

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU.

FACULTE : DES SCIENCES  
DEPARTEMENT : MATHEMATIQUES

**MEMOIRE DE MAGISTER (ECOLE DOCTORALE)**

SPECIALITE : MATHEMATIQUES

OPTION : STATISTIQUE

Présenté par :  
*M<sup>elle</sup>* **BELKACEM Nadia**

Sujet :

**Modèles d'incertitude appliqués au problème  
de management de l'eau**

Devant le jury d'examen composé de :

M <sup>r</sup> HAMADOUCHE Djamel	; Professeur	; UMMTO	;Président
M <sup>r</sup> FELLAG Hocine	; Professeur	; UMMTO	;Rapporteur
M <sup>r</sup> BOUDIBA Mohand Arezki	; Maître de conférence A	; UMMTO	;Examineur
M <sup>elle</sup> ATIL Lynda	; Maître de conférence B	; UMMTO	;Examinatrice

Soutenu le : .. /.. /2012

## *Remerciements*

*Je tiens à exprimer toute ma reconnaissance à M<sup>r</sup> FELLAG Hocine et à M<sup>r</sup> RIOS David Insua pour l'honneur qu'ils m'ont fait en assurant la direction et le suivi scientifique et technique du présent mémoire. Je les remercie pour leur grande contribution à l'aboutissement de ce travail, et pour leur disponibilité.*

*Je remercie vivement M<sup>r</sup> HAMADOUCHE Djamel pour l'honneur qu'il me fait en acceptant de présider le jury de ce mémoire.*

*Mes remerciements chaleureux s'adressent également à M<sup>r</sup> BOUDIBA Mohand Arezki, M<sup>elle</sup> ATIL Lynda et M<sup>me</sup> BELKACEM Cherifa pour avoir accepté d'examiner ce travail.*

*Je tiens également à remercier tout le personnel de l'ADE pour nous avoir facilité cette tâche au sein de leur entreprise.*

*Comme je remercie nos collaborateurs Espagnols pour leurs contributions efficaces et leurs orientations à l'aboutissement de ce travail.*

*Je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.*

# Table des matières

<b>Introduction générale</b>	<b>2</b>
<b>1 Les éléments de la décision avec l'approche Bayésienne</b>	<b>6</b>
1.1 Introduction . . . . .	6
1.2 La décision de l'approche Bayésienne . . . . .	6
1.2.1 Le choix Bayésien . . . . .	6
1.2.2 Notions de base . . . . .	8
1.2.3 Théorème de Bayes . . . . .	8
1.3 Les bases de la Théorie de la Décision . . . . .	15
1.3.1 Estimateur de Bayes . . . . .	15
1.3.2 Fonction de perte et risque . . . . .	16
1.3.3 Fonctions de coût usuelles . . . . .	17
1.3.4 Admissibilité et minimaxité . . . . .	20
1.4 Choix des lois a priori . . . . .	23
1.4.1 Approche partiellement informative . . . . .	23
1.4.2 Approche non informative . . . . .	27
1.5 Méthodes de calcul Bayésien . . . . .	31
1.5.1 Méthodes classiques d'approximation . . . . .	31
1.5.2 Méthodes de Monte Carlo par chaînes de Markov (MCMC) . . . . .	32
1.6 Avantages de l'approche Bayésienne . . . . .	33
1.7 Conclusion . . . . .	35
<b>2 Modèles dynamiques linéaires</b>	<b>37</b>
2.1 Introduction . . . . .	37
2.2 Généralités . . . . .	38
2.3 Modèle à espace-état . . . . .	40
2.4 Modèle dynamique linéaire . . . . .	41

---

2.5	Estimation et prévision . . . . .	43
2.5.1	Filtrage . . . . .	44
2.5.2	Le lissage . . . . .	46
2.5.3	Prévision . . . . .	47
2.6	Analyse des séries chronologiques . . . . .	48
2.6.1	Les modèles <i>ARMA</i> . . . . .	49
2.6.2	Modèle dynamique linéaire pour l'analyse des séries chronologiques	57
2.7	Modèles avec des paramètres inconnus . . . . .	66
2.7.1	Estimation avec le maximum de vraisemblance . . . . .	67
2.7.2	Inférence Bayésienne . . . . .	67
2.7.3	Inférence Bayésienne conjuguée . . . . .	68
<b>3</b>	<b>Modèles dynamiques linéaires pour la demande d'eau</b>	<b>73</b>
3.1	Description du réseau de distribution d'eau dans la région de Kabylie . . .	73
3.1.1	La chaîne de Tassadort . . . . .	76
3.1.2	Description physique du réseau . . . . .	78
3.1.3	Alimentation équitable en eau potable . . . . .	79
3.1.4	Les variables et les contraintes . . . . .	80
3.1.5	La fonction objectif . . . . .	82
3.1.6	Distribution équitable de l'eau avec un coût efficace . . . . .	84
3.2	Problématique du management de l'eau . . . . .	85
3.3	Analyse exploratrice . . . . .	87
3.4	Modèle pour le nombre de compteurs . . . . .	90
3.4.1	Spécification de $W_t$ par le facteur d'escompte . . . . .	91
3.4.2	Représentation du modèle sous forme d'un DLM + une composante <i>ARMA</i> . . . . .	94
3.5	Modèle pour la demande d'eau . . . . .	97
	<b>Conclusion générale</b>	<b>112</b>
	<b>Annexes</b>	<b>112</b>
	<b>Bibliographie</b>	<b>112</b>

# Introduction générale

L'eau est essentielle à la vie, c'est une ressource naturelle. L'eau est le constituant essentiel des êtres vivants : un homme ne peut vivre que 2 jours sans eau, la quantité moyenne d'eau contenue dans le corps d'un adulte est de 70%. Il y a encore quelques années, l'eau n'était qu'une molécule chimique que l'on mettait en bouteille, ressource inépuisable à laquelle on ne prêtait guère attention, mais, aujourd'hui les mentalités commencent à évoluer. l'homme utilise, divers moyens pour en satisfaire ses besoins. L'alimentation en eau potable s'étend dans les villes et les agglomérations parallèlement à la croissance démographique et socio-industrielle.

Le comptage de l'eau est la façon la plus équitable et correcte pour rémunérer le service de distribution d'eau potable, selon le principe élémentaire que plus on consomme, plus on paie. Il s'agit aussi d'un système efficace de limitation des consommations et des gaspillages d'une ressource de plus en plus chère. Les compteurs d'eau sont des machines hydrauliques qui permettent la mesure automatique du volume d'eau traversant en un temps donné une section déterminée d'un courant liquide. Les compteurs normalement utilisés pour la facturation de l'eau sont de type mécanique : l'organe de mesure est actionné directement par la force hydrodynamique de l'eau et met en rotation les mécanismes d'un dispositif d'affichage grâce à l'énergie mécanique transmise par le courant d'eau. Pour une bonne gestion de ces compteurs une prévision est envisageable bien sur la prise en compte de l'incertitude sur la quantité d'eau demandée est indispensable.

Les modèles dynamiques linéaires ont été développés en ingénierie au début des années 1960, pour surveiller et contrôler les systèmes dynamiques, même si les résultats des pionniers peuvent être trouvés dans la littérature statistique, bien avant 1960, le précurseur étant Thiele (1880). Les premières célèbres applications sont Apollon et Polaris (programmes aérospatiaux) (voir, par exemple, Hutchinson, 1984), mais dans les dernières décennies des modèles dynamiques linéaires, et plus généralement des modèles espace-état, ont reçu une impulsion énorme, avec des applications dans un très vaste éventail de do-

maines, de la biologie à l'économie, de l'ingénierie et de contrôle de qualité pour les études environnementales, de la science géophysique à la génétique.

Cette croissance impressionnante des applications est largement due à la possibilité de résoudre des difficultés de calcul en utilisant des méthodes de Monte Carlo modernes dans un cadre Bayésien. L'analyse statistique des séries temporelles en utilisant des modèles dynamiques linéaires a été largement développée dans les années 1970-1980, et des modèles espace-état sont aujourd'hui un centre d'intérêt. En fait, on peut trouver l'approche espace-état pour l'analyse de séries chronologiques par rapport aux modèles ARMA et modèle de spécification Box- Jenkins, un peu difficile au premier abord, mais le cadre puissant offert par les modèles dynamiques linéaires se révèle être un atout gagnant. Les modèles ARMA peuvent être utiles quand ils sont considérés comme des modèles dynamiques linéaires. Mais les modèles dynamiques linéaires offrent beaucoup plus de flexibilité dans le traitement de séries temporelles non stationnaires ou la modélisation des changements structurels, et sont souvent plus facilement interprétables.

Il y a, bien sûr, des approches différentes pour estimer les modèles dynamiques linéaires, en passant par les moindres carrés généralisés ou le maximum de vraisemblance, mais nous pensons qu'une approche Bayésienne a plusieurs avantages, tant méthodologiques que informatiques. Kalman (1960) souligne déjà quelques notions de base des modèles dynamiques linéaires qui sont propres à l'approche Bayésienne. Une première étape est de passer d'un système déterministe à un système stochastique ; l'incertitude, qui est toujours présente en raison de variables non observées, les erreurs de mesure, ou des imperfections, est décrite à travers la probabilité.

Par conséquent, l'estimation des quantités d'intérêt (en particulier, l'état du système au temps  $t$ ) est résolue par le calcul de leur distribution conditionnelle, compte tenu des informations disponibles, c'est un concept de base dans l'inférence Bayésienne.

L'objectif du présent travail est d'évaluer l'intérêt de l'utilisation des modèles dynamiques linéaires en hydraulique par leur application à des données qui représentent le nombre de compteurs d'eau. Le premier chapitre introduit les notions de base concernant l'approche Bayésienne. Le chapitre 2 introduit les modèles dynamiques linéaires. Les modèles ayant été appliqués à des séries temporelles représentant le nombre de compteurs d'eau dans les différents villages de la Kabylie, les résultats sont discutés dans le troisième chapitre. L'intérêt de l'approche (modèles dynamiques linéaires) et ses perspectives sont discutés en conclusion.

Cette étude vise à approcher une bonne maîtrise du monde de distribution de l'eau

afin de la rendre plus équitable. Toutefois, cette problématique est initiée dans le cadre d'un projet de coopération Algéro-Espagnole du type PCI méditerrané qui lie notre université (Université Mouloud Mammeri) et celle de Madrid (Rey Juan Carlos). La réalisation n'a été possible qu'avec l'accord institutionnel de l'Algérienne des eaux (zone de Tizi-ouzou) qui nous a ouvert toutes ses portes et facilité nos investigations. A travers cette collaboration, l'ADE a montré un professionnalisme actif qui peut servir d'exemple aux partenariats futurs. La contribution des partenaires espagnols qui ont acquis une longue expérience dans le domaine, tant sur le plan de la modélisation que sur l'usage des logiciels est une très grande importance. Par conséquent, cette étude n'est qu'une suite et une étape d'un processus plus approfondi qui se développera plus tard dans ce projet.

# Chapitre 1

## Les éléments de la décision avec l'approche Bayésienne

### 1.1 Introduction

Le troisième millénaire sera, dit-on, celui de l'information. Aussi, la statistique y sera-t-elle appelée à jouer un rôle important et le paradigme Bayésien plus que tout autre, puisqu'il offre un cadre de raisonnement bien adapté à l'intégration des opinions et des faits de toutes provenances qui interviennent dans la gestion des risques et la prise de décision en contexte d'incertitude. De la collecte de données à la prévision, l'analyse statistique pose plusieurs défis. L'élaboration du modèle représente sans doute la phase la plus délicate de l'exercice, car elle doit répondre à un double impératif de réalisme et de parcimonie. Hormis quelques cas de figure, une démarche Bayésienne n'est envisageable qu'à charge de disposer d'outils efficaces pour la quantification et la mise à jour de l'information.

### 1.2 La décision de l'approche Bayésienne

#### 1.2.1 Le choix Bayésien

La statistique est un art interdisciplinaire de la quantification sous incertitudes, utilisée par les physiciens, les économistes, les ingénieurs, les biologistes, les assureurs, les psychologues, les météorologues, etc. Tous les praticiens sont soucieux de bâtir, sur des fondations solides, un pont entre théorie et données expérimentales. Depuis un siècle, la statistique s'est considérablement développée, initiant une révolution dans les modes de pensée, car elle porte un langage de représentation du monde et de ses incertitudes. C'est aujourd'hui une science mathématique dont l'objectif est de décrire ce qui s'est produit et

de faire des projections quant à ce qu'il peut advenir dans le futur. Parfois, la situation peut être simplement décrite par quelques représentations graphiques d'analyse élémentaire des données. Bien souvent, le problème est beaucoup plus compliqué car de multiples facteurs d'influence doivent être pris en compte. Schématiquement, on construit deux ensembles avec ces facteurs. Un premier paquet contient les facteurs dits explicatifs, bien identifiés, ceux dont on souhaite étudier l'influence en détail. En ce qui concerne le second paquet de facteurs, on ne sait, ou on ne veut pas, représenter leur effet perturbateur au cas par cas et, de ce fait, le jargon des modélisateurs le baptise sous le terme bruit, décrit alors de façon plus grossière par ses caractéristiques statistiques générales. Dans tous les cas, l'étude de la variabilité est au centre des débats : il s'agit d'abord de caractériser l'influence des facteurs identifiés et ensuite de représenter et d'évaluer le bruit résiduel dû à ces autres facteurs non pris en compte dans l'analyse de façon explicite. Dans une telle situation, le statisticien classique utilise à la fois un raisonnement déterministe par l'absurde, afin de proposer des valeurs acceptables pour les paramètres décrivant les effets des facteurs explicatifs et un raisonnement probabiliste, pour traduire la variabilité des résultats observés due au bruit. Ce mode de pensée s'appuie sur l'hypothèse de la réalité objective des paramètres ainsi que sur l'interprétation de la probabilité comme limite des fréquences de résultats observés. C'est cette conception, dite fréquentiste, qui est généralement. Par contre, le statisticien Bayésien utilise le même cadre de pensée pour traiter par le pari probabiliste l'interaction de ces deux niveaux d'incertitudes : ignorance quant aux valeurs possibles des paramètres et aléa des bruits entachant les résultats expérimentaux.

Choisir la piste Bayésienne paraîtra à certains inutilement trop sophistiqué si on se limite aux modèles élémentaires (binomial, normal, etc.), pour ces cas d'écoles simples, l'approche fréquentiste est facile (nombreux logiciels), et offre au praticien des résultats souvent très proches de ceux que donnerait une analyse Bayésienne avec une distribution a priori peu informative. Mais pour peu que l'analyste souhaite prendre à bras le corps des problèmes plus proches de son réel quotidien, apparaissent variables multiples, données manquantes, effets aléatoires, grandeurs latentes..., la structure des modèles de la vie scientifique moderne se présente sous une forme où des couches successives de conditionnement s'emboîtent... et pour lesquels l'approche Bayésienne affirme sa véritable pertinence.

### 1.2.2 Notions de base

Un système est caractérisé par un certain nombre de variables définissant son état. Les valeurs de ces variables, les mesurands, sont obtenues par le biais d'un système de mesure. Les résultats de la mesure, que nous appellerons aussi observations, ne permettent qu'une estimation de l'état, car interviennent dans le processus des phénomènes de nature aléatoire non maîtrisés par l'observateur. On obtient une correspondance entre l'état et l'observation qui est de nature statistique, associant à un état fixé une répartition des observations (gaussienne, poissonnienne ou autre). Le but de la mesure est alors d'inverser cette relation, en ce sens que l'observateur ayant obtenu un résultat, doit en inférer l'état.

La relation donnant la répartition des observations pour un état donné se nomme *le modèle*. Elle est objective, extérieure à l'observateur, car elle n'est dépendante que de l'objet mesuré et de l'instrument utilisé. En répétant l'expérience en face du même état, l'observateur verra ses résultats se distribuer selon le modèle. On trouve une probabilité au sens fréquentiel.

Comme un état peut donner des résultats de mesure différents, une observation peut très bien correspondre à plusieurs états. Quel est le bon ? Ou mieux, quelle répartition peut-on imaginer sur les états ayant en main cette observation ? L'incertitude passe du domaine des résultats dans celui des états. Cette incertitude est subjective, en ce sens qu'elle est propre à l'observateur. La répartition des états n'existe que dans la description interne qu'a l'observateur du système à un moment donné, mais nullement dans la réalité qui lui est extérieure.

En statistique classique, le modèle est donné par une fonction dite de vraisemblance  $\mathcal{L}(\theta|x)$ , où  $\theta$  désigne l'état du système et  $x$  l'observation, Cette fonction est normalisée et considérée comme densité de probabilité.

### 1.2.3 Théorème de Bayes

Soient  $A$  et  $B$  deux événements aléatoires. La probabilité de  $B$ , conditionnellement à la réalisation de  $A$ , est par définition exprimée (si la probabilité de l'événement  $A$  est

non-nulle) par la relation suivante :

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

où  $P[B, A]$  est la probabilité que les deux événements  $A$  et  $B$  aient lieu simultanément. Puisque  $P[B, A] = P[A, B]$  si dans l'expression de  $P[A|B]$  :

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

on remplace  $P[A, B]$  par  $P[B, A] = P[B|A].P[A]$ , on en déduit la relation entre les deux probabilités conditionnelles  $P[A|B]$  et  $P[B|A]$  :

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

Cette équation est une conséquence triviale de la définition de la probabilité conditionnelle, est appelée *Formule de Bayes* (ou aussi Théorème de Bayes) en l'honneur du Révérend Thomas Bayes (1702–1761) à qui l'on doit un des ouvrages qui sont à la base de l'inférence statistique : "*Essay Towards Solving a Problem in the Doctrine of Chances*" publié en 1763 dans les *Philosophical Transactions* de la Royal Society de Londres.

Thomas Bayes, pieux serviteur de Dieu<sup>1</sup> et grand mathématicien, n'est pas passé à l'histoire pour la découverte de la formule qui porte son nom, mais pour son interprétation et son application à un problème d'estimation.

Le problème du Rév. Bayes (Bernier et al., 2000) est le suivant : une balle est lancée sur une table parfaitement horizontale et s'arrête à un certain point  $P_0$ . Ensuite, une deuxième balle est lancée  $n$  fois et on s'intéresse au nombre de fois  $x$ , qu'elle s'est arrêtée à la droite de la première (appelons ces événements "succès"). Comment estimer la probabilité de succès  $\theta$  ?

Bayes imagine que  $\theta$  est une variable aléatoire définie sur l'intervalle  $[0, 1]$  et décrit avec une distribution de probabilité donnée  $\pi(\theta)$  sa connaissance concernant cette variable, préalablement à l'observation des données.

---

1. Il est auteur en 1731 de l'essai : "Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures".

En utilisant la formule de Bayes, il est possible d'écrire l'expression de la distribution de probabilité de  $\theta$ , conditionnellement à l'observation des données  $x$  :

$$\pi(\theta|x) = \frac{\pi(\theta)P(x|\theta)}{M(x)} = \frac{P(x|\theta)\pi(\theta)}{\int_{\theta} P(x|\theta)\pi(\theta)}$$

Les différents termes de cette formule peuvent être interprétés de la manière suivante :

$\pi(\theta)$  est la distribution de probabilité a priori du paramètre inconnu  $\theta$ . L'appellation *a priori* exprime le fait qu'elle a été établie préalablement à l'observation des données  $x$ . Elle peut être issue de l'opinion personnelle du statisticien, ou établie sur la base de l'analyse d'autres données similaires ou de l'avis d'expert. Des exemples de mise en oeuvre de ces méthodes sont fournis par Kadane et Wolfson (1998), O'Hagan (1998), Garthwaite et O'Hagan (2000) et Parent et Prevost (2003). En revanche, son indépendance des données est obligatoire pour éviter d'utiliser deux fois la même information (Berry, 1996).

$P(x|\theta)$  est la probabilité des observations conditionnellement à la valeur  $\theta$  du paramètre du modèle statistique qu'on utilise pour leur description. Il s'agit de la vraisemblance des données, sous le modèle paramétré par  $\theta$ .

$\pi(\theta|x)$  est la distribution de probabilité a posteriori du paramètre du modèle, sur la base de la connaissance a priori et de l'information apportée par les données. L'appellation *a posteriori* vient du fait que, logiquement, elle suit l'observation des données.

Le dénominateur, indépendant de  $\theta$ , est uniquement une constante de normalisation.

Le passage de la distribution a priori à la distribution a posteriori des paramètres du modèle statistique, exprimé par la formule de Bayes, peut être alors interprété comme une mise à jour de la connaissance, sur la base des observations (figure 1.1).

Cette lecture de la formule de Bayes est à la base de l'origine de la distinction entre les statisticiens dits fréquentistes et les Bayésiens.

La différence fondamentale entre les deux approches est que, pour les fréquentistes, toute technique statistique (inférence, test, choix de modèle...) doit être fondée uniquement sur les données et aucune information externe à l'échantillon observé ne peut être intro-

duite dans les calculs.

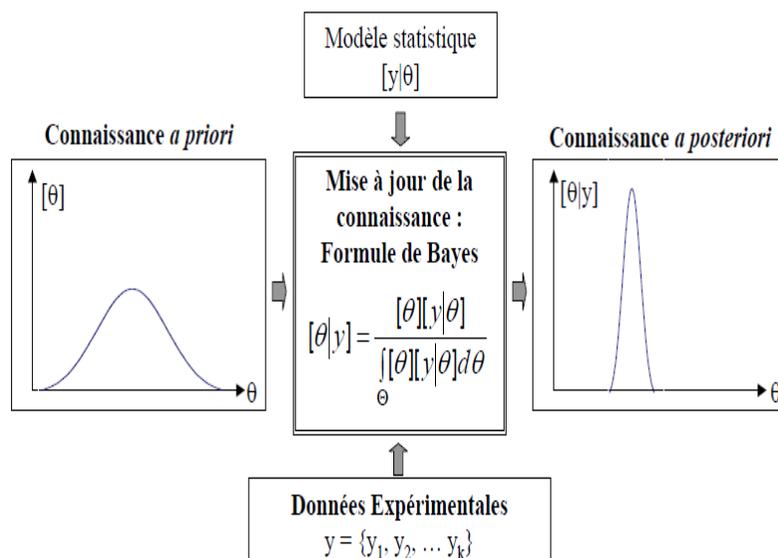


FIGURE 1.1 – Mise à jour de la connaissance avec la formule de Bayes.

L'école fréquentiste, née au début du  $XX^{\text{ème}}$  siècle des travaux des fondateurs de la statistique moderne, notamment W.S. Gosset<sup>2</sup> "Student", K. Pearson, R.A. Fisher, N.J. Neyman, est aussi appelée classique, parce qu'elle a longtemps prévalu parmi les statisticiens même si historiquement l'approche Bayésienne de T. Bayes, qui fut aussi, comme le fait remarquer Sivia (1996), celle de Jakob Bernoulli (1654 – 1705) et P. de Laplace (1749 – 1827) la précède d'environ un siècle et demi.

**Exemple 1.1.** Revenons au problème du Rev. Bayes. Si  $x$  est le nombre de succès observés et  $\theta$  la probabilité de succès, le modèle naturel pour décrire le phénomène est le modèle binomial (annexe A). C'est à dire que, conditionnellement à  $\theta$ , la probabilité d'observer  $x$  succès s'écrit :

2. W.S. Gosset (1856 – 1937) est universellement connu avec son pseudonyme "Student" qu'il a été contraint d'adopter parce que son employeur, le fabricant de bière Guinness, ne permettait pas, à l'époque, à ses employés de publier.

$$P(x|\theta) = C_n^\theta \theta^x (1 - \theta)^{n-x}$$

Concernant la distribution a priori de  $\theta$ , un bon choix est une loi de la famille Bêta (annexe A). Ces lois, à deux paramètres  $\alpha$  et  $\beta$ , sont définies dans l'intervalle  $[0, 1]$  et, selon les valeurs de  $\alpha$  et  $\beta$  peuvent avoir des formes très différentes. Cette souplesse rend la famille Bêta très adaptée à la formalisation de la connaissance préliminaire sur une variable bornée. En vertu de ce choix, l'expression de la loi a priori de  $\theta$  est :

$$\pi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

D'après la formule de Bayes la loi a posteriori  $\pi(\theta|x)$ , est proportionnelle au produit entre loi a priori et vraisemblance :

$$\pi(\theta|x) \propto \theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1}$$

est alors encore une loi Bêta de paramètres :  $x + \alpha$  et  $n - x + \beta$ .

Dans ce cas l'expression du numérateur de la formule de Bayes nous a permis directement de reconnaître que la loi a priori et la loi a posteriori appartiennent à la même famille (Bêta). On exprime cette circonstance heureuse en disant que les lois Bêta et binomiales sont conjuguées.

Malheureusement, sauf quelques autres cas d'écoles, dans la pratique courante de la modélisation, la propriété de conjugaison n'est pas vérifiée et alors le calcul du dénominateur de la formule de Bayes s'impose.

Or, le calcul analytique de cette intégrale, souvent multidimensionnelle, est infaisable dans la plupart des cas. C'est pour cette raison que l'approche Bayésienne a été longtemps mise à l'écart, à l'avantage de l'approche classique qui offre, elle, des solutions simples à de nombreux problèmes statistiques.

La découverte et la possibilité de mettre en oeuvre des algorithmes de simulation capables d'obtenir des tirages aléatoires dans la loi a posteriori des paramètres a libéré les Bayésiens des fardeaux du calcul intégral et a permis l'estimation de modèles à structures complexes. L'expression mathématique de la loi a posteriori restera inconnue à jamais mais,

avec des échantillons aléatoires de cette loi de taille significative, on peut en calculer empiriquement la moyenne, la variance, les percentiles et toutes les autres grandeurs statistiques qui la décrivent, ce qui est d'ailleurs plus intéressant en pratique que d'avoir la formule mathématique de la loi jointe des paramètres du modèle (figure 1.2).

