Le Bayésien hiérarchique représente le juste milieu entre l'ajustement indépendant des données à des modèles μ_i pour différents groupes de données X_i*

$$\begin{array}{ccc}
 \mu_1 & \mu_2 & \mu_3 \\
 \downarrow & \downarrow & \downarrow \\
 X_1 & X_2 & X_3
 \end{array}$$

et l'ajustement simultané de ce même modèle à un ensemble complet de données.



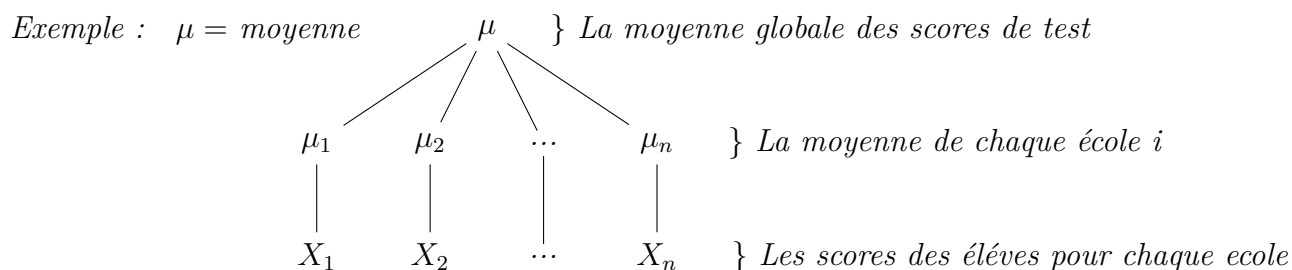
Les modèles hiérarchiques tentent d'écrire des modèles qui décrivent la variabilité des paramètres de nos modèles, Considérons par exemple une autorité éducative étatique qui collecte les scores de tests standardisés en mathématiques pour des élèves (du même âge) sélectionnés au hasard dans 100 écoles de la région.

Dans ce cas nous sommes confrontés à des données structurées de manière hiérarchique, où les individus (les élèves) sont regroupés en classes, les classes sont regroupées dans des écoles, et les écoles sont situées dans différentes régions ou états.

Nous pouvons raisonnablement supposer que les performances aux tests varient non seulement entre les élèves individuels, mais aussi entre les classes et les écoles. Par exemple, les différences dans la qualité de l'enseignement, les ressources disponibles et les caractéristiques socio-économiques des quartiers peuvent influencer les performances des élèves.

Implicitement à cet exemple est la supposition que les individus dans la même classe sont plus similaires que les individus dans la même école (mais pas dans la même classe), qui, à leur tour, sont plus similaires que deux individus choisis au hasard dans le même état.

Le modèle Bayésien hiérarchique offre une solution pour prendre en compte cette structure hiérarchique des données. Plutôt que de simplement estimer un modèle distinct pour chaque école, le modèle hiérarchique tient compte de la variabilité intra-groupe (entre les élèves d'une même classe...) et la variabilité entre les groupes (par exemple, entre les classes d'une même école ou entre les écoles de différentes régions).



En utilisant cette approche, le modèle peut fournir des estimations plus fiables de la moyenne des scores aux tests pour un élève sélectionné au hasard dans la région. De plus, en intégrant la structure hiérarchique des données, le modèle est moins sensible aux valeurs aberrantes car ils prennent en compte la structure hiérarchique des données et ne se contentent pas de modéliser chaque groupe individuellement de manière isolée.

2.3 La Statistique Bayésien Empirique

La méthode Bayésienne empirique est utilisée pour estimer les paramètres d'une distribution lorsque les observations sont des variables aléatoires indépendantes et identiquement distribuées suivant cette distribution. Il convient de noter que cette technique n'est pas purement Bayésienne car elle utilise des approximations fréquentistes lorsque l'information

a priori est insuffisante ou indisponible. Néanmoins, il a été démontré que cette approche peut fournir des résultats asymptotiquement équivalents à ceux du modèle Bayésien hiérarchique.

L'analyse Bayésienne empirique peut être considérée comme une combinaison efficace des méthodes fréquentiste et Bayésienne. Elle complète le modèle hiérarchique Bayésien en permettant d'estimer la densité a posteriori lorsque les hyperparamètres sont inconnus.

2.3.1 Le Concept Du Bayésien Empirique

Dans la méthode Bayésienne traditionnelle, la distribution a priori est définie avant d'obtenir les données, puis on utilise les données observées pour modifier la distribution a priori afin d'obtenir une probabilité a posteriori.

En revanche, la distribution a priori dans la méthode Bayésienne empirique n'est pas donnée à l'avance, mais est estimée à partir des données d'observation obtenues, ce qui confirme le mot « empirique » dans son nom.

Exemple 18. *Je savais avant de prendre l'avion que la probabilité d'un crash d'avion était très faible, et cette probabilité était dans mon esprit avant de prendre l'avion, mais si j'avais observé les tensions internationales récentes (données d'observation), la probabilité d'un crash d'avion aurait pu augmenter (a posteriori corriger).*

On considère X , un vecteur d'observations indépendantes et identiquement distribuées de vraisemblance $f(x|\theta)$ avec θ un vecteur de paramètres inconnus de fonction de densité a priori $\pi(\theta|\lambda)$ et λ un vecteur des hyperparamètres, $\lambda \in \Upsilon$ l'espace des hyperparamètres. La densité a posteriori de θ est donnée par :

$$\pi(\theta|x, \lambda) = \frac{f(x|\theta)\pi(\theta|\lambda)}{m(x|\lambda)}, \quad (2.12)$$

où $m(x|\lambda)$ est la distribution marginale de x sachant λ

$$m(x|\lambda) = \int f(x|\theta)\pi(\theta|\lambda) d\theta \quad (2.13)$$

Si λ est connu, Nous sommes dans le cas Bayésien standard. Donc la distribution a posteriori de θ serait simplement écrite comme

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}, \quad (2.14)$$

où $m(x)$ est la distribution marginale de x

$$m(x) = \int f(x|\theta)\pi(\theta) d\theta \quad (2.15)$$

Si λ est inconnu, dans l'approche Bayésienne empirique on considère la distribution hyperpriori qui est la loi a priori de l'hyperparamètre λ notée $\pi(\lambda)$. Nous définissons la distribution a posteriori de θ de la manière suivante :

$$\pi(\theta|x) = \frac{\int f(x|\theta)\pi(\theta|\lambda)\pi(\lambda) d\lambda}{\int \int f(x|\theta)\pi(\theta|\lambda)\pi(\lambda) d\theta d\lambda} \quad (2.16)$$

En principe, il n'y a aucune raison pour que l'hyperpriori ne dépende pas elle-même d'une collection de paramètres inconnus u , ce qui conduit à une généralisation de l'équation (2.16) comprenant un deuxième stade a priori $\pi(\lambda|u)$ et un troisième stade a priori $\pi(u)$. Cette démarche de spécification d'un modèle sur plusieurs niveaux est appelée modélisation hiérarchique, chaque nouvelle distribution formant un nouveau niveau dans la hiérarchie. Le nombre approprié de niveaux varie selon le problème.

Continuant avec le modèle à deux niveaux nous observons que l'équation (2.12) peut être exprimée en notation abrégée pratique :

$$\pi(\theta | x) \propto f(x | \theta)\pi(\theta|\lambda). \quad (2.17)$$

L'analyse Bayésienne empirique propose d'estimer les hyperparamètres à partir de la loi marginale :

$$m(x|\lambda) = \int_{\Theta} f(x|\theta)\pi(\theta|\lambda)d\theta$$

pour obtenir $\hat{\lambda}(x)$ et d'utiliser $\pi(\theta|\hat{\lambda}(x), x)$ en tant que pseudo-a posteriori.

Un inconvénient de l'analyse Bayésienne empirique est qu'elle repose sur des méthodes fréquentistes pour l'estimation des hyperparamètres de $m(x|\lambda)$, alors qu'on pourrait tout aussi bien employer des techniques Bayésiennes. Une conséquence de ce choix est qu'un grand nombre de méthodes est utilisable : par exemple, l'estimateur de λ peut être choisi par la méthode des moments ou par la méthode du maximum de vraisemblance. L'analyse Bayésienne empirique est alors souvent employée comme un outil pour justifier a posteriori des estimateurs déjà existants. L'approche la plus répandue est d'utiliser des estimateurs du maximum de vraisemblance pour des raisons à la fois pratiques et théoriques.

2.3.2 Estimation Bayésienne empirique pour la loi de Poisson

Nous considérons le modèle de Poisson suivant :

$$\begin{cases} x|\theta & \sim P(\theta) \\ \theta|\lambda & \sim \exp(\lambda) \\ \lambda & \text{inconnu.} \end{cases} \quad (2.18)$$

L'objectif est d'estimer θ , la moyenne Poissonnienne basée sur n observations indépendantes et identiquement distribuées selon $Poisson(\theta)$.

On suppose qu'on a n observations, $(x_1, \theta_1), \dots, (x_n, \theta_n)$ du même modèle et une observation récente (x_{n+1}, θ_{n+1}) pour laquelle on veut estimer son paramètre θ_{n+1} , pour cela on estime d'abord l'hyperparamètre λ par la méthode du maximum de vraisemblance.

La densité marginale

La densité marginale est donc :

$$\begin{aligned}
 m(x_i|\lambda) &= \int_0^\infty f(x_i|\theta) \pi(\theta|\lambda) d\theta \\
 &= \int_0^\infty e^{-\theta} \frac{\theta^{x_i}}{x_i!} \lambda e^{-\lambda\theta} d\theta \\
 &= \frac{\lambda}{x_i!} \int_0^\infty \theta^{x_i} e^{-\theta-\lambda\theta} d\theta \\
 &= \frac{\lambda}{x_i!} \frac{\Gamma(x_i+1)}{(1+\lambda)^{x_i+1}} \int_0^\infty \frac{(1+\lambda)^{x_i+1}}{\Gamma(x_i+1)} \theta^{(x_i+1)-1} e^{-\theta(1+\lambda)} d\theta \\
 &= \frac{\lambda}{x_i!} \frac{\Gamma(x_i+1)}{(1+\lambda)^{x_i+1}} \int_0^\infty G(x_i+1, 1+\lambda) d\theta
 \end{aligned}$$

On voit bien que la fonction G (qui représente aussi la distribution a priori de $\theta, \pi(\theta)$) est la densité de la loi gamma(α, β) de paramètres $\alpha = x_i + 1$ et $\beta = 1 + \lambda$, et donc son intégrale vaut 1 sur $[0, +\infty[$, et on sait bien que $\Gamma(n+1) = n!$, donc :

$$\begin{aligned}
 m(x_i|\lambda) &= \frac{\lambda}{(1+\lambda)^{x_i+1}} \\
 &= \left(\frac{1}{1+\lambda}\right)^{x_i} \frac{\lambda}{1+\lambda}
 \end{aligned}$$

Donc $x_i|\lambda$ est une géométrique de moyenne $\frac{1+\lambda}{\lambda}$

L'estimateur de maximum de vraisemblance

Le logarithme de la fonction de vraisemblance pour les n premières observations (x_1, x_2, \dots, x_n) est donné par :

$$\begin{aligned}
 \log\left(\prod_{i=0}^n m(x_i|\lambda)\right) &= \log\left(\left(\frac{1}{1+\lambda}\right)^{\sum x_i} \left(\frac{\lambda}{1+\lambda}\right)^n\right) \\
 &= \log\left(\left(\frac{1}{1+\lambda}\right)^{\sum x_i}\right) + \log\left(\left(\frac{\lambda}{1+\lambda}\right)^n\right) \\
 &= -\sum x_i \log(1+\lambda) + n \log\left(\frac{\lambda}{1+\lambda}\right) \\
 &= -\sum x_i \log(1+\lambda) + n \log(\lambda) - n \log(1+\lambda)
 \end{aligned}$$

On dérive par λ , on obtient :

$$\frac{\partial}{\partial \lambda} \log \left(\prod_{i=0}^n m(x_i | \lambda) \right) = \frac{-\sum x_i}{1 + \lambda} + \frac{n}{\lambda} - \frac{n}{1 + \lambda} = 0$$

ce qui est équivalent à

$$-\sum x_i + \frac{n(1 + \lambda)}{\lambda} - n = 0$$

$$-\sum x_i + \frac{n}{\lambda} + n - n = 0$$

$$\frac{n}{\lambda} - \sum x_i = 0$$

$$\frac{n}{\lambda} = \sum x_i$$

$$\Rightarrow \hat{\lambda}(x) = \frac{n}{\sum x_i} \quad (2.19)$$

$$\Rightarrow \hat{\lambda}(x) = \frac{1}{\bar{x}} \quad \text{qui est l'EMV de } \lambda \quad (2.20)$$

Estimateur Bayésien empirique

La distribution a posteriori de θ sachant $\hat{\lambda}(x)$ et x , d'après (2.17) est estimée par :

$$\pi(\theta | \hat{\lambda}, x_{n+1}) \propto f(x_{n+1} | \theta) \times \pi(\theta)$$

$$\pi(\theta | \hat{\lambda}, x_{n+1}) \propto \left(e^{-\theta} \frac{\theta^{x_{n+1}}}{x_{n+1}!} \right) \hat{\lambda} e^{-\theta \hat{\lambda}}$$

$$\propto \theta^{x_{n+1}} e^{-\theta(\hat{\lambda}+1)}$$

$$\propto \Gamma(x_{n+1} + 1, \hat{\lambda} + 1).$$

L'estimateur Bayésien empirique paramétrique de θ sous la quadratique d'après (2.6) est donné par :

$$\begin{aligned} \delta^E(X_{n+1}) &= E^{\theta | \hat{\lambda}, x_{n+1}}(\theta) = \frac{x_{n+1} + 1}{\hat{\lambda} + 1} \\ &= \frac{(x_{n+1} + 1)\bar{X}}{1 + \bar{X}}. \end{aligned}$$

Comparaison de l'estimation Bayésienne empirique avec l'estimation Bayésienne traditionnelle :

D'autre part, si on considère que λ est connu on aura :

$$\delta^B(X_{n+1}) = \frac{1 + x_{n+1}}{1 + \lambda}$$

Et le risque intégré de l'estimateur Bayésien est :

$$\begin{aligned} r(\pi, \delta^B) &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta^B(x)) f(x|\theta) \pi(\theta) dx d\theta \\ &= \int_0^{\infty} \sum_{x=0}^{\infty} \left(\frac{x+1}{\lambda+1} - \theta \right)^2 e^{-\theta} \frac{\theta^x}{x!} \lambda e^{-\lambda\theta} d\theta \\ &= \frac{1}{\lambda(1+\lambda)} \end{aligned}$$

Ainsi d'après (2.4), le risque intégré de l'estimateur Bayésien empirique est :

$$\begin{aligned} r(\pi, \delta^E) &= r(\pi, \delta^B) + \int (\delta(x)^E - \delta^B(x))^2 f(x) dx, \\ &= r(\pi, \delta^B) + E^{X_{n+1}|\hat{\lambda}}((\delta^E - \delta^B)^2) \\ &= r(\pi, \delta^B) + E \left(\left(\frac{(1 + X_{n+1})\bar{X}}{1 + \bar{X}} - \frac{(1 + X_{n+1})}{1 + \lambda} \right)^2 \right) \\ &= r(\pi, \delta^B) + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda} \right)^2 E((1 + X_{n+1})^2) \\ &= r(\pi, \delta^B) + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda} \right)^2 (Var((1 + X_{n+1})^2) + E^2(1 + X_{n+1})) \\ &= r(\pi, \delta^B) + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda} \right)^2 \left(\frac{1}{\lambda^2} + \frac{1}{\lambda} + \frac{1}{\lambda^2} + \frac{4}{\lambda} + 4 \right) \\ &= \frac{1}{\lambda(1 + \lambda)} + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda} \right)^2 \left(\frac{2}{\lambda^2} + \frac{5}{\lambda} + 4 \right) \end{aligned}$$

Maintenant, on fait une étude de simulation en donnant une valeur à λ et en estimant les paramètres selon le modèle de Poisson décrit ci-dessus.