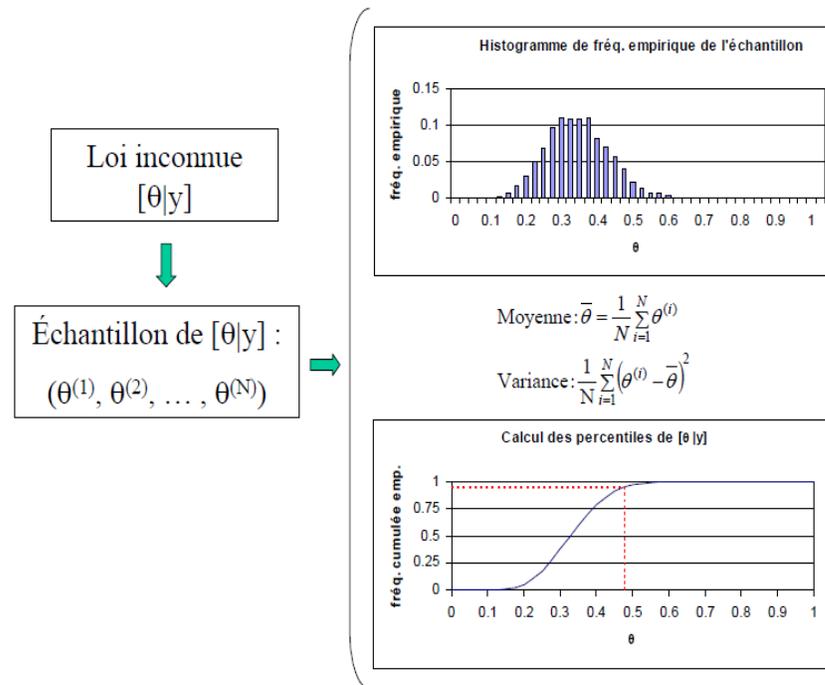


FIGURE 1.2 – Approximation de la loi a posteriori avec des échantillons aléatoires.

**Exemple 1.2.** (L.Larbi, 2011) *Numérique traité avec R*

Une étude est réalisée sur un échantillon de petits garçons, le but de cette étude est d'analyser les effets de l'exposition de petits garçons au plomb (rencontrés dans plusieurs types de nourriture), dans le développement cérébral de ses petits gamins.

Les chercheurs analysent la teneur en plomb des dents de lait des petits après les avoir perdus. On prend un échantillon de 29 petits garçons, qui ont une teneur en plomb supérieur à 22.22 ppm(part par million), 22 parmi ses petits ont terminé leur éducation secondaire, et les 7 autres non.

Supposons une distribution a priori  $p \sim \mathcal{B}e(1, 1)$  pour les élèves qui ont terminé leurs études secondaires.

**Analyse du problème :**

1. Notre échantillon suit une loi de bernouilli  $P(x/\theta) = B(p)$ ,
2. Nous prenons comme loi a priori une bêta,  $\pi(\theta) \sim \mathcal{B}e(1, 1)$
3. En faisant les mêmes calculs que dans l'exemple théorique passé, on obtient la distribution a posteriori qui est une  $\pi(\theta/x) = \mathcal{B}e(1 + 22, 1 + 29 - 22)$

**Caractéristiques de la loi a posteriori :**

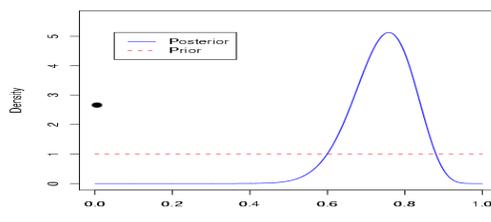
- . `densite = dbeta(x, 23, 8)`
- . `summary(densite)`

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.06994	2.86800	4.10500	3.69400	4.88500	5.12800

**Représentation graphique de la densité a priori et a posteriori :**

Nous commencerons par télécharger le package *Bolstad*, nous utiliserons la fonction "binogcp" qui évalue et dessine la densité a posteriori et a priori qui est utilisée seulement dans le cas d'un échantillon binomial et une distribution a priori continue.

```
results <- binogcp(22, 29, density = "beta", params = c(1, 1), n : pi = 1e4)
```

FIGURE 1.3 – Distribution a priori  $\mathcal{Be}(2, 4)$  et a posteriori  $\mathcal{Be}(8, 13)$ 

## 1.3 Les bases de la Théorie de la Décision

### 1.3.1 Estimateur de Bayes

Soit une fonction de coût  $L(\theta, \delta)$ , et une loi de probabilité a priori (ou une loi impropre)  $\pi$ , pour trouver l'estimateur de Bayes  $\delta^\pi(x)$ , on applique la règle suivante :

$$\delta^\pi(x) = \min_{\delta} \mathbb{E}^\pi[L(\theta, \delta)/x]$$

L'estimateur  $\delta^\pi(x)$  sera déterminé analytiquement ou numériquement ceci dépendra de la fonction de perte, de sa nature et complexité.

Généralement, les solutions associées à des coûts classiques sont formellement connues et correspondent aux caractéristiques usuelles d'une distribution (moyenne, médiane, fractiles, etc.).

Par exemple, l'estimateur de Bayes associé au coût quadratique est la moyenne a posteriori. Cette construction formelle des estimateurs de Bayes classiques n'évite pas toujours le recours à une approximation numérique, particulièrement dans des cas multidimensionnels.

**Lemme 1.1.** (*Christian P. Robert 2006*)

Soit  $f(x/\theta)$  une distribution de probabilité appartenant à une famille exponentielle. Pour

toute loi a priori  $\pi$  la moyenne a posteriori de  $\theta$  est donnée par

$$\delta^\pi(x) = \nabla \log m_\pi(x) \nabla \log h(x)$$

où  $\nabla$  est l'opérateur gradient et  $m_\pi$  est la loi marginale associée à  $\pi$ .

**Preuve.** L'espérance a posteriori est donnée par

$$\begin{aligned} \mathbb{E}^\pi[\theta_i/x] &= \frac{\int_{\Theta} \theta_i h(x) e^{\theta \cdot x - \psi(\theta)} \pi(\theta) d\theta}{m_\pi(x)} \\ &= \left( \frac{\partial}{\partial x_i} \int_{\Theta} h(x) e^{\theta \cdot x - \psi(\theta)} \pi(\theta) d\theta \right) \frac{1}{m_\pi(x)} - \left( \frac{\partial}{\partial x_i} \right) \frac{1}{h(x)} \\ &= \frac{\partial}{\partial x_i} [\log m_\pi(x) - \log h(x)]. \end{aligned}$$

### 1.3.2 Fonction de perte et risque

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

- si le but est d'estimer  $\theta$  alors  $D = \Theta$
- pour un test,  $D = \{0, 1\}$

La fonction de perte est une fonction mesurable de  $(\Theta \times D)$  à valeurs réelles positives :  $L : \Theta \times D \rightarrow \mathbb{R}_+$ . Elle est définie selon le problème étudié et constitue l'armature du problème statistique.

**Définition 1.1.** Risque fréquentiste

Pour  $(\theta, \delta)$ , le risque fréquentiste est défini par :

$$\begin{aligned} R(\theta, \delta) &= E_\theta[L(\theta, \delta(x))] \\ &= \int_X L(\theta, \delta(x)) f(x|\theta) dx \end{aligned}$$

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

**Définition 1.2.** Risque a posteriori

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

$$\rho(\pi, \delta|x) = E^\pi[L(\theta, \delta)|x] \\ \int_{\Theta} L(\theta, \delta)\pi(\theta|x)d\theta$$

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

**Définition 1.3.** Risque intégré

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

$$r(\pi, \delta) = E^\pi[R(\theta, \delta)] \\ \int_{\Theta} R(\theta, \delta)d\pi(\theta)$$

**Définition 1.4.** Estimateur Bayésien

Un estimateur Bayésien est un estimateur vérifiant :

$$r(\pi, \delta^\pi) = \inf_{\delta \in D} r(\pi, \delta) < \infty \quad (1.1)$$

La valeur  $r(\pi, \delta^\pi)$  est alors appelée risque de Bayes.

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

**Théorème 1.1.** *Méthode de calcul (Rousseau, (2009))*

*Si  $\exists \delta \in D, r(\pi, \delta) < \infty$  et  $\forall x \in \chi \delta^\pi(x) = \arg \min_{\delta} \rho(\pi, \delta|x)$ , alors  $\delta^\pi(x)$  est un estimateur Bayésien.*

### 1.3.3 Fonctions de coût usuelles

#### 1.3.3.1 Perte quadratique

Introduit par Légende (1805) et Gauss (1810), ce coût est sans conteste le critère d'évaluation le plus commun. Fondant sa validité sur l'ambiguïté de la notion d'erreur dans un contexte statistique (soit erreur de mesure, soit variation aléatoire), il a aussi

donné lieu à de nombreuses critiques, la plus fréquente étant sans doute le fait que le coût quadratique

$$L(\theta, \delta) = (\theta - \delta)^2 \quad (1.2)$$

pénalise trop fortement les grandes erreurs.

les estimateurs de Bayes associés au coût quadratique sont les moyennes a posteriori. Cependant, notons que le coût quadratique n'est pas le seul coût à avoir cette caractéristique. Les fonctions de coût conduisant à la moyenne a posteriori comme estimateur de Bayes sont appelées *fonctions de coût propres* et ont été identifiées par Lindley (1985).

**Proposition 1.1.** (Christian P. Robert 2006)

*L'estimateur de Bayes  $\delta^\pi$  associé à la loi a priori  $\pi$  et au coût quadratique est la moyenne a posteriori*

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta|x] = \frac{\int_\theta \theta f(x|\theta)\pi(\theta)d\theta}{\int_\theta f(x|\theta)\pi(\theta)d\theta} \quad (1.3)$$

**Corollaire 1.1.** (Christian P. Robert 2006)

*Quand  $\Theta \in \mathbb{R}^p$ , l'estimateur de Bayes  $\delta^\pi$  associé à  $\pi$  et au coût quadratique,*

$$L(\theta, \delta) = (\theta - \delta)^t Q (\theta - \delta)$$

*est la moyenne a posteriori,  $\delta^\pi(x) = \mathbb{E}^\pi[\theta|x]$ , pour toute matrice  $Q$  ( $p \times p$ ) symétrique définie positive.*

Le coût quadratique est particulièrement intéressant lorsque l'espace des paramètres est borné et le choix d'un coût plus subjectif est impossible.

Le tableau ci-dessus représente quelques estimateurs de Bayes du paramètre  $\theta$  sous coût quadratique pour les lois a priori conjuguées des familles exponentielles usuelles.

Loi de x	Loi conjuguée	Moyenne a posteriori
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\frac{\mu\sigma^2 + \tau^2x}{\sigma^2 + \tau^2}$
Poisson $P(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + x}{\beta + 1}$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + \nu}{\beta + x}$
Binomiale $B(n, \theta)$	Beta $Be(\alpha, \beta)$	$\frac{\alpha + x}{\alpha + \beta + n}$
Binomiale Négative $Neg(m, \theta)$	Bêta $Be(\alpha, \beta)$	$\frac{\alpha + n}{\alpha + \beta + x + n}$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichlet $D(\alpha_1, \dots, \alpha_k)$	$\frac{\alpha_i + x_i}{(\sum_j \alpha_j) + n}$
Normale $\mathcal{N}(\mu, 1/\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + 1}{\beta + (\mu - x)}$ <sup>2</sup>

Tab 1.1-Quelques estimateurs de Bayes usuels.

### 1.3.3.2 L'erreur de coût absolu

Une solution alternative au coût quadratique en dimension une est d'utiliser le coût absolu,

$$L(\theta, d) = |\theta - d|$$

déjà considéré par Laplace (1773) ou, plus généralement, une fonction linéaire par morceaux

$$L_{k_1, k_2}(\theta, d) = \begin{cases} k_2(\theta - d) & \text{si } \theta > d \\ k_1(d - \theta) & \text{sinon} \end{cases}$$

**Proposition 1.2.** (Christian P. Robert 2006)

L'estimateur de Bayes associé à la loi a priori  $\pi$  et à la fonction de coût linéaire par morceaux est le fractile  $(k_2/(k_1 + k_2))$  de  $\pi(\theta|x)$ .

En particulier, si  $k_1 = k_2$ , dans le cas du coût absolu, l'estimateur de Bayes est la médiane a posteriori, qui est l'estimateur obtenu par Laplace.

### 1.3.3.3 Le coût 0 – 1

Ce coût est surtout utilisé dans l'approche classique des tests d'hypothèse, proposée par Neyman et Pearson. Plus généralement, c'est un exemple typique d'un coût non quantitatif. En effet, pour ce coût, la pénalité associée à un estimateur  $\delta$  est 0 si la réponse est correcte et 1 sinon.

**Exemple 1.3.** Soit le test de  $H_0 : \theta \in \Theta_0$  contre  $H_1 : \text{sinon}$ . Alors  $D = \{0, 1\}$ , où 1 représente l'acceptation de  $H_0$  et 0 son rejet, pour la fonction de coût 0 – 1, qui vaut

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

le risque associé est

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

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

**Proposition 1.3.** (Christian P. Robert 2006)

L'estimateur de Bayes associé à  $\pi$  et au coût 0 – 1 est

$$\delta^\pi(x) = \begin{cases} 1 & \text{si } P(\theta \in \Theta_0|x) > P(\theta \notin \Theta_0|x) \\ 0 & \text{sinon} \end{cases}$$

donc  $\delta^\pi(x)$  vaut 1 si et seulement si  $P(\theta \in \Theta_0|x) > 1/2$ .

### 1.3.4 Admissibilité et minimaxité

**Définition 1.5.** Estimateur randomisé

Pour le modèle  $X \in \{\mathcal{X}, \beta, \{P_\theta, \theta \in \Theta\}\}$ , un ensemble de décisions  $D$ , on définit  $D^*$  comme l'ensemble des probabilités sur  $D$ .  $\delta^* \in D^*$  est appelé estimateur randomisé.

L'idée à l'origine de cette notion est de rendre  $D$  convexe pour pouvoir maximiser facilement.

**Théorème 1.2.** (Christian P. Robert 2006)

Pour toute distribution a priori  $\pi$  sur  $\Theta$ , le risque de Bayes pour l'ensemble des estimateurs randomisés est le même que celui pour l'ensemble des estimateurs non randomisés, soit

$$\inf_{\delta \in D} r(\pi, \delta) = \inf_{\delta^* \in D^*} r(\pi, \delta^*) = r(\pi)$$

### 1.3.4.1 Minimaxité

Le critère de minimaxité apparaît comme une assurance contre le pire, car il vise à minimiser le coût moyen dans le cas le moins favorable. Il représente aussi un effort fréquentiste pour éviter de recourir au paradigme Bayésien, tout en engendrant un ordre (faible) sur  $D^*$ .

**Définition 1.6.** On appelle risque minimax

$$\bar{R} = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta) = \inf_{\delta \in D^*} \sup_{\theta} \mathbb{E}_{\theta}[L(\theta, \delta(x))] \quad (1.4)$$

et estimateur minimax tout estimateur  $\delta_0$  tel que

$$\bar{R} = \sup_{\theta} R(\theta, \delta_0) \quad (1.5)$$

L'estimateur minimax correspond au point de vue de faire le mieux dans le pire des cas, c'est-à-dire à s'assurer contre le pire. Il est utile dans des cadres complexes mais trop conservateur dans certains cas où le pire est très peu probable. Il peut être judicieux de voir l'estimation comme un jeu entre le statisticien (choix de  $\delta$ ) et la Nature (choix de  $\theta$ ), l'estimation minimax rejoint alors celle de la Théorie des Jeux.

#### *règle minimax et stratégie maximin*

Une difficulté importante liée à la notion de minimaxité est que les estimateurs minimax n'existent pas nécessairement. En particulier, il existe une stratégie minimax quand  $\Theta$  est fini et la fonction de coût est continue. Plus généralement, Brown (1976) (voir aussi Le Cam, 1986, et Strasser, 1985) considère l'espace de décision  $D$  comme plongé dans un autre espace de manière telle que l'ensemble des fonctions de risque sur  $D$  est compact dans ce grand espace. Dans cette perspective et sous des hypothèses supplémentaires, il est alors possible de construire des estimateurs minimax lorsque la fonction de coût est continue.

**Théorème 1.3.** (Christian P. Robert 2006)

*Si  $D \subset \mathbb{R}^k$  est convexe et compact et si  $L(\theta, d)$  est continue et convexe en tant que fonction de  $d$ , pour chaque  $\theta \in \Theta$ , alors, il existe un estimateur minimax non randomisé.*

**Lemme 1.2.** (Rousseau (2009))

*Le risque de Bayes est toujours plus petit que le risque minimax,*

$$\underline{R} = \sup_{\pi} r(\pi) = \sup_{\pi} \inf_{\delta \in D} r(\pi, \delta) \leq \bar{R} = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta)$$

La première valeur est dite *risque maximin* et une distribution  $\pi^*$  telle que  $r(\pi^*) = \underline{R}$  est appelée *distribution a priori la moins favorable*, quand de telles distributions existent. En général, la borne supérieure  $r(\pi^*)$  est atteinte plutôt par une distribution impropre pouvant s'exprimer comme une limite de distributions a priori propre  $\pi_n$ . Mais ce phénomène n'empêche pas nécessairement la construction d'estimateurs minimax. Quand elles existent, les distributions les moins favorables sont celles qui ont le risque de Bayes le plus grand, donc aussi les moins intéressantes en terme de coût lorsqu'elles ne sont pas suggérées par l'information a priori disponible. Le résultat ci-dessus est assez logique au sens où l'information a priori ne peut qu'améliorer l'erreur d'estimation, même dans le pire des cas.

**Définition 1.7.** Un problème d'estimation est dit *admettre une valeur* si  $\underline{R} = \overline{R}$ , c'est-à-dire quand

$$\sup_{\pi} \inf_{\delta \in D} r(\pi, \delta) = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta)$$

Quand le problème admet une valeur, certains estimateurs minimax sont des estimateurs de Bayes correspondant aux lois a priori les moins favorables. Cependant, ils peuvent être randomisés. Par conséquent le principe minimax ne fournit pas toujours des estimateurs acceptables.

**Lemme 1.3.** (Christian P. Robert 2006)

Si  $\delta_0$  est un estimateur de Bayes pour  $\pi_0$  et si  $R(\theta, \delta_0) \leq r(\pi_0)$  pour tout  $\theta$  dans le support de  $\pi_0$ ,  $\delta_0$  est minimax et  $\pi_0$  est la distribution la moins favorable.

### 1.3.4.2 Admissibilité

**Définition 1.8.** Estimateur admissible

Soit  $X \in \{\chi, \beta, \{P_\theta, \theta \in \Theta\}\}$  un modèle paramétrique et  $L$  une fonction de perte sur  $\Theta \times D$  où  $D$  est l'ensemble des décisions. On dit que  $\delta \in \Theta$  est inadmissible si et seulement si  $\exists \delta_0 \in D, \forall \theta \in \Theta, R(\theta, \delta) \geq R(\theta, \delta_0)$  et  $\exists \theta_0 \in \Theta, R(\theta_0, \delta) > R(\theta_0, \delta_0)$ . Dans le cas contraire,  $\delta$  est admissible.

**Proposition 1.4.** (Christian P. Robert 2006)

S'il existe un unique estimateur minimax, cet estimateur est admissible.

Notons que la réciproque de ce résultat est fausse, car il peut exister plusieurs estimateurs minimax admissibles. Par exemple, dans le cas  $N_p(\theta, I_p)$ , il existe des estimateurs de Bayes réguliers minimax pour  $p \geq 5$ . Quand la fonction de coût  $L$  est absolument convexe (en d), la caractérisation suivante est aussi possible.

**Proposition 1.5.** (Rousseau (2009))

Si  $\delta_0$  est admissible de risque constant,  $\delta_0$  est l'unique estimateur minimax.

**Théorème 1.4.** Estimateurs Bayésiens admissibles, (Rousseau (2009))

Si l'estimateur Bayésien  $\delta^\pi$  associé à une fonction de perte  $L$  et une loi a priori  $\pi$  est unique, alors il est admissible

**Proposition 1.6.** (Christian P. Robert 2006)

Si un estimateur de Bayes,  $\delta^\pi$ , associé à une loi a priori (propre ou impropre)  $\pi$ , est tel que le risque de Bayes,

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

soit fini,  $\delta^\pi$  est admissible.

**Définition 1.9.**  $\pi$ -admissibilité

Un estimateur  $\delta_0$  est  $\pi$ -admissible si et seulement si

$$\forall(\delta, \theta), R(\theta, \delta) \leq R(\theta, \delta_0) \Rightarrow \pi(\{\theta \in \Theta, R(\theta, \delta) < R(\theta, \delta_0)\}) = 0$$

**Propriété 1.1.** (Christian P. Robert 2006)

Tout estimateur Bayésien tel que  $r(\pi) < \infty$  est  $\pi$ -admissible

**Théorème 1.5.** Continuité et  $\pi$ -admissibilité, (Rousseau (2009))

Si  $\pi > 0$  sur  $\Theta$ ,  $r(\pi) < \infty$  pour une fonction de perte  $L$  donnée, si  $\delta^\pi$  estimateur Bayésien correspondant existe et si  $\theta \mapsto R(\theta, \delta)$  est continu, alors  $\delta^\pi$  est admissible.

## 1.4 Choix des lois a priori

Un grand intérêt de la théorie Bayésienne est sa grande cohérence et sa méthodologie unifiée. Ainsi, données les lois a priori et a posteriori ainsi que la fonction de perte suffit pour déterminer, entre autres, un estimateur optimal. Le choix de la loi a priori  $\pi$  est donc crucial. Avec beaucoup d'observations, le comportement asymptotique peut guider ce choix mais sinon il est nécessaire de le justifier avec précision.

### 1.4.1 Approche partiellement informative

#### 1.4.1.1 Maximum d'entropie

Si l'on possède des informations partielles du type  $\mathbb{E}^\pi[g_k(\theta)] = \mu_k$  où pour chaque  $k = 1, \dots, n$ ,  $g_k$  est une fonction donnée.

Pour  $\theta \in \{1, \dots, n\}$  et  $\pi(\theta) = (\pi_1, \dots, \pi_n)$  tel que  $\pi_i > 0$  et  $\sum_{i=1}^n \pi_i = 1$ , l'entropie de la loi est définie par

$$Ent(\pi) = - \sum_{i=1}^n \pi_i \log(\pi_i) \leq - \sum_{i=1}^n \frac{1}{n} \log\left(\frac{1}{n}\right) = \log n \quad (1.6)$$

Ce dernier terme correspond à une répartition uniforme. Pour la masse de Dirac  $\delta(j)$  (telle que  $\pi_j = 1$  et  $\forall i \neq j, \pi_i = 0$ ),  $Ent(\delta(j)) = 0$  ce qui correspond à l'intuition puisqu'alors il n'y a plus d'incertitude et l'information est totale. Une entropie petite s'interprète comme une loi concentrée et informative. La maximisation de l'entropie sous les contraintes permet de chercher la loi qui apporte le moins d'information. Le principe à la base de cette méthode est donc de chercher à calculer :

$$\arg \max_{\pi} Ent(\pi) \quad \text{sous la contrainte} \quad \mathbb{E}^{\pi}[g_k(\theta)] = \mu_k$$

La solution de ce problème est alors donnée par :

$$\pi^* \propto e^{\sum_{k=1}^n \lambda_k g_k(\theta)}$$

où les  $\lambda_k$  sont les multiplicateurs de Lagrange associés. Dans la pratique, on détermine ces valeurs  $\lambda$  à partir des contraintes (systèmes d'équations) comme l'indique l'exemple à suivre.

**Exemple 1.4.** *Un cas dénombrable*

Ici,  $\Theta = \mathbb{N}$  et  $\mathbb{E}^{\pi}[\theta] = x > 1$ , c'est-à-dire qu'ici  $g(\theta) = \theta$  et  $\mu = x$ . On sait que  $\pi^* \propto e^{\lambda\theta}$  et que  $\lambda$  est déterminé par :

$$\frac{\sum_{\theta \in \mathbb{N}} \theta e^{\lambda\theta}}{\sum_{\theta \in \mathbb{N}} e^{\lambda\theta}} = x$$

Cela conduit à résoudre :

$$\frac{x}{1 - e^{\lambda}} = \frac{1}{e^{\lambda}} \frac{e^{\lambda}}{(1 - e^{\lambda})^2}$$

$$e^{\lambda} = \frac{x - 1}{x}$$

Par exemple si  $x = \frac{12}{11}$  alors  $\lambda = -\log(12)$

En continu, il n'est pas possible de définir l'entropie comme ci-dessus puisqu'on ne peut dénombrer les états (pas de mesure de comptage) en l'absence de mesure de référence. Dans le cas continu, on définit alors l'équivalent de l'entropie par rapport à une mesure  $\pi_0$  :

$$Ent(\pi | \pi_0) = \int_{\Theta} \pi(\theta) \log\left(\frac{\pi(\theta)}{\pi_0(\theta)}\right) d\theta$$

C'est en fait la divergence de Kullback. Dans l'idée  $\pi_0$  est la plus plate possible, la plus proche de la répartition uniforme. L'objectif est donc de maximiser  $Ent(\pi | \pi_0)$  sous les contraintes  $\mathbb{E}^\pi[g_k(\theta)] = \mu_k$ . Là encore, la solution générale est connue :

$$\pi^*(\theta) \propto e^{\sum_{k=1}^n \lambda_k g_k(\theta)} \pi_0(\theta)$$

#### 1.4.1.2 Familles conjuguées

On considère une variable  $x$  suivant une fonction de densité  $f(x | \theta)$

**Définition 1.10.** Famille conjuguée

Une famille  $F$  de distributions de probabilité sur  $\Theta$  est dite conjuguée (ou fermée par échantillonnage) par une fonction de vraisemblance  $f(x|\theta)$  si, pour tout  $\pi \in F$ , la distribution a posteriori  $\pi(\cdot|x)$  appartient également à  $F$ .