Pour une valeur de $\lambda = 0.2$, on génère 1000 échantillons aléatoires de taille n pour tout n entre 2 et 1000 selon le modèle (2.18).

Par la suite, on calcule δ^E , l'estimateur Bayésien empirique de θ_{n+1} ainsi que le risque intégré correspondant ($r(\pi, \delta^E)$), sachant l'observation X_{n+1} afin de voir la performance de l'estimateur Bayésien empirique de θ_{n+1} , c'est-à-dire, la variation du risque intégré en fonction de la taille de l'échantillon n .

On obtient un graphique de $r(\pi, \delta^E)$ en fonction de n à la figure (2.1), qui est obtenu avec le programme A.3.

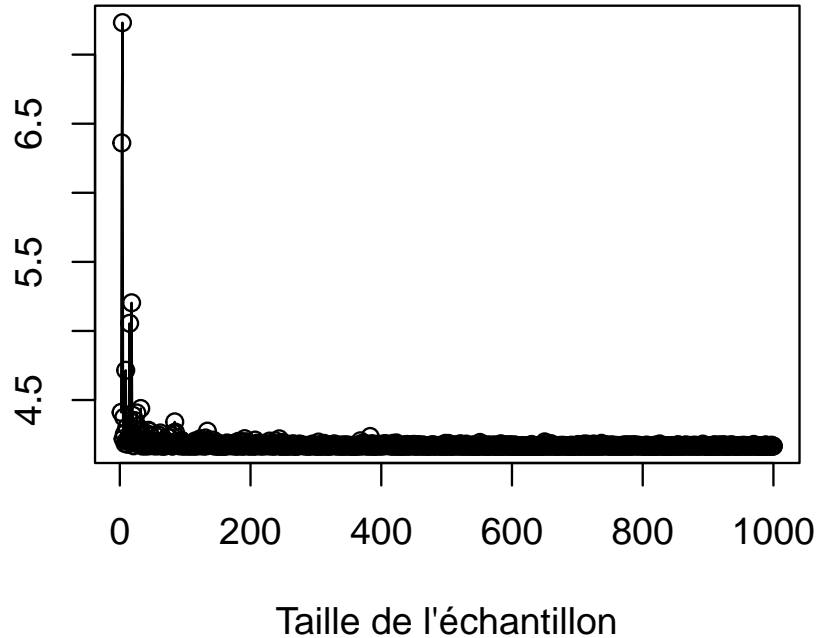


FIGURE 2.1 – Risque intégré de l'estimateur Bayésien empirique

Elle nous donne le comportement asymptotique de l'estimateur. Ainsi on remarque que la convergence du risque intégré de l'estimateur Bayésien empirique $r(\pi, \delta^E)$ asymptotique est très forte vers le risque de l'estimateur Bayésien $r(\pi, \delta^B)$ (qui est égal à $\frac{1}{\lambda(1+\lambda)} = 4.16$).

Le deuxième point d'intérêt est la comparaison de l'estimateur Bayésien empirique avec la vraie valeur de θ_{n+1} en fonction de la taille de l'échantillon. On présente cette comparaison dans la figure(2.2).

```
# Graphique de l'estimateur Bayésien empirique et theta
x <- seq(1, 50)
plot(x, BE10(50, 0.2)$b7, type = "l", xlab = "Taille de l'échantillon", ylab = " ")
points(x, BE10(50, 0.2)$be2, type = "l", lwd = 3)
```

L'estimateur Bayésien empirique est un trait foncé et celui de la vraie valeur de θ_{n+1} une ligne plus fine. Cette figure nous montre que les deux graphiques se rapproche quand la taille

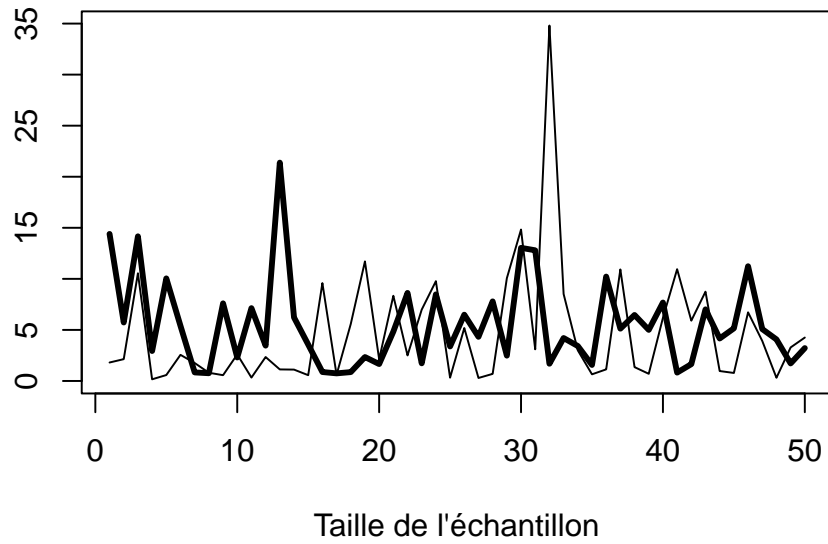


FIGURE 2.2 – L'estimateur Bayésien empirique et la vraie valeur de θ

le l'échantillons s'agrandit, un résultat souhaitable qui montre l'efficacité de l'estimateur Bayésien empirique.

Quelques résultats importants dans le cas poissonnien

Résultat de Fisher - problème des espèces manquantes

Alexander Corbet est un naturaliste qui a consacré deux ans en plein 2^{nde} guerre mondiale à capturer des spécimens d'espèces de papillons en Malaisie. Le tableau suivant présente certaines des données expérimentales qu'il a recueillies : Où x représente le nombre de fois qu'un certain papillon est capturé, et y_x le nombre d'espèces de papillons capturé x fois.

x	1	2	3	4	5	6	7	8	9	10	11	12
y	118	74	44	24	29	22	20	19	20	15	12	14
x	13	14	15	16	17	18	19	20	21	22	23	24
y	6	12	6	9	9	6	10	10	11	5	3	3

TABLE 2.2 – Données sur les papillons ; y nombre d'espèces observées x fois chacune au cours de deux années de capture.

D'après le tableau ci-dessus, il est évident qu'en deux ans, il y a 118 espèces de papillons qui sont extrêmement rares, n'ayant été capturées qu'une seule fois. En tout, 74 espèces de papillons ont été capturées deux fois au cours de cette période. En revanche, seulement 3 espèces de papillons sont très communes, ayant été capturées 24 fois au total sur deux ans. La différence la plus frappante ici est l'absence de la catégorie $x = 0$ de données, qui représente les espèces de papillons non capturées et non observées, ce que l'on appelle les "espèces manquantes".

Après deux ans de collecte de données, Corbet était curieux de savoir combien de papillons qui n'avaient pas encore été observés dans la nature. Il a donc posé cette question à Fisher, l'un des statisticiens les plus renommés du XXe siècle : "Si je reste en Malaisie pendant un an de plus, combien de nouvelles espèces de papillons pourrais-je capturer que je n'aurais jamais vues auparavant ?" À première vue, cette question peut sembler absurde et dépourvue de réponse.

Cependant, Fisher avait une réponse, qui, avec le recul, correspondrait indéniablement à la méthode Bayésienne empirique paramétrique.

Réponse de Fisher

La réponse de Fisher est la suivante :

Tout d'abord, nous utilisons i pour numéroter les espèces de papillons, $i = 1, 2, \dots, S$, où S représente le nombre total d'espèces de papillons observées et non observées. \mathbf{x}_i , est le nombre de fois où l'espèce i est capturée en une unité de temps t , suit une distribution de **Poisson** de paramètre θ_i

$$P\{x_i = x\} = e^{-\theta_i} \frac{\theta_i^x}{x!} \tag{2.21}$$

Le paramètre θ_i représente le **taux auquel l'espèce de papillon i est capturée par unité de temps** (dans ce cas, sur une période de deux ans).

Pour $x = 0, 1, 2, \dots$ la valeur estimé e_x de y_x est la somme des probabilités d'être vues exactement x fois dans la période initiale,

$$e_x = E\{y_x\} = \sum_{i=1}^S e^{-\theta_i} \frac{\theta_i^x}{x!} \quad (2.22)$$

$$= S \int_0^\infty \left[e^{-\theta} \frac{\theta^x}{x!} \right] \pi(\theta) d\theta. \quad (2.23)$$

En particulier, pour une nouvelle période de capture de longueur t , le nombre de fois que l'espèce i sera capturée (x_i) suit une distribution de Poisson $\mathcal{P}(\theta_i t)$, indépendante de son nombre de captures initial $x_i \sim \mathcal{P}(\theta_i)$.

$$P_i(t) = P\{x_i = 0 \text{ et } x_i(t) > 0\} = e^{-\theta_i} [1 - e^{-\theta_i t}] \quad (2.24)$$

Cette dernière équation représente la probabilité que l'espèce i n'ait pas été capturée lors de la phase initiale de capture, mais qu'elle le soit lors d'une phase de capture future t , devenant ainsi la nouvelle "espèce manquante" recherchée par Corbet.

La distribution a priori $\pi(\theta)$ peut être considérée comme une distribution empirique¹, où chaque paramètre θ_i a la même probabilité $1/S$.

Dans ce cas, nous pouvons exprimer le nombre moyen de nouvelles espèces qui peuvent être capturée dans les t unités de temps supplémentaires comme suit :

$$E(t) = \sum_i P_i(t) = S \int_0^\infty e^{-\theta} [1 - e^{-\theta t}] \pi(\theta) d\theta \quad (2.25)$$

Pour résoudre ce problème, Fisher a posé l'hypothèse que les paramètres du processus de Poisson θ_i de la distribution a priori $\pi(\theta)$ suivent une loi gamma

$$\begin{cases} x_i | \theta_i & \sim \text{Poisson}(\theta_i) \\ \theta_i | \nu, \sigma & \sim \text{Gamma}_{\nu, \sigma} \end{cases} \quad (2.26)$$

Avec :

$$\pi(\theta | \nu, \sigma) = \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} \quad \text{pour } \theta > 0 \quad (2.27)$$

On a $E(t)$ de (2.25) qui peut s'écrire comme suite :

$$E(t) = e_1 \left\{ 1 - (1 + \gamma t)^{-\nu} \right\} / (\gamma \nu) \quad (2.28)$$

où

$$\gamma = \frac{\sigma}{1 + \sigma}.$$

1. $\pi(\theta)$ est basée sur des observations empiriques plutôt que sur des connaissances théoriques préalables.

Démonstration.

pour $x = 1$ dans (formule.2.22)(espèces capturé une seule fois) en aura :

$$e_1 = S \int_0^\infty e^{-\theta} \pi(\theta) d\theta. \quad (2.29)$$

en combinant (2.29) avec (2.25) on obtient :

$$\begin{aligned} E(t) &= S \int_0^\infty e^{-\theta} [1 - e^{-\theta t}] \pi(\theta) d\theta \\ &= S \left[\int_0^\infty e^{-\theta} \pi(\theta) d\theta - \int_0^\infty e^{-\theta(1+t)} \pi(\theta) d\theta \right] \\ &= \frac{S \int_0^\infty e^{-\theta} \pi(\theta) d\theta \cdot \left[\int_0^\infty e^{-\theta} \pi(\theta) d\theta - \int_0^\infty e^{-\theta(1+t)} \pi(\theta) d\theta \right]}{\int_0^\infty \theta e^{-\theta} \pi(\theta) d\theta} \\ E(t) &= \frac{e_1 \left[\int_0^\infty e^{-\theta} \pi(\theta) d\theta - \int_0^\infty e^{-\theta(1+t)} \pi(\theta) d\theta \right]}{\int_0^\infty \theta e^{-\theta} \pi(\theta) d\theta} \quad (\star) \end{aligned}$$

En utilisant l'a priori $Gamma(\nu, \sigma)$ donnée dans (2.27) pour $\pi(\theta)$:

$$\pi(\theta) = \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)}$$

On obtient :

$$E(t) = \frac{e_1 \left[\int_0^\infty e^{-\theta} \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} d\theta - \int_0^\infty e^{-\theta(1+t)} \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} d\theta \right]}{\int_0^\infty \theta e^{-\theta} \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} d\theta}$$

D'abord, simplifions le numérateur :

$$\int_0^\infty e^{-\theta} \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} d\theta = \frac{1}{\sigma^\nu \Gamma(\nu)} \int_0^\infty \theta^{\nu-1} e^{-\theta(1+1/\sigma)} d\theta$$

Utilisons le changement de variable $\theta' = \theta(1 + 1/\sigma)$, d'où $d\theta = \frac{d\theta'}{1+1/\sigma}$:

$$\begin{aligned} \int_0^\infty \theta^{\nu-1} e^{-\theta(1+1/\sigma)} d\theta &= \int_0^\infty \left(\frac{\theta'}{1+1/\sigma} \right)^{\nu-1} e^{-\theta'} \frac{d\theta'}{1+1/\sigma} \\ &= \frac{1}{(1+1/\sigma)^\nu} \int_0^\infty \theta'^{\nu-1} e^{-\theta'} d\theta' \\ &= \frac{1}{(1+1/\sigma)^\nu} \Gamma(\nu) \int_0^\infty \frac{\theta'^{\nu-1} e^{-\theta'}}{\Gamma(\nu)} d\theta' \\ &= \frac{1}{(1+1/\sigma)^\nu} \Gamma(\nu) \end{aligned}$$

Donc,

$$\int_0^\infty e^{-\theta} \pi(\theta) d\theta = \frac{1}{(1 + 1/\sigma)^\nu} \frac{\Gamma(\nu)}{\sigma^\nu \Gamma(\nu)} = \frac{1}{\sigma^\nu (1 + 1/\sigma)^\nu}$$

Pour la deuxième intégrale du numérateur :

$$\int_0^\infty e^{-\theta(1+t)} \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} d\theta = \frac{1}{\sigma^\nu \Gamma(\nu)} \int_0^\infty \theta^{\nu-1} e^{-\theta(1+t+1/\sigma)} d\theta$$

Utilisons le changement de variable $\theta' = \theta(1 + t + 1/\sigma)$:

$$\int_0^\infty \theta^{\nu-1} e^{-\theta(1+t+1/\sigma)} d\theta = \frac{1}{(1 + t + 1/\sigma)^\nu} \Gamma(\nu)$$