Un exemple trivial d'une famille conjuguée est l'ensemble  $F_0$  de toutes les lois de probabilité sur  $\Theta$ . L'avantage des familles conjuguées est avant tout de simplifier les calculs. Avant l'essor du calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire aboutir des calculs.

Les lois a priori conjuguées sont généralement associées à un type particulier de lois d'échantillonnage qui permet toujours leur obtention ; il est même caractéristique des lois a priori conjuguées comme nous le verrons ci-dessous. Ces lois constituent ce qu'on appelle des *familles exponentielles*.

**Définition 1.11.** Familles exponentielles

Soient  $\mu$  une mesure  $\sigma$ -finie sur  $\chi$ ,  $\Theta$  l'espace des paramètres,  $C$  et  $h$  des fonctions respectivement de  $\chi$  et  $\Theta$  dans  $\mathbb{R}_+$ , et  $R$  et  $T$  des fonctions de  $\Theta$  et  $\chi$  dans  $\mathbb{R}^k$ . La famille des distributions de densité (par rapport à  $\mu$ )

$$f(x|\theta) = C(\theta)h(x) \exp\{R(\theta).T(x)\} \quad (1.7)$$

est dite famille exponentielle de dimension  $k$ . Dans le cas particulier où  $\Theta \subset \mathbb{R}^k, \chi \subset \mathbb{R}^k$  et

$$f(x|\theta) = C(\theta)h(x) \exp\{\theta.x\} \quad (1.8)$$

la famille est dite *naturelle*.

D'un point de vue analytique, les familles exponentielles ont certaines caractéristiques intéressantes, pour tout échantillon de (1.8), en particulier, il existe une statistique exhaustive de dimension constante, en effet, si  $x_1, \dots, x_n \sim f(x|\theta)$ , avec  $f$  satisfaisant (1, 8),

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \in \mathbb{R}^k$$

est exhaustive pour tout  $n$ . La réciproque de ce résultat a été aussi établie par Koopman (1936) et Pitman (1936).

**Théorème 1.6.** (*Lemme de Pitman-Koopman*) (Christian P. Robert 2006)

*Si une famille de lois  $f(\cdot|x)$  à support constant est telle que, à partir d'une taille d'échantillon suffisamment grande, il existe une statistique exhaustive de taille fixe, la famille est exponentielle.*

**Exemple 1.5.** Soit  $x \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$  alors,

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sigma^p} \frac{1}{(2\pi)^{p/2}} \exp\left\{-\sum_{i=1}^p (x_i - \theta_i)^2 / 2\sigma^2\right\} \\ &= C(\theta, \sigma) h(x) \exp\{x \cdot (\theta/\sigma^2) + \|x\|^2 (-1/2\sigma^2)\} \end{aligned}$$

et la distribution normale appartient à une famille exponentielle de paramètres naturels  $\theta/\sigma^2$  et  $-1/2\sigma^2$ . De la même façon, si  $x_1, \dots, x_n \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ , la distribution jointe satisfait

$$f(x_1, \dots, x_n) = C'(\theta, \sigma) h'(x_1, \dots, x_n) \times \exp\{n\bar{x} \cdot (\theta/\sigma^2) + \sum_{i=1}^n \|x_i - \bar{x}\|^2 (-1/2\sigma^2)\}$$

et la statistique  $\bar{x}, \sum_i \|x_i - \bar{x}\|^2$  est exhaustive pour tout  $n \geq 2$ .

**Définition 1.12.** Soit  $f(x|\theta) = C(\theta)h(x) \exp\{\theta \cdot x\}$ , une famille exponentielle naturelle. L'espace naturel des paramètres est

$$N = \left\{ \theta; \int_{\mathcal{X}} e^{\theta \cdot x} h(x) d\mu(x) < +\infty \right\}$$

La famille est dite régulière si  $N$  est un ensemble ouvert et minimale si  $\dim(N) = \dim(K) = k$ , où  $K$  est la clôture de l'enveloppe convexe du support de  $\mu$ .

**Remarque 1.1.** Les familles exponentielles naturelles peuvent aussi être réécrites sous la forme

$$f(x|\theta) = h(x) e^{\theta \cdot x - \psi(\theta)} \quad (1.9)$$

et  $\psi(\theta)$  est dite *fonction cumulante des moments*.

### Lois conjuguées des familles exponentielles

Soit  $f(x|\theta) = h(x)e^{\theta \cdot x - \psi(\theta)}$ , loi générique d'une famille exponentielle. Cette loi admet alors une famille conjuguée.

**Proposition 1.7.** (Christian P. Robert 2006)

Une famille conjuguée pour  $f(x|\theta)$  est donnée par

$$\pi(\theta|\mu, \lambda) = K(\mu, \lambda)e^{\theta \cdot \mu - \lambda\psi(\theta)} \quad (1.10)$$

où  $K(\mu, \lambda)$  est la constante de normalisation de la densité. La loi a posteriori correspondante est  $\pi(\theta|\mu + x, \lambda + 1)$ .

Le tableau ci-dessus représente quelques lois a priori conjuguées naturelles pour quelques familles exponentielles usuelles.

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\varrho(\sigma^2\mu + \tau^2x), \varrho\sigma^2\tau^2)$
Poisson $P(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + x, \beta + 1)$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + \nu, \beta + x)$
Binomiale $B(n, \theta)$	Beta $Be(\alpha, \beta)$	$Be(\alpha + x, \beta + n - x)$
Binomiale Négative $Neg(m, \theta)$	Bêta $Be(\alpha, \beta)$	$Be(\alpha + m, \beta + x)$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichlet $D(\alpha_1, \dots, \alpha_k)$	$D(\alpha_1 + x_1, \dots, \alpha_k + x_k)$
Normale $\mathcal{N}(\mu, 1/\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + 0.5, \beta + (\mu - x)^2/2)$

Tab 1.2-Lois a priori conjuguées usuelles.

### 1.4.2 Approche non informative

Lorsque aucune information a priori n'est disponible, le choix de la loi a priori est analytique, puisqu'elles donnent des expressions exactes pour quelques quantités a posteriori. Dans de telles situations, il est impossible de justifier le choix d'une loi a priori sur des bases subjectives. Plutôt que de revenir aux alternatives classiques, comme l'estimation par

maximum de vraisemblance, ou d'utiliser les données pour approcher ces hyperparamètres, comme dans une analyse Bayésienne empirique, il est préférable de faire appel à des techniques Bayésiennes, ne serait-ce que parce qu'elles sont à la base des critères classiques d'optimalité. Dans un tel cas, ces lois a priori particulières doivent être construites à partir de la distribution d'échantillonnage, puisque c'est la seule information disponible. Pour des raisons évidentes, de telles lois sont dites *non informatives*.

#### 1.4.2.1 Les lois a priori de Laplace

Laplace fut le premier à utiliser des techniques non informatives puisque, bien que ne disposant pas d'information, il munit ces paramètres d'une loi a priori qui prend en compte son ignorance en donnant la même vraisemblance à chaque valeur du paramètre, soit donc en utilisant une loi uniforme. Son raisonnement, appelé plus tard principe de la raison insuffisante, se fondait sur l'équiprobabilité des événements élémentaires.

Trois critiques ont été plus tard avancées sur ce choix. Premièrement, les lois résultantes sont impropres quand l'espace des paramètres n'est pas compact et certains statisticiens se refusent à utiliser de telles lois, car elles mènent à des difficultés comme le paradoxe de marginalisation. Deuxièmement, le principe des événements équiprobables de Laplace n'est pas cohérent en termes de partitionnement : si  $\Theta = \{\theta_1, \theta_2\}$ , la règle de Laplace donne  $\pi(\theta_1) = \pi(\theta_2) = 1/2$  mais, si la définition de  $\Theta$  est plus détaillée, avec  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , la règle de Laplace mène à  $\pi(\theta_1) = 1/3$ , ce qui évidemment n'est pas cohérent avec la première formulation, cette cohérence n'est pas un problème important : il peut être évacué en argumentant que le niveau de partitionnement doit être fixé à un certain stade de l'analyse et que l'introduction d'un degré plus fin dans le partitionnement modifie le problème d'inférence.

La troisième critique est plus fondamentale, car elle concerne le problème de l'invariance par reparamétrisation. Si on passe de  $\theta \in \Theta$  à  $\eta = g(\theta)$  par une transformation bijective  $g$ , l'information a priori reste totalement inexistante et ne devrait pas être modifiée. Cependant, si  $\pi(\theta) = 1$ , la loi a priori sur  $\eta$  est :

$$\pi^*(\eta) = \left| \frac{d}{d\eta} g^{-1}(\eta) \right|$$

### 1.4.2.2 Loi a priori de Jeffreys

les lois a priori non informatives de Jeffreys sont fondées sur l'information de Fisher, donnée par

$$I(\theta) = \mathbb{E}_\theta \left[ \left( \frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right]$$

dans le cas unidimensionnel.

Sous certaines conditions, cette information est aussi égale à

$$I(\theta) = -\mathbb{E}_\theta \left[ \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right]$$

La loi a priori de Jeffreys est

$$\pi_j \propto \sqrt{|I(\theta)|}$$

Ainsi la loi a priori de Jeffreys est invariante par reparamétrisation. Malgré l'immense intérêt d'une telle propriété, il faut savoir que la loi a priori de Jeffreys n'a de bonnes propriétés que dans le cas des petites dimensions et en particulier de la dimension 1. Vérifions que c'est effectivement une loi de probabilité invariante par reparamétrisation : si  $\eta = g(\theta)$  avec  $g \in \mathcal{C}^1$ , alors  $\pi_j(\eta) = \sqrt{|I(\eta)|}$ . En outre,  $\tilde{I}(\eta) = \nabla(\theta)' I(\theta) \nabla g(\theta)$ , il s'en suit que dans le cas de la dimension 1, on a :

$$\pi_j(\eta) = \sqrt{|I(\eta)|} = \sqrt{|I(\theta)|} \left| \frac{d\theta}{d\eta} \right| = \pi_j(\theta) \left| \frac{d\theta}{d\eta} \right|$$

Il existe une variante de l'approche précédente : la loi de référence de Bernardo dont l'idée est de reposer sur un critère d'objectivité. On regarde ici la loi a priori qui amène le moins d'information par rapport à ce que pouvait fournir les données. On s'intéresse donc à la fonction de  $\pi$ , la loi a priori, suivante :

$$I_n(\pi) = \int K(\pi(\theta|x^n); \pi(\theta)) m_\pi(x^n) dx^n$$

En outre, on peut vérifier que :

$$I_n(\pi) \sim_{n \rightarrow \infty} \frac{d}{2} \log \frac{n}{2\pi} - \int \pi(\theta) \log \frac{\pi}{\sqrt{|I(\theta)|}} d\theta + o_p(1)$$

Dans le cas d'un modèle régulier  $I_n(\pi)$  est donc maximal lorsque  $n$  tend vers  $+\infty$  si  $\pi(\theta) \propto \sqrt{|I(\theta)|}$ .

Bernardo introduit la stratégie suivante avec un paramètre  $\theta$  de la forme  $\theta = (\theta_1, \theta_2)$  où  $\theta_1$  est un paramètre d'intérêt et  $\theta_2$  est un paramètre de nuisance. On raisonne alors

séquentiellement en écrivant :  $\pi(\theta_1, \cdot) = \pi(\cdot|\theta_1)\pi(\theta_1)$  et en prenant pour  $\theta_1$  la loi de Jeffreys. Cette manière de faire peut se généraliser si  $\theta = (\theta_1, \dots, \theta_n)$  où l'on a ordonné sans perte de généralité les  $\theta_i$  par intérêt croissant. Un algorithme consiste à prendre pour  $\theta_1$  la loi de Jeffreys, puis de calculer  $\pi(\theta_2|\theta_1)$  la plus objective puis  $\pi(\theta_3|\theta_1, \theta_2)$  la plus objective, etc. Cependant, il est clair ce raisonnement n'est pas purement objectif parce que donner plus d'importance à un paramètre qu'à un autre relève une fois encore d'un choix.

**Exemple 1.6.** Soit  $x \sim \mathcal{B}(n, p)$ ,

$$\begin{aligned} f(x/p) &= C_n^x p^x (1-p)^{n-x} \\ \frac{\partial^2 f(x/p)}{\partial^2 p^2} &= \frac{x}{p^2} + \frac{n-x}{(1-p)^2} \\ I(p) &= n \left[ \frac{1}{p} + \frac{1}{p-1} \right] = \frac{n}{p(1-p)} \end{aligned}$$

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  $\mathcal{Be}(1/2, 1/2)$ .

### 1.4.2.3 Loi a priori de concordance (matching priors)

Le but est de trouver une loi a priori concernant le paramètre  $\theta$  qui se rapproche le plus possible de la méthode de choix fréquentiste, cela revient à faire en sorte que le tirage  $x$  n'influence pas le résultat.

On rappelle dans un premier temps des exemples d'une région de confiance ou  $\alpha$ -crédible. Elle peut être par exemple un intervalle unilatéral  $\{\theta \leq \theta_d^{(x)}\}$  ou bien bilatéral  $\{\theta_{\alpha,1} \leq \theta \leq \theta_{\alpha,2}\}$ . Il peut s'agir aussi de région HPD,  $\theta \in \mathcal{C}_\alpha^\pi$  avec par exemple  $\{\log(\hat{\theta}) - \log(\theta) \leq h_\alpha\}$  tel que  $P^\pi(\theta \in \mathcal{C}|x) = 1 - \alpha$ .

On cherche  $\pi$  tel que  $\forall \theta; P_\theta(\theta \in \mathcal{C}) = 1 - \alpha$ , appelé la parfaite concordance. C'est en général impossible. On va alors chercher  $r_n$  le plus petit possible tel que :

$$\forall \theta \in \Theta; \forall \alpha \in ]0, 1[, P_\theta(\theta \in \mathcal{C}) = 1 - \alpha + o(r_n)$$

La loi a priori est alors dite concordante à l'ordre  $r_n$ .

## 1.5 Méthodes de calcul Bayésien

La simplicité ultime de l'approche Bayésienne est que, pour une fonction de coût et une loi a priori données, l'estimation Bayésienne associée à une observation  $x$  est la décision  $d$  minimisant le coût a posteriori, dans la pratique, minimiser un tel coût peut être difficile car le calcul explicite de la loi a posteriori peut être impossible des fois.

### 1.5.1 Méthodes classiques d'approximation

#### 1.5.1.1 Intégration numérique

A partir de la simple méthode de Simpson, plusieurs approches ont été conçues en mathématiques appliquées pour l'approximation numérique d'intégrales. Par exemple, la quadrature polynômiale est censée approcher les intégrales liées à des distributions proches de la loi normale. L'approximation de base est donnée par

$$\int_{-\infty}^{+\infty} e^{-t^2/2} f(t) dt \approx \sum_{i=1}^n \omega_i f(t_i)$$

où

$$\omega_i = \frac{2^{n-1} n! \sqrt{n}}{n^2 [H_{n-1}(t_i)]^2}$$

et  $t_i$  est le  $i$ -ième zéro du  $n$ -ième polynôme d'Hermite  $H_n(t)$ .

D'autres approximations d'intégrales reliées à la méthode précédente sont disponibles, qui reposent sur différentes bases orthogonales classiques (voir Abramowitz et Stegun, 1964) mais ces méthodes requièrent généralement des hypothèses de régularité sur la fonction  $f$ , ainsi que des études préliminaires pour déterminer quelle base est la plus adéquate et à quel point cette approximation est précise. Par exemple, des transformations du modèle peuvent être nécessaires pour mettre en pratique l'approximation d'Hermite (voir Naylor et Smith, 1982, et Hills et Smith, 1992).

**Remarque 1.2.** Quelle que soit la méthode d'intégration numérique utilisée, sa précision diminue dramatiquement lorsque la dimension de  $\Theta$  augmente. De façon plus spécifique, l'erreur associée aux méthodes numériques se comporte comme une puissance de la dimension de  $\Theta$ . En pratique, une règle empirique est que la plupart des méthodes standard ne devraient pas être utilisées pour l'intégration en dimension supérieure à 4. En effet, la

taille de la partie de l'espace non pertinente pour le calcul d'une intégrale donnée augmente considérablement avec la dimension de l'espace. Ce problème est appelé fléau de la dimension.

### 1.5.1.2 Les méthodes de Monte Carlo

Dans un problème statistique, l'approximation de l'intégrale

$$\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) d\theta \quad (1.11)$$

doit tirer avantage de la nature particulière, à savoir le fait que  $\pi$  soit une densité de probabilité ou plutôt, que  $f(x|\theta)\pi(\theta)$  soit proportionnel à une densité. Une conséquence naturelle de cette perspective est d'utiliser la méthode de Monte Carlo, introduite par Metropolis et Ulam (1949) et von Neumann (1951). Par exemple, s'il est possible de produire des variables aléatoires  $\theta_1, \dots, \theta_m$  de loi  $\pi(\theta)$ , la moyenne

$$\frac{1}{m} \sum_{i=1}^m g(\theta_i) f(x|\theta_i) \quad (1.12)$$

converge (presque sûrement) vers (1.11) lorsque  $m$  tend vers  $\infty$ , selon la Loi des Grands Nombres. De la même façon, si un échantillon iid de  $\theta_i$  de  $\pi(\theta|x)$  peut être simulé, la moyenne

$$\frac{1}{m} \sum_{i=1}^m g(\theta_i) \quad (1.13)$$

converge vers

$$\frac{\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) d\theta}{\int_{\Theta} f(x|\theta) \pi(\theta) d\theta} \quad (1.14)$$

## 1.5.2 Méthodes de Monte Carlo par chaînes de Markov (MCMC)

Les méthodes de Monte-Carlo par chaîne de Markov (MCMC) génèrent une suite de variables aléatoires  $(\theta^1, \dots, \theta^n, \dots)$  et, hormis la première à laquelle on donne une valeur arbitraire, chacune d'entre elles dépend uniquement de celle qui la précède. Les calculs sont ensuite poursuivis en appliquant à cette séquence une loi des grands nombres pour les chaînes markoviennes ergodiques de forme identique.

### 1.5.2.1 Algorithme Hasting-Metropolis

Pour  $\theta^{(0)}$  est une valeur initiale, on définit par récurrence les valeurs de  $\theta^{(t)}$ . A l'étape  $t$ , à partir de  $\theta^{(t-1)}$ ,  $\theta^{(t)}$  est construit en tirant un  $\theta'$  à l'aide d'une distribution

de probabilité instrumentale :  $\theta' \sim q(\cdot|\theta^{(t-1)})$ .  $\theta^{(t)}$  est alors donné par :

$$\theta^{(t)} = \begin{cases} \theta' & \text{avec une probabilité } \alpha(\theta', \theta^{(t-1)}) \\ \theta^{(t-1)} & \text{avec une probabilité } 1 - \alpha(\theta', \theta^{(t-1)}) \end{cases}$$

où  $\alpha(\theta', \theta^{(t-1)}) = \min\left(\frac{\pi(\theta')q(\theta^{(t-1)}|\theta')}{\pi(\theta^{(t-1)})q(\theta'|\theta^{(t-1)})}, 1\right)$ .

La loi de densité  $\pi(\theta)$  est souvent appelée *loi cible* ou loi objet, tandis que la loi de densité  $q(\cdot|\theta)$  est dite *loi de proposition*. Une propriété stupéfiante de cet algorithme est d'autoriser un nombre infini de lois de proposition produisant toute une chaîne de Markov convergeant vers la loi d'intérêt.

**Remarque 1.3.** Notons qu'il est possible suivant cette construction de rester au même endroit après une itération. On peut alors montrer en écrivant la condition de balance, que pour ce choix de  $\alpha$ , on obtient une chaîne de Markov de loi stationnaire  $\pi$ .

Cette chaîne de Markov est ergodique si et seulement si  $(\theta^{(t)})_t$  est irréductible et apériodique.

### 1.5.2.2 L'échantillonnage de Gibbs

Le second groupe de méthodes de Monte Carlo par chaînes de Markov (MCMC) est encore appelé échantillonnage de Gibbs. Plus intuitif pour certains praticiens, il ne demande pas de mettre en place une fonction d'exploration de l'espace des états de la nature. De plus, algorithme de Gibbs pour l'estimation et construction de modèle par conditionnement probabiliste. De fait, les méthodes de Gibbs utilisent plus complètement que ne le font les méthodes de Metropolis-Hastings, les structures conditionnelles des modèles.

Pour  $\theta = (\theta_1, \dots, \theta_p)$ , on veut simuler  $\pi(\theta)$  à partir de  $\pi_i(\theta_i|\theta_{(-i)}) = \pi_i(\theta_i|\theta_j, j \neq i)$  pour tout  $i$ . On initialise avec  $\theta^{(0)}$  et à l'instant  $t$ , on écrit :

$$\begin{aligned} (\theta_1^{(t)}|\theta^{(t-1)}) &\sim \pi_1(\theta_1^{(t)}|\theta_{(-1)}^{(t-1)}) \\ (\theta_2^{(t)}|\theta^{(t-1)}, \theta_1^{(t)}) &\sim \pi_2(\theta_2^{(t)}|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}) \\ (\theta_p^{(t)}|\theta^{(t-1)}, \theta_{(-p)}^{(t)}) &\sim \pi_p(\theta_p^{(t)}|\theta_{(-p)}^{(t)}) \end{aligned}$$

Dans le cas où une telle loi  $\pi$  existe,  $\theta^{(t)}$  issu de cet algorithme est une chaîne de Markov ergodique de loi stationnaire  $\pi$ .

## 1.6 Avantages de l'approche Bayésienne

– La vision Bayésienne des probabilités, nous semble plus cohérente que les théories fréquentistes qui interprètent la probabilité comme une proportion limite déterminée sur

la base d'une séquence infinie d'expériences (Von Mises et Geiringer, 1964). D'ailleurs on associe souvent une probabilité à des événements qui par définition ne sont pas répétables, (par exemple l'argumentaire de Hartigan (1983) autour de l'assertion, qui n'est heureusement plus d'actualité : "*Je pense que la probabilité qu'un conflit nucléaire éclate entre Etats Unis et Union Soviétique avant l'année 2000 est de 0.05*").

L'interprétation Bayésienne de la probabilité (De Finetti, 1937, 1974) est associée à une notion de pari rationnel : la probabilité attribuée à un événement est définie par les conditions auxquelles un individu rationnel est prêt à parier sur la réalisation de tel événement. La rationalité de l'individu, est nécessaire pour éviter que la définition de la probabilité soit arbitraire et est décrite par certaines règles de comportement, face à l'incertitude (Savage, 1972).

– La question de choisir une interprétation Bayésienne ou fréquentiste de la probabilité n'est pas uniquement philosophique. Il est possible de combiner les informations objectives, apportées par les données, avec des informations extérieures à l'échantillon observé et la formule de Bayes est l'instrument qui rend possible ce couplage dans un cadre cohérent basé uniquement sur le calcul des probabilités (Bernardo et Smith, 1994). Le rôle de la formule de Bayes comme trait d'union entre la vision du modélisateur et l'évidence des résultats expérimentaux est souligné, entre autres, par Box et Tiao (1973), Berry et al.(1996), Press et Tanur (2000).

Cette possibilité d'intégrer des éléments extérieurs aux échantillons est très adaptée à la pratique technique. Dans le monde réel, le statisticien est normalement confronté à des données peu représentatives ou incohérentes mais en revanche il dispose de l'avis technique des experts qui, sur la base de leur expérience et savoir-faire, sont capables de donner des informations complémentaires de grande utilité, qu'il serait dommage de ne pas prendre en compte (Cullen et Frey, 1999), (Bernier et al., 2000), (Perreault, 2000).

– L'analyse Bayésienne fournit des résultats d'interprétation plus directe que ceux de la statistique classique. L'exemple le plus flagrant est la définition de l'intervalle de confiance du paramètre  $\theta$  d'un modèle. Pour les Bayésiens, qui préfèrent parler plutôt d'intervalle de crédibilité<sup>3</sup>, il s'agit de l'intervalle qui contient le paramètre avec une probabilité donnée. Par exemple l'intervalle de crédibilité a posteriori à 95% est typiquement celui

---

3. Autres terminologies d'usage moins courant : Région à plus grande probabilité ou HDR, (Lee, 1997), intervalle Bayésien de confiance (Lindley, 1965)

délimite inférieurement par le percentile d'ordre 2.5% et supérieurement par le percentile d'ordre 97.5%. Dans l'inférence classique cette assertion n'est plus vraie parce que le paramètre (inconnu) du modèle n'est pas une variable aléatoire mais une grandeur constante. L'interprétation correcte de l'intervalle de confiance est que, si on imagine l'ensemble des échantillons aléatoires pouvant être obtenus à partir du modèle, paramétré par  $\theta$ , 95% des intervalles de confiance calculés (sur la base des différents échantillons) contiennent la vraie valeur du paramètre. L'interprétation Bayésienne, décidément plus naturelle, est d'ailleurs celle de la plupart des praticiens qui font de l'inférence Bayésienne... sans le savoir (Leconte et Poitevineau, 1996).

– Les résultats de l'inférence bayésienne sont plus riches que les estimateurs fournis par les techniques classiques d'inférences (Berger, 1985). Les techniques Bayésiennes permettent d'obtenir la loi jointe des paramètres du modèle et donc de prendre en compte simultanément l'effet de l'incertitude globale sur l'ensemble des paramètres inconnus sur les prévisions futures du comportement du système étudié et sur les décisions suggérées par ce comportement (Krzysztofowicz, 1983).

– Enfin, les techniques d'inférence par méthode MCMC, relativement faciles à mettre en oeuvre, sont très adaptées à l'estimation de modèles complexes à plusieurs paramètres ou à structure hiérarchique. L'estimation de ces modèles avec la technique usuelle du Maximum de Vraisemblance se révèle parfois nettement plus compliquée. L'argument de la commodité opérationnelle qui a longtemps joué contre l'approche Bayésienne commence aujourd'hui à peser dans le sens opposé dans la vieille querelle entre fréquentistes et Bayésiens (Robert, 1992).

## 1.7 Conclusion

La pertinence et l'efficacité de l'approche Bayésienne, comme guide du raisonnement scientifique face à l'incertitude, sont reconnues depuis longtemps (de Finetti, 1937; Savage, 1954; etc.). La mise en oeuvre des principes Bayésiens, en dehors de cas d'école, s'est longtemps heurtée aux difficultés pratiques de calcul. Les moyens informatiques, dont on disposait avant les années 1990, étaient insuffisamment puissants et trop centralisés. Les problèmes réels, avec leurs dimensions et leurs complexités importantes, faisaient alors la part belle aux méthodes statistiques classiques. La situation, depuis lors, a subi une

véritable révolution. Maintenant on peut affirmer qu'il n'existe, au moins a priori, aucun contre-argument justifie a l'emploi des méthodes Bayésiennes quelle que soit la complexité du cas envisagé. On doit ce nouveau paysage scientifique au développement de nouveaux outils de calcul : les méthodes MCMC (simulations Monte Carlo par Chaînes de Markov) et à l'amélioration des anciens (échantillonnage pondéré ou importance sampling et méthodes des particules), et a leur relance par la puissance nouvelle de la micro-informatique décentralisée. Les fondements conceptuels de ces méthodes de calcul sont solidaires des modes de raisonnements conditionnels de la modélisation Bayésienne et le paradigme Bayésien apparaît comme une démarche rationnelle, efficace et solidement intégrée du programme complet : modélisation  $\rightarrow$  calcul  $\rightarrow$  décision.

# Chapitre 2

## Modèles dynamiques linéaires

### 2.1 Introduction

Ces dernières années il y a eu un intérêt croissant pour l'application des modèles espace-état dans l'analyse de série chronologique ; voir, par exemple, Harvey (1989), West et Harrison (1997), Durbin et Koopman (2001), Künsch (2001) et Migon (2005). Les modèles espace-état considèrent une série chronologique comme rendement d'un système dynamique perturbé par des perturbations aléatoires. Ils permettent une interprétation normale d'une série chronologique comme combinaison de plusieurs composantes, telles que les composantes tendancielle, saisonnières ou régressives. En même temps, ils ont une structure probabiliste, offrant un cadre flexible pour une étendue des applications très large. Des calculs peuvent être mis-en application par des algorithmes récursifs, les problèmes de l'estimation et des prévisions sont résolus en calculant périodiquement la distribution conditionnelle des quantités d'intérêt, dans ce sens, ils sont tout à fait naturellement traités dans un cadre Bayésien. Les modèles espace-état peuvent être employés pour modéliser des séries chronologiques univariées ou multivariées, aussi en présence de la non-stationnarité, des changements structurels, et des modèles irréguliers.

Dans ce chapitre nous discutons les notions de base au sujet des modèles espace-état et leur utilisation dans l'analyse des séries chronologiques. Le modèle dynamique linéaire est présenté comme un cas spécial d'un modèle espace-état général, étant linéaire et gaussien. Pour les modèles dynamiques linéaires, l'évaluation et les prévisions peuvent être obtenues périodiquement par le filtre de Kalman.

## 2.2 Généralités

Dans le domaine de la statistique dénommée analyse des séries temporelles, la dimension temporelle des observations devient primordiale. Une série temporelle est définie comme une suite d'observations indexées par le temps. L'attention va se focaliser sur les propriétés évolutives d'une variable aléatoire, tant pour sa prévision que dans sa relation avec son passé. Comme exemple de série temporelle, viennent immédiatement à l'esprit toutes les séries macroéconomiques, mais aussi les séries financières.

**Définition 2.1.** Un processus stochastique est une suite de variables aléatoires réelles qui sont indexées par le temps :

$$Y_t, \quad t \in \mathbb{Z}$$

Ici  $t$  appartient à un espace discret, ce qui définit un processus en temps discret. Un processus stochastique est donc une famille de variables aléatoires  $Y$  dont on va observer des valeurs réelles issues de l'espace  $S$  des échantillons selon une certaine loi de probabilité. Pour chaque point  $s$  de l'espace des échantillons  $S$ , la fonction qui à  $t$  associe  $Y_t(s)$  est appelée la trajectoire du processus. Les observations successives forment l'histoire du processus. On peut les noter :  $Y_0^t$  pour désigner l'histoire du processus entre 0 et  $t$ .

La connaissance d'un processus équivaut à connaître la loi de tout vecteur  $(Y_{t+1}, \dots, Y_{t+h})$ ,  $h$  entier. La notion de stationnarité au sens strict, analogue à celle de régime permanent en physique, est donnée par la :

**Définition 2.2.**  $(Y_t)$  est stationnaire au sens strict si et seulement si la loi de  $(Y_{t+1}, \dots, Y_{t+h})$  dépend seulement de  $h$ .

Cette notion est parfois trop restrictive et l'on préfère alors la stationnarité au second ordre :

**Définition 2.3.**  $(Y_t)$  est stationnaire au second ordre ou simplement : stationnaire, si et seulement si pour tout entier  $h$ ,  $E(Y_{t+h})$  et  $cov(Y_t, Y_{t+h})$  ne dépendent que de  $h$ .

Notons que ces définitions sont équivalentes lorsque  $(Y_t)$  est un processus gaussien (c'est-à-dire lorsque la loi de tout vecteur  $(Y_{t+1}, \dots, Y_{t+h})$  est gaussienne).

Pour un processus stationnaire, les moments d'ordre 1 et 2,  $E(Y_t)$  et  $var(Y_t)$ , sont donc constants au cours du temps. Toutefois cette propriété n'est pas suffisante. Quitte à retrancher  $\mu = E(Y_t)$ , on peut toujours se ramener à un processus centré. Dans la suite tous les processus sont centrés, et la notion de stationnarité doit être comprise au sens faible.

### Fonction d'autocovariance et d'autocorrélation.

La connaissance d'un processus centré stationnaire se ramène entièrement à l'étude de la fonction d'autocovariance  $\gamma(h) = cov(Y_t, Y_{t+h})$ , où bien à la connaissance de la variance  $\sigma^2 = \gamma(0) = var(Y_t)$ , et à la fonction d'autocorrélation  $\rho(h) = corr(Y_t, Y_{t+h}) = \frac{\gamma(h)}{\gamma(0)}$ .

Notons que ces notions n'ont de sens que pour les processus stationnaires.

Les propriétés sont semblables pour  $\gamma$  et  $\rho$ . Dans le cas de la fonction d'autocorrélation, on a :

- \*  $\rho(0) = 1$
- \*  $|\rho(h)| \leq 1$  (d'après l'inégalité de Cauchy-Schwartz).
- \*  $\rho(h) = \rho(-h)$ ,  $\rho$  est une fonction paire.

### Autocorrélations partielles

**Définition 2.4.** Si  $Y_t$  est un processus stationnaire, l'autocorrélation partielle d'ordre  $h$ , notée  $\pi(h)$ , est définie par

$$\pi(h) = corr(Y_{t+h} - E_L(Y_{t+h} | Y_{t+1}, \dots, Y_{t+h-1}), Y_t - E_L(Y_t | Y_{t+1}, \dots, Y_{t+h-1})) \quad \text{pour } h \geq 2$$

où  $E_L$  désigne l'espérance conditionnelle linéaire,  $E_L(\cdot | Y_{t+1}, \dots, Y_{t+h-1})$  est donc l'opérateur de projection orthogonale sur l'espace vectoriel engendré par  $Y_{t+1}, \dots, Y_{t+h-1}$ . On convient que  $\pi(0) = 1$  et  $\pi(1) = \rho(1)$ .

Le coefficient  $\pi(h)$  jouit des mêmes propriétés qu'un coefficient de corrélation ordinaire.

Le résultat suivant donne un autre moyen de calcul de  $\pi(h)$  :

**Proposition 2.1.** (Roustant O., 2008)

Considérons la projection de  $Y_{t+h}$  sur l'espace vectoriel engendré par  $Y_t, Y_{t+1}, \dots, Y_{t+h-1}$  :

$$E_L(Y_{t+h} | Y_t, Y_{t+1}, \dots, Y_{t+h-1}) = \phi_{h,1}Y_{t+h-1} + \dots + \phi_{h,h-1}Y_{t+1} + \phi_{h,h}Y_t$$

Alors  $\pi(h) = \phi_{h,h}$

Le calcul pratique des coefficients  $\pi(h)$  est basé sur ce résultat et utilise astucieusement le fait que lorsque  $h$  augmente la projection sur l'espace augmente de la nouvelle variable se déduit de la projection précédente. C'est l'algorithme de Durbin-Levinson.

## 2.3 Modèle à espace-état

Considérons la série chronologique  $(Y_t)_{t \geq 1}$ , en particulier, dans des applications des séries chronologiques l'acceptation de l'indépendance est rarement justifiée. La dépendance markovienne est la forme la plus simple de la dépendance dans laquelle le temps a un rôle défini.

$(Y_t)_{t \geq 1}$  est une chaîne de Markov si pour tout  $t > 1$  on a :

$$\pi(y_t | y_{1:t-1}) = \pi(y_t | y_{t-1})$$

Ceci signifie que les informations sur  $Y_t$  portées par toutes les observations jusqu'au  $t - 1$  sont exactement identiques que l'information portée par  $y_{t-1}$ . Une autre manière de dire la même chose est que  $Y_t$  et  $Y_{t-2}$  sont conditionnellement indépendants en donnant  $y_{t-1}$ . Pour une chaîne de Markov la distribution jointe peut être écrite sous la forme assez simple :

$$\pi(y_{1:t}) = \pi(y_1) \prod_{j=2}^t \pi(y_j | y_{j-1})$$

Assumer une structure markovienne pour les observations n'est pas facile dans beaucoup d'applications. Dans un modèle espace-état nous supposons qu'il y a une chaîne de Markov inobservable  $(\theta_t)$ , appelée *le processus d'état*, et que  $Y_t$  est une mesure imprécise de  $\theta_t$ . Dans des applications de technologie le  $\theta_t$  décrit habituellement l'état d'un système physiquement observable qui a produit le rendement  $Y_t$ .

un modèle espace-état se compose de deux séries chronologiques  $\theta_t, t = 0, 1, \dots$  et  $Y_t, t = 1, 2, \dots$  satisfaisantes les conditions suivantes :

(A.1)  $(\theta_t)$  est une chaîne de Markov, Ainsi, la loi de probabilité du processus  $(\theta_t, t = 0, 1, \dots)$  est spécifiée par la densité initiale  $\pi(\theta_0)$  et les densités de transition  $\pi(\theta_t|\theta_{t-1})$ .

(A.2) conditionnellement à  $(\theta_t)$  Les  $(Y_t)$  sont indépendants et ils dépendent que de  $\theta$ .

La conséquence de ces conditions est qu'un modèle espace-état est complètement spécifié par la loi initiale  $\pi(\theta_0)$  et les densités conditionnelles  $\pi(\theta_t|\theta_{t-1})$  et  $\pi(y_t|\theta_t)$ ,  $t \geq 1$ , en fait  $\forall t > 0$

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \cdot \prod_{j=1}^t \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j) \quad (2.1)$$

## 2.4 Modèle dynamique linéaire

La première classe importante des modèles espace-état est donnée par les modèles espace-état linéaires gaussiens, également appelés les modèles dynamiques linéaires. Un modèle dynamique linéaire (DLM) est spécifié par une loi a priori normale pour le vecteur d'état de dimension  $p$ .

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0) \quad (2.2)$$

et par le système d'équations suivant :

$$\begin{aligned} Y_t &= F_t \theta_t + v_t, & v_t &\sim \mathcal{N}_m(0, V_t) \\ \theta_t &= G_t \theta_{t-1} + \omega_t, & \omega_t &\sim \mathcal{N}_p(O, W_t) \end{aligned} \quad (2.3)$$

où

$\theta_t$  est l'état du système à l'instant  $t$  ;

$Y_t$  est le vecteur des observations ;

$\omega_t$  est le vecteur des innovations à la date  $t$  ;

$v_t$  est le vecteur des erreurs de mesure au temps  $t$  ;

$G_t$  est la matrice de transition d'ordre  $(p \times p)$  ;

$F_t$  est la matrice de mesure d'ordre  $(m \times p)$  ;

La première équation c'est *l'équation des observations*, et la seconde c'est *l'équation d'état*, de plus on suppose que  $\theta_0$  est indépendant de  $v_t$  et  $\omega_t$ . Comme on peut montrer qu'un DLM satisfait les conditions (A.1) et (A.2) avec  $Y_t|\theta_t \sim \mathcal{N}(F_t \theta_t, V_t)$  et  $\theta_t|\theta_{t-1} \sim \mathcal{N}(G_t \theta_{t-1}, W_t)$ .

Un modèle espace-état général peut être spécifié par une loi a priori pour  $\theta_0$ , ainsi que les équations d'observation et d'état

$$\begin{aligned} Y_t &= h_t(\theta_t, v_t) \\ \theta_t &= g(\theta_{t-1}, \omega_t) \end{aligned}$$

$g$  et  $h$  sont deux fonctions arbitraires. la linéarité des modèles espace-état vient du fait que les fonctions  $g$  et  $h$  sont linéaires, et le cas gaussien vient du fait que la loi a priori est gaussienne. la supposition de la normalité est sensible dans beaucoup d'applications, et elle peut être justifiée par le théorème centrale limite.

**Exemple 2.1.** *Le modèle le plus simple pour une série temporelle univariée ( $Y_t : t = 1, 2, \dots$ ) est le modèle marche aléatoire plus un bruit, définie par :*

$$\begin{cases} Y_t = \mu_t + v_t, & v_t \sim \mathcal{N}(0, V) \\ \mu_t = \mu_{t-1} + \omega_t & \omega_t \sim \mathcal{N}(0, W) \end{cases} \quad (2.4)$$

où les séquences d'erreur  $v_t$  et  $\omega_t$  sont indépendantes, il s'agit d'un DLM avec  $m = p = 1$ ,  $\theta_t = \mu_t$  et  $F_t = G_t = 1$ . Intuitivement, il est approprié pour les séries chronologiques ne présentant pas de tendance claire ou des variations saisonnières : les observations ( $Y_t$ ) sont modélisées comme des observations bruitées d'un niveau  $\mu_t$  qui, à son tour, est soumis à des changements aléatoires dans le temps, décrite par une marche aléatoire.

Si  $W = 0$ , nous sommes de retour au modèle de moyenne constante. Notez que la marche aléatoire ( $\mu_t$ ) est non stationnaire. En effet, un DLM peut être utilisé pour la modélisation des séries temporelles non stationnaires, ce qui est le contraire pour les modèles habituels ARMA qui exigent une transformation préliminaire des données pour assurer la stationnarité.

**Exemple 2.2.** *Modèle de régression linéaire dynamique*

Un DLM peut être considéré comme une généralisation d'un modèle de régression linéaire. Un modèle de régression linéaire simple décrit la relation entre une variable  $y$  et une variable explicative  $x$  :

$$Y_t = \theta_1 + \theta_2 x_t + \epsilon_t, \quad \epsilon_t \text{ iid } \sim \mathcal{N}(0, \sigma^2)$$

Un modèle de régression linéaire dynamique est présenté comme suite

$$Y_t = \theta_{1,t} + \theta_{2,t} x_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$

avec une autre équation pour décrire l'évolution de système

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}(0, W_t)$$

c'est un DLM avec  $F_t = (1, x_t)$  et l'état  $\theta_t = (\theta_{1,t}, \theta_{2,t})'$ . Si  $G_t = I$  alors  $\sigma_t^2 = \sigma^2$  et  $\omega_t = 0$ , nous sommes de retour au modèle de régression linéaire simple.

## 2.5 Estimation et prévision

Pour un modèle espace-état donné, la tâche principale c'est de faire une inférence sur les états non observés ou de prédire les observations futures reposant sur une partie de la séquence d'observation. Estimation et prévision sont résolues par le calcul des distributions conditionnelles des quantités d'intérêt, compte tenu des informations disponibles.

Pour estimer le vecteur d'état, nous calculons la densité conditionnelle  $\pi(\theta_s | y_{1:t})$ . On distingue entre les problèmes de filtrage (si  $s = t$ ), la prédiction de l'état ( $s > t$ ) et de lissage ( $s < t$ ), Il faut bien souligner la différence entre le filtrage et le lissage.

Dans le problème de filtrage, les données sont censées arriver séquentiellement dans le temps, pour résoudre ce problème de filtrage, nous calculons la densité conditionnelle  $\pi(\theta_t | y_{1:t})$ . Dans les modèles dynamiques linéaires le filtre de Kalman fournit les formules de mise à jour de notre inférence en cours sur le vecteur d'état que de nouvelles données deviennent disponibles, c'est pour passer de la densité de filtrage  $\pi(\theta_t | y_{1:t})$  à  $\pi(\theta_{t+1} | y_{1:t+1})$ .

Le problème de lissage, consiste au contraire à l'estimation de la séquence d'état à l'instant  $1, \dots, t$ , en sachant les données  $y_1, \dots, y_t$ . Dans de nombreuses applications, on a des observations sur une série de temps pendant une certaine période, et on veut étudier le comportement a posteriori du système sous-jacent des observations. Le problème de lissage est résolu par le calcul de la distribution conditionnelle de  $\theta_{1:t}$  sachant  $y_{1:t}$ .

En fait, dans les séries chronologiques, la prévision est souvent la tâche principale, l'estimation d'état est alors simplement une étape pour la prédiction des futures observations. Pour la prévision en une étape en avant, c'est-à-dire prédire la prochaine observation  $Y_{t+1}$  en se basant sur les données  $y_{1:t}$ , on estime d'abord la valeur  $\theta_{t+1}$  du vecteur d'état, et sur la base de cette estimation, on calcule les prévisions pour  $Y_{t+1}$ . La densité de prédiction

d'état en une étape en avant est  $\pi(\theta_{t+1}|y_{1:t})$  et elle est basée sur la densité de filtrage de  $\theta_t$ , sur cette base, on obtient la densité de prévision de la première étape en avant  $\pi(y_{t+1}|y_{1:t})$ .

On pourrait être intéressé à regarder un peu plus loin, en estimant l'évolution du système, représentée par le vecteur d'état  $\theta_{t+k}$  pour  $k \geq 1$ , et de faire des prévisions des  $k$ -étapes en avant pour  $Y_{t+k}$ . La prédiction d'état est résolue par le calcul de la densité de prévision des  $k$ -étapes en avant d'état  $\pi(\theta_{t+k}|y_{1:t})$ , en se basant sur cette densité on peut calculer la densité prédictive des  $k$ -étapes en avant  $\pi(y_{t+k} | y_{1:t})$  pour les futures observations à l'instant  $t+k$ , bien sûr quand  $t+k$  s'éloigne dans le temps les prévisions deviennent de plus en plus incertaines, mais nous pouvons tout de même quantifier l'incertitude par une densité de probabilité c'est-à-dire la densité prédictive de  $Y_{t+1}$  sachant  $y_{1:t}$ . En particulier, la moyenne conditionnelle  $E(Y_{t+1}|y_{1:t})$  fournit une prévision optimale de  $Y_{t+1}$ .

### 2.5.1 Filtrage

Nous décrirons d'abord les étapes récursives nécessaires pour calculer les densités de filtrage  $\pi(\theta_t|y_{1:t})$  dans un modèle espace-état. Même si nous n'allons pas faire un large usage de ces formules, il est utile de voir maintenant la récursion générale afin de mieux comprendre le rôle des hypothèses d'indépendance conditionnelle qui ont été introduites. Ensuite, nous passons au cas d'un DLM, pour lesquels le problème de filtrage est résolu par le filtre de Kalman.

L'avantage des modèle espace-état est que la structure markovienne de l'état dynamique, et l'hypothèse de l'indépendance conditionnelle sur les observations, les densités de filtrage et de prévision sont obtenues en appliquant un algorithme récursif.

A partir de  $\theta_0$ , on peut calculer de manière récursive, pour  $t = 1, 2, \dots$

- (i) La distribution prédictive en une étape en avant pour  $\theta_t$  sachant  $y_{1:t-1}$  est basée sur la densité de filtrage  $\pi(\theta_{t-1}|y_{1:t-1})$ , et la distribution conditionnelle de  $\theta_t$  sachant  $\theta_{t-1}$ ;
- (ii) La distribution prédictive en une étape en avant pour la prochaine observation;
- (iii) La densité de filtrage  $\pi(\theta_t|y_{1:t})$ , en utilisant la règle de Bayes avec  $\pi(\theta_t|y_{1:t-1})$  comme loi a priori et  $\pi(y_t|\theta_t)$  comme fonction de vraisemblance.

La proposition suivante contient une présentation formelle des récursions de filtrage pour un modèle espace-état général.

**Proposition 2.2.** (Petris et al. 2009)

Pour un modèle espace-état général, les propositions suivantes sont vérifiées :

(i) La densité prédictive pour l'état en une étape en avant est calculée à partir de la densité de filtrage  $\pi(\theta_{t-1}|y_{1:t-1})$  comme suite

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})d\theta_{t-1} \quad (2.5)$$

(ii) La densité prédictive pour l'observation en une étape en avant est calculée à partir de la densité prédictive des états

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})d\theta_t \quad (2.6)$$

(iii) La densité de filtrage est calculée comme suite :

$$\pi(\theta_t|y_{1:t}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|y_{1:t-1})} \quad (2.7)$$

**Remarque 2.1.** A partir des densités prédictives en une étape en avant calculées avec la proposition précédente, On peut calculer les densités prédictives en  $k$  étapes en avant pour les états et les observations avec les formules suivantes :

$$\pi(\theta_{t+k}|y_{1:t}) = \int \pi(\theta_{t+k}|\theta_{t+k-1})\pi(\theta_{t+k-1}|y_{1:t})d\theta_{t+k-1}$$

et

$$\pi(y_{t+k}|y_{1:t}) = \int \pi(y_{t+k}|\theta_{t+k})\pi(\theta_{t+k}|y_{1:t})d\theta_{t+k}$$

### Filtre de Kalman pour les modèles dynamiques linéaires

En général le calcul des distributions conditionnelles n'est pas une tâche facile. Dans un modèle dynamique linéaire en utilisant les propriétés de la distribution gaussienne multivariante, on montre facilement que le vecteur aléatoire  $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$  suit une loi gaussienne  $\forall t$ , il suit que les distributions marginales et conditionnelles sont également gaussiennes. Puisque toutes les distributions sont gaussiennes, elles sont complètement déterminées par une moyenne et une variance. Le problème de filtrage pour les DLMS est résolu par le filtre de Kalman.

**Proposition 2.3.** (Petris et al. 2009), *Filtrage de Kalman*

Pour un modèle dynamique linéaire et

$$\theta_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$$

on a les propriétés suivantes :

(i) La distribution prédictive en une étape en avant de  $\theta_t$  sachant  $y_{1:t-1}$  est gaussienne, de paramètres

$$a_t = E(\theta_t|y_{1:t-1}) = G_t m_{t-1} \quad (2.8)$$

$$R_t = \text{Var}(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t$$

(ii) La distribution prédictive en une étape en avant de  $Y_t$  sachant  $y_{1:t-1}$  est gaussienne, de paramètres :

$$f_t = E(Y_t|y_{1:t-1}) = F_t a_t \quad (2.9)$$

$$Q_t = \text{Var}(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t$$

(iii) La distribution de filtrage de  $\theta_t$  sachant  $y_{1:t}$  est gaussienne, de paramètres :

$$m_t = E(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t \quad (2.10)$$

$$C_t = \text{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t$$

où  $e_t = Y_t - f_t$  est l'erreur de prévision.

La matrice de gain de Kalman est définie par

$$K_t = R_t F_t' Q_t^{-1}$$

Comme exemple, considérons l'exemple (2.1). Le filtre de Kalman nous donne :

$$\mu_t|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, R_t = C_{t-1} + W)$$

$$Y_t|y_{1:t-1} \sim \mathcal{N}(f_t = m_{t-1}, Q_t = R_t + V)$$

$$\mu_t|y_{1:t} \sim \mathcal{N}(m_t = m_{t-1} + K_t e_t, C_t = K_t V)$$

où  $K_t = R_t/Q_t$  et  $e_t = Y_t - f_t$ . Notons que le processus  $(Y_t)$  est généralement influencé par le rapport d'erreur  $r = W/V$

## 2.5.2 Le lissage

L'une des caractéristiques les plus importantes des modèles espace-état est que l'estimation et la prévision peuvent être appliquées séquentiellement, pendant que de nouvelles données deviennent disponibles. Cependant, l'analyse d'une série chronologique a souvent des observations sur  $Y_t$  pendant une certaine période,  $t = 1, \dots, T$ , et on veut étudier le comportement du système. Pour cela, on peut appliquer un algorithme arrière récursif pour calculer les distributions conditionnelles de  $\theta_t$  sachant  $y_{1:T}$ , pour tout  $t < T$ , à partir de la distribution de filtrage  $\pi(\theta_T|y_{1:T})$  et l'estimation de tous les états.

**Proposition 2.4.** (Petris et al. 2009)

Pour un modèle espace-état donné on a les propriétés suivantes :

(i) Conditionnellement à  $y_{1:T}$ , la succession d'état  $(\theta_0, \dots, \theta_T)$  a des probabilités de transition en arrière données par :

$$\pi(\theta_t | \theta_{t+1}, y_{1:T}) = \frac{\pi(\theta_{t+1} | \theta_t) \pi(\theta_t | y_{1:t})}{\pi(\theta_{t+1} | y_{1:t})}$$

(ii) Les distributions de lissage de  $\theta_t$  sachant  $y_{1:T}$  sont calculées avec la récursion en arrière suivante

$$\pi(\theta_t | y_{1:T}) = \pi(\theta_t | y_{1:t}) \int \frac{\pi(\theta_{t+1} | \theta_t)}{\pi(\theta_{t+1} | y_{1:t})} \pi(\theta_{t+1} | y_{1:T}) d\theta_{t+1}$$

Pour un DLM les récursions de lissage peuvent être décrites en termes de moyenne et de variance.