Donc,

$$\int_0^\infty e^{-\theta(1+t)} \pi(\theta) d\theta = \frac{1}{(1 + t + 1/\sigma)^\nu} \frac{\Gamma(\nu)}{\sigma^\nu \Gamma(\nu)} = \frac{1}{\sigma^\nu (1 + t + 1/\sigma)^\nu}$$

Le dénominateur devient :

$$\int_0^\infty \theta e^{-\theta} \pi(\theta) d\theta = \int_0^\infty \theta e^{-\theta} \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^\nu \Gamma(\nu)} d\theta = \frac{1}{\sigma^\nu \Gamma(\nu)} \int_0^\infty \theta^\nu e^{-\theta(1+1/\sigma)} d\theta$$

Utilisons à nouveau $\theta' = \theta(1 + 1/\sigma)$:

$$\int_0^\infty \theta^\nu e^{-\theta(1+1/\sigma)} d\theta = \frac{1}{(1 + 1/\sigma)^{\nu+1}} \Gamma(\nu + 1)$$

Sachant que $\Gamma(\nu + 1) = \nu \Gamma(\nu)$,

$$\int_0^\infty \theta e^{-\theta} \pi(\theta) d\theta = \frac{1}{(1 + 1/\sigma)^{\nu+1}} \frac{\nu \Gamma(\nu)}{\sigma^\nu \Gamma(\nu)} = \frac{\nu}{\sigma^\nu (1 + 1/\sigma)^{\nu+1}}$$

En substituant les résultats des intégrales dans l'équation (*), nous obtenons :

$$E(t) = \frac{e_1 \left[\frac{1}{\sigma^\nu (1+1/\sigma)^\nu} - \frac{1}{\sigma^\nu (1+t+1/\sigma)^\nu} \right]}{\frac{\nu}{\sigma^\nu (1+1/\sigma)^{\nu+1}}}$$

En Remarque que σ^ν s'annule dans le numérateur et le dénominateur :

$$\begin{aligned} E(t) &= e_1 \frac{(1 + 1/\sigma)^{\nu+1} \left[\frac{1}{(1+1/\sigma)^\nu} - \frac{1}{(1+t+1/\sigma)^\nu} \right]}{\nu} \\ &= \frac{e_1}{\nu} \left[(1 + 1/\sigma)^{\nu+1} \left[(1 + 1/\sigma)^{-\nu} - (1 + t + 1/\sigma)^{-\nu} \right] \right] \\ &= \frac{e_1}{\nu} \left[(1 + 1/\sigma) - (1 + 1/\sigma)^{\nu+1} (1 + t + 1/\sigma)^{-\nu} \right] \end{aligned}$$

En utilisant $\gamma = \frac{\sigma}{1+\sigma}$, nous obtenons la formule(2.28) :

$$E(t) = e_1 \left\{ 1 - (1 + \gamma t)^{-\nu} \right\} / (\gamma \nu) \quad (2.30)$$

Certification. □

En prenant $e_1 = y_1$, l'estimation du maximum de vraisemblance donne

$$\hat{\nu} = 0.104 \text{ et } \hat{\sigma} = 89.79$$

La Figure (2.3) obtenu avec le programme A.6 montre que l'estimation paramétrique de Fisher $E(t)$ (formule.2.28), donne des résultats raisonnables pour $t > 1$, par exemple, $E(2) = 123$, pour une période future de capture de 2 unités (4 ans).

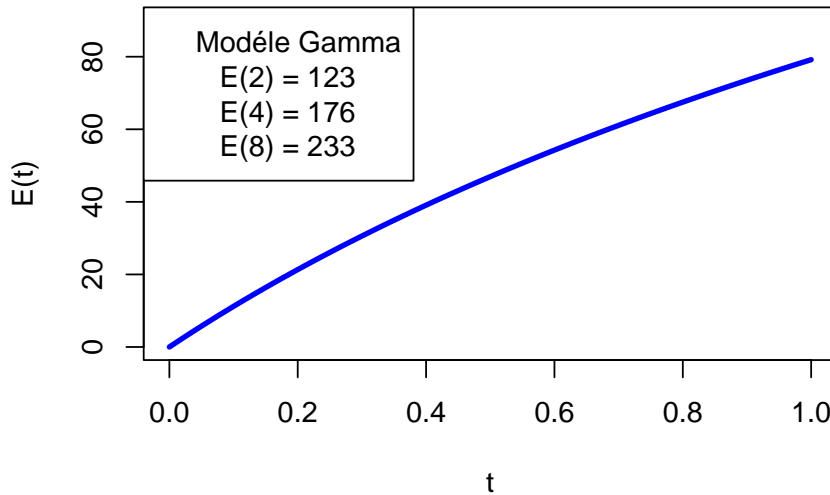


FIGURE 2.3 – Données des papillons ; nombre attendu de nouvelles espèces en fonction du temps supplémentaire de piégeage t

Formule de Robbins - Robbins' Poisson formula (Cas non-paramétrique)

Nous souhaitons estimer les paramètres $\theta_1, \theta_2, \dots, \theta_N$ à partir des observations x_1, x_2, \dots, x_N , qui suivent une loi de poisson associées à ces paramètres. La méthode du maximum de vraisemblance est souvent utilisée pour cette tâche, où chaque paramètre θ_i est estimé en utilisant uniquement l'observation x_i correspondante. Cependant, cette estimation ne prend en compte que les observations x_i , sans considérer les informations contenues dans les autres observations x_j pour $j \neq i$.

Cependant, dans son article [19], Robbins soulève un point historique crucial : les autres observations x_j pour ($j \neq i$) peuvent également être utilisées pour améliorer l'estimation du paramètre θ_i . Cela signifie que, plutôt que de baser l'estimation de θ_i uniquement sur x_i , nous pouvons utiliser l'ensemble des observations x_1, x_2, \dots, x_N pour obtenir une estimation plus précise de chaque θ_i .

Définition 22. Soit X_i des variables aléatoires de Poisson indépendants mais non identiquement distribués

$$X_i \stackrel{\text{ind}}{\sim} \text{Poi}(\theta_i) \quad i = 1, 2, \dots, N,$$

sont observés, avec les θ_i supposés tirés d'une distribution a priori $\pi(\theta)$ inconnue. En appliquant la règle de Bayes avec la densité de Poisson $e^{-\theta}\theta^x/x!$, Robbins a montré que l'espérance a posteriori est donné par

$$E\{\theta_i|x_i\} = \frac{(x_i + 1)y_{x_i+1}}{y_{x_i}}, \tag{2.31}$$

où y_{x_i} représente la fréquence ou le nombre d'occurrences de la valeur x_i dans l'ensemble des données observées.

Démonstration.

Commençons par ce qui a été mentionné précédemment : la distribution $f(x_i|\theta_i)$ est de type Poisson et peut s'écrire sous la forme $f(x_i|\theta_i) = e^{-\theta_i} \frac{\theta_i^{x_i}}{x_i!}$.

Avec $\theta \sim \pi(\theta)$ et $x | \theta \sim f(x|\theta)$, selon le théorème de Bayes (1), nous avons :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)}$$

où $f(x) = \int_{\Theta} f(x|\theta)\pi(\theta) d\theta$ est la densité marginale, Θ représentant l'ensemble de tous les paramètres possibles θ . En substituant la densité de Poisson $f(x|\theta) = e^{-\theta} \frac{\theta^x}{x!}$, nous obtenons :

$$f(x) = \int_0^{\infty} e^{-\theta} \frac{\theta^x}{x!} \pi(\theta) d\theta$$

$$\pi(\theta|x) = \frac{(e^{-\theta}\theta^x/x!)\pi(\theta)}{f(x)}$$

Donc $E\{\theta|x\}$ est donné par :

$$\begin{aligned} E\{\theta|x\} &= \int_0^\infty \theta \pi(\theta|x) d\theta = \frac{\int_0^\infty (e^{-\theta} \theta^{x+1}/x!) \pi(\theta) d\theta}{f(x)} \\ &= (x+1) \frac{\int_0^\infty (e^{-\theta} \theta^{x+1}/(x+1)!) \pi(\theta) d\theta}{f(x)} \\ &= (x+1) \frac{f(x+1)}{f(x)} \end{aligned}$$

Dans l'expérience Bayésienne elle-même, les valeurs de $f(x)$ et $f(x+1)$ sont inconnues, mais elles peuvent être estimées : ici, $f(x)$ et $f(x+1)$ sont estimés par y_x/N et y_{x+1}/N^2 respectivement, avec $N = \sum_x y_x$;

En substituant ces valeurs dans l'équation ci-dessus, nous obtenons naturellement la formule de Robbins que nous avons mentionnée précédemment :

$$E\{\theta|x\} = (x+1) \frac{y_{x+1}}{y_x}$$

Certification. □

Cela peut sembler étrange au premier abord, mais voici un exemple qui illustre le point de vue de Robbins :

Exemple 19. *Une compagnie d'assurance automobile Européenne a examiné le nombre de réclamations par client parmi ses 9461 clients de l'année dernière, enregistrant le nombre de réclamations et le nombre de clients associés à chaque niveau de réclamation. Comme le montrent les première et deuxième lignes du tableau ci-dessous, il est clair que sur 9461 conducteurs, 7840 ont conduit très prudemment l'année dernière sans faire aucune réclamation, tandis qu'à l'autre extrême, un conducteur a fait 7 réclamations en une seule année.*

Nous supposons que x_i est le nombre de réclamations à faire en une seule année par

Réclamations x_i	0	1	2	3	4	5	6	7
nbr clients y_{x_i}	7840	1317	239	42	14	4	4	1
Formule 2.31	0.168	0.363	0.527	1.33	1.43	6.00	1.75	
Gamma EMV	0.165	0.376	0.633	0.87	1.10	1.34	1.57	

TABLE 2.3 – Ligne1 : Nombre de réclamations au cours de la dernière année, Ligne2 : Nombre de clients correspondant au nombre de réclamations, Ligne3 : Formule de Robbins, Ligne4 : Utilisation de la distribution Gamma comme résultat d'une distribution a priori

l'assuré i , $i \in \{1, 2, 3, \dots, 9461\}$, suit une distribution de Poisson avec le paramètre θ_i .

$$P\{x_i = x\} = e^{-\theta_i} \frac{\theta_i^x}{x!} \tag{2.32}$$

2. Selon la loi des grands nombres, à mesure que le nombre d'observations N augmente, la fréquence relative y_x/N d'un événement x converge vers la probabilité réelle de cet événement.

Pour $x \in \{0, 1, 2, 3, \dots\}$, θ_i est la valeur attendue de x_i . Un bon client, du point de vue de la compagnie, a une petite valeur de θ_i , même si dans une année donnée, son nombre réel d'accidents x_i variera de manière aléatoire selon la densité de probabilité (2.32).

Afin de planifier les primes d'assurance automobile pour l'année à venir, la compagnie d'assurance souhaite être en mesure de prédire le nombre approximatif de réclamations que ses clients feront au cours de la prochaine année. Cela permettra d'ajuster les primes d'assurance à l'avance pour garantir le rendement souhaité de la compagnie.

La question est la suivante : pour les 7 840 conducteurs qui n'ont pas fait de réclamation l'année dernière, quel est le nombre moyen de réclamations qu'ils feront au cours de la nouvelle année ? Et pour les 1 317 conducteurs qui ont fait une réclamation l'année dernière, quel est le nombre moyen de réclamations qu'ils feront au cours de la nouvelle année ?

Notre objectif est d'estimer correctement $E\{\theta_i|x_i\}$ pour faire une prédiction, pour ce faire, On utilise la formule de Robbins (2.31).

Prenons un exemple simple : pour les 7 840 conducteurs qui n'ont pas fait de réclamation l'année dernière, selon la formule de Robbins (2.31), le nombre moyen de réclamations qu'ils feront au cours de la prochaine année est de

$$E\{\theta_1 | x_1 = 0\} = \frac{(0 + 1)1317}{7840} = \frac{1317}{7840} = 0.168$$

et ainsi de suite pour obtenir les données de la troisième ligne du tableau 2.3 .

REMARQUE.

1. Dans cette exemple, une simple ligne de données (y_x) fournit tous les éléments nécessaires pour estimer une statistique Bayésienne ($E\{\theta_i|x_i\}$). Cette méthode d'estimation, appelée estimation Bayésienne empirique, ne repose pas sur une distribution a priori, mais plutôt sur la fréquence des données observées. En un sens, cette approche est une méthode d'estimation complète basée sur la fréquence, confirmant ainsi le nom de cette méthode : Bayésienne empirique.
2. La formule de Robbins $E\{\theta|x\} = (x + 1)y_{x+1}/y_x$, montre que l'estimation d'un conducteur est influencée non seulement par sa propre expérience passée, mais aussi par celle d'autres conducteurs. Par exemple, pour un conducteur prudent (0 réclamation), son estimation dépend de l'expérience d'un conducteur légèrement moins prudent (1 réclamation).
3. La formule de Robbins est essentiellement non paramétrique, ce qui signifie qu'elle ne repose sur aucune forme paramétrique de distributions a priori. Dans la dérivation précédente, aucune estimation de cette distribution a priori $\pi(\theta)$ n'a été nécessaire.
4. Il est important de noter que dans l'exemple, l'estimation basée sur la formule de Robbins (y_x) peut poser des problèmes lorsque la valeur de x est petite (sur le côté droit du tableau (2.3). En revanche, une approche paramétrique donnera des résultats plus fiables en supposons que la densité a priori $\pi(\theta)$ des valeurs θ_k des

clients a une forme gamma, dont le details de l'obtention de la ligne 3 est bien expliqué dans [8] (pages 77,78, 89)

2.3.3 Estimation Bayésienne empirique pour la loi Normale

Les méthodes Bayésienne empiriques qu'on a vu ci-dessus sont principalement basées sur l'hypothèse que les observations sont distribuées par Poisson, tandis que l'estimateur de James-Stein est basé sur l'hypothèse que les observations sont normalement distribuées.

Estimateur de James–Stein

Une conséquence de la proposition(1) est que s'il existe un seul estimateur minimax, il est admissible. Réciproquement, si un estimateur minimax δ_0 du risque constant n'est pas admissible, il existe d'autres estimateur minimax δ qui le dominant uniformément, c'est-à-dire, que $R_\delta(\theta) \leq R_{\delta_0}(\theta)$ pour toute θ de Θ . Avant 1955, on pensait que si $X \sim N(\theta, I)$, alors l'estimateur des moindres carrés $\delta_0(X) = X$ était admissible, puisque c'est le seul estimateur minimax de risque constant. Stein (1955) a montré que ce résultat ne peut être vrai pour un vecteur de plus de deux composantes. Plus précisément, le risque de l'estimateur du type James-Stein $\hat{\mu}_i^{JS}$ qui peut être écrit aussi $\delta^{JS}(X)$ est plus petit ou égal au risque de $\delta_0(X) = X$ pour tout θ , ç-à-d :

$$R_{\delta^{JS}}(X) < R_{\delta_0}(X).$$

Définition 23.