**Proposition 2.5.** (Petris et al. 2009), Lissage de Kalman

Pour un DLM donné, si  $\theta_{t+1} | y_{1:T} \sim \mathcal{N}(s_{t+1}, S_{t+1})$  alors,  $\theta_t | y_{1:T} \sim \mathcal{N}(s_t, S_t)$ , où :

$$s_t = m_t + C_t G'_{t+1} R_{t+1}^{-1} (s_{t+1} - a_{t+1})$$

$$S_t = C_t - C_t G'_{t+1} R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t$$

### 2.5.3 Prévision

Ce qui nous intéresse dans l'analyse des séries chronologiques, est de faire des prévisions sur les observations,  $Y_{t+k}$ , ou bien sur le vecteur d'état  $\theta_{t+k}$ , en se basant sur les données disponibles  $y_{1:t}$ . Pour les modèles espace-état, la forme récursive des calculs facilite le calcul des prévisions en une étape en avant et de les mettre à jour séquentiellement pendant que de nouvelles données deviennent disponibles.

Le calcul des densités de prévision pour l'état et l'observation sont donnés dans la proposition suivante

**Proposition 2.6.** (Petris et al. 2009)

Pour un modèle espace-état général, on a les propriétés suivantes pour chaque  $k > 0$

(i) La distribution de prévision en  $k$ -étapes en avant pour l'état est

$$\pi(\theta_{t+k} | y_{1:t}) = \int \pi(\theta_{t+k} | \theta_{t+k-1}) \pi(\theta_{t+k-1} | y_{1:t}) d\theta_{t+k-1}$$

(ii) La distribution de prévision en  $k$ -étapes en avant pour l'observation est :

$$\pi(y_{t+k} | y_{1:t}) = \int \pi(y_{t+k} | \theta_{t+k}) \pi(\theta_{t+k} | y_{1:t}) d\theta_{t+k}$$

Pour un DLM, la proposition précédente prend une forme plus spécifique, toutes les intégrales peuvent être calculées explicitement, comme pour le filtrage et le lissage, toutes les distributions de prévision sont gaussiennes, ce qui nous reste c'est le calcul de leur moyennes et variances. La proposition suivante nous permet de les calculées, pour cela nous devons présenter certaines notations, pour  $k \geq 1$ , on définit :

$$a_t(k) = E(\theta_{t+k}|y_{1:t}) \quad (2.11)$$

$$R_t(k) = Var(\theta_{t+k}|y_{1:t})$$

$$f_t(k) = E(Y_{t+k}|y_{1:t})$$

$$Q_t(k) = Var(Y_{t+k}|y_{1:t})$$

**Proposition 2.7.** (Petris et al. 2009)

Pour un modèle dynamique linéaire, soient  $a_t(0) = m_t$ ,  $R_t(0) = C_t$ , pour tous  $k \geq 1$ , on a les propriétés suivantes :

(i) La distribution de  $\theta_{t+k}$  sachant  $y_{1:t}$  est gaussienne de paramètres :

$$a_t(k) = G_{t+k}a_t(k-1)$$

$$R_t(k) = G_{t+k}R_t(k-1)G'_{t+k} + W_{t+k}$$

(ii) La distribution de  $Y_{t+k}$  sachant  $y_{1:t}$  est gaussienne de paramètres :

$$f_t(k) = F_{t+k}a_t(k)$$

$$Q_t(k) = F_{t+k}R_t(k)F'_{t+k} + V_t$$

## 2.6 Analyse des séries chronologiques

L'examen graphique d'une série temporelle montre que la valeur prise au temps  $t$  dépend fortement de la valeur prise au temps  $t-1$ . Le processus qui les engendre est dynamique. On voudra en construisant un modèle, acquérir de l'information sur ce processus théorique que l'on appelle Processus de Génération des Données ou *PGD*. Le problème est alors de trouver le modèle pratique qui approchera le plus possible le processus théorique et ensuite de l'estimer. Une fois cette étape franchie, on pourra faire de la prévision ou du contrôle avec ce modèle. Les types de modèles que l'on peut considérer sont nombreux. En statistique on va s'intéresser à modéliser une série univariée au moyen d'un modèle *ARMA*, où bien un modèle dynamique linéaire.

### 2.6.1 Les modèles *ARMA*

La branche de la statistique mathématique qui s'intéresse aux séries temporelles a développé plusieurs modèles de représentation des séries temporelles dont nous allons très brièvement rappeler les plus simples. Il s'agira de préciser quelques notions sur les modèles *AR*, *MA* et *ARMA* et quelques outils mathématiques qui leurs sont reliés.

#### Les processus *MA*

Le processus  $(y_t, t \in \mathbb{Z})$  satisfait une représentation *MA*( $q$ ), si et seulement si :

$$y_t = m_t + \Theta(B)\epsilon_t$$

avec  $B$  c'est l'opérateur retard de box-jenkins, et  $E(y_t) = m$  en effet

$$E(y_t) = E\left[m + \sum_{j=0}^q \theta_j E(\epsilon_{t-j})\right] = m$$

avec,  $\Theta(B) = \sum_{j=0}^q \theta_j B^j$

$$\theta_0 = 1$$

$\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$  sont *iid*

**Exemple 2.3.**  $y_t = \epsilon_t + 2\epsilon_{t-3}$  est un processus *MA*(3) de moyenne nulle.

$y_t = 4 + \epsilon_t + 0.8\epsilon_{t-1}$  est un processus *MA*(1) de moyenne 4.

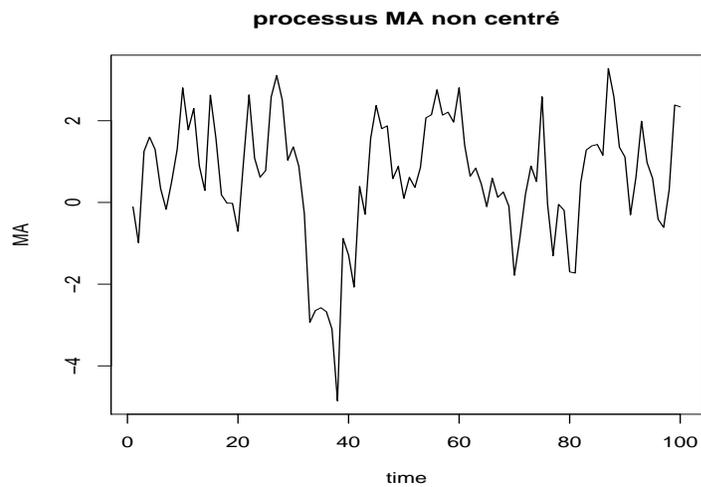
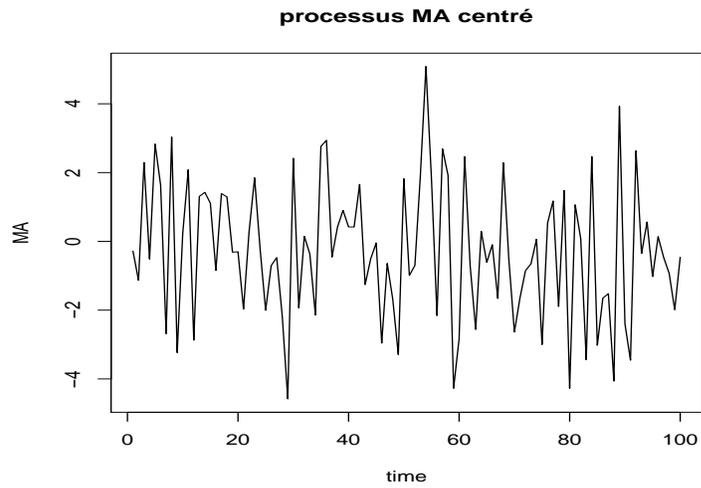
Une simulation de taille  $T = 100$  de ces deux processus pour  $\epsilon_t \sim \mathcal{N}(0, 1)$  est représentée sur les figures suivantes à l'aide du *R*

```
MA = arima.sim(list(ma = c(0, 0, 2)), 100, innov = rnorm(200, 0, 1))
```

```
ma = arima.sim(list(ar = c(0.8)), 100, innov = rnorm(200, 0, 1))
```

```
plot(MA, main="processus MA centré", xlab="time", ylab="MA")
```

```
plot(ma, main="processus MA non centré", xlab="time", ylab="MA")
```



### Les processus $AR$

Le processus  $(y_t, t \in \mathbb{Z})$  satisfait une représentation  $AR(p)$ , si et seulement si :

$$\Phi(B)y_t = c_t + \epsilon_t$$

avec :  $B$  c'est l'opérateur retard de box-jenkins.

$c \in \mathbb{R}$

$$\Phi(B) = \sum_{j=0}^p \phi_j B^j$$

$$\phi_0 = 1$$

$\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$  sont *iid*

Contrairement au cas du *MA*; la constante  $c$  de l'équation précédente ne correspond pas à l'espérance du processus  $(y_t, t \in \mathbb{Z})$ . Cette espérance, dans le cas d'un processus stationnaire, est déterminée par la relation suivante si  $\phi_1 = 0$  :

$$E(y_t) = \Phi(B)^{-1}c = \Phi(1)^{-1}c$$

en effet

$$E(y_t) = c - \phi_1 E(y_{t-1}) - \dots - \phi_p E(y_{t-p}) + E(\epsilon_t)$$

Par définition  $E(\epsilon_t) = 0$ , Si l'on suppose maintenant que l'espérance de  $(y_t, t \in \mathbb{Z})$ , que l'on notera  $m$ , ne dépend pas du temps :

$$E(y_{t-j}) = m \quad \forall j \in \mathbb{Z}$$

Donc on a la formule qui s'écrira sous la forme suivante :

$$E(y_t) = m = \frac{c}{1 + \phi_1 + \dots + \phi_p}$$

$$\Phi(1)^{-1} = \left[ \sum_{j=0}^p \phi_j \right]^{-1} = \frac{1}{\phi_0 + \dots + \phi_p} \quad \text{avec } \phi_0 = 1$$

**Exemple 2.4.**  $y_t = -0.5y_{t-1} + 0.2y_{t-3} + \epsilon_t$  est un *AR(3)* de moyenne nulle.

$y_t = 0.8 + 0.7y_{t-1} + 0.25y_{t-2} + \epsilon_t$  est un *AR(2)* d'espérance  $\frac{0.8}{1-0.7-0.25} = 16$

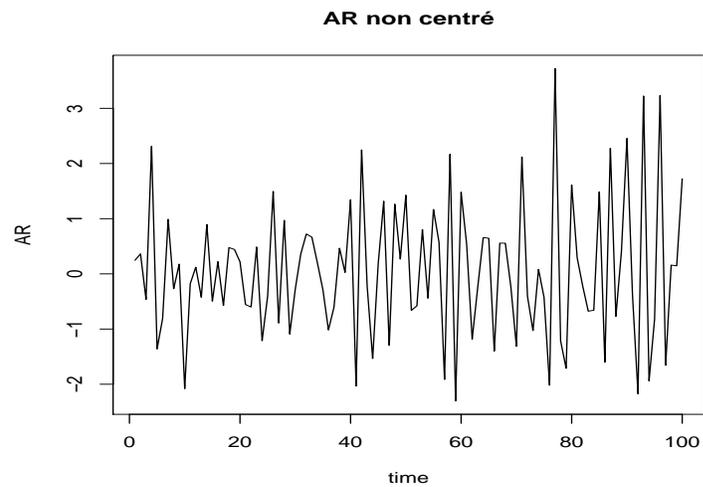
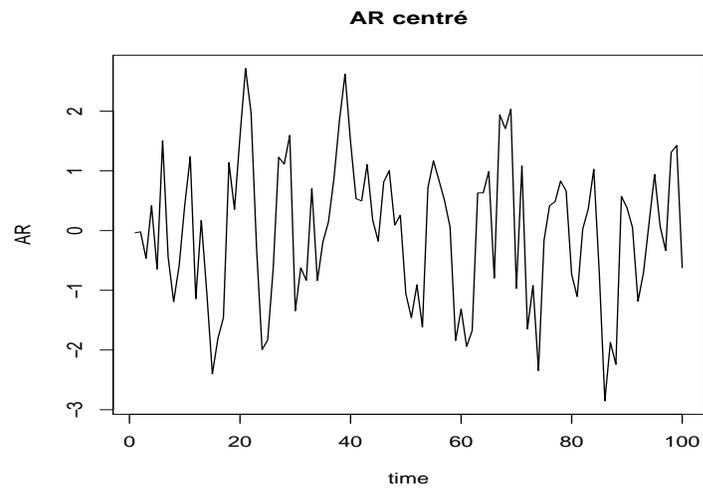
Une simulation de taille  $T = 100$  de ces deux processus pour  $\epsilon_t \sim \mathcal{N}(0, 1)$  est représentée sur les figures suivantes à l'aide du *R*

```
AR = arima.sim(list(ar = c(0.5, 0, -0.2)), 100, innov = rnorm(200, 0, 1))
```

```
ar = arima.sim(list(ar = c(-0.7, -0.25)), 100, innov = rnorm(200, 0, 1))
```

```
plot(AR, main="AR centré", xlab="time", ylab="AR")
```

```
plot(ar, main="AR non centré", xlab="time", ylab="AR")
```



### Les processus *ARMA*

Le processus  $(y_t, t \in \mathbb{Z})$  satisfait une représentation  $ARMA(p, q)$ , si et seulement si :

$$\Phi(B)y_t = c + \Theta(B)\epsilon_t$$

avec :  $c \in \mathbb{R}$

$$\Theta(B) = \sum_{j=0}^q \theta_j B^j$$

$$\Phi(B) = \sum_{j=0}^p \phi_j B^j$$

$$\theta_0 = \phi_0 = 1, (\theta_j, \phi_j) \in \mathbb{R}^{*2}$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2) \text{ sont } iid$$

les différentes conditions sur les paramètres  $\phi_j$  et  $\theta_j$  signifient que dans un  $ARMA(p, q)$  tous les paramètres du polynôme  $\Phi(B)$  peuvent être nuls à l'exception du paramètre correspondant au  $p^{eme}$  retard et que tous les paramètres du polynôme  $\Theta(B)$  peuvent être nuls à l'exception du paramètre correspondant au  $q^{eme}$  retard.

En effet, si  $\phi_p = 0$  alors le processus  $(y_t, t \in \mathbb{Z})$  satisfait une représentation  $ARMA(p-1, q)$ , de la même façon, si  $\theta_q = 0$  alors le processus  $(y_t, t \in \mathbb{Z})$  satisfait une représentation  $ARMA(p, q-1)$ .

**Remarque 2.2.** Ainsi, on constate que les processus  $AR$  et  $MA$  ne sont que des cas particuliers des processus  $ARMA$ . Un  $AR(p)$  correspond à un  $ARMA(p, 0)$ , de la même façon un  $MA(q)$  correspond à un  $ARMA(0, q)$ .

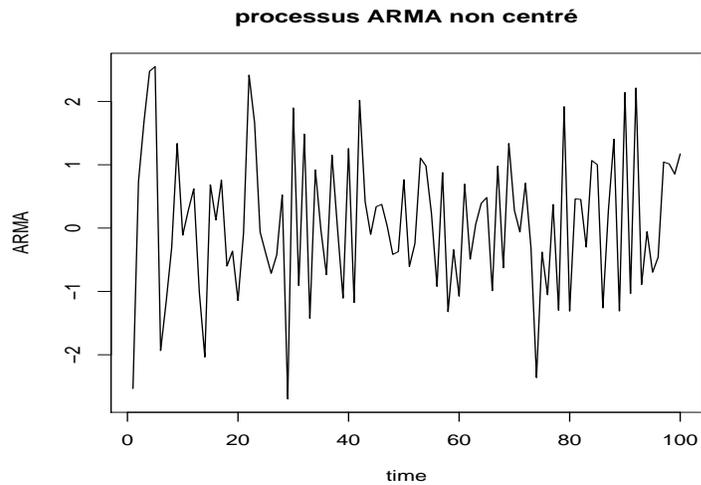
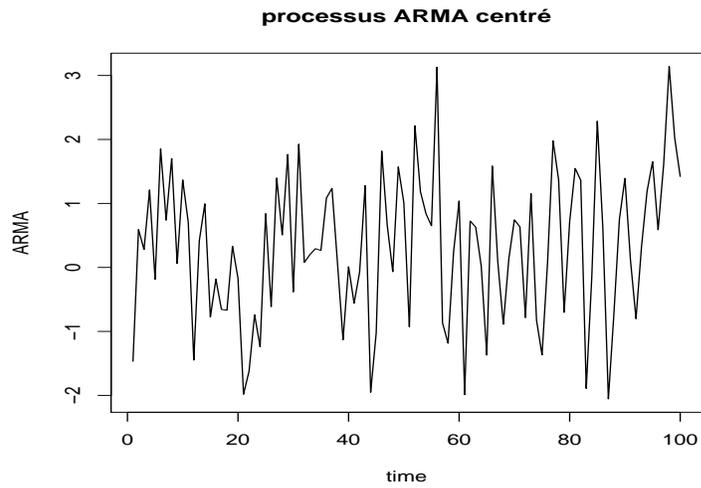
**Exemple 2.5.**  $y_t = 2 - 0.5y_{t-1} + 0.2y_{t-3} + \epsilon_t - 0.5\epsilon_{t-1}$  est un processus  $ARMA(3, 1)$  de moyenne  $\frac{2}{1+0.5-0.2} = 1.5385$ .

$y_t - 0.2y_{t-1} = \epsilon_t + 0.5\epsilon_{t-1} + 0.6\epsilon_{t-4}$  est un  $ARMA(1, 4)$  de moyenne nulle.

Une simulation de taille  $T = 100$  de ces deux processus pour  $\epsilon_t \sim N(0, 1)$  est représentée sur les figures suivantes à l'aide du R

```
ARMA = arima.sim(list(ar = c(0.5, 0, -0.2), ma = -0.5), 100)
```

```
arma = arima.sim(list(ar = c(-0.2), ma = c(0.5, 0, 0, 0.6)), 100)
```



## Prédiction

Toute l'analyse qu'on fait sur les séries chronologiques (identifications du modèle, estimation des coefficients,...etc.), n'est en fait qu'un passage obligatoire pour arriver à l'étape de la prédiction des valeurs futures  $Y_{t+1}, \dots$  nous supposons donc dans cette étape que les observations proviennent d'un processus  $ARMA(p, q)$ , dont on connaît l'ordre  $(p, q)$  ainsi que les coefficients.

### Equations linéaires de prédiction

La prédiction linéaire de la série chronologique, signifie que si on observe  $y_1, \dots, y_t$ , on cherche à prédire la valeur  $y_{t+s}$  (où  $s$  est un nombre positif fixé). Ainsi on note  $\hat{y}_{t+s}$  la valeur prévue au temps  $t + s$ .

L'étude se fera alors sur l'équation suivante :

$$\hat{y}_{t+s} = c_0 + c_1 y_t + \dots + c_t y_1$$

les coefficients  $c_0, c_1, \dots$  devraient être choisis de telle sorte que la prédiction  $\hat{y}_{t+s}$  soit le plus proche possible de la vraie réalisation  $y_{t+s}$  du processus.

Toute fois une erreur de mesure ne peut être évitée, et pour mesurer l'exactitude de notre prédiction on se doit de calculer cette dernière et essayer de la minimiser autant que possible, on considère alors l'erreur quadratique moyenne :

$$S(C) = E[y_{t+s} - c_0 - c_1 y_t - \dots - c_t y_1]^2$$

Les coefficients  $C = (c_1, c_2, \dots, c_t)'$  seront choisis de manière à minimiser l'erreur, ce qui nous amène à traiter donc un problème de minimisation à  $t$  variables qui s'écrit sous la forme suivante :

$$E(y_{t+s} - c_0 - \sum_{i=1}^t c_i y_{t+1-i}) = 0$$

$$E[(y_{t+s} - c_0 - \sum_{i=1}^t c_i y_{t+1-i}) y_{t+1-j}] = 0 \quad \text{tel que } j = 1, \dots, t$$

La première équation pour  $i = 0$ , donne  $c_0 = 0$ , quant aux  $t - 1$  équations suivantes, elles peuvent se réécrire (avec  $c_0 = 0$ ) :

$$r_{s-1+j}^y - \sum_{i=1}^t c_i r_{|i-j|}^y = 0$$

qui donne avec la notation matricielle

$$R_t^y c_t = r_t^c(S)$$

$R_t^y$  est la matrice  $t \times t$  de covariance dont l'élément  $(i, j)$  est :

$$(R_t^y)_{ij} = r_{|i-j|}^y$$

Cette dernière équation résout en principe le problème de la prédiction des valeurs futures, après avoir estimé la fonction d'autocovariance du processus (fonction d'autocovariance empirique), la prédiction de  $\hat{y}_{t+s}$  est donnée par la relation qui suit :

$$C_t = (R_t^y)^{-1} r_t^y(s)$$

à condition que l'inverse de la matrice existe.

### Algorithme de Durbin-Levinson

Théoriquement, les équations de prédictions ont été trouvées, il reste juste à inverser une matrice  $t \times t$  pour trouver les coefficients  $C(i)$  pour  $i = 1, \dots, t$ . cependant, cette inversion est théoriquement toujours possible mais toutes fois coûteuse en temps quand le rang de la matrice dépasse un certain rang, c'est pour ceci qu'on cherche à développer une perspective algorithmique d'inversion plus rapide.

### Description de l'algorithme Durbin-Levinson

L'idée de cet algorithme est que pour prédire la valeur de la série au temps  $t + s$ , on pourrait utiliser une prédiction de la série au temps  $t + s - 1$ . En remontant ainsi dans l'intervalle de temps de prédiction, la démarche à suivre est donc la suivante : à partir des  $t$  observations, on évalue  $\hat{y}_{t+1}$ , à partir des  $t$  observations auxquelles on ajoute  $\hat{y}_{t+1}$  on évalue  $\hat{y}_{t+2}$ , et ainsi de suite jusqu'à  $\hat{y}_{t+s}$  pour le  $s$  souhaité.

Pour la première étape de cette démarche, on trouve en posant  $c_i = \phi_{ti}$  :

$$\hat{y}_{t+1} = \phi_{t1}y_t + \dots + \phi_{tt}y_1$$

l'algorithme de Durbin-Levinson permet d'exprimer chaque coefficients  $\phi_t$  en fonction des coefficients  $\phi_{t-i}$ , ainsi pour prédire  $\hat{y}_{t+1}$ , on utilisera les coefficients  $\phi_t$  trouvés pour cet algorithme. Récursivement, pour trouver  $\hat{y}_{t+2}$ , on aura besoin des coefficients  $\phi_{t+1}$  que l'on calcule à partir de  $\phi_t$ , ainsi de suite.

Si on note  $K(i, j) = E(y_i y_j)$ , et on note  $\hat{y}_j$  c'est la prévision d'ordre 1 de  $y_j$  se basant sur les  $(y_{j-1}, y_{j-2}, \dots)$  passés de  $y_j$ , alors :

$$\hat{y}_{t+1} = \sum_{j=1}^t c_j (y_{t+1-j} - \hat{y}_{t+1-j})$$

Et on pose l'erreur quadratique moyenne est :

$$\epsilon_t = E(y_{t+1} - \hat{y}_{t+1})^2$$

**Proposition 2.8.** (*Shumway et al (2000)*)

si  $(y_t, t \in \mathbb{Z})$  est un processus stationnaire de moyenne zéro, d'autocovariance  $r_h^y$  telle que  $r_0^y > 0$  et  $r_h^y \rightarrow 0$  lorsque  $h \rightarrow \infty$ , alors les coefficients  $c_j$  et les erreurs quadratiques moyennes de prédiction  $\epsilon_t$  définis dans la description de l'algorithme de Durbin-Levinson satisfont l'algorithme suivant :

Etape 1 :  $\phi_{11} = \frac{r_1^y}{r_0^y}$ ,  $\epsilon_0 = r_0^y$

Etape  $t$  : (i)  $\phi_{tt} = [r_t^y - \sum_{j=1}^{t-1} \phi_{t-1,j} r_{t-j}^y] \epsilon_{t-1}^y$

(ii)  $(\phi_{t1}, \dots, \phi_{tt})' = (\phi_{t-1,1}, \dots, \phi_{t-1,t+1})' - \phi_{tt}(\phi_{t-1,t-1}, \dots, \phi_{t-1,1})'$

(iii)  $\epsilon_t = \epsilon_{t-1}(1 - \phi_{tt}^2)$

## 2.6.2 Modèle dynamique linéaire pour l'analyse des séries chronologiques

Comme nous avons vu dans les sections précédentes, le filtre de Kalman nous fournit les formules d'estimation et de prévision pour un DLM complètement spécifié, c-à-d, un DLM où les matrices  $F_t$ ,  $G_t$  et les matrices de covariance  $V_t$  et  $W_t$  sont connues. En pratique, la spécification d'un modèle peut être une tâche difficile, l'approche générale est d'imaginer la série chronologique est obtenue en combinant les composantes élémentaires, telle que la tendance, la saisonnalité, et la régression, chaque composante est représentée par un DLM, et les différentes composantes sont alors combinées dans un DLM unique, produisant un modèle pour la série chronologique donnée. Pour être précis, les composantes sont combinées d'une façon additive ; pour les séries pour laquelle une décomposition multiplicative est plus appropriée, elles peuvent être modélées en utilisant une décomposition additive après une transformation.

Nous détaillons au-dessous la technique additive de décomposition dans le cas univarié, bien que la même approche reportée à la série chronologique multivariée avec des

modifications évidentes.