Soit

$$x_i \mid \mu_i \stackrel{\text{ind}}{\sim} N(\mu_i, 1)$$

L'estimateur de James-Stein de μ_i est donné par :

$$\hat{\mu}_i^{JS} = \bar{x} + \hat{B}(x_i - \bar{x}) \quad \text{pour } i = 1, 2, \dots, N, \quad (2.33)$$

Avec \hat{B} le facteur de rétrécissement donné par :

$$\hat{B} = 1 - \frac{(N-3)}{\sum (x_i - \bar{x})^2} \quad \text{pour } N > 3$$

où N est le nombre de moyennes inconnues.

Démonstration.

Supposons que nous souhaitons estimer un seul paramètre μ à partir des observations x dans la modèle Bayésien suivant :

$$\begin{cases} x \mid \mu \sim \mathcal{N}(\mu, 1) \\ \mu \sim \mathcal{N}(M, A) \end{cases} \quad (2.34)$$

Dans ce cas, d'après la proposition (2.1.5), μ a une distribution a posteriori donnée par

$$\mu \mid x \sim \mathcal{N}(M + B(x - M), B), \quad B = \frac{A}{A + 1} \quad (2.35)$$

L'estimateur Bayésien de μ selon la proposition (2.6),

$$\hat{\mu}^{\text{Bayes}} = M + B(x - M) \quad (2.36)$$

En utilisons la propriété que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $E\{(X - \mu)^2\} = \sigma^2$. Dans ce cas, $\mu \mid x$ a une variance de B . Ainsi $\hat{\mu}^{\text{Bayes}}$ a une erreur quadratique moyenne attendue

$$E\{(\hat{\mu}^{\text{Bayes}} - \mu)^2\} = B, \quad (2.37)$$

comparé à 1 pour l'estimateur du maximum de vraisemblance (EMV) $\hat{\mu}^{\text{EMV}} = x$,

$$E\{(\hat{\mu}^{\text{EMV}} - \mu)^2\} = 1. \quad (2.38)$$

Si, par exemple, $A = 1$ dans (2.34), alors $B = 1/2$ et $\hat{\mu}^{\text{Bayes}}$ a seulement la moitié du risque de l'EMV.

Le même calcul s'applique à une situation où nous avons N versions indépendantes de (2.34), disons

$$\mu = (\mu_1, \mu_2, \dots, \mu_N)' \quad \text{et} \quad x = (x_1, x_2, \dots, x_N)', \quad (2.39)$$

avec

$$\begin{cases} x_i \mid \mu_i \sim \mathcal{N}(\mu_i, 1) \\ \mu_i \sim \mathcal{N}(M, A) \end{cases} \quad (2.40)$$

indépendamment pour $i = 1, 2, \dots, N$. Indiquons $\hat{\mu}_i^{\text{Bayes}}$ comme le vecteur des estimations individuelles de Bayes

$$\hat{\mu}_i^{\text{Bayes}} = M + B(x_i - M), \quad M = (M, M, \dots, M)', \quad (2.41)$$

et de même

$$\hat{\mu}_i^{\text{EMV}} = x_i.$$

En utilisant (2.37) le risque quadratique total de $\hat{\mu}_i^{\text{Bayes}}$ est

$$E\left\{\left\|\hat{\mu}_i^{\text{Bayes}} - \mu\right\|^2\right\} = E\left\{\sum_{i=1}^N (\hat{\mu}_i^{\text{Bayes}} - \mu_i)^2\right\} = N \cdot B \quad (2.42)$$

comparé à

$$E\left\{\left\|\hat{\mu}_i^{\text{EMV}} - \mu\right\|^2\right\} = N. \quad (2.43)$$

Encore une fois, $\hat{\mu}_i^{\text{Bayes}}$ a seulement B fois le risque de $\hat{\mu}_i^{\text{EMV}}$.

Ceci est bien si nous connaissons M et A (ou équivalentement M et B) dans (2.34). Si ce n'est pas le cas, nous pourrions essayer de les estimer à partir de $x = (x_1, x_2, \dots, x_N)$. Marginalement, d'après (2.10) Le modèle (2.40) donne

$$x_i \stackrel{\text{ind}}{\sim} \mathcal{N}(M, A + 1), \quad (2.44)$$

et

$$\mu_i | x_i \stackrel{\text{ind}}{\sim} \mathcal{N}(M + B(x_i - M), B), \quad B = \frac{A}{A + 1} \quad (2.45)$$

Alors $\hat{M} = \bar{x}$ est une estimation sans biais de M . De plus,

$$\hat{B} = 1 - \frac{(N - 3)}{S} \left[S = \sum_{i=1}^N (x_i - \bar{x})^2 \right] \quad (2.46)$$

estime sans biais B , tant que $N > 3$. L'estimateur de James-Stein est la version plug-in³ de (2.36),

$$\hat{\mu}_i^{\text{JS}} = \hat{M} + \hat{B}(x_i - \hat{M}) \quad \text{pour } i = 1, 2, \dots, N, \quad (2.47)$$

ou équivalamment

$$\hat{\mu}^{\text{JS}} = \hat{M} + \hat{B}(x - \hat{M}), \quad \text{avec } \hat{M} = (\hat{M}, \hat{M}, \dots, \hat{M})'.$$

À ce point, la terminologie "Bayes empirique" semble particulièrement appropriée : le modèle Bayésien (2.40) conduit à l'estimateur de Bayes (2.41), qui lui-même est estimé empiriquement (c'est-à-dire fréquemment) à partir de toutes les données x , puis appliqué aux cas individuels. Bien sûr, $\hat{\mu}^{\text{JS}}$ ne peut pas performer aussi bien que la règle Bayésienne réelle $\hat{\mu}^{\text{Bayes}}$, mais l'augmentation du risque est étonnamment modeste. Le risque quadratique moyen de $\hat{\mu}^{\text{JS}}$ sous le modèle (2.40) est

$$E \left\{ \left\| \hat{\mu}^{\text{JS}} - \mu \right\|^2 \right\} = NB + 3(1 - B). \quad (2.48)$$

Si, par exemple, $N = 20$ et $A = 1$, alors (2.48) égale 11.5, comparé avec le vrai risque Bayésien 10 de (2.42), beaucoup moins que le risque 20 pour $\hat{\mu}^{\text{EMV}}$. □

REMARQUE. Une version plus générale de (2.34) commence par

$$\begin{cases} x_i | \mu_i \sim \mathcal{N}(\mu_i, \sigma_0^2) \\ \mu_i \sim \mathcal{N}(M, A) \end{cases} \quad (2.49)$$

pour $i = 1, 2, \dots, N$, où M et A sont la moyenne et la variance de la distribution a priori. Ensuite, (2.44) et (2.45) deviennent

$$x_i \sim \mathcal{N}(M, A + \sigma_0^2) \quad \text{et} \quad \mu_i | x_i \sim \mathcal{N}(M + B(x_i - M), B\sigma_0^2) \quad (2.50)$$

pour $i = 1, 2, \dots, N$, où

3. un "plug-in estimator" est un estimateur qui utilise des valeurs estimées pour les paramètres inconnus dans une formule pour calculer une estimation finale.

$$B = \frac{A}{A + \sigma_0^2} \quad (2.51)$$

Maintenant, la règle de Bayes $\hat{\mu}_i^{(\text{Bayes})} = M + B(x_i - M)$ a l'estimateur Bayésien empirique de James-Stein

$$\hat{\mu}_i^{(\text{JS})} = \bar{x} + \left(1 - \frac{(N-3)\sigma_0^2}{S}\right)(x_i - \bar{x}), \quad (2.52)$$

avec $\bar{x} = \sum x_i/N$ et $S = \sum(x_i - \bar{x})^2$. Le théorème ci-dessous reste vrai, tel qu'énoncé, sauf que nous exigeons pour ce cas que $N \geq 4$.

Théorème 4.

Supposons que

$$x_i | \mu_i \sim N(\mu_i, 1) \quad (2.53)$$

indépendamment pour $i=1,2,\dots,N$, l'estimateur de James-Stein pour $N \geq 3$ domine partout l'estimateur du maximum de vraisemblance (EMV) en termes d'erreur quadratique moyenne, c'est-à-dire que pour tout choix de $\mu \in \mathbb{R}^N$,

$$E_\mu \left\{ \|\hat{\mu}^{JS} - \mu\|^2 \right\} < N = E_\mu \left\{ \|\hat{\mu}^{EMV} - \mu\|^2 \right\} \quad (2.54)$$

En termes de théorie des décisions, l'équation (2.54) indique que μ^{EMV} est inadmissible : son risque d'erreur quadratique totale dépasse celui de μ^{JS} quelle que soit la valeur de μ . C'est une forme fréquente forte de défaite pour μ^{EMV} , qui ne dépend pas des hypothèses Bayésiennes.

Démonstration.

Une preuve rapide du théorème commence par l'identité suivante :

$$(\hat{\mu}_i - \mu_i)^2 = (x_i - \hat{\mu}_i)^2 - (x_i - \mu_i)^2 + 2(\hat{\mu}_i - \mu_i)(x_i - \mu_i). \quad (2.55)$$

En sommant (2.55) pour $i = 1, 2, \dots, N$ et en prenant les espérances, on obtient

$$\mathbb{E}_\mu \left\{ \|\hat{\mu} - \mu\|^2 \right\} = \mathbb{E}_\mu \left\{ \|x - \hat{\mu}\|^2 \right\} - N + 2 \sum_{i=1}^N \text{cov}_\mu(\hat{\mu}_i, x_i), \quad (2.56)$$

où cov_μ indique la covariance sous $x \sim \mathcal{N}_N(\mu, I)$.

$$\text{cov}_\mu(\hat{\mu}_i, x_i) = \mathbb{E}_\mu \left\{ \frac{\partial \hat{\mu}_i}{\partial x_i} \right\} \quad (2.57)$$

Cela réduit (2.56) à

$$\mathbb{E}_\mu \left\{ \|\hat{\mu} - \mu\|^2 \right\} = \mathbb{E}_\mu \left\{ \|x - \hat{\mu}\|^2 \right\} - N + 2 \sum_{i=1}^N \mathbb{E}_\mu \left\{ \frac{\partial \hat{\mu}_i}{\partial x_i} \right\}. \quad (2.58)$$

Appliquer (2.58) à $\hat{\mu}^{(JS)}$ (2.33) donne

$$\mathbb{E}_{\mu} \left\{ \|\hat{\mu}^{(JS)} - \mu\|^2 \right\} = N - \mathbb{E}_{\mu} \left\{ \frac{(N-2)^2}{S} \right\} \quad (2.59)$$

avec $S = \sum x_i^2$. Le dernier terme de (2.59) est positif si N dépasse 2, prouvant le théorème. □

Exemple 20. *Joueur de Baseball*

Le tableau (2.4) montre les moyennes au bâton p_i (nombre de coups réussis divisé par le nombre d'essais) présentées pour 18 joueurs de baseball de la ligue majeure au début de la saison 1970. Les valeurs réelles μ_i sont considérées comme leurs moyennes pour le reste de la saison, comprenant environ 370 autres 'au bâton' chacun. Nous pouvons imaginer essayer de prédire les valeurs réelles à partir des résultats initiaux, en utilisant soit $\hat{\mu}_i^{(EMV)} = p_i$, soit les estimations de James-Stein (2.33).

TABLE 2.4 – Moyennes au bâton $p_i = \hat{\mu}_i^{(EMV)}$ pour 18 joueurs majeurs de la ligue au début de la saison 1970 ; μ_i sont les moyennes sur le reste de la saison. \hat{p}_i^{JS} Les estimations de James-Stein ; (Par coïncidence, $\hat{\mu}_i$ et μ_i ont tous deux une moyenne de 0.265 ; la moyenne de \hat{p}_i^{JS} doit être égale à celle de $\hat{\mu}_i^{(EMV)}$).

Nom	Coups/45	$\hat{\mu}_i^{(EMV)}$	μ_i	\hat{p}_i^{JS}
Clemente	18/45	.400	.346	.294
F. Robinson	17/45	.378	.298	.289
F. Howard	16/45	.356	.276	.285
Johnstone	15/45	.333	.222	.280
Berry	14/45	.311	.273	.275
Spencer	14/45	.311	.270	.275
Kessinger	13/45	.289	.263	.270
L. Alvarado	12/45	.267	.210	.266
Santo	11/45	.244	.269	.261
Swoboda	11/45	.244	.230	.261
Unser	10/45	.222	.264	.256
Williams	10/45	.222	.256	.256
Scott	10/45	.222	.303	.256
Petrocelli	10/45	.222	.264	.256
E. Rodriguez	10/45	.222	.226	.256
Campaneris	9/45	.200	.286	.252
Munson	8/45	.178	.316	.247
Alvis	7/45	.156	.200	.242
Moyenne Générale		.265	.265	.265

$\hat{\mu}_i^{(EMV)}$ indique le nombre moyen de coups sûrs réussis par les 18 frappeurs dans leurs

45 premières apparitions au bâton de la saison (45 apparitions représentent environ 1/10 des apparitions totales d'un frappeur dans une saison). Prenons le premier frappeur, Clemente, par exemple : sur un total de 45 apparitions au bâton, il a réussi 18 coups sûrs, donc sa moyenne au bâton est de $18/45 = 0,4$. Dans ce cas, la moyenne au bâton observée sur les 45 premières apparitions peut être utilisée comme une estimation très probable de la moyenne au bâton de ce frappeur pour le reste de la saison.

μ_i , sont les moyennes sur le reste de la saison (après leurs 45 premières apparitions au bâton) comprenant environ 370 autres apparitions au bâton chacun.

La dernière colonne du tableau présente les estimations des moyennes au bâton des joueurs obtenues à l'aide de l'estimateur de James-Stein \hat{p}_i^{JS} (formule.2.33).

On suppose que p_i (la moyenne au bâton lors des 45 premiers essais) suit une loi binomial,

$$p_i \sim Bi(45, \mu_i)/45^4. \quad (2.60)$$

Ce qui signifie que p représente la proportion moyenne de succès dans 45 essais indépendants, chaque essai ayant une probabilité μ_i de succès.

Pour procéder dans ce cas on utilise une approximation normale de (2.60),

$$p_i \sim \mathcal{N}(\mu_i, \sigma_0^2), \quad (2.61)$$

où σ_0^2 est la variance binomiale

$$\sigma_0^2 = \frac{\bar{p}(1 - \bar{p})}{45}, \quad (2.62)$$

avec $\bar{p} = 0.265$, la moyenne des valeurs p_i . En posant $x_i = p_i/\sigma_0$, en appliquant (2.47), et en transformant en arrière pour obtenir $\hat{p}_i^{JS} = \sigma_0 \hat{\mu}_i^{JS}$, on obtient les estimations de James-Stein

$$\hat{p}_i^{JS} = \bar{p} + \left[1 - \frac{(N-3)\sigma_0^2}{\sum (p_i - \bar{p})^2} \right] (p_i - \bar{p}).$$

Le code utiliser pour obtenir la dernière colonne du tableau est :

```
# Valeurs observées des moyennes au bâton (EMV)
p <- c(0.400, 0.378, 0.356, 0.333, 0.311, 0.311, 0.289, 0.267,
      0.244, 0.244, 0.222, 0.222, 0.222, 0.222, 0.222, 0.200,
      0.178, 0.156)
# Nombre de moyennes inconnues
N <- length(p) # = 18
```

4. En divisant par 45, on obtient la proportion moyenne de succès sur ces essais plutôt qu'un nombre entier de succès

2.4. PRINCIPALES DIFFÉRENCES ENTRE LES DEUX MÉTHODES BAYÉSIENNES ET BAYÉSIENNES EMPIRIQUE

```
# estimation pour sigma^2 :
s2 <- mean(p)*(1 - mean(p))/45
B=1 - ((N-3)*s2)/sum((x-mean(p))^2)
est_js= mean(p)+B*(p-mean(p))
```

Sur la base des données ci-dessus, nous pouvons simplement calculer l'erreur de prédiction de l'estimateur du maximum de vraisemblance et de l'estimateur de James-Stein, comme suit :

$$\sum_{i=1}^{18} (\hat{\mu}_i^{EMV} - \mu_i)^2 = 0.075545, \quad \sum_{i=1}^{18} (\hat{p}_i^{JS} - \mu_i)^2 = 0.02137776.$$

Nous voyons bien que l'estimateur de James-Stein \hat{p}_i^{JS} donne de meilleures estimations que $\hat{\mu}_i^{(EMV)}$.

2.3.4 L'importance des méthodes Bayésiennes empiriques en statistiques

On ne saurait trop insister sur la valeur des méthodes de Bayes empiriques lorsqu'il s'agit d'applications pratiques dans le domaine des statistiques. Elles sont particulièrement avantageuses pour faire des prédictions sur les observations futures et pour améliorer la précision des estimations des paramètres dans les modèles complexes. En incorporant des a priori basés sur des données, les analystes peuvent affiner leurs déductions, ce qui les rend plus robustes et plus fiables. Un autre avantage important est la réduction de la complexité informatique. Les approches Bayésienne traditionnelles peuvent parfois être coûteuses en termes de calcul, en particulier pour les grands ensembles de données. Les méthodes empiriques de Bayes, en estimant l'a priori à partir des données, simplifient souvent les calculs nécessaires, ce qui rend les analyses statistiques sophistiquées plus accessibles.

2.4 Principales différences Entre les deux méthodes Bayésiennes et Bayésiennes empirique

Pour comprendre en quoi la méthode Bayes empirique diffère de la méthode Bayes traditionnelle, il est utile de comparer leurs approches en matière de traitement des informations a priori :

1. **Spécification de l'a priori** : La méthode Bayésienne traditionnelle exige que l'information a priori soit entièrement spécifié par l'analyste, souvent sur la base de connaissances externes. Les estimateur Bayésien empiriques, en revanche, estiment l'information a priori à l'aide des données disponibles.
2. **Dépendance à l'égard des données** : L'analyse Bayésienne traditionnelle est indépendante des données utilisées pour estimer l'information a priori, alors que le Bayésien empirique utilise directement les données actuelles pour cette estimation.

2.4. PRINCIPALES DIFFÉRENCES ENTRE LES DEUX MÉTHODES BAYÉSIENNES ET BAYÉSIENNES EMPIRIQUE

3. **Flexibilité et application** : Les méthodes Bayésiennes empiriques offrent plus de flexibilité dans les applications où les connaissances préalables sont incertaines ou indisponibles. Elle permet également d'utiliser des méthodes adaptatives qui peuvent traiter de grands ensembles de données de manière plus efficace.

Bien que les deux méthodes s'inscrivent dans le cadre Bayésien, leurs approches distinctes du traitement des antécédents les distinguent et les rendent adaptées à différents types de problèmes statistiques.

Chapitre 3

ESTIMATION BAYÉSIENNE EMPIRIQUE DANS UN PROCESSUS AUTORÉGRESSIF D'ORDRE UN (AR(1))

3.1 Introduction

Les processus autorégressifs d'ordre 1 (AR(1)) sont des modèles largement utilisés en statistique pour modéliser des données temporelles, c'est-à-dire des séries chronologiques où les observations sont corrélées dans le temps. Dans un processus AR(1), chaque valeur de la série dépend linéairement de la valeur précédente, avec un terme aléatoire ajouté. Le coefficient d'autocorrelation est un paramètre clé qui mesure la dépendance entre les valeurs successives de la série. L'estimation précise de ce coefficient est cruciale, car il fournit un "bond d'information" qui permet de mieux comprendre la structure temporelle des données.

Dans ce chapitre, basé sur l'article [14], nous nous intéressons à l'estimation du coefficient d'autocorrelation en utilisant la méthode Bayésienne empirique. Cette approche permet de combiner les avantages de l'inférence Bayésienne, tout en estimant les hyperparamètres de manière adaptative à partir des données. Cela donne une meilleure précision dans l'estimation des paramètres du modèle AR(1), surtout dans les situations où les données sont limitées ou présentent une forte incertitude.

3.2 Présentation du modèle

Nous nous intéressons à un processus autorégressif d'ordre un (AR(1)), qui s'écrit comme suit :

$$y_t(j) = \phi_j y_{t-1}(j) + \varepsilon_t(j), \quad (3.1)$$

où $y_t(j)$ est l'observation sur la j -ème série temporelle au temps t , pour $t = 1, \dots, T$ et $j = 1, \dots, n$. Avec

$$\phi_j \stackrel{\text{ind}}{\sim} \mathcal{N}(x_j' \beta, \tau^2) \quad (3.2)$$

où β est un vecteur de paramètres de dimension $(p \times 1)$, et x_j sont des vecteurs de covariables connues de dimension $(p \times 1)$.

De plus, les erreurs $\varepsilon_t(j)$ sont supposées indépendantes et normalement distribuées :

$$\varepsilon_t(j) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2) \quad (3.3)$$

3.3 Estimation Bayésienne Empirique

L'approche Bayésienne empirique nécessite des hyperparamètres, comme $\mu = x_j' \beta$ et τ^2 , qui sont souvent inconnus. L'estimation Bayésienne empirique consiste à estimer ces hyperparamètres à partir des données, puis à calculer l'estimation de ϕ_j en remplaçant ces hyperparamètres par leurs estimations.

Souvent on débute par l'estimation des hyperparamètres à l'aide d'une méthode classique, suivie de la détermination de la loi a posteriori, puis de l'estimateur Bayésien empirique. D'autres auteurs, Commencent par identifier d'abord la loi a posteriori du paramètre ϕ , puis à calculer son estimateur Bayésien, enfin, ils procèdent à l'estimation fréquentiste de l'hyperparamètre β , ce qui a conduit à l'obtention de l'estimateur Bayésien empirique de ϕ . Il convient de noter que ces approches aboutissent à des résultats équivalents.

Étant donné les observations $\{y_t(j), t = 1, \dots, T, j = 1, \dots, n\}$, nous cherchons à estimer le paramètre ϕ .

Soit $y(j)$ le vecteur des observations pour la j -ème série, c'est-à-dire $y(j) = (y_1(j), \dots, y_T(j))'$, et soit $y(n+1)$ noté y , et x_{n+1} noté x . Nous supposons que $y_0(j) = 0$ pour chaque j .

3.3.1 La fonction de Vraisemblance

Vu que $\varepsilon_t(j) \sim \mathcal{N}(0, \sigma^2)$, et que $y_t(j) = \phi_j y_{t-1}(j) + \varepsilon_t(j)$, on obtient :

$$y_t(j) \mid \phi_j \sim \mathcal{N}(\phi_j y_{t-1}(j), \sigma^2)$$

la vraisemblance est donc donnée par :

$$\begin{aligned} p(y \mid \phi) &= \prod_{t=1}^T p(y_t(j) \mid \phi_j) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t(j) - \phi_j y_{t-1}(j))^2 \right\} \end{aligned}$$

3.3.2 La loi a priori

On sait que ϕ_j est une variable aléatoire qui suit une distribution normale :

$$\phi_j \sim \mathcal{N}(x'_j\beta, \tau^2)$$

ce qui donne la loi a priori :

$$\pi(\phi) \propto \exp \left\{ -\frac{1}{2\tau^2} (\phi_j - x'_j\beta)^2 \right\}.$$

3.3.3 Loi a posteriori

La procédure de Bayes est basée sur la densité a posteriori $\pi(\phi | y)$ qui est proportionnel à la vraisemblance est la loi a priori,

$$\pi(\phi|y) = \frac{p(y|\phi)\pi(\phi | \beta)}{p(y(1), \dots, y(n); \beta)}.$$

Avec $\pi(\phi|y)$ est la loi posteriori, $p(y|\phi)$ présente la vraisemblance, $\pi(\phi)$ est La loi a priori et $p(y(1), \dots, y(n); \beta)$ est la loi marginale conjointe.

Pour obtenir la loi a posteriori on procède de la manière suivante :

$$\begin{aligned} \pi(\phi|y) &\propto p(y|\phi)\pi(\phi) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t(j) - \phi_j y_{t-1}(j))^2 \right\} \cdot \exp \left\{ -\frac{1}{2\tau^2} (\phi_j - x'_j\beta)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\frac{1}{\sigma^2} \sum_{t=1}^T (y_t(j) - \phi_j y_{t-1}(j))^2 + \frac{1}{\tau^2} (\phi_j - x'_j\beta)^2 \right) \right\} \end{aligned}$$

Développons le premier terme :

$$\begin{aligned} \frac{1}{\sigma^2} \sum_{t=1}^T (y_t(j) - \phi_j y_{t-1}(j))^2 &= \frac{1}{\sigma^2} \sum_{t=1}^T (y_t(j)^2 - 2\phi_j y_t(j) y_{t-1}(j) + \phi_j^2 y_{t-1}(j)^2) \\ &= \frac{1}{\sigma^2} \left(\sum_{t=1}^T y_t(j)^2 - 2\phi_j \sum_{t=1}^T y_t(j) y_{t-1}(j) + \phi_j^2 \sum_{t=1}^T y_{t-1}(j)^2 \right) \end{aligned}$$

Insérons ceci dans la dernière formule de $\pi(\phi|y)$:

$$\pi(\phi | y) \propto \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{t=1}^T y_t(j)^2 - 2\phi_j \sum_{t=1}^T y_t(j) y_{t-1}(j) + \phi_j^2 \sum_{t=1}^T y_{t-1}(j)^2 \right) + \frac{1}{2\tau^2} (\phi_j - x'_j\beta)^2 \right) \quad (3.4)$$

Développons le deuxième terme :

$$\frac{1}{2\tau^2} (\phi_j - x'_j\beta)^2 = \frac{1}{2\tau^2} (\phi_j^2 - 2\phi_j x'_j\beta + (x'_j\beta)^2)$$

En combinant les deux termes, nous avons :

$$\pi(\phi | y) \propto \exp\left(-\frac{1}{2}\left(A\phi_j^2 - 2B\phi_j + C\right)\right) \quad (3.5)$$

avec :

$$\begin{aligned} A &= \frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}(j)^2 + \frac{1}{\tau^2} \\ B &= \frac{1}{\sigma^2} \sum_{t=1}^T y_t(j)y_{t-1}(j) + \frac{1}{\tau^2} x'\beta \\ C &= \frac{1}{\sigma^2} \sum_{t=1}^T y_t(j)^2 + \frac{1}{\tau^2} (x'\beta)^2 \end{aligned}$$

Développons l'équation (3.21)

$$\begin{aligned} \pi(\phi | y) &\propto \exp\left(-\frac{1}{2}\left(A\phi^2 - 2B\phi + C\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left(A\phi^2 - 2B\phi\right) - \frac{1}{2}C\right) \\ &\propto \exp\left(-\frac{1}{2}\left(A\phi^2 - 2B\phi\right)\right) \\ &\propto \exp\left(-\frac{1}{2}A\left(\phi^2 - 2\frac{B}{A}\phi\right)\right) \\ &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi^2 - 2\frac{B}{A}\phi + \left(\frac{B}{A}\right)^2 - \left(\frac{B}{A}\right)^2\right)\right) \\ &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi^2 - 2\frac{B}{A}\phi + \left(\frac{B}{A}\right)^2\right) + \frac{1}{2}A\left(\frac{B}{A}\right)^2\right) \\ &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi^2 - 2\frac{B}{A}\phi + \left(\frac{B}{A}\right)^2\right)\right) \\ \pi(\phi | y) &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi - \frac{B}{A}\right)^2\right) \end{aligned}$$

qui est la densité d'une loi normale de moyenne

$$\frac{B}{A} = \frac{\frac{1}{\sigma^2} \sum_{t=1}^T y_t(j)y_{t-1}(j) + \frac{1}{\tau^2} x'\beta}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}(j)^2 + \frac{1}{\tau^2}},$$

et de variance

$$\frac{1}{A} = \frac{1}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}(j)^2 + \frac{1}{\tau^2}}.$$

3.3.4 Estimateur Bayésien

l'estimation Bayésienne de ϕ par rapport à la fonction coût quadratique est donnée par la moyenne a posteriori.

La moyenne a posteriori

On voit bien que la moyenne a posteriori δ est la suivante

$$\delta = \frac{B}{A} = \frac{\frac{1}{\sigma^2} \sum_{t=1}^T y_t(j)y_{t-1}(j) + \frac{1}{\tau^2}(x'\beta)}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}(j)^2 + \frac{1}{\tau^2}}$$

On multiplie et on divise par $\sigma^2\tau^2$, on obtient :

$$\begin{aligned} \delta &= \frac{\frac{\sigma^2\tau^2}{\sigma^2} \sum_{t=1}^T y_t(j)y_{t-1}(j) + \frac{\sigma^2\tau^2}{\tau^2}(x'\beta)}{\frac{\sigma^2\tau^2}{\sigma^2} \sum_{t=1}^T y_{t-1}(j)^2 + \frac{\sigma^2\tau^2}{\tau^2}} \\ &= \frac{\tau^2 \sum_{t=1}^T y_t(j)y_{t-1}(j) + \sigma^2(x'\beta)}{\tau^2 \sum_{t=1}^T y_{t-1}(j)^2 + \sigma^2} \\ &= \frac{\sigma^2}{\tau^2 \sum_{t=1}^T y_{t-1}(j)^2 + \sigma^2} (x'\beta) + \frac{\tau^2 \sum_{t=1}^T y_t(j)y_{t-1}(j)}{\tau^2 \sum_{t=1}^T y_{t-1}(j)^2 + \sigma^2} \end{aligned}$$

Développons le deuxième terme, en passons de la série générale (j) à la série nouvellement observée ($n+1$)

$$\frac{\tau^2 \sum_{t=1}^T y_t(n+1)y_{t-1}(n+1)}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)} = \frac{\tau^2}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)} \cdot \sum_{t=1}^T y_t(n+1)y_{t-1}(n+1).$$

d'autre part, on a cette égalité

$$1 = \frac{\sigma^2}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)} + \frac{\tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)}.$$

$$\Rightarrow \frac{\tau^2}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)} = \left(1 - \frac{\sigma^2}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)}\right) \cdot \frac{1}{\sum_{t=1}^T y_{t-1}^2(n+1)}.$$

En utilisant cette relation, nous pouvons exprimer le 2^{ème} terme comme suit :

$$\frac{\tau^2 \sum_{t=1}^T y_t(n+1)y_{t-1}(n+1)}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)} = \left(1 - \frac{\sigma^2}{\sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(n+1)}\right) \cdot \frac{\sum_{t=1}^T y_t(n+1)y_{t-1}(n+1)}{\sum_{t=1}^T y_{t-1}^2(n+1)} \quad (3.6)$$

On pose

$$a = \sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}(n+1)^2,$$

on obtient

$$\delta = \left(\frac{\sigma^2}{a} \right) (x' \beta) + \left(1 - \frac{\sigma^2}{a} \right) \hat{\phi} \quad (3.7)$$

Avec

$$\hat{\phi} = \frac{\sum_{t=1}^T y_t(n+1)y_{t-1}(n+1)}{\sum_{t=1}^T y_{t-1}^2(n+1)}. \quad (3.8)$$

REMARQUE. Les résultats (3.7) et (3.8) et (a) peuvent être obtenus en utilisant des résultats standards de régression linéaire Bayésienne avec une reformulation mineure du modèle.