Considérons une série univariante  $(Y_t)$ . On peut supposer que la série peut être écrite comme la somme de composantes indépendantes :

$$Y_t = Y_{1,t} + \dots + Y_{h,t} \quad (2.12)$$

où  $Y_{1,t}$  pourrait représenter la tendance,  $Y_{2,t}$  la composante saisonnière, et ainsi de suite. La  $i^{eme}$  composante  $Y_{i,t}, i = 1, \dots, h$ , pourrait être décrite par un DLM comme suit :

$$\begin{aligned} Y_{i,t} &= F_{i,t}\theta_{i,t} + v_{i,t} & v_{i,t} &\sim \mathcal{N}(0, V_{i,t}) \\ \theta_{i,t} &= G_{i,t}\theta_{i,t-1} + \omega_{i,t} & \omega_{i,t} &\sim \mathcal{N}(0, W_{i,t}) \end{aligned}$$

où les séries  $(Y_{i,t}, \theta_{i,t})$  et  $(Y_{j,t}, \theta_{j,t})$  sont indépendantes pour tous  $(i \neq j)$ . En supposant l'indépendance des composantes, sa sera facile de montrer que  $Y_t = \sum_{i=1}^h Y_{i,t}$  est représenté par le DLM suivant :

$$\begin{aligned} Y_t &= F_t\theta_t + v_t & v_t &\sim \mathcal{N}(0, V_t) \\ \theta_t &= G_t\theta_{t-1} + \omega_t & \omega_t &\sim \mathcal{N}(0, W_t) \end{aligned}$$

$$\text{où : } \theta_t = \begin{pmatrix} \theta_{1,t} \\ \cdot \\ \cdot \\ \cdot \\ \theta_{h,t} \end{pmatrix}, \quad F_t = [F_{1,t} | \dots | F_{h,t}]$$

$G_t$  et  $W_t$  sont les éléments diagonaux des matrices suivantes :

$$G_t = \begin{pmatrix} G_{1,t} & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & G_{h,t} \end{pmatrix} \quad W_t = \begin{pmatrix} W_{1,t} & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & W_{h,t} \end{pmatrix}$$

et  $V_t = \sum_{i=1}^h V_{i,t}$ . Dans ce qui suit nous supposons que toutes les matrices définissant un DLM sont connues.

### Modèle de la tendance

Les modèles dynamiques linéaires polynômiaux sont les modèles les plus utilisés généralement pour décrire la tendance d'une série chronologique, où la tendance est vue comme développement du lissage de la série avec le temps.

A l'instant  $t$  la tendance prévue pour la série chronologique peut être considérée comme le comportement prévu de  $Y_{t+k}$  pour  $k \geq 1$ , en d'autres termes, la tendance prévue représente la fonction de prévision  $f_t(k) = E(Y_{t+k}|y_{1:t})$ . Le modèle polynômial d'ordre  $n$  est un DLM avec les matrices constantes  $F_t = F$  et  $G_t = G$ , et la fonction de prévision est de la forme :

$$f_t(k) = E(Y_{t+k}|y_{1:t}) = a_{t,0} + a_{t,1}k + \dots + a_{t,n-1}k^{n-1}, \quad k \geq 1 \quad (2.13)$$

où  $a_{t,0}, \dots, a_{t,n-1}$  sont des fonctions linéaires de  $m_t = E(\theta_t|y_{1:t})$  et sont indépendantes de  $k$ . Ainsi la fonction de prévision est un polynôme d'ordre  $n - 1$ . En général, n'importe quelle forme raisonnable de la fonction de prévision peut être décrite ou étroitement rapprochée par un polynôme, en choisissant  $n$  suffisamment grand.

Le modèle polynômial le plus connu est la marche aléatoire plus un bruit, qui est un modèle polynômial d'ordre  $n = 1$ , la distribution prédictive en  $k$ -étapes en avant pour ce modèle est

$$Y_{t+k}|y_{1:t} \sim \mathcal{N}(m_t, Q_t(k)), \quad k \geq 1 \quad (2.14)$$

où  $Q_t(k) = C_t + \sum_{j=1}^k W_{t+j} + V_{t+k} = C_t + kW + V$ . Nous voyons bien que la fonction de prévision  $f_t(k) = E(Y_{t+k}|y_{1:t}) = m_t$  est constante (fonction de  $k$ ). L'incertitude sur les futures observations est résumée par la variance  $Q_t(k) = C_t + kW + V$ .

Les matrices de contrôlabilité et d'observabilité pour ce modèle sont :

$$\mathcal{C} = [W^{1/2}]$$

$$\mathcal{O} = F = [1]$$

Le filtre de Kalman pour ce modèle est asymptotiquement stable, avec  $R_t$ ,  $C_t$  et la matrice  $K_t$  convergeant aux valeurs limites  $R$ ,  $C$  et  $K$ , respectivement (voir West et Harrison, 1997). Comme on peut montrer que :

$$K = \frac{r}{2} \left( \sqrt{1 + \frac{4}{R}} - 1 \right) \quad (2.15)$$

ce qui nous donne  $C = K.V$ . Ceci nous donne une limite supérieure à la précision possible en estimant la valeur courante de  $\mu_t$ . ainsi, nous obtenons la forme limite des prévisions en une étape en avant :

$$f_{t+1} = E(Y_{t+1}|y_{1:t}) = m_t = m_{t-1} + K_t(Y_t - m_{t-1}) = m_{t-1} + K_t e_t$$

Pour  $t$  assez grand,  $K_t \approx K$  asymptotiquement, la prévision en une étape en avant est donnée par :

$$f_{t+1} = m_{t-1} + K e_t \quad (2.16)$$

Cette fonction de prévision est employée dans beaucoup de modèles, elle correspond au facteur prédictif de point de Holt.

### Modèle saisonnier

Dans cette section nous présentons une manière de modeler une série chronologique qui montre un comportement cyclique, ou le « caractère saisonnier ».

Supposons que nous avons des données trimestrielles ( $Y_t, t = 1, 2, \dots$ ), qui montrent un comportement cyclique. Pour simplifier on suppose que la série a une moyenne zéro, une moyenne différente de zéro, ou une composante tendancielle peut être modelé séparément, pour le moment nous considérons la série purement saisonnière. Nous pourrions décrire la série en présentant des déviations saisonnières de la moyenne, exprimées par différents coefficients  $\alpha_i$  pour les différents quarts,  $i = 1, \dots, 4$ . Ainsi, si  $Y_{t-1}$  se rapporte au premier trimestre et  $Y_t$  au deuxième, on suppose

$$Y_{t-1} = \alpha_1 + v_{t-1} \quad (2.17)$$

$$Y_t = \alpha_2 + v_t$$

et ainsi de suite. Ce modèle peut être écrit comme un modèle dynamique linéaire, soit  $\theta_{t-1} = (\alpha_1, \alpha_4, \alpha_3, \alpha_2)'$  et  $F_t = F = (1, 0, 0, 0)$ , alors l'équation d'observation du DLM est donnée par

$$Y_{t-1} = F\theta_{t-1} + v_{t-1}$$

L'équation d'état doit tourner les composantes de  $\theta_{t-1}$  dans le vecteur  $\theta_t = (\alpha_2, \alpha_1, \alpha_4, \alpha_3)'$  de sorte que  $Y_t = F\theta_t + v_t = \alpha_2 + v_t$ . La permutation du vecteur d'état peut être obtenue par la matrice de permutation  $G$  définie par :

$$G = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

alors l'équation d'état peut être écrite comme suit

$$\theta_t = G\theta_{t-1} + \omega_t = (\alpha_2, \alpha_1, \alpha_4, \alpha_3)' + \omega_t$$

En général, une série chronologique saisonnière avec une période  $s$  peut être modélisée par un vecteur d'état  $\theta_t$  de dimension  $s$  des déviations saisonnières, en spécifiant un DLM avec  $F = (1, 0, \dots, 0)$  et  $G$  est obtenue par les permutations. Des contraintes d'identifiabilité doivent être imposées aux facteurs saisonniers  $\alpha_1, \dots, \alpha_s$ , tel que  $\sum_{i=1}^s \alpha_i = 0$ . la contrainte linéaire sur les  $s$  facteurs saisonniers implique qu'en réalité il y a seulement  $s - 1$  facteurs saisonniers, et ceci suggère une alternative, une représentation plus parcimonieuse qui emploie un vecteur d'état de dimension  $(s - 1)$ . Par exemple si les données sont trimestrielles, on considère  $\theta_{t-1} = (\alpha_1, \alpha_4, \alpha_3)'$  et  $\theta_t = (\alpha_2, \alpha_1, \alpha_4)'$  alors  $F = (1, 0, 0)$ . Pour passer de  $\theta_{t-1}$  à  $\theta_t$  supposons pour le moment un modèle statique sans erreurs avec la contrainte  $\sum_{i=1}^4 \alpha_i = 0$ , et on doit appliquer la transformation linéaire donnée par la matrice suivante :

$$G = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Généralement pour un modèle saisonnier de période  $s$ , on considère l'espace-état est de dimension  $(s - 1)$ , avec  $F = (1, 0, \dots, 0)$  et

$$G = \begin{pmatrix} -1 & -1 & \dots & \dots & -1 & -1 \\ 1 & 0 & & & 0 & 0 \\ 0 & 1 & & & 0 & 0 \\ & & \ddots & & & \\ & & & \ddots & & \\ 0 & 0 & & & 1 & 0 \end{pmatrix}$$

Une variation dynamique des composantes saisonnières peut être présentée par un système d'évolution d'erreur avec une variance  $W = \text{diag}(\sigma_\omega^2, 0, \dots, 0)$ .

### Représentation des modèles *ARIMA* avec un DLM

N'importe quel modèle *ARIMA* peut être représenté comme un DLM, avec plus de précision, pour n'importe quel processus *ARIMA*, il est possible de trouver un DLM dont le processus de mesure ( $Y_t$ ) a la même distribution que l'*ARIMA* donné. L'espace-état avec sa dynamique n'est pas déterminé : plusieurs représentations ont été proposées dans la littérature. Ici nous présenterons seulement une d'entre elles, qui est probablement la plus utilisée. Pour les différentes représentations voir Gourieroux et Monfort (1997).

On va Commencer par le cas stationnaire, considérons le processus *ARMA*( $p, q$ )

$$Y_t = \sum_{j=1}^r \phi_j Y_{t-j} + \sum_{j=1}^{r-1} \psi_j \epsilon_{t-j} + \epsilon_t$$

avec  $r = \max\{p, q + 1\}$ ,  $\phi_j = 0$  pour  $j > p$ , et  $\psi_j = 0$  pour  $j > q$ , Définissons les matrices suivante :

$$F = ( 1 \ 0 \ \dots \ 0 )$$

$$G = \begin{pmatrix} \phi_1 & 1 & 0 & \dots & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \\ \vdots & \vdots & \vdots & & \ddots & \\ \phi_{r-1} & 0 & \dots & \dots & 0 & 1 \\ \phi_r & 0 & \dots & \dots & 0 & 0 \end{pmatrix} \tag{2.18}$$

$$R = ( 1 \ \psi_1 \ \dots \ \psi_{r-2} \ \psi_{r-1} )'$$

Si on introduit le vecteur d'état  $\theta_t = (\theta_{1,t}, \dots, \theta_{r,t})'$  alors le modèle *ARMA* aura une représentation du DLM suivant :

$$\begin{cases} Y_t = F\theta_t \\ \theta_{t+1} = G\theta_t + R\epsilon_t \end{cases} \quad (2.19)$$

c'est un modèle dynamique linéaire avec  $V = 0$  et  $W = RR'\sigma^2$ , où  $\sigma^2$  est la variance de la séquence d'erreur  $(\epsilon_t)$ . Pour vérifier cette équivalence, notons que l'équation d'observation donne  $y_t = \theta_{1,t}$ , et l'équation d'état est :

$$\begin{aligned} \theta_{1,t} &= \phi_1\theta_{1,t-1} + \theta_{2,t-1} + \epsilon_t \\ \theta_{2,t} &= \phi_2\theta_{1,t-1} + \theta_{3,t-1} + \psi_1\epsilon_t \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ \theta_{r-1,t} &= \phi_{r-1}\theta_{1,t-1} + \theta_{r,t-1} + \psi_{r-2}\epsilon_t \\ \theta_{r,t} &= \phi_r\theta_{1,t-1} + \psi_{r-1}\epsilon_t \end{aligned} \quad (2.20)$$

En remplaçant l'expression de  $\theta_{2,t-1}$  obtenue à partir de la deuxième équation, dans la première équation, on aura

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \phi_2\theta_{1,t-2} + \theta_{3,t-2} + \psi_1\epsilon_{t-1} + \epsilon_t$$

et en procédant successivement on obtient :

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \dots + \phi_r\theta_{1,t-r} + \psi_1\epsilon_{t-1} + \dots + \psi_{r-1}\epsilon_{t-r-1} + \epsilon_t$$

Rappelons que  $r = \max\{p, q + 1\}$  et  $y_t = \theta_{1,t}$ , on voit bien que c'est un modèle *ARMA*.

Prenons le cas le plus simple des modèles auto-régressifs, par exemple un *AR(2)*

$$Y_t = \phi_1Y_{t-1} + \phi_2Y_{t-2} + \epsilon_t \quad \epsilon_t \sim^{iid} \mathcal{N}(0, \sigma^2) \quad (2.21)$$

On peut bien penser à une représentation plus simple pour l'*AR(2)* comme un DLM avec  $F_t = [Y_{t-1}, Y_{t-2}]$  et  $\theta_t = [\phi_{1,t}, \phi_{2,t}]'$ , la matrice  $F_t$  ne dépend pas des observations passées : dans notre cas sa implique

$$Y_t | y_{t-1}, y_{t-2}, \theta_t \sim \mathcal{N}(\phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2)$$

c'est-à-dire,  $Y_t$  n'est pas indépendant des valeurs passées  $y_{t-1}$ ,  $y_{t-2}$  sachant  $\theta_t$ , de ce fait on n'a pas l'hypothèse (A.2) dans la définition du modèle espace-état. Afin de définir un tel modèle on va prolonger la définition pour inclure les modèles espace-état conditionnellement gaussiens (Lipster et Shiryaev ; 1972). On va augmenter le vecteur d'état par une deuxième composante  $\theta_{2,t}$ , et on va choisir  $G$  et  $W$  tels que

$$\begin{pmatrix} \theta_{1,t} \\ \theta_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} \theta_{1,t-1} \\ \theta_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}$$

et on a

$$\theta_{1,t} = \phi_1 \theta_{1,t-1} + \phi_2 \theta_{1,t-2} + \epsilon_t$$

De cette façon, nous obtenons une représentation DLM pour le modèle  $AR(p)$  dans le cadre des modèles espace-état. Une autre étape est nécessaire pour exprimer la composante  $MA(q)$ , Par exemple, pour un modèle  $ARMA(1, 1)$

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t + \psi_1 \epsilon_{t-1} \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (2.22)$$

$r = q + 1 = 2$  et les matrices du DLM sont :

$$\begin{aligned} F &= \begin{pmatrix} 1 & 0 \end{pmatrix} & V &= 0 \\ G &= \begin{pmatrix} \phi_1 & 1 \\ 0 & 0 \end{pmatrix} & & (2.23) \\ W &= \begin{pmatrix} 1 & \psi_1 \\ \psi_1 & \psi_1^2 \end{pmatrix} \sigma^2 \end{aligned}$$

**Remarque 2.3.** La représentation d'un modèle  $ARMA$  comme un DLM est utile car la composante  $ARMA$  dans un DLM peut expliquer l'autocorrélation résiduel non expliqué par d'autres composantes tels que la tendance et la saisonnalité

La représentation du modèle  $ARIMA(p, d, q)$ , avec  $d > 0$  comme un DLM, peut être considérée comme prolongation du cas stationnaire. En fait, si on considère  $Y_t^* = \Delta^d Y_t$  alors  $Y_t^*$  est un modèle  $ARMA$  stationnaire, pour lequel sa représentation autant qu'un DLM est donnée ci-dessus. Supposons que  $Y_t^*$  est un  $AR(2)$  alors La représentation de  $Y_t$  autant qu'un DLM est donnée par le système

$$\begin{cases} Y_t = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \theta_{t-1} \\ \theta_t = \begin{pmatrix} 1 & 1 & 0 \\ 0 & \phi_1 & 0 \\ 0 & \phi_2 & 1 \end{pmatrix} \theta_{t-1} + \omega_t \quad \omega_t \sim \mathcal{N}(0, W) \end{cases} \quad (2.24)$$

avec

$$\theta_t = \begin{pmatrix} Y_{t-1} \\ Y_t^* \\ \phi_2 Y_{t-1}^* \end{pmatrix} \quad (2.25)$$

et  $W = \text{diag}(0, \sigma^2, 0)$ . Pour un  $d$  quelconque  $Y_t^* = \Delta^d Y_t$ , on peut montrer que la formule suivante est vraie

$$\Delta^{d-j} Y_t = Y_t^* + \sum_{i=1}^j \Delta^{d-i} Y_{t-i} \quad j = 1, \dots, d \quad (2.26)$$

et le vecteur d'état est

$$\begin{pmatrix} Y_{t-1} \\ \Delta Y_{t-1} \\ \vdots \\ \Delta^{d-1} Y_{t-1} \\ Y_t^* \\ \phi_2 Y_{t-1}^* + \dots + \phi_r Y_{t-r+1}^* + \psi_1 \epsilon_t = \dots + \psi_{r-1} \epsilon_{t-r+2} \\ \phi_3 Y_{t-1}^* + \dots + \phi_r Y_{t-r+2}^* + \psi_2 \epsilon_t = \dots + \psi_{r-1} \epsilon_{t-r+3} \\ \vdots \\ \phi_r Y_{t-1}^* + \psi_{r-1} \epsilon_t \end{pmatrix} \quad (2.27)$$

Les matrices du système sont définies comme suit

$$F = \begin{pmatrix} 1 & 1 & \dots & \dots & 1 & 0 & \dots & \dots & 0 \end{pmatrix}$$

$$\begin{aligned}
G &= \begin{pmatrix} 1 & 1 & \dots & \dots & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots \\ \dots & \dots & \dots & 1 & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & \phi_1 & 1 & 0 & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \phi_2 & 0 & 1 & \dots & \dots & \dots & 0 \\ \dots & \dots \\ \dots & \dots \\ \dots & \dots \\ \dots & \dots & \dots & \dots & \phi_{r-1} & 0 & \dots & \dots & \dots & 0 & 1 \\ 0 & \dots & \dots & 0 & \phi_r & 0 & \dots & \dots & \dots & 0 & 0 \end{pmatrix} \\
R &= ( 0 \quad \dots \quad 0 \quad 1 \quad \psi_1 \quad \dots \quad \psi_{r-2} \quad \psi_{r-1} )' \\
W &= RR'\sigma^2
\end{aligned} \tag{2.28}$$

Avec les définitions ci-dessus le modèle dynamique linéaire pour l'*ARIMA* ( $Y_t$ ) est

$$\begin{cases} Y_t = F\theta_t \\ \theta_t = G\theta_{t-1} + \omega_t \end{cases} \quad \omega_t \sim \mathcal{N}(0, W) \tag{2.29}$$

## 2.7 Modèles avec des paramètres inconnus

Dans la section précédente on a présenté certain DLM de base pour l'analyse des séries chronologiques, en supposant que les matrices du système  $F_t$ ,  $G_t$ ,  $V_t$  et  $W_t$  sont connues, ce qui n'est pas le cas en pratique, dans cette section on va s'intéresser aux modèles où les matrices dépendent d'un vecteur de paramètres inconnus  $\psi_t$ .

Dans un cadre classique on commence par estimer  $\psi_t$ , habituellement par la méthode du maximum de vraisemblance, si le chercheur est seulement intéressé par les paramètres inconnus, l'analyse se termine ici; mais s'il est intéressé au lissage ou à la prédiction, la manière usuelle de procéder c'est de considérer l'estimation de  $\psi_t$  comme une constante connue, et appliquer les techniques appropriées dans les sections précédentes pour des prévisions ou le lissage. D'un point de vue Bayésien, les paramètres inconnus sont plutôt des quantités aléatoires. Même si le principe de l'inférence Bayésienne est simple, mais sa implique des calculs qui ne sont pas analytiques; cependant, les méthodes de Monte Carlo par chaîne de Markov peuvent être efficaces en fournissant une approximation des distributions.

### 2.7.1 Estimation avec le maximum de vraisemblance

Supposons qu'on a  $n$  vecteurs aléatoires,  $Y_1, \dots, Y_n$ , dont leurs distributions dépendent d'un paramètre inconnu  $\psi$ . On note la densité jointe des observations pour une valeur particulière du paramètre par  $P(y_1, \dots, y_n; \psi)$ , la fonction de vraisemblance est définie jusqu'à un facteur constant, la densité de probabilité des données observées est en fonction de  $\psi$ , on peut écrire  $L(\psi) = \text{const.} P(y_1, \dots, y_n; \psi)$ . Pour un modèle dynamique linéaire la densité jointe des observations peut être écrite sous la forme suivante

$$P(y_1, \dots, y_n; \psi) = \prod_{t=1}^n P(y_t | y_{1:t-1}; \psi) \quad (2.30)$$

où  $P(y_t | y_{1:t-1}; \psi)$  est la densité conditionnelle de  $y_t$  sachant les données jusqu'au temps  $t - 1$ , supposons que  $\psi$  est la valeur du paramètre inconnu. Cette densité est gaussienne de moyenne  $f_t$  et de variance  $Q_t$ , Par conséquent nous pouvons écrire le log de la vraisemblance :

$$l(\psi) = -\frac{1}{2} \sum_{t=1}^n \log |Q_t| - \frac{1}{2} \sum_{t=1}^n (y_t - f_t)' Q_t^{-1} (y_t - f_t) \quad (2.31)$$

où  $f_t$  et  $Q_t$  dépendent implicitement de  $\psi$ , la formule précédente peut être numériquement maximisée pour obtenir le MLE de  $\psi$  :

$$\hat{\psi} = \arg \max_{\psi} l(\psi) \quad (2.32)$$

On note par  $H$  la matrice Hessian de  $-l(\psi)$ , évaluée a  $\psi = \hat{\psi}$ , La matrice  $H^{-1}$  nous donne l'estimation de la variance du MLE,  $Var(\hat{\psi})$  (Caines (1988), Hannan et Deistler (1988), et Shumway et Stoffer (2000)). L'optimisation numérique de la fonction de vraisemblance d'un DLM peut présenter plusieurs maximum.

### 2.7.2 Inférence Bayésienne

En pratique, l'utilisation du principe de maximum de vraisemblance n'est pas efficace vu l'incertitude sur le paramètre  $\psi$ , l'approche Bayésienne offre une formulation plus cohérente au problème, car le paramètre inconnu est considéré comme un vecteur aléatoire, la connaissance a priori sur  $\psi$  est exprimée par la probabilité  $\pi(\psi)$ . Ainsi, pour tous  $n \geq 1$  on suppose que :

$$(\theta_0, \theta_1, \dots, \theta_n, Y_1, \dots, Y_n, \psi) \sim \pi(\theta_0|\psi)\pi(\psi) \prod_{t=1}^n \pi(y_t|\theta_t, \psi)\pi(\theta_t|\theta_{t-1}, \psi) \quad (2.33)$$

En donnant les données  $y_{1:t}$ , l'inférence sur l'état inconnu  $\theta_s$  à l'instant  $s$  et sur les paramètres est résolue en calculant la distribution jointe a posteriori

$$\pi(\theta_s, \psi|y_{1:t}) = \pi(\theta_s|\psi, y_{1:t})\pi(\psi|y_{1:t}) \quad (2.34)$$

En général, on est intéressé au problème de filtrage ( $s = t$ ), la prédiction d'état  $s > t$ , et au lissage  $s < t$ . La densité conditionnelle marginale de  $\theta_s$  est obtenue à partir de l'équation précédente, par exemple la densité de filtrage est donnée par

$$\pi(\theta_t|y_{1:t}) = \int \pi(\theta_t|\psi, y_{1:t})\pi(\psi|y_{1:t})d\psi$$

L'inférence sur  $\theta_{0:t}$  et  $\psi$ , sachant les données  $y_{1:t}$ , est exprimée par leur densité a posteriori jointe :

$$\pi(\theta_{0:t}, \psi|y_{1:t}) = \pi(\theta_{0:t}|\psi, y_{1:t})\pi(\psi|y_{1:t}) \quad (2.35)$$

En principe, cette densité a posteriori est obtenue en appliquant la règle de Bayes. MCMC et, en particulier, l'échantillonnage de Gibbs est employé pour l'approximation de la loi jointe a posteriori  $\pi(\theta_{0:t}, \psi|y_{1:t})$ .

### 2.7.3 Inférence Bayésienne conjuguée

Dans certains modèles simples où les matrices du système  $F_t$  et  $G_t$  sont connues, mais les matrices de covariance  $V_t$  et  $W_t$  sont rarement connues, alors, un problème de base est d'estimer  $V_t$  et  $W_t$ . Ici on va considérer un cas simple où les matrices de covariances sont connues jusqu'à un facteur en commun ; c'est-à-dire,  $V_t = \sigma^2\tilde{V}_t$  et  $W_t = \sigma^2\tilde{W}_t$ , avec  $\sigma^2$  est inconnue.

#### 2.7.3.1 Matrices de covariance inconnues : inférence conjuguée

Soit  $((Y_t, \theta_t) : t = 1, 2, \dots)$  définit par le modèle dynamique linéaire suivant :

$$V_t = \sigma^2\tilde{V}_t, \quad W_t = \sigma^2\tilde{W}_t, \quad C_0 = \sigma^2\tilde{C}_0 \quad (2.36)$$

Ici on suppose que toutes les matrices  $\tilde{V}_t$ ,  $\tilde{W}_t$ , ainsi que  $\tilde{C}_0$  sont connues, et  $\sigma^2$  est inconnue, dans l'inférence Bayésienne il est intéressant de travailler avec l'inverse  $\phi = 1/\sigma^2$ .