Tout d'abord, supposons que σ^2 et τ^2 sont connus. Comme δ dépend du paramètre inconnu β , nous procédons à le remplacer par une estimation classique pour obtenir l'estimateur Bayésien empirique $\hat{\delta}_n$ de ϕ .

3.3.5 Estimateur Bayésien empirique

La densité marginale conjointe des n premiers vecteurs d'échantillons $\{y(j), j = 1, \dots, n\}$ est donnée par :

$$p(y(1), \dots, y(n); \beta) = \prod_{j=1}^n p(y(j); \beta), \quad (3.9)$$

où :

$$\begin{aligned} p(y(j); \beta) &= \left(\frac{1}{2\pi\sigma^2} \right)^{T/2} \left(\frac{\sigma^2}{a_j} \right)^{1/2} \exp \left\{ -\frac{1}{2} a_j^{-1} \sum_{t=1}^T [y_t(j) - (x'_j \beta) y_{t-1}(j)]^2 \right\} \\ &\times \exp \left\{ -\frac{\tau^2}{2\sigma^2 a_j} \left[\sum_{t=1}^T y_t^2(j) y_{t-1}^2(j) - \left(\sum_{t=1}^T y_t(j) y_{t-1}(j) \right)^2 \right] \right\}. \end{aligned} \quad (3.10)$$

Estimateur du maximum de vraisemblance

L'estimation du maximum de vraisemblance $\hat{\beta}_n$ de β à partir de l'équation (3.9) est donnée par :

$$\hat{\beta}_n = \frac{\sum_{j=1}^n \sum_{t=1}^T y_t(j) y_{t-1}(j) x_j x'_j a_j^{-1}}{\sum_{j=1}^n \sum_{t=1}^T y_{t-1}^2(j) (x_j x'_j) a_j^{-1}}, \quad (3.11)$$

où $a_j = \sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2(j)$.

On suppose que l'inverse de la matrice utilisée dans l'équation (3.11) existe.

Dans le cas d'une estimation de maximum de vraisemblance, la vraisemblance totale pour tous les $j = 1, \dots, n$ est le produit des vraisemblances pour chaque j , ce qui est donné par :

$$L(\beta) = \prod_{j=1}^n p(y(j); \beta).$$

En prenant le logarithme de $L(\beta)$, on obtient :

$$\log L(\beta) = \sum_{j=1}^n \log p(y(j); \beta).$$

En utilisant l'expression de $p(y(j); \beta)$, cela donne :

$$\begin{aligned} \log L(\beta) = & \sum_{j=1}^n \left[-\frac{T}{2} (\log(2\pi\sigma^2)) + \frac{1}{2} (\log(\sigma^2 a_j^{-1})) \right] \\ & - \sum_{j=1}^n \frac{1}{2} a_j^{-1} \sum_{t=1}^T [y_t(j) - (x'_j \beta) y_{t-1}(j)]^2 - \sum_{j=1}^n \frac{\tau^2}{2\sigma^2 a_j} \left[\sum_{t=1}^T y_t^2(j) \sum_{t=1}^T y_{t-1}^2(j) - \left(\sum_{t=1}^T y_t(j) y_{t-1}(j) \right)^2 \right] \end{aligned}$$

Pour trouver l'estimateur du maximum de vraisemblance, il faut maximiser la log-vraisemblance par rapport à β . Cela revient à résoudre l'équation suivante, obtenue en dérivant $\log L(\beta)$ par rapport à β et en l'égalisant à zéro :

$$\frac{\partial}{\partial \beta} \log L(\beta) = 0.$$

Concentrons-nous sur la partie dépendant de β dans la log-vraisemblance :

$$- \sum_{j=1}^n \frac{1}{2} a_j^{-1} \sum_{t=1}^T [y_t(j) - (x'_j \beta) y_{t-1}(j)]^2 \quad (3.12)$$

En dérivant (3.25) par rapport à β , on obtient :

$$\begin{aligned} \frac{\partial}{\partial \beta} \log L(\beta) = & - \sum_{j=1}^n \frac{1}{2} a_j^{-1} \sum_{t=1}^T \frac{\partial}{\partial \beta} [y_t(j) - (x'_j \beta) y_{t-1}(j)]^2 \\ = & - \frac{1}{2} \sum_{j=1}^n a_j^{-1} \sum_{t=1}^T \frac{\partial}{\partial \beta} [y_t^2(j) - 2y_t(j)(x'_j \beta) y_{t-1}(j) + (x'_j \beta)^2 y_{t-1}^2(j)] \\ = & - \frac{1}{2} \sum_{j=1}^n a_j^{-1} \sum_{t=1}^T [-2y_t(j) y_{t-1}(j) x_j + 2(x'_j \beta) y_{t-1}^2(j) x_j] \\ = & \sum_{j=1}^n a_j^{-1} \sum_{t=1}^T [y_t(j) y_{t-1}(j) x_j - (x'_j \beta) y_{t-1}^2(j) x_j] \\ = & \sum_{j=1}^n a_j^{-1} \sum_{t=1}^T [y_t(j) - (x'_j \beta) y_{t-1}(j)] y_{t-1}(j) x_j \end{aligned}$$

En l'égalisant à zéro, on obtient l'équation suivante :

$$\begin{aligned}
 & \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} \left[y_t(j) - (x'_j \beta) y_{t-1}(j) \right] y_{t-1}(j) x_j = 0. \\
 \implies & \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} \left[y_t(j) y_{t-1}(j) x_j - (x'_j \beta) y_{t-1}^2(j) x_j \right] = 0. \\
 \implies & \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} y_t(j) y_{t-1}(j) x_j - \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} (x'_j \beta) y_{t-1}^2(j) x_j = 0. \\
 \implies & \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} (x'_j \beta) y_{t-1}^2(j) x_j = \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} y_t(j) y_{t-1}(j) x_j. \\
 \implies & \beta \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} y_{t-1}^2(j) (x_j x'_j) = \sum_{j=1}^n \sum_{t=1}^T a_j^{-1} y_t(j) y_{t-1}(j) x_j. \\
 \implies & \hat{\beta}_n = \frac{\sum_{j=1}^n \sum_{t=1}^T a_j^{-1} y_t(j) y_{t-1}(j) x_j}{\sum_{j=1}^n \sum_{t=1}^T a_j^{-1} y_{t-1}^2(j) (x_j x'_j)}.
 \end{aligned}$$

Donc l'estimateur Bayésienne empirique $\hat{\delta}_n$ de ϕ Ai donné en remplaçant β par $\hat{\beta}_n$ dans l'équation (3.7) par la formule :

$$\hat{\delta}_n = \left(\frac{\sigma^2}{a} \right) x' \hat{\beta}_n + \left(1 - \frac{\sigma^2}{a} \right) \hat{\phi}, \quad (3.13)$$

Avec

$$a = \sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}(n+1)^2,$$

et

$$\hat{\phi} = \frac{\sum_{t=1}^T y_t(n+1) y_{t-1}(n+1)}{\sum_{t=1}^T y_{t-1}^2(n+1)}. \quad (3.14)$$

3.4 Propriétés asymptotiques

L'étude explore les propriétés asymptotiques de l'estimateur Bayésien empirique. Elle montre que l'estimateur converge en probabilité vers l'estimation Bayésienne optimale et qu'il est asymptotiquement efficace, c'est-à-dire qu'il minimise le risque moyen (erreur quadratique).

Bien que l'approche Bayésienne empirique présente des avantages pratiques pour l'estimation de processus AR(1), elle repose principalement sur des méthodes fréquentistes pour l'estimation des hyperparamètres, ce qui peut être considéré comme un compromis entre la rigueur théorique Bayésienne et les exigences pratiques des applications.

L'article de YOUNG-WON KIM et I.V. BASAWA [14] conclut que l'estimation Bayésienne empirique offre une alternative efficace aux méthodes d'estimation classiques, avec une convergence asymptotique robuste, tout en intégrant les caractéristiques spécifiques des séries temporelles étudiées.

3.5 Estimation du coefficient d'un AR(1) avec diverses approches

Nous nous intéressons à l'estimation du coefficient autorégressif d'un processus autorégressif d'ordre (1) par diverses approches : Estimateur de maximum de vraisemblance, estimateur Bayésien et estimateur Bayésien empirique.

3.5.1 Présentation du modèle

Nous nous intéressons à un processus autorégressif d'ordre un (AR(1)), qui s'écrit comme suit :

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad (3.15)$$

où y_t est l'observation de la série temporelle au temps t , pour $t = 1, \dots, T$, et les erreurs ε_t sont supposées indépendantes et normalement distribuées :

$$\varepsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2). \quad (3.16)$$

et ϕ le coefficient autorégressif que l'on souhaite estimer.

3.5.2 Estimateur de maximum de vraisemblance

Pour trouver l'estimateur du maximum de vraisemblance de ϕ dans le modèle (3.15) nous allons calculer la fonction de vraisemblance.

Sachant que chaque y_t suit une loi normale conditionnelle :

$$y_t | y_{t-1}, \phi \sim \mathcal{N}(\phi y_{t-1}, \sigma^2).$$

La fonction de vraisemblance jointe pour les T observations est alors :

$$L(Y | \phi) = \prod_{t=1}^T f(y_t | y_{t-1}, \phi),$$

où chaque $f(y_t | y_{t-1}, \phi)$ est une densité normale donnée par :

$$f(y_t | y_{t-1}, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \phi y_{t-1})^2}{2\sigma^2}\right).$$

La log-vraisemblance correspondante est :

3.5. ESTIMATION DU COEFFICIENT D'UN AR(1) AVEC DIVERSES APPROCHES

$$\log L(Y | \phi) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \phi y_{t-1})^2.$$

En maximisant cette log-vraisemblance par rapport à ϕ , nous avons

$$\begin{aligned} \sum_{t=1}^T (y_t - \phi y_{t-1})^2 &= \sum_{t=1}^T (y_t^2 - 2\phi y_t y_{t-1} + \phi^2 y_{t-1}^2). \\ \frac{d}{d\phi} \left(\sum_{t=1}^T (y_t - \phi y_{t-1})^2 \right) &= -2 \sum_{t=1}^T y_t y_{t-1} + 2\phi \sum_{t=1}^T y_{t-1}^2. \\ -2 \sum_{t=1}^T y_t y_{t-1} + 2\phi \sum_{t=1}^T y_{t-1}^2 &= 0. \end{aligned}$$

Ce qui donne l'estimateur du maximum de vraisemblance de ϕ par :

$$\hat{\phi}^{EMV} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2}. \quad (3.17)$$

3.5.3 Estimateur Bayésien

Loi a priori

Dans ce cas on suppose que ϕ est une variable aléatoire dont la loi a priori est

$$\phi \stackrel{\text{ind}}{\sim} \mathcal{N}(\mu, \tau^2), \quad (3.18)$$

où μ et τ^2 sont connus.

Ce qui donne la densité a priori :

$$\pi(\phi) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2} (\phi - \mu)^2 \right\} \quad (3.19)$$

Vraisemblance

Étant donné que $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, et que $y_t = \phi y_{t-1} + \varepsilon_t$, on a :

$$y_t | \phi \sim \mathcal{N}(\phi y_{t-1}, \sigma^2)$$

La vraisemblance $f(y | \phi)$ est donc donnée par :

$$\begin{aligned} f(y | \phi) &= \prod_{t=1}^T f(y_t | \phi) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \phi y_{t-1})^2 \right\} \end{aligned}$$

Loi a posteriori

Pour obtenir la loi a posteriori, nous combinons la vraisemblance et la loi a priori :

$$\begin{aligned}\pi(\phi|y) &\propto f(y|\phi)\pi(\phi) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^T(y_t - \phi y_{t-1})^2\right\} \cdot \exp\left\{-\frac{1}{2\tau^2}(\phi - \mu)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^T(y_t - \phi y_{t-1})^2 + \frac{1}{\tau^2}(\phi - \mu)^2\right)\right\}\end{aligned}$$

Développons le premier terme :

$$\begin{aligned}\frac{1}{\sigma^2}\sum_{t=1}^T(y_t - \phi y_{t-1})^2 &= \frac{1}{\sigma^2}\sum_{t=1}^T(y_t^2 - 2\phi y_t y_{t-1} + \phi^2 y_{t-1}^2) \\ &= \frac{1}{\sigma^2}\left(\sum_{t=1}^T y_t^2 - 2\phi \sum_{t=1}^T y_t y_{t-1} + \phi^2 \sum_{t=1}^T y_{t-1}^2\right)\end{aligned}$$

Insérons ceci dans la dernière formule de la loi a posteriori :

$$\pi(\phi | y) \propto \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{t=1}^T y_t^2 - 2\phi \sum_{t=1}^T y_t y_{t-1} + \phi^2 \sum_{t=1}^T y_{t-1}^2\right) - \frac{1}{2\tau^2}(\phi - \mu)^2\right) \quad (3.20)$$

Développons le deuxième terme :

$$\frac{1}{2\tau^2}(\phi - \mu)^2 = \frac{1}{2\tau^2}(\phi^2 - 2\phi\mu + \mu^2)$$

En combinant les deux termes, nous avons :

$$\pi(\phi | y) \propto \exp\left(-\frac{1}{2}(A\phi^2 - 2B\phi + C)\right) \quad (3.21)$$

avec :

$$\begin{aligned}A &= \frac{1}{\sigma^2}\sum_{t=1}^T y_{t-1}^2 + \frac{1}{\tau^2} \\ B &= \frac{1}{\sigma^2}\sum_{t=1}^T y_t y_{t-1} + \frac{1}{\tau^2}\mu \\ C &= \frac{1}{\sigma^2}\sum_{t=1}^T y_t^2 + \frac{1}{\tau^2}\mu^2\end{aligned}$$

Développons l'équation (3.21) :

$$\begin{aligned}
 \pi(\phi | y) &\propto \exp\left(-\frac{1}{2}\left(A\phi^2 - 2B\phi + C\right)\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(A\phi^2 - 2B\phi\right) - \frac{1}{2}C\right) \\
 &\propto \exp\left(-\frac{1}{2}\left(A\phi^2 - 2B\phi\right)\right) \\
 &\propto \exp\left(-\frac{1}{2}A\left(\phi^2 - 2\frac{B}{A}\phi\right)\right) \\
 &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi^2 - 2\frac{B}{A}\phi + \left(\frac{B}{A}\right)^2 - \left(\frac{B}{A}\right)^2\right)\right) \\
 &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi^2 - 2\frac{B}{A}\phi + \left(\frac{B}{A}\right)^2\right)\right) \\
 &\propto \exp\left(-\frac{1}{2\frac{1}{A}}\left(\phi - \frac{B}{A}\right)^2\right)
 \end{aligned}$$

On reconnaît la densité d'une loi normale de moyenne $\frac{B}{A}$ et de variance $\frac{1}{A}$. Par conséquent, la loi a posteriori est :

$$\phi | y \sim \mathcal{N}\left(\frac{\frac{1}{\sigma^2} \sum_{t=1}^T y_t y_{t-1} + \frac{1}{\tau^2} \mu}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}^2 + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}^2 + \frac{1}{\tau^2}}\right)$$

L'estimateur Bayésien sous la fonction coût quadratique et la moyenne a posteriori qui est donné par la formule suivante :

$$\hat{\phi}^B = \frac{B}{A} = \frac{\frac{1}{\sigma^2} \sum_{t=1}^T y_t y_{t-1} + \frac{1}{\tau^2} \mu}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}^2 + \frac{1}{\tau^2}} \quad (3.22)$$

3.5.4 L'estimateur Bayésien Empirique

Dans ce cas nous supposons que ϕ est une variable aléatoire dont on assigne une loi à priori gaussienne avec une moyenne inconnue μ et une variance τ^2 connue.

$$\phi \sim \mathcal{N}(\mu, \tau^2).$$

Pour cela nous devons estimer μ par la méthode d'estimation du maximum de vraisemblance en utilisant la densité marginale $m(y_t | \mu)$.

Estimateur du maximum de vraisemblance

En prenant

$$\begin{aligned} f(y | \phi) &= \prod_{t=1}^T f(y_t | \phi) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \phi y_{t-1})^2\right\} \end{aligned}$$

et

$$\pi(\phi) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\phi - \mu)^2\right\}, \quad (3.23)$$

nous avons

$$\begin{aligned} m(y_t | \mu) &= \int f(y | \phi) \pi(\phi) d\phi \\ &= \int \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \phi y_{t-1})^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\phi - \mu)^2\right\} d\phi \end{aligned}$$

En développant cela donnera

$$\begin{aligned} m(y_t | \mu) &= \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} \left(\frac{\sigma^2}{a}\right)^{1/2} \exp\left\{-\frac{1}{2}a^{-1} \sum_{t=1}^T [y_t - (\mu)y_{t-1}]^2\right\} \\ &\quad \times \exp\left\{-\frac{\tau^2}{2\sigma^2 a} \left[\sum_{t=1}^T y_t^2 y_{t-1}^2 - \left(\sum_{t=1}^T y_t y_{t-1}\right)^2\right]\right\}, \end{aligned} \quad (3.24)$$

où $a = \sigma^2 + \tau^2 \sum_{t=1}^T y_{t-1}^2$.

En prenant le logarithme de (3.24), on obtient :

$$\begin{aligned} \log m(y_t | \mu) &= \left[-\frac{T}{2} (\log(2\pi\sigma^2)) + \frac{1}{2} (\log(\sigma^2 a^{-1}))\right] \\ &\quad - \frac{1}{2} a^{-1} \sum_{t=1}^T [y_t - \mu y_{t-1}]^2 - \frac{\tau^2}{2\sigma^2 a} \left[\sum_{t=1}^T y_t^2 y_{t-1}^2 - \left(\sum_{t=1}^T y_t y_{t-1}\right)^2\right] \end{aligned}$$

Pour trouver l'estimateur du maximum de vraisemblance de μ , il faut maximiser cette dernière équation par rapport à μ . Cela revient à résoudre l'équation suivante :

$$\frac{\partial}{\partial \mu} \log m(y_t | \mu) = 0.$$

Concentrons-nous sur la partie dépendante de μ :

$$-\frac{1}{2}a^{-1} \sum_{t=1}^T [y_t - \mu y_{t-1}]^2 \quad (3.25)$$

En dérivant (3.25) par rapport à μ , on obtient :

$$\begin{aligned} \frac{\partial}{\partial \mu} \log m(y_t | \mu) &= -\frac{1}{2}a^{-1} \sum_{t=1}^T \frac{\partial}{\partial \mu} [y_t - \mu y_{t-1}]^2 \\ &= -\frac{1}{2}a^{-1} \sum_{t=1}^T \frac{\partial}{\partial \mu} [y_t^2 - 2y_t \mu y_{t-1} + \mu^2 y_{t-1}^2] \\ &= -\frac{1}{2}a^{-1} \sum_{t=1}^T [-2y_t y_{t-1} + 2\mu y_{t-1}^2] \\ &= a^{-1} \sum_{t=1}^T [y_t y_{t-1} - \mu y_{t-1}^2] \\ &= a^{-1} \sum_{t=1}^T [y_t - \mu y_{t-1}] y_{t-1} \end{aligned}$$

En l'égalisant à zéro, on obtient l'équation suivante :

$$\begin{aligned} a^{-1} \sum_{t=1}^T [y_t - \mu y_{t-1}] y_{t-1} &= 0. \\ \implies \sum_{t=1}^T y_t y_{t-1} a^{-1} - \mu \sum_{t=1}^T y_{t-1}^2 a^{-1} &= 0. \\ \implies \mu \sum_{t=1}^T y_{t-1}^2 a^{-1} &= \sum_{t=1}^T y_t y_{t-1} a^{-1}. \\ \implies \hat{\mu} &= \frac{\sum_{t=1}^T y_t y_{t-1} a^{-1}}{\sum_{t=1}^T y_{t-1}^2 a^{-1}}, \quad \text{qui est l'EMV de } \mu. \end{aligned}$$

l'estimateur Bayésien empirique ϕ^E est obtenu en remplaçant μ par $\hat{\mu}$ dans l'équation (3.22), ce qui donne ;

$$\hat{\phi}^E = \frac{\frac{1}{\sigma^2} \sum_{t=1}^T y_t y_{t-1} + \frac{1}{\tau^2} \hat{\mu}}{\frac{1}{\sigma^2} \sum_{t=1}^T y_{t-1}^2 + \frac{1}{\tau^2}} \quad (3.26)$$

3.6 Application numérique

Nous allons simuler le AR(1) suivant

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2). \quad (3.27)$$

Le programme A.4 donne :

n	$\phi = 0.2$		$\phi = 0.5$		$\phi = 0.8$	
	BE	Bayésien	BE	Bayésien	BE	Bayésien
10	0.281	0.324	0.479	0.506	0.690	0.700
50	0.203	0.212	0.489	0.494	0.763	0.764
100	0.174	0.179	0.483	0.485	0.7934	0.7939
1000	0.166	0.166	0.4610	0.4612	0.77401	0.77405
10000	0.1852	0.1853	0.48910	0.48912	0.79372	0.79373
100000	0.197252	0.197256	0.49710	0.49711	0.796338	0.796339

TABLE 3.1 – Comparaison des estimations Bayésiennes empiriques et estimations Bayésiennes classiques pour différentes valeurs de n et ϕ .

Le tableau présente une comparaison entre les estimations Bayésiennes empiriques (BE) et les estimations Bayésiennes classiques pour différentes valeurs de ϕ (0.2, 0.5, 0.8) et différents échantillons n (de 10 à 100,000). On peut tirer de ce tableau que :

Lorsque ϕ est petit (0.2), on observe que la différence entre les deux méthodes est plus prononcée pour les petits échantillons ($n=10$). À mesure que n augmente, cette différence disparaît.

Pour des valeurs de ϕ plus élevées (0.5 et 0.8), les deux méthodes fournissent des estimations proches même pour des tailles d'échantillon modérées.

Pour de petites tailles d'échantillons, l'estimation Bayésienne empirique peut légèrement différer de l'estimation Bayésienne classique, mais elle tend à converger vers la même valeur que l'approche classique lorsque n augmente.

3.6.1 Calcul du MSE pour chaque estimation

Formule du MSE

Pour chaque valeur de ϕ et chaque taille d'échantillon n , le MSE est donné par :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{\phi}_i - \phi_{\text{vraie}})^2$$

Ici, $\hat{\phi}_i$ représente l'estimation (Bayésienne empirique ou Bayésienne classique) pour une taille d'échantillon donnée, et ϕ_{vraie} est la vraie valeur de ϕ .

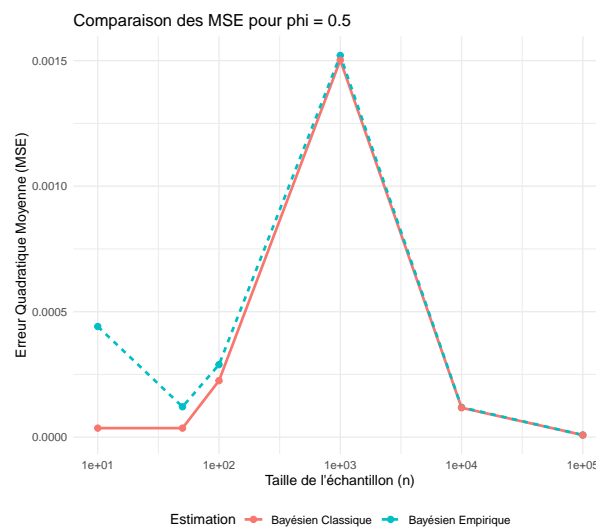
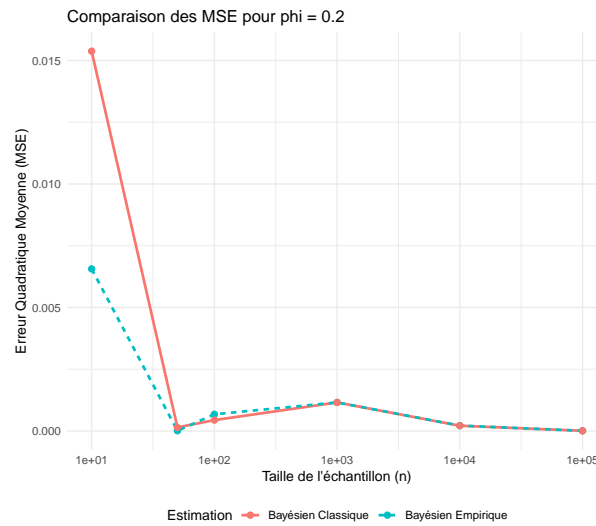
On applique cette formule pour chaque valeur de ϕ (0.2, 0.5, 0.8) et chaque taille d'échantillon n . Nous obtenons le tableau suivant

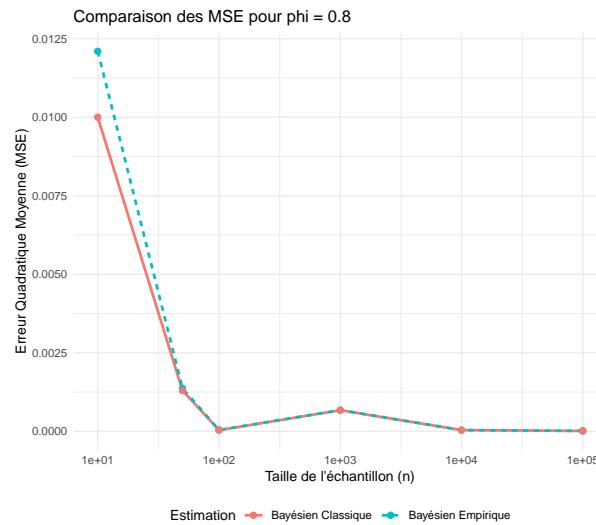
n	$\phi = 0.2$ (BE)	$\phi = 0.2$ (Bayésien)	$\phi = 0.5$ (BE)	$\phi = 0.5$ (Bayésien)	$\phi = 0.8$ (BE)	$\phi = 0.8$ (Bayésien)
10	0.006561	0.015376	0.000441	0.000036	0.012100	0.010000
50	0.000009	0.000144	0.000121	0.000036	0.001369	0.001296
100	0.000676	0.000441	0.000289	0.000225	0.00004356	0.00003721
1000	0.001156	0.001156	0.001521	0.001505	0.0006748	0.0006731
10000	0.00021904	0.00021316	0.0001188	0.0001183	0.0000386	0.0000385
100000	0.00000755	0.00000753	0.00000841	0.00000835	0.00001341	0.00001342

TABLE 3.2 – MSE pour les estimations Bayésiennes empiriques et classiques pour différentes valeurs de ϕ et n .

Graphes de comparaison

Le programme A.5 donne les graphes suivants





Les 3 graphiques montre comment le MSE change en fonction de la taille de l'échantillon n pour les différentes valeurs de ϕ , en comparant les estimations Bayésiennes empiriques et classiques.

Conclusion de l'étude MSE

Pour de petites tailles d'échantillon, on remarque que l'estimateur Bayésien empirique et l'estimateur Bayésien classique sont un peut proche et ils se rapproche de plus en plus au fur et à mesure que l'échantillon s'agrandit. Pour $n \geq 1000$, les MSE deviennent pratiquement identiques pour les deux approches, indiquant que l'estimateur Bayésien empirique est aussi performant que l'estimateur classique sur de grands échantillons.

Les différences entre les méthodes sont plus visibles pour de petites tailles d'échantillon, mais elles tendent à s'estomper avec des échantillons plus larges. D'autres termes nous pouvons dire que les deux estimateurs se comporte de la même manière, ce qui veut dire que si j'ai suffisamment d'information a priori nous pouvons faire appel aux Bayésiens empirique.

CONCLUSION GÉNÉRALE

L'analyse Bayésienne empirique, telle que présentée dans ce mémoire, offre un cadre méthodologique puissant pour l'estimation et la prédiction dans les séries chronologiques, notamment les modèles autorégressifs d'ordre 1 (AR(1)). En combinant les forces des approches Bayésiennes et fréquentistes, elle permet de mieux gérer l'incertitude dans l'estimation des paramètres et améliore la qualité des prévisions. Cette approche est particulièrement utile dans les contextes où les informations a priori fiables sont limitées ou inexistantes, comme dans les phénomènes économiques ou climatiques.

En intégrant les données observées pour ajuster les hypothèses, le Bayésien Empirique réduit l'arbitraire du choix de la distribution a priori, rendant les résultats plus robustes et précis. Les applications variées dans des domaines tels que la finance, la météorologie ou la biologie montrent que cette méthode s'adapte à un large éventail de problématiques.

Comme perspective il serait intéressant de reprendre un AR(1) Avec un bruit exponentiel et estimer son paramètre avec la méthode Bayésienne empirique, ou encore à utiliser d'autres fonctions coût que le coût quadratique, et pourquoi pas faire une comparaison entre l'estimateur Bayésienne empirique en utilisons l'estimateur maximum vraisemblance d'une part et la méthode des moments d'autre part.