Toute l'incertitude est, donc, dans les vecteurs d'état et dans le paramètre  $\phi$ . Pour tous  $t \geq 1$  on suppose que :

$$\begin{aligned} Y_t | \theta_t, \phi &\sim \mathcal{N}_m(F_t \theta_t, \phi^{-1} \tilde{V}_t) \\ \theta_t | \theta_{t-1}, \phi &\sim \mathcal{N}_p(G_t \theta_{t-1}, \phi^{-1} \tilde{W}_t) \end{aligned}$$

comme loi a priori pour  $(\phi, \theta_0)$  il est convenant de choisir la loi conjuguée gamma normale :

$$\phi \sim \mathcal{G}(\alpha_0, \beta_0) \quad \theta_0 | \phi \sim \mathcal{N}(m_0, \phi^{-1} \tilde{C}_0)$$

alors  $(\theta_0, \phi) \sim \mathcal{NG}(m_0, \tilde{C}_0, \alpha_0, \beta_0)$ , et on peut avoir les formules récursives suivantes pour le filtrage.

**Proposition 2.9.** (Petris et al. 2009)

Pour un modèle dynamique linéaire donné, si

$$(\theta_{t-1}, \phi) | y_{1:t-1} \sim \mathcal{NG}(m_{t-1}, \tilde{C}_{t-1}, \alpha_{t-1}, \beta_{t-1})$$

où  $t \geq 1$ , alors :

(i) La densité prédictive en une étape en avant de  $(\theta_t, \phi) | y_{1:t-1}$  est gamma normale de paramètres  $(a_t, \tilde{R}_t, \alpha_{t-1}, \beta_{t-1})$ , où

$$a_t = G_t m_{t-1}, \quad \tilde{R}_t = G_t \tilde{C}_{t-1} G_t' + \tilde{W}_t \quad (2.37)$$

(ii) La densité prédictive en une étape en avant de  $Y_t | y_{1:t-1}$  est  $t$ -student de paramètres  $(f_t, \tilde{Q}_t \beta_{t-1} / \alpha_{t-1}, 2\alpha_{t-1})$ , où

$$f_t = F_t a_t \quad \tilde{Q}_t = F_t \tilde{R}_t F_t' + \tilde{V}_t \quad (2.38)$$

(iii) La densité de filtrage de  $(\theta_t, \phi | y_{1:t})$  est gamma normale de paramètres :

$$\begin{aligned} m_t &= a_t + \tilde{R}_t F_t \tilde{Q}_t^{-1} (y_t - f_t) \\ \tilde{C}_t &= \tilde{R}_t - \tilde{R}_t F_t' \tilde{Q}_t^{-1} \tilde{R}_t \\ \alpha_t &= \alpha_{t-1} + \frac{m}{2} \\ \beta_t &= \beta_{t-1} + \frac{1}{2} (y_t - f_t)' \tilde{Q}_t^{-1} (y_t - f_t) \end{aligned} \quad (2.39)$$

De (iii), la densité marginale de filtrage de  $\sigma^2 = \phi^{-1}$  sachant  $y_{1:t}$  est inverse gamma de paramètres  $(\alpha_t, \beta_t)$ , tel que pour  $\alpha_t > 2$  :

$$E(\sigma^2|y_{1:t}) = \frac{\beta_t}{\alpha_t - 1} \quad \text{Var}(\sigma^2|y_{1:t}) = \frac{\beta_t^2}{(\alpha_t - 1)^2(\alpha_t - 2)}$$

La densité marginale de filtrage des états est  $t$ -student :

$$\theta_t|y_{1:t} \sim \mathcal{T}(m_t, \tilde{C}_t\beta_t/\alpha_t, 2\alpha_t)$$

avec

$$E(\theta_t|y_{1:t}) = m_t \quad \text{Var}(\theta_t|y_{1:t}) = \frac{\beta_t}{\alpha_t - 1} \tilde{C}_t \quad (2.40)$$

Si  $\sigma^2$  est connue, le filtre de Kalman nous donnera la même chose,  $E(\theta_t|y_{1:t}) = m_t$ , avec la matrice de covariance  $\text{Var}(\theta_t|y_{1:t}) = \sigma^2 \tilde{C}_t$ .

En ce qui concerne le lissage, on a :

$$(\theta_T, \phi|y_{1:T}) \sim \mathcal{NG}(s_T, \tilde{S}_T, \alpha_T, \beta_T) \quad (2.41)$$

avec  $s_T = m_t$ ,  $\tilde{S}_T = \tilde{C}_T$ , et on écrit :

$$\pi(\theta_t, \phi|y_{1:T}) = \pi(\theta_t|\phi, y_{1:T})\pi(\phi|y_{1:T}) \quad t = 1, \dots, T \quad (2.42)$$

### 2.7.3.2 Spécification des paramètres par le facteur d'escompte

#### *Spécification de $W_t$ par le facteur d'escompte*

L'analyse conjuguée Bayésienne présentée dans la section précédente s'applique si les matrices  $\tilde{V}_t$ ,  $\tilde{W}_t$  et  $\tilde{C}_t$  sont connues, ce qui n'est pas évident. Cependant, un cas intéressant est quand  $\tilde{V}_t = I_m$  et  $\tilde{W}_t$  est spécifié par le facteur d'escompte (West et Harrison (1997)). De grandes valeurs pour la diagonale de  $W_t$  implique une incertitude élevée dans l'évolution de l'état, de sorte que beaucoup d'information sont perdues en passant de  $\theta_{t-1}$  à  $\theta_t$ , les observations passées  $y_{1:t-1}$  fournissent des informations sur  $\theta_{t-1}$ , qui, devient de peu d'importance dans la prévision de  $\theta_t$ . Dans les récursions du filtre de Kalman, l'incertitude sur  $\theta_{t-1}$  sachant  $y_{1:t-1}$  est résumée dans la matrice de covariance  $\text{Var}(\theta_{t-1}|y_{1:t-1}) = C_{t-1}$ . En passant de  $\theta_{t-1}$  à  $\theta_t$  par l'équation d'état  $\theta_t = G_t\theta_{t-1} + W_t$ , l'incertitude augmente, et on

a  $Var(\theta_t|y_{1:t-1}) = R_t = G'_t C_{t-1} G_t + W_t$ , si  $W_t = 0$  il n'y a aucune erreur dans l'équation d'état,  $R_t = Var(G_t \theta_{t-1} | y_{1:t-1}) = P_t$ , et on a  $R_t = P_t + W_t$ . Dans ce sens,  $W_t$  exprime la perte d'information en passant de  $\theta_{t-1}$  à  $\theta_t$  dû à la composante d'erreur stochastique dans l'évolution d'état, on peut exprimer  $W_t$  comme proportion de  $P_t$  :

$$W_t = \frac{1 - \delta}{\delta} P_t$$

où  $\delta \in (0, 1]$ , il suit que  $R_t = 1/\delta P_t$  avec  $1/\delta > 1$ , le paramètre  $\delta$  s'appelle facteur d'escompte.

**Remarque 2.4.** Si  $\delta = 1$ ,  $W_t = 0$  on a aucune perte d'information de  $\theta_{t-1}$  à  $\theta_t$  :  $Var(\theta_t|y_{1:t}) = Var(G_t \theta_{t-1} | y_{1:t-1})$ .

Pour  $\delta < 1$  par exemple  $\delta = 0.8$  on a  $Var(\theta_t|y_{1:t}) = (1/0.8)Var(G_t \theta_{t-1} | y_{1:t-1}) = 1.25P_t$  et cela représente une grande incertitude. Dans la pratique, la valeur du facteur d'escompte est habituellement fixée entre 0.9 et 0.99.

La spécification du facteur d'escompte peut être utilisée avec les prioris conjugués :

$$\tilde{W}_t = \frac{1 - \delta}{\delta} G'_t \tilde{C}_{t-1} G_t \tag{2.43}$$

sachant  $\tilde{C}_0$  et  $\tilde{V}_t$  ( $\tilde{V}_t = I_m$ ) la valeur de  $\tilde{W}_t$  peut être calculée récursivement pour chaque  $t$ .

### **Modèle de facteur d'escompte pour $V_t$ variant dans le temps**

Dans la section précédente le paramètre inconnu  $\phi$  est supposé constant dans le temps Soit le modèle dynamique linéaire définie dans (2.36), et supposons qu'à l'instant  $t - 1$ , on a

$$\phi_{t-1} | y_{1:t-1} \sim \mathcal{G}(\alpha_{t-1}, \beta_{t-1})$$

maintenant  $\phi$  évolue de  $t - 1$  à  $t$  par conséquent l'incertitude sur  $\phi_t$  sachant les données  $y_{1:t-1}$  est grande :  $Var(\phi_t | y_{1:t-1}) > Var(\phi_{t-1} | y_{1:t-1})$ . Supposons pour le moment que la densité de  $\phi_t | y_{1:t-1}$  est gamma, et en particulier supposons que :

$$\phi_t | y_{1:t-1} \sim \mathcal{G}(\delta^* \alpha_{t-1}, \delta^* \beta_{t-1}) \tag{2.44}$$

où  $0 < \delta^* < 1$ , notons que  $E(\phi_t|y_{1:t-1}) = E(\phi_{t-1}|y_{1:t-1}) = \alpha_{t-1}/\beta_{t-1}$ , tandis que la variance est grande :  $Var(\phi_t|y_{1:t-1}) = 1/\delta^* Var(\phi_{t-1}|y_{1:t-1})$ , avec  $1/\delta^* > 1$ , soit  $\alpha_t^* = \delta^* \alpha_{t-1}$ ,  $\beta_t^* = \delta^* \beta_{t-1}$ , et on obtient :

$$Y_t|y_{1:t-1} \sim \mathcal{T}(f_t, \tilde{Q}_t \frac{\beta_t^*}{\alpha_t^*}, 2\alpha_t^*)$$

et

$$\theta_t|y_{1:t} \sim \mathcal{T}(m_t, \tilde{C}_t \frac{\beta_t}{\alpha_t}, 2\alpha)$$

où

$$\alpha_t = \alpha_t^* + \frac{m}{2}, \quad \beta_t = \beta_t^* + \frac{1}{2}(y_t - f_t)' \tilde{Q}_t^{-1} (y_t - f_t) \quad (2.45)$$

# Chapitre 3

## Modèles dynamiques linéaires pour la demande d'eau

### 3.1 Description du réseau de distribution d'eau dans la région de Kabylie

La région de Tizi-ouzou se situe dans le nord de l'Algérie à  $103km$  à l'est de la capitale, limitée au nord par la mer méditerranée, à l'est par la wilaya de Bejaia, à l'ouest par la wilaya de Boumerdes, et au sud par la wilaya de Bouira.

La superficie totale de la wilaya est de  $2957km^2$ , dont plus de 80% se trouve sur une pente supérieure à 12%. L'espace montagneux est donc prédominant dans le territoire. Ce dernier est constitué par le massif Kabylie, le Djurdjura. Cette région est traversée de l'est à l'ouest par l'Oued-Sebaou, qui recueille, à travers ses affluents, l'essentiel des eaux en provenance du Djurdjura.

La population de la wilaya est alimentée par des eaux souterraines, puisées à partir des nappes alluviales des Oued-Sebaou, Ait-Khellil, Rebta, Bouguedoura, Takhoukt, Djemaa, Sidi-Khelifa et Ouassif. Des eaux superficielles mobilisées à partir du barrage de Taksebt, de prise d'eaux sur oued, de sources naturelles, et récemment, par des eaux desalées (Tigzirt). La pluviométrie, qui est un facteur déterminant dans l'alimentation en eau potable de la wilaya, tourne autour d'une moyenne de 800 à 1000  $mm/an$ .

Les eaux souterraines constituent à l'heure actuelle, la source principale de la production en eau potable dans la wilaya. La production annuelle atteint 63 millions  $m^3$  répartie

par origine des eaux comme suit :

Eaux souterraines : 72%

Eaux superficielles : 19%

Eaux de sources : 08%

Eaux dessalées : 01%

Sur les 67 communes de la wilaya, l'unité de l'ADE (Algérienne Des Eaux) de Tizi-ouzou gère actuellement 59, soit une population desservie de 1146611 habitants implantés essentiellement dans les villages. La majeure partie de ceux-ci est située en zone montagneuse sur des altitudes dépassant parfois les 900m.

De part sa configuration géographique de la wilaya, l'alimentation en eau potable se fait par un système complexe de refoulement fait généralement de trois à quatre stations de pompage qui refoulent l'eau d'une altitude à une autre jusqu'à atteindre l'abonné.

L'éparpillement des zones habitables (1380 villages) ajoute à la complexité d'exploitation, une difficulté supplémentaire, la nécessité de multiplier les infrastructures de mobilisation, de transfert et de distribution.

L'algérienne des eaux de Tizi-ouzou gère actuellement :

171 forages et puits.

24 sources.

06 stations de traitement.

01 station de dessalement de l'eau de mer.

123 stations de pompage.

698 réservoirs.

1600km de réseaux d'adductions soit 12% du parc national.

Ceci entraîne des coûts en énergie électrique considérable, le coût spécifique de l'énergie par mètre cube d'eau produit à Tizi-ouzou est trois fois supérieur au coût spécifique moyen national. Le fonctionnement simultané des infrastructures électromécaniques nécessite une puissance énergétique équivalente à 85000KVA, ce qui place l'ADE de loin le premier client de la sonelgaz de la wilaya.

Le service des eaux de Tizi-ouzou veille, à travers son laboratoire central et ses annexes, à mettre à la disposition de ses clients une eau répondant aux normes de la qualité. A cet effet, un grand nombre de contrôles sont effectués dans la wilaya, un suivi quotidien et permanent est assuré et la consommation moyenne annuelle en produits chimiques (hypochlorite et sulfate d'alumine) avoisine les 450 tonnes.

La distribution des eaux de Tizi-ouzou est assurée par 18 centres opérationnels repartis sur les principaux chefs lieux de dairas de la wilaya, ces centres gèrent en plus des activités de productions, plus de 165 milles abonnées, facturé à près de 98% au réel (au compteur) le management des services publics va certainement marquer de son empreinte l'ensemble des actions de réforme et de modernisation qui seront entreprises dans le cadre de l'amélioration des prestations envers les usagers (clients).

De l'analyse de la conduite de certaines activités au niveau des services publics, il ressort généralement un cloisonnement entre différents centres d'interventions. Ceci induit à l'intérieur de l'entreprise des fonctionnements de gestion tels qu'on perd de vue l'objectif essentiel : apporter des réponses concrètes aux questions que se posent les clients.

Une approche de management par la méthode du processus semble, aujourd'hui, l'approche la mieux indiquée pour répondre aux exigences de qualité. Manager par processus consiste donc, a partir du point de vue du client exprimé sous forme d'un cahier des charges du service à lui rendre, à définir les actions à réaliser par les services chargés des différentes activités du processus permettant d'assurer le service attendu de façon permanente. C'est sur ce principe de base que l'ADE s'est engagée pour piloter les chaîne d'AEP (Alimentation en Eau Potable) qui reste la colonne vertébrale de la production et la distribution des eaux potables.

Les chaînes d'AEP sont des systèmes complexes, constitués généralement d'une batterie de forage, de conduite de refoulement et d'adduction de différents matériaux et diamètres, de stations de pompes équipées par des infrastructures éclectiques et hydromécaniques et de réservoirs.

### 3.1.1 La chaîne de Tassadort

La source essentielle d'alimentation en eau potable de la région par la chaîne de Tassadort est un champ captant situé au lieu dit Bouaid à 04km à l'ouest de la ville de Tizi-ouzou, il comporte 10 forages dont 09 opérationnels, implantés sur la rive gauche d'Oued-Sebaou, à une altitude de 50m NGA (Nivellement Général Algérien), totalisent un débit maximal de 438.5l/s, soit  $37886.4m^3/j$ .

La distribution d'eau est effectuée par l'ADE, qui dispose de 18 centres opérationnels dont 04 qui gèrent la chaîne de Tassadort, l'ADE offre à son tour de l'eau potable de haute qualité, car cette eau est extraite des aquifères en premier lieu (eau de source), ensuite se traite dans les laboratoires suréquipés d'un matériel très sophistiqué et avec une disponibilité à jour des produits chimiques afin de rendre l'eau consommable à zéro (%) de risque. Dans cette région, l'eau est aussi utilisée pour des besoins ménagers, et aussi pour l'élevage des animaux.

Notre région d'étude est alimentée par des eaux souterraines captées par des forages qui se situent dans les nappes d'Oued-Sebaou :

09 forages au champ captant de Bouaid avec une capacité de pompage de  $1040m^3/h$ .

05 forages au champ captant de Takhoukt avec une capacité de pompage de  $140m^3/h$ .

Des eaux souterraines captées à partir du barrage Taksebt (source alternative), on note qu'au premier trimestre 2010, on a un transfert de  $295347m^3$  de ce barrage.



### 3.1.2 Description physique du réseau

La majorité de l'eau provient des aquifères (forages) et on l'introduit dans le réseau par le moyen de pompage. Le réseau de distribution dans la région Kabyle est composé de 38 réservoirs, 06 stations de pompages, ainsi qu'un réseau de tuyauteries qui s'étend sur 25000 mètres linéaires en communiquant tous les points du réseau. Les stations de pompage poussent l'eau depuis des points moins importants jusqu'à des points plus importants, par exemple Tassadort pompe de l'eau d'une altitude de 92m vers Mezdata 712m d'altitude.

Les canalisations qui envoient l'eau depuis les gares de pompages jusqu'aux réservoirs supportent des pressions assez importantes, depuis les réservoirs de têtes (les réservoirs les plus importants auxquels arrive l'eau depuis les stations de pompage) l'eau est envoyée directement vers d'autres réservoirs moins importants ou aux villes (villages). Cette eau est transportée par le réseau de canalisations mais elle se déplace d'une altitude à une autre par gravité, donc il n'est pas nécessaire de la pomper (pas de coût électrique) et la pression que supportent ces canalisations est petite.

La capacité de stockage dans les réservoirs de tête est très importante dans le fonctionnement de cette chaîne, car plus la capacité augmente plus les risques de non approvisionnement des populations diminue et ceci est dû au fait qu'on peut pomper suffisamment d'eau pendant les heures où l'eau est moins chère afin de la consommer tout au long de la journée.

Les histogrammes suivants indiquent les différentes capacités et hauteurs des réservoirs :

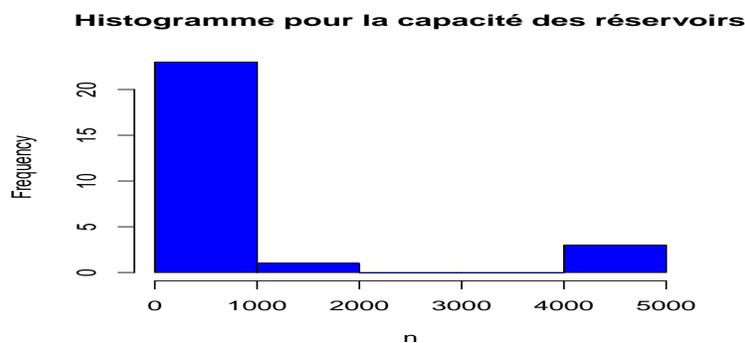


FIGURE 3.2 – Les différentes capacités des réservoirs

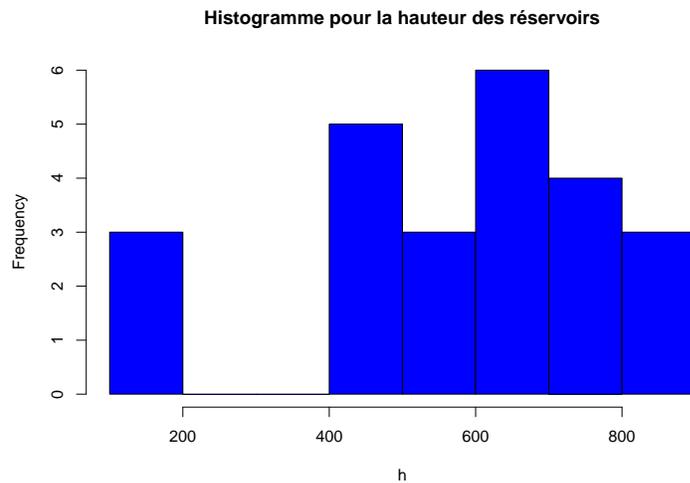


FIGURE 3.3 – Les différentes hauteurs des réservoirs

### 3.1.3 Alimentation équitable en eau potable

Des problèmes d'alimentation d'eau peuvent être décrits en termes relativement simples, voir par exemple Soncini-Sessa(2007). Nous avons un certain nombre de sources d'eau, habituellement les puits et les réservoirs, qui fournissent ensemble l'eau. Nous avons également un certain nombre de points de consommation, qui, selon la granularité du problème, pourraient se rapporter à des maisons ou à des villages. L'ADE vise à satisfaire des besoins d'approvisionnement en eau, tenant compte d'un certain nombre de contraintes physiques, qui reflètent la structure du réseau de distribution, typiquement avec les dépôts et les stations de pompage intermédiaires. L'ADE dit qu'elle ne peut pas satisfaire complètement la demande, et qu'elle recherche, ainsi, des manières de rendre cette demande aussi équilibrée que possible auprès des consommateurs, c'est ce que nous appelons dans ce travail : alimentation équitable en eau potable. Le schéma suivant fournit un arrangement simple, qui reflète la structure du problème utilisé avec seulement des puits comme points de source, et les dépôts en tant que points intermédiaires.

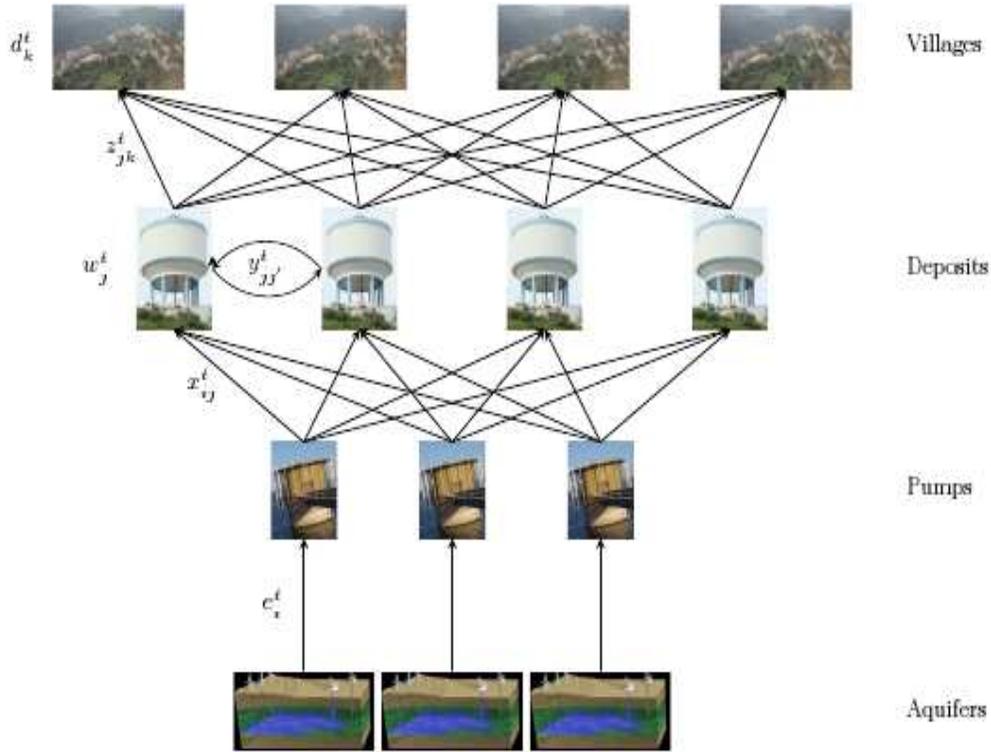


FIGURE 3.4 – Réseau de distribution d'eau

### 3.1.4 Les variables et les contraintes

On suppose qu'on a  $N_p$  points source, avec  $i \in P$ ;  $N_d$  points intermédiaires,  $j \in D$ ;  $N_v$  points de consommation,  $k \in V$ ;  $N_\omega$  pompes,  $\omega \in W$ ; et  $N_s$  stations de pompage,  $s \in S$ .

A chaque temps  $t \in \{1, 2, \dots, T\}$ , on définit les variables :

- $I_\omega^t$  et  $I_s^t$  qui décrivent l'état de la pompe,

$$I_{\omega,s}^t = \begin{cases} 1, & \text{si la pompe } \omega \in W \text{ (} s \in S \text{) fonctionne à l'instant } t \\ 0, & \text{sinon} \end{cases}$$

- On décide d'extraire une quantité d'eau  $e_i^t$  du  $i^{\text{ème}}$  point source, avec  $0 \leq e_i^t \leq \epsilon_i^t$ , où  $\epsilon_i^t > 0$  est la capacité maximale qu'on peut extraire à partir du  $i^{\text{ème}}$  point source à l'instant  $t$ .

- La quantité stockée dans le  $j^{\text{ème}}$  point intermédiaire est  $\omega_j^t$ , avec  $0 \leq \omega_j^t \leq \omega_j$ , où  $\omega_j$  est la capacité maximale stockée dans le  $j^{\text{ème}}$  point intermédiaire.

- On décide de transférer une quantité d'eau  $x_{ij}^t$  du  $i^{\text{ème}}$  point source vers le  $j^{\text{ème}}$  point intermédiaire, avec  $0 \leq x_{ij}^t \leq \phi_{ij}$ , où  $\phi_{ij}$  est la capacité maximale d'eau qu'on peut transférer entre les deux points.

- $Q_\omega$ ,  $\omega \in W$  est la quantité d'eau qu'on peut pomper a partir de la pompe  $N_\omega$ , elle ne varie pas avec le temps, mais la condition suivante est vérifiée :

$$\sum_{i \in P} x_{ij}^t = \sum_{\omega \in W} I_\omega^t \cdot Q_\omega, \forall j \in D, \forall t$$

- On décide de transférer une quantité d'eau  $y_{jj'}^t$  du point intermédiaire  $j$  vers  $j'$ , avec  $0 \leq y_{jj'}^t \leq \xi_{jj'}$ , où  $\xi_{jj'}$  est la quantité d'eau maximale qu'on peut transférer entre  $j$  et  $j'$ .