Annexe A

PROGRAMMATION

A.1 Programme 1

```
ts2 = read.table("D:/ADDITIVE2.txt", header = F, sep = "",dec = ".")

graphique = ggplot(ts2, aes(x = V1, y = V2)) +
  geom_line() +
  geom_point() +
  geom_line(data = ts2[c(9, 21, 32, 44, 56, 81), ], aes(x = V1, y = V2),
    color = "red", linetype = "dashed") +
  geom_line(data = ts2[c(1, 13, 25, 37, 61, 73), ], aes(x = V1, y = V2),
    color = "blue", linetype = "dashed") +

  labs(title = "Graphe des données avec lignes spécifiques",
    x = "Mois",
    y = "concentration de CO2")

print(graphique)
```

A.2 Programme 2

```
# Convertire la série chronologique en un data frame
df <- data.frame(Year = as.numeric(time(ts2)), Value = as.numeric(ts(ts2)))

graphique = ggplot(df, aes(x = Year, y = Value)) +
  geom_line() +
  geom_point() +
  geom_line(data = df[c(3, 7, 19, 31, 44, 56, 67, 79, 91, 103, 116, 128, 139 ), ],
  aes(x = Year, y = Value), color = "red", linetype = "dashed") +
```

```

geom_line(data = df[c(11, 23, 35, 59, 71, 83, 95, 107, 119, 131, 143), ],
aes(x = Year, y = Value), color = "blue", linetype = "dashed") +

labs(title = "Graphe des données avec lignes spécifiques",
      x = "Temps",
      y = "Valeur")

print(graphique)

```

A.3 Programme 3

On définit quelques fonctions dont on a besoin pour notre estimateur

```

# Indicateur J
J1 <- function(x, y) {
  a <- 0
  if (-0.5 <= (x - y) & (x - y) < 0.5) a <- 1
  a
}

J <- function(u, y) {
  k <- length(y)
  h <- rep(0, k) #Vecteur vide de longueur y Pour stocker les résultats de la fonction
  for (i in 1:k) {
    h[i] <- J1(u, y[i])
  }
  h
}

# Indicateur K
K1 <- function(x, y, z) {
  a <- 0
  if (x <= y & y <= z) a <- 1
  if (z <= y & y <= x) a <- 1
  a
}

K <- function(u, y, z) {
  k <- length(y)
  h <- rep(0, k) #Vecteur vide de longueur y Pour stocker les résultats de la fonction
  for (i in 1:k) {
    h[i] <- K1(u, y[i], z)
  }
}

```

```

h
}

# Calculs des différents estimateur Bayésiens empiriques et l'estimateur de Bayes
BE10 <- fonction(n, a) {
  b1 <- rep(0, n)
  be2 <- rep(0, n)
  m3 <- rep(0, n)
  wb4 <- rep(0, n)
  wbe5 <- rep(0, n)
  bp6 <- rep(0, n)
  b7 <- rep(0, n)
  b8 <- rep(0, n)
  ben9 <- rep(0, n)
  bn10 <- rep(0, n)
  eqmp11 <- 0
  eqmn12 <- 0
  benh <- rep(0, n)
  eqmah <- rep(0, n)
  eqmh <- 0
  eqmax <- rep(0, n)
  eqmx <- 0

  for (i in 1:n) {
    theta <- rexp(i, a)
    x <- rpois(length(theta), theta)
    t <- rexp(1, a)
    xi <- rpois(1, t)
    m <- max(x, 1)
    xb <- mean(x) #x bar
    pn <- sum(J(xi, x)) #Le nombre de fois où xi et les valeurs de x sont proches
    pm <- sum(J(xi + 1, x)) #le nombre de fois où xi + 1 et les valeurs de x sont pro
    hn <- sum(K(xi, x, xb)) #le nombre de fois où xi et les valeurs de x sont en
      relation avec xb selon la definition K1
    hm <- sum(K(xi + 1, x, xb)) #le nombre de fois où xi+1 et les valeurs de x sont e
      relation avec xb selon la definition K1

    b1[i] <- (xi + 1) / (1 + a) # sigma^B
    be2[i] <- (1 + xi) * xb / (1 + xb) # sigma^E
    m3[i] <- xb #vecteur des x bar
    # le risque intégré de l'estimateur de Bayes
    wb4[i] <- 1 / (a * (1 + a))
    # le risque intégré de l'estimateur de Bayes empirique

```

```

wbe5[i] <- 1 / (a * (1 + a)) + (2 / (a^2) + 5 / a + 4) * (xb / (1 + xb) - 1 / (1 + a))
bp6[i] <- (t - be2[i])^2 #erreurs quadratiques moyennes
b7[i] <- t
b8[i] <- xi
ben9[i] <- (1 + xi) * pm / (1 + pn)
bn10[i] <- (t - ben9[i])^2
eqmax[i] <- (t - xi)^2
benh[i] <- (1 + xi) * hm / (1 + hn)
eqmah[i] <- (t - benh[i])^2
}

w <- list(
  b1 = b1, be2 = be2, m3 = m3, wb4 = wb4, wbe5 = wbe5, bp6 = bp6,
  b7 = b7, b8 = b8, ben9 = ben9, bn10 = bn10, eqmp11 = mean(bp6),
  eqmn12 = mean(bn10), eqmx = mean(eqmax), benh = benh, eqmh = mean(eqmah)
)
w
}

# Graphique de l'estimateur en fonction de la taille de l'échantillon
x <- seq(2, 1000)
plot(x, BE10(1000, 0.2)$wbe5[-1], type = "o", xlab = "Taille de l'échantillon", ylab = "Estimateur")

```

A.4 Programme 4

```

library(MASS)

set.seed(123)
##### valeurs qui change ###
T <- 1000 # Number of time points
phi_true <- 0.5 # True value of phi
#####
sigma2 <- 1 # Variance of the noise
y <- numeric(T)
y[1] <- 0

# Généré un AR(1)
for (t in 2:T) {
  y[t] <- phi_true * y[t - 1] + rnorm(1, 0, sqrt(sigma2))
}

```

```

# Définir les paramètres a priori
mu_prior <- 0 #Moyenne a priori de phi
tau2_prior <- 1 #variance a priori de phi

# Empirical Bayésien estimation
yt_lagged <- y[1:(T - 1)] # y_t-1 (lagged)
yt <- y[2:T] # y_t

A <- sum(yt_lagged^2) / sigma2 + (1 / tau2_prior)
B_classic <- sum(yt * yt_lagged) / sigma2 + mu_prior / tau2_prior

# Moyenne a posteriori de phi
phi_post_mean <- B_classic / A #delta^B

# estimation maximum de vraisemblance de mu
mu_emv = sum(yt*yt_lagged)/ sum(yt_lagged^2)

B <- sum(yt * yt_lagged) / sigma2 + mu_emv / tau2_prior
est_Bemp <- B/ A #delta^E

# Results
cat("l'estimateur Bayésien de phi:", phi_post_mean, "\n")
cat("l'estimateur Bayésien empirique de phi:", est_Bemp, "\n")

```

A.5 Programme 5

```

library(reshape2)
library(ggplot2)

# Données : valeurs MSE pour différentes valeurs de phi et tailles d'échantillon (n)
data <- data.frame(
  n = c(10, 50, 100, 1000, 10000, 100000),
  phi_0_2_BE = c(0.006561, 0.000009, 0.000676, 0.001156, 0.00021904, 0.00000759),
  phi_0_2_Bayes = c(0.015376, 0.000144, 0.000441, 0.001156, 0.00021316, 0.00000753),
  phi_0_5_BE = c(0.000441, 0.000121, 0.000289, 0.001521, 0.0001188, 0.00000841),
  phi_0_5_Bayes = c(0.000036, 0.000036, 0.000225, 0.001502, 0.0001171, 0.00000837),
  phi_0_8_BE = c(0.012100, 0.001369, 0.00004356, 0.0006748, 0.0000386, 0.00001341),
  phi_0_8_Bayes = c(0.010000, 0.001296, 0.00003721, 0.0006731, 0.0000385, 0.00001342)
)

# Transformer les données en format long pour ggplot2
data_long <- melt(data, id.vars = "n", variable.name = "Type", value.name = "MSE")

```

```

# Extraire des informations sur phi et le type d'estimation
data_long$phi <- ifelse(grepl("phi_0_2", data_long$Type), "phi = 0.2",
                      ifelse(grepl("phi_0_5", data_long$Type), "phi = 0.5", "phi = 0.8"))

data_long$Estimation <- ifelse(grepl("BE", data_long$Type), "Bayésien Empirique", "Bayésien")

# Fonction pour créer un graphe pour une valeur spécifique de phi
create_plot <- function(data, phi_value) {
  ggplot(data, aes(x = n, y = MSE, color = Estimation, linetype = Estimation)) +
    geom_line(size = 1) + # Graphe avec des lignes
    geom_point(size = 2) + # Ajouter des points
    scale_x_log10() + # Échelle logarithmique pour l'axe des x
    labs(
      title = paste("Comparaison des MSE pour", phi_value),
      x = "Taille de l'échantillon (n)",
      y = "Erreur Quadratique Moyenne (MSE)"
    ) +
    theme_minimal() + # Thème minimaliste
    theme(legend.position = "bottom") # Placer la légende en bas
}

# Créer les graphes pour chaque valeur de phi
plot_phi_0_2 <- create_plot(subset(data_long, phi == "phi = 0.2"), "phi = 0.2")
plot_phi_0_5 <- create_plot(subset(data_long, phi == "phi = 0.5"), "phi = 0.5")
plot_phi_0_8 <- create_plot(subset(data_long, phi == "phi = 0.8"), "phi = 0.8")

# Afficher les graphes individuellement
print(plot_phi_0_2)
print(plot_phi_0_5)
print(plot_phi_0_8)

```

A.6 Programme 6

```

# Définition des paramètres
e1 <- 118
nu <- 0.104
sigma <- 89.79
gamma <- sigma / (1 + sigma)

# Fonction E(t)
E <- function(t) {

```

```
e1 / (gamma * nu) * (1 - (1 + gamma * t)^(-nu))
}

# Génération des données pour le graphee
t <- seq(0, 1, length.out = 100) # Axe des t de 0 à 1
E_values <- E(t) # Valeurs de E(t)

# Tracé du graphe
plot(t, E_values, type = "l", col = "blue",
      xlab = "t", ylab = "E(t)",
      xlim = c(0, 1), ylim = c(0, 90),
      lwd=3)

legend("topleft",c("Modèle Gamma", " E(2) = 123", " E(4) = 176", " E(8) = 233"))
```

Bibliographie

- [1] Atil.L (2022-2023) Introduction aux statistiques Bayésiennes. Cours de 2^{ème} année Master, U.M.M.T.O Tizi ousou.
- [2] Aragon, Y. (2011). *Séries temporelles avec R, méthodes et cas*. France : Springer-Verlag.
- [3] Brockwell, P.J., and Davis, R.A. (1991). *Time Series : Theory and Methods (Second ed.)*. New York : Springer.
- [4] Carlin, B.P.,Louis, T.A. (2000). Bayes and empirical Bayes methods for data analysis, 2nd edn. London/Boca Raton : Chapman & Hall/CRC
- [5] Christain P.Robert. (2006) Le choix Bayésien, France,Paris : Springer-Verlag
- [6] Cryer, J.D., and Chan, K.S. (2008). *Time Series Analysis With Applications in R*. Springer.
- [7] Efron.B(2010).Empirical Bayes modeling, computation, and accuracy.Stanford University.
- [8] Efron.B(2017) , Hastie.T.Computer Age Statistical Inference Algorithms, Evidence, and Data Science.Stanford University.
- [9] Efron.B(2010) ,Large-Scale Inference : Empirical Bayes Methods for Estimation, Testing, and Prediction. Stanford University, California.
- [10] El-Habti ahmed (2007), Sur Estimation Bayésienne Empirique Pour Les Plans D'expérience Non Equilibrés, Mémoire Comme Exigence Partielle De La maitrise En Mathématiques.
- [11] Fuller, W.A. (1995). *Introduction to Statistical Time Series, 2nd ed.*. New York : Wiley.
- [12] Gourieroux, C., and Monfort, A. (1995). *Time Series and Dynamic Models*. New York : Springer.
- [13] Shumway, R.H., and Stoffer, D.S. (2006). *Time series analysis and its applications with R examples*. New York : Springer-Verlag.
- [14] Kim, Y. W., and Basawa, I. V. (1992). EMPIRICAL BAYES ESTIMATION FOR FIRST-ORDER AUTOREGRESSIVE PROCESSES. Australian Journal of Statistics, 34(1), 105-114.
- [15] Lagnoux, Agnès. *Séries Chronologiques*. France : Université Toulouse-Jean-Jaurès.
- [16] Lambert,B. (2018) *A Student's Guide to Bayesian Statistics*.SAGE Publications

- [17] Loll, Tina (2012). *Forecasting Economic Time Series using Locally Stationary Processes*. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien : Peter Lang.
- [18] Maritz, J. S. (2018). *Empirical Bayes methods with applications*. Chapman and Hall/-CRC.
- [19] Robbins, H. (1956). An empirical Bayes approach to statistics. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I, Berkeley and Los Angeles, pp. 157–163. University of California Press.
- [20] Roussel, Arnaud. Chapitre 7 : *Séries chronologiques*.
- [21] Rubenthaler, S., Vasilieiadis, A. (2023-2024) *Séries chronologiques (avec R) (Cours et exercices) M1 IM*. France : Université Nice-Sophia-Antipolis

Site web :

- ★ Till Tantau. TikZ&PGF Manual. Download from CTAN : [//graphics/pgf/base/doc/generic/pgf/pgfmanual.pdf](://graphics/pgf/base/doc/generic/pgf/pgfmanual.pdf)
- ★ <http://foucart.thierry.free.fr/StatPC>. Chapitre 8 : *ANALYSE DES SÉRIES CHRONOLOGIQUES*.
- ★ <https://www.itl.nist.gov/div898/handbook/datasets/MLCO2MON.DAT>

Résumé

Ce mémoire propose une application de la méthode Bayésienne Empirique à l'analyse des séries chronologiques, en se focalisant particulièrement sur les modèles autorégressifs d'ordre un (AR(1)).

Il commence par présenter les concepts fondamentaux des séries chronologiques et des principales méthodes de modélisation. Ensuite, l'étude met l'accent sur la méthode Bayésienne Empirique, qui permet d'estimer les paramètres des modèles directement à partir des données observées.

Cette approche est validée à travers des simulations numériques, où les résultats des estimations Bayésiennes Empiriques sont comparés à ceux obtenus par l'estimation Bayésienne classique.

Ce travail met en évidence l'efficacité de la méthode Bayésienne Empirique pour l'analyse des séries chronologiques, en particulier dans les situations où les connaissances a priori sont limitées ou incertaines.

Mots clés : Bayésien, Bayésien Empirique, Série chronologique, Auto-Régressif, Prédiction.

Abstract

This work presents an application of the Empirical Bayesian method to the analysis of time series, focusing particularly on first-order autoregressive models (AR(1)).

It begins by introducing the fundamental concepts of time series and the main modeling methods. The study then emphasizes the Empirical Bayesian method, which allows for the estimation of model parameters directly from the observed data.

This approach is validated through numerical simulations, where the results of the Empirical Bayesian estimates are compared to those obtained using the classical Bayesian estimation.

This work highlights the effectiveness of the Empirical Bayesian method in analyzing time series, particularly in contexts where prior knowledge is limited or uncertain.

Keywords : Bayesian, Empirical Bayesian, Time Series, Autoregressive, Forecasting.