- La quantité d'eau au niveau de la station de pompage  $N_s$  est  $Q_s$ ,  $s \in S$ , elle ne dépend pas du temps, et elle doit vérifier la condition suivante :

$$\sum_{j \in D} y_{jj'}^t = \sum_{s \in S} I_s^t \cdot Q_s, \forall j' \in D, \forall t$$

- $z_{jk}^t$  est la quantité d'eau transférée du point intermédiaire  $j$  vers le point de consommation  $k$ , avec  $0 \leq z_{jk}^t \leq \zeta_{jk}$ , où  $\zeta_{jk}$  c'est la quantité maximale qu'on peut transférer de  $j$  vers  $k$ .

- $d_k^t$  est la quantité d'eau demandée par le point de consommation  $k$  qui est supposée incertaine dans notre cas.

Pour la généralité de la notation, on va décrire un réseau de distribution entièrement relié avec des liens entre tous les points intermédiaires; des liens entre chaque point de source et chaque point intermédiaire et, en conclusion, des liens entre chaque point intermédiaire et chaque point de consommation. Des liens qui ne sont pas disponibles peuvent être décrits par une limite supérieure à zéro.

Comme on va prendre en considération Les contraintes suivantes :

- La quantité d'eau pompée de chaque point source ne doit pas dépasser la quantité maximale offerte :

$$e_i^t = \sum_{j \in D} x_{ij}^t \leq \epsilon_i, \forall i \in P, \forall t$$

- L'équation de masse de continuité pour les points intermédiaires, décrivant que la variation de l'eau stockée au  $j^{\text{ème}}$  point intermédiaire entre la période  $t - 1$  et  $t$  est égale à

l'équilibre entre les apports venant des points de source et les dégagements vers les points de consommation :

$$\omega_j^t = \omega_j^{t-1} + \sum_{i \in P} x_{ij}^t + \sum_{j' \in D} y_{j'j}^t - \sum_{j \in D} y_{jj'}^t - \sum_{k \in v} z_{jk}^t, \forall j \in D, \forall t$$

• La demande est satisfaite à chaque point de consommation. Notre but est de réaliser une demande cible. Pour ce faire on suppose la contrainte suivante :

$$d_k^t = \sum_{j \in D} z_{jk}^t - s_k^t + \delta_k^t$$

où  $s_k^t$  et  $\delta_k^t$  représentent respectivement le surplus et le déficit d'eau au point de consommation  $k$  à l'instant  $t$ , Les deux valeurs ne peuvent pas être positives simultanément, et pour assurer cela on introduit la contrainte suivante :  $s_k^t \cdot \delta_k^t = 0$ , bien que cette contrainte puisse être enlevée, selon la fonction objectif présentée.

### 3.1.5 La fonction objectif

Nous décrivons maintenant de diverses fonctions objectif se référant à des capitaux propres dans l'alimentation en eau, au début on va considérer la fonction objectif proposée par la compagnie.

#### Solution égalitaire

La fonction objectif initiale proposée par la compagnie ADE a visé à infliger le même déficit de l'eau à tous les habitants, réduisant au minimum un tel déficit. Le déficit au  $k^{\text{ème}}$  point de consommation au cours de toute la période de planification est  $\Delta_k = \sum_{t=1}^T \delta_k^t$ . Si  $N_k$  est le nombre de population au  $k^{\text{ème}}$  point de consommation, le déficit par habitant est  $L_k = \Delta_k / N_k$ . Selon les exigences de la compagnie le système à résoudre donc :

$$\begin{cases} \min L_k, \\ (x, y, z) \in \mathcal{C}, \\ L_k = L_{k'}, \forall k, k' \in v \end{cases}$$

### Solution de Smorodinsky-Kalai

En considérant la fonction objectif proposée par Smorodinsky et Kalai (Kalai (1977), Alexander (1992)), le problème aura la forme d'un problème min – max qui est reformulé sous la forme d'un problème à minimiser, en effet le problème est :

$$\begin{cases} \min \max L_k \\ (x, y, z) \in \mathcal{C} \end{cases}$$

qui est transformé en

$$\begin{cases} \min \mu \\ (x, y, z) \in \mathcal{C} \\ L_k \leq \mu, \forall k \in v \end{cases}$$

### Solution de Nash

Un autre concept important est la solution de Nash, voir Nydegger et Owen (1974), Alexandre et Ledermann (1994) ou Mariotti (1999), ce qui dans notre contexte serait formulé de deux manières équivalentes, la seconde étant plus efficace d'un point de vue informatique

$$\begin{cases} \max \prod_{k \in v} (-\alpha_k L_k + \beta_k) \\ (x, y, z) \in \mathcal{C} \end{cases}$$

ou, d'une manière équivalente

$$\begin{cases} \max \sum_{k \in v} \log(-\alpha_k L_k + \beta_k) \\ (x, y, z) \in \mathcal{C} \end{cases}$$

où  $\alpha_k$  et  $\beta_k$  sont des constantes.

### Solution utilitaire

Le quatrième concept important d'arbitrage est l'utilitaire, voir Ponsati et Watson (1997) ce qui vise à maximiser la somme des utilités atteintes (dans ce cas, réduisant au minimum la somme des déficits)

$$\begin{cases} \min L_k \\ (x, y, z) \in \mathcal{C} \end{cases}$$

D'une manière équivalente, nous essayons de maximiser l'utilité moyenne,

$$\begin{cases} \min \frac{1}{N_v} \sum_{k \in v} L_k \\ (x, y, z) \in \mathcal{C} \end{cases}$$

### 3.1.6 Distribution équitable de l'eau avec un coût efficace

les divers modèles décrits dans la section précédente menant à une distribution d'eau d'une façon équitable c'est-à-dire ils réduisent au minimum les déficits. Une fois le modèle est choisi, ce qui est intéressant c'est de trouver la méthode de distribution, tenant compte des coûts de pompage varions selon un programme quotidien, cette question est importante, car le pompage coûte la quantité la plus grande des coûts de distribution. Pour faire ainsi, il faut diviser la période  $t$  en  $m$  périodes. En supposant qu'à la période  $l$ , le coût de distribution est  $c_l$ , le système à une période donnée  $t$  est alors  $(x_{ij}^t, y_{jj'}^t, z_{jk}^t)$  ce qui peut être exprimé par

$$(x_{ij}^{tl}, y_{jj'}^{tl}, z_{jk}^{tl})_{l=1}^m$$

Le coût de pompage à minimiser à chaque période  $t$  est

$$\Phi(x, y, z) = \sum_{i \in P} \sum_{j \in D} \sum_{l=1}^m x_{ij}^{tl} c_l + \sum_{j, j' \in D} \sum_{l=1}^m y_{jj'}^{tl} c_l + \sum_{j \in D} \sum_{k \in V} \sum_{l=1}^m z_{jk}^{tl} c_l$$

avec les contraintes suivantes :

$$\sum_{l=1}^m x_{ij}^{tl} = x_{ij}^t$$

$$\sum_{l=1}^m y_{jj'}^{tl} = y_{jj'}^t$$

$$\sum_{l=1}^m z_{jk}^{tl} = z_{jk}^t$$

mais aussi avec les contraintes suivantes :

$$\sum_{j \in D} x_{ij}^t \leq \epsilon_i$$

$$x_{ij}^t \leq \phi_{ij}$$

$$y_{jj'}^t \leq \varepsilon_{jj'}$$

$$z_{jk}^t \leq \xi_{jk}$$

$$\sum_{i \in P} x_{ij}^t + \sum_{j' \in D} y_{j'j}^t - \sum_{j \in D} y_{jj'}^t - \sum_{k \in V} z_{jk}^t = \omega_j^t - \omega_j^{t-1}$$

$\forall i \in P, \forall j, j' \in D, \text{ et } \forall k \in V.$

## 3.2 Problématique du management de l'eau

Au commencement, le problème peut être décrit comme un problème réel d'alimentation en eau, avec de l'eau venant de plusieurs puits, avec plusieurs points intermédiaires, des stations de pompage et des points de consommation ayant lieu dans divers villages. Particulièrement pendant l'été, quand il y a un manque d'eau et la population a presque doublé, il y a une pénurie considérable en eau, qui est aggravée par des pertes significatives dans le réseau. Un deuxième dispositif d'intérêt pour ce problème provient des traditions culturelles dans la région. Traditionnellement, les Kabyles préfèrent la vie en haut de montagne, et ceci crée des problèmes importants, avec une proportion très élevée de coûts de distribution dus à la consommation d'électricité pendant le pompage.

L'intérêt de la recherche opérationnelle pour une bonne gestion du réseau de distribution de l'eau a été prouvé dans le précédent travail fait par A.L. Udias, D.Rios Insua, J.Cano, H.Fellag (2011), qui ont effectué une analyse détaillée des données et de la conception du réseau courant, où ils ont conclu que le problème n'était pas uniquement un problème de disponibilité de l'eau mais aussi produit par des pertes d'eau, incluant les piratages, et des programmes inadéquats. Le problème posé est modélisé par un modèle d'optimisation bi-critère en deux phases ; le premier objectif (maximiser les capitaux propres) semble beaucoup plus important pour la gestion courante que le second (réduire au minimum le coût). Dans la première phase, l'objectif est de déterminer un programme équitable d'alimentation en eau pour la région, tenant compte de diverses contraintes de distribution.

Ceci suggère la possibilité de rapprocher une frontière de Pareto avec deux critères, l'un en se référant au capital propre, l'autre par coût. Cela se fait en ajustant le niveau de la mesure du capital propre, puis l'exécution du programme le moins coûteux. Par exemple, dans le cas où ils ont considéré la solution de Smorodinsky-Kalai, ils ont résolu pour plusieurs valeurs de  $\rho k$  le problème  $P(\rho k)$  :

$$\begin{cases} \min \mu \\ (x, y, z) \in \mathcal{C} \\ L_k \leq \mu, \quad \forall k \in V \\ \mu \geq \rho k \end{cases}$$

ce qui a donné un programme optimal  $(x^{\rho k}, y^{\rho k}, z^{\rho k})$  avec la mesure du capital  $\mu_{\rho k}$ , ce qui mènerait alors à un coût optimal  $\Phi(x^{\rho k}, y^{\rho k}, z^{\rho k}) = \Phi(\rho k)$ . L'utilisation d'une telle frontière de Pareto est un outil de gestion puissant pour le distributeur d'eau. Donc, pour un coût de distribution donné, la compagnie peut trouver le programme le plus équitable, et vice-versa.

La période de simulation est de 48h en considérant deux jours consécutifs avec une variation de la demande (avec une demande moyenne de 150 litres par jour par personne). Ceci implique, que pour chaque variable de décision, il y a 48 variables additionnelles, correspondantes au volume de l'eau transporté à chaque période.

La courbe ci-dessous représente la quantité moyenne d'eau demandée chaque jour (en supposant que c'est la même pour les différents villages)

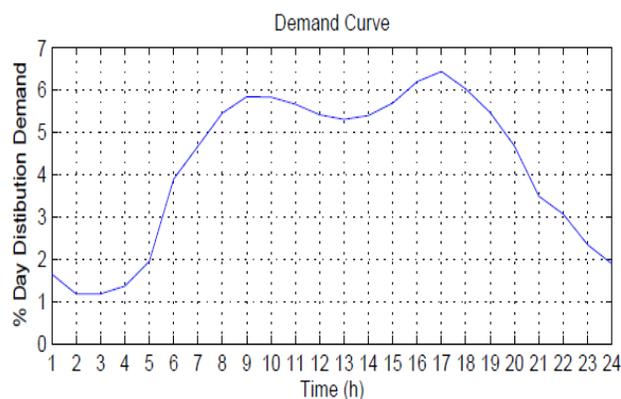


FIGURE 3.5 – La quantité moyenne de l'eau demandée

Dans cette étude l'incertitude sur la demande n'est pas prise en considération, car la quantité demandée est supposée fixe. Ce qui n'est pas le cas dans la réalité, c'est pour ça que cette incertitude doit être traitée avec des méthodes statistiques, voir par exemple

Birge et Louveaux (1997). Par conséquent, nous devons l'estimer car elle est inconnue et non maîtrisée.

### 3.3 Analyse exploratrice

L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps, c'est le cas de nos séries qui représentent le nombre de compteurs d'eau au niveau des différents villages. Notre région d'étude se compose de quatre villages : Beni-Douala, Beni-Zmenzer, Maatkas, et Bouhinoun, Betrouna, Ihasnouene, ces trois derniers sont considérés comme un seul village.

La première étape consiste à tracer les données, ce qui est fait sur la figure ci-après.

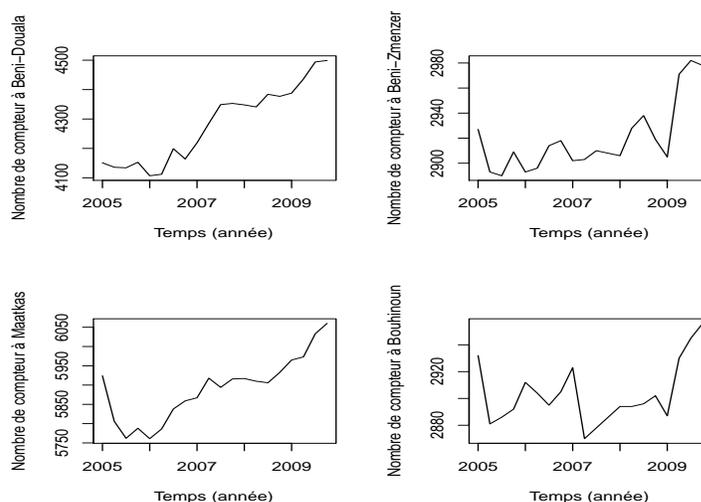


FIGURE 3.6 – séries représentant le nombre de compteurs dans les différents villages

L'analyse des séries chronologiques considère la décomposition en trois types de variations : la tendance qui représente l'évolution à long terme de la série, elle se représente comme une fonction du temps, la composante cyclique qui indique une relation trigonométrique du temps, et les variations aléatoires qui reflètent l'ensemble des perturbations apportées par des éléments nouveaux au temps  $t$ .

Les graphes suivants illustrent cette décomposition pour les quatre séries :

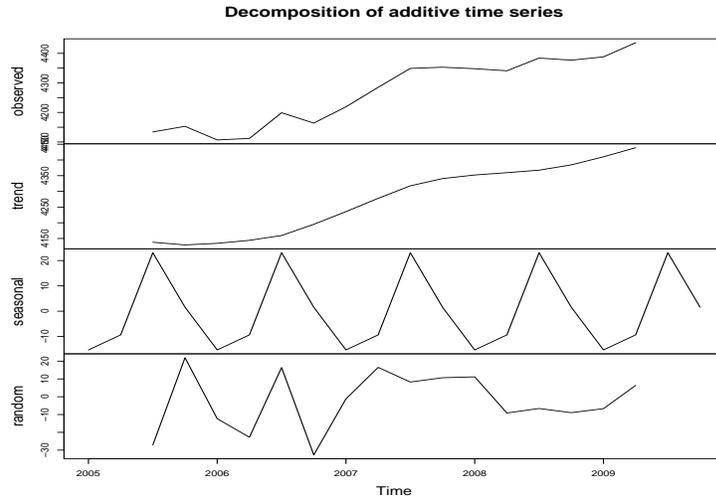


FIGURE 3.7 – Décomposition de la série de Beni-Douala

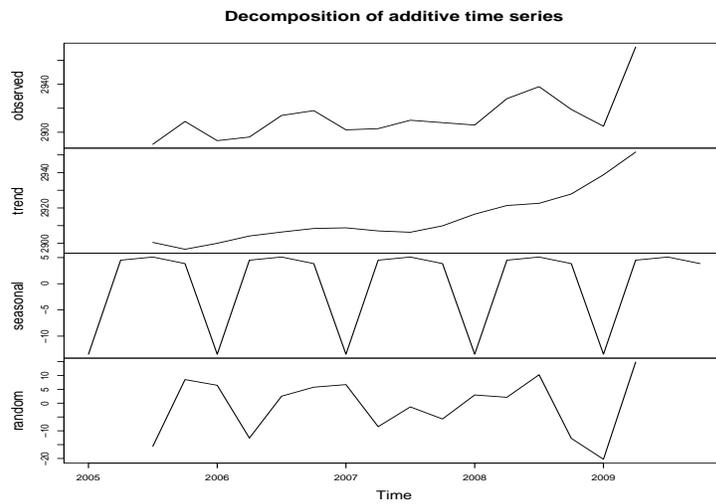


FIGURE 3.8 – Décomposition de la série de Beni-Zmenzer

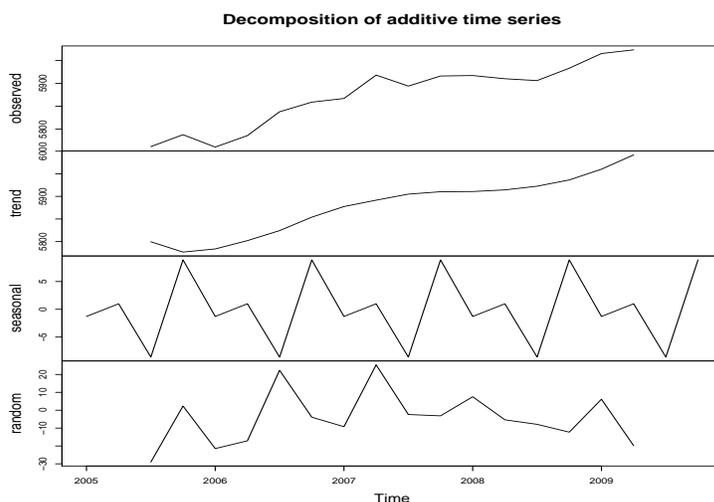


FIGURE 3.9 – Décomposition de la série de Maatkas

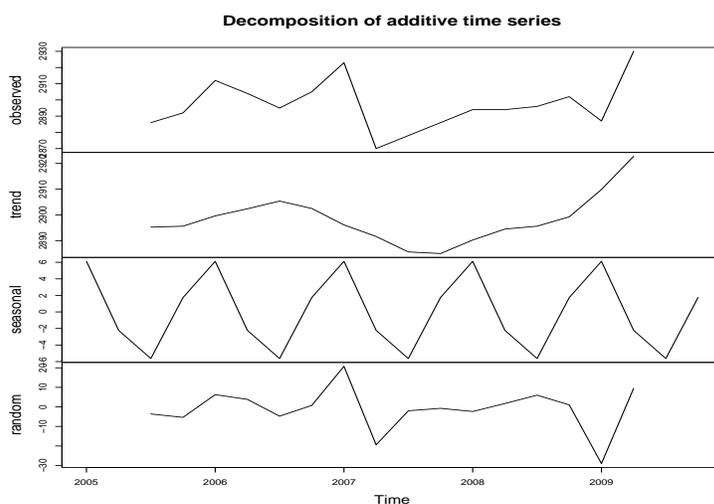


FIGURE 3.10 – Décomposition de la série de Bouhinoun (Betrouna, Ihasnouene)

On constate que le nombre de compteurs à Beni-Douala augmente de façon régulière, ce qui est bien visible à partir de l'année 2006 ce qui n'est pas le cas pour les deux autres villages Beni-Zmenzer et Maatkas, où on constate une forte chute le premier trimestre de l'année 2005, suivi d'une augmentation au fil des années. Par contre le nombre de compteurs à Bouhinoun (Bouhinoun, Betrouna et Ihasnouene) augmente jusqu'à 2007, et à partir de là on remarque une chute bien visible suivi d'une augmentation à partir de 2008.

Comme on peut bien constater la présence d'une composante saisonnière au niveau de tous les villages vu qu'on a des données trimestrielles .

### 3.4 Modèle pour le nombre de compteurs

Le cadre puissant offert par les modèles dynamiques linéaires se révèle être un atout gagnant. Les modèles *ARMA* peuvent être utiles quand ils sont considérés comme des modèles dynamiques linéaires. Mais les modèles dynamiques linéaires offrent beaucoup plus de flexibilité dans le traitement de séries temporelles non stationnaires ou la modélisation des changements structurels, et sont souvent plus facilement interprétables, c'est pour ça que la série de nombre de compteurs est modélisée comme combinaison de deux modèles dynamiques linéaires, avec  $Y_t^1$  c'est la composante tendancielle et  $Y_t^2$  la composante saisonnière.

$$Y_t^1 = ( 1 \ 0 ) ( \theta_t^1 \ \theta_t^2 ) + v_t^1$$

$$(\theta_t^1, \theta_t^2) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_{t-1}^1 \\ \theta_{t-1}^2 \end{pmatrix} + \begin{pmatrix} \omega_t^1 \\ \omega_t^2 \end{pmatrix}$$

et

$$Y_t^2 = ( 1 \ 0 \ 0 \ 0 ) ( \theta_t^3 \ \theta_t^4 \ \theta_t^5 \ \theta_t^6 ) + v_t^2$$

$$(\theta_t^3 \ \theta_t^4 \ \theta_t^5 \ \theta_t^6) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \theta_{t-1}^3 \\ \theta_{t-1}^4 \\ \theta_{t-1}^5 \\ \theta_{t-1}^- \end{pmatrix} + \begin{pmatrix} \omega_t^3 \\ \omega_t^4 \\ \omega_t^5 \\ \omega_t^6 \end{pmatrix}$$

avec  $Y_t = Y_t^1 + Y_t^2$ , on obtient le modèle suivant :

$$Y_t = ( 1 \ 0 \ 1 \ 0 \ 0 \ 0 ) \theta_t + v_t \tag{3.1}$$

$$\theta_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \theta_{t-1}^1 \\ \theta_{t-1}^2 \\ \theta_{t-1}^3 \\ \theta_{t-1}^4 \\ \theta_{t-1}^5 \\ \theta_{t-1}^6 \end{pmatrix} + \begin{pmatrix} \omega_t^1 \\ \omega_t^2 \\ \omega_t^3 \\ \omega_t^4 \\ \omega_t^5 \\ \omega_t^6 \end{pmatrix}$$

Dans notre modèle les matrices du système  $F_t$  et  $G_t$  sont connues, mais les matrices de covariance  $V_t$  et  $W_t$  sont inconnues, alors, le problème de base c'est d'estimer  $V_t$  et  $W_t$ , dans un premier temps on suppose que  $V_t = I$  et constant et  $W_t$  sera spécifier par le facteur d'escompte. Puis on représente le modèle comme combinaison d'un modèle DLM et d'une composante *ARMA*.

### 3.4.1 Spécification de $W_t$ par le facteur d'escompte

Le problème de base c'est d'estimer  $V_t$  et  $W_t$ , Ici on va considérer les matrices de covariances sont connues jusqu'à un facteur en commun ; c'est-à-dire,  $V_t = \sigma^2 \tilde{V}_t$  et  $W_t = \sigma^2 \tilde{W}_t$ , avec  $\sigma^2$  est inconnue. Dans l'inférence Bayésienne il est intéressant de travailler avec l'inverse  $\phi = 1/\sigma^2$ . Toute l'incertitude est, donc, dans les vecteurs d'état et dans le paramètre  $\phi$ . Pour tous  $t \geq 1$  on suppose que :  $\tilde{V}_t = I_m$  et  $\tilde{W}_t$  sera spécifier par le facteur d'escompte. Dans les récursions du filtre de Kalman, l'incertitude sur  $\theta_{t-1}$  sachant  $y_{1:t-1}$  est résumée dans la matrice de covariance  $Var(\theta_{t-1}|y_{1:t-1}) = C_{t-1}$ . En passant de  $\theta_{t-1}$  a  $\theta_t$  par l'équation d'état  $\theta_t = G_t \theta_{t-1} + W_t$ , l'incertitude augmente, et on a  $Var(\theta_t|y_{1:t-1}) = R_t = G'_t C_{t-1} G_t + W_t$ , ce qui est équivalent à  $R_t = P_t + W_t$ . Dans ce sens,  $W_t$  exprime la perte d'information en passant de  $\theta_{t-1}$  a  $\theta_t$  dû a la composante d'erreur stochastique dans l'évolution d'état, on peut exprimer  $W_t$  comme proportion de  $P_t$  :

$$W_t = \frac{1 - \delta}{\delta} P_t$$

avec le facteur d'escompte  $\delta \in (0, 1]$ , il suit que  $R_t = 1/\delta P_t$  avec  $1/\delta > 1$ .

Pour tous  $t \geq 1$ , on a

$$\begin{aligned} Y_t | \theta_t, \phi &\sim \mathcal{N}(F_t \theta_t, \phi^{-1} \tilde{V}_t) \\ \theta_t | \theta_{t-1}, \phi &\sim \mathcal{N}(G_t \theta_{t-1}, \phi^{-1} \tilde{W}_t) \end{aligned}$$

comme loi a priori pour  $(\phi, \theta_0)$  il est convenant de choisir la loi conjuguée gamma normale :

$$\phi \sim \mathcal{G}(\alpha_0, \beta_0) \quad \theta_0 | \phi \sim \mathcal{N}(m_0, \phi^{-1} \tilde{C}_0)$$

alors  $(\theta_0, \phi) \sim \mathcal{NG}(m_0, \tilde{C}_0, \alpha_0, \beta_0)$ .

Les paramètres  $a_t$ ,  $f_t$ ,  $m_t$ ,  $\tilde{C}_t$  et  $\tilde{R}_t$  sont calculés avec le filtre de Kalman avec les matrices de covariance  $\tilde{C}_0$  et  $V$  connues. En fait, l'évolution de  $\tilde{W}_t$  est connue à l'instant

$t = 1$ , puis pour  $t > 1$  le  $\tilde{W}_t$  sera calculer récursivement avec la formule (2.43), c'est-à-dire pour chaque  $t = 1, 2, \dots$ , on calcule  $\tilde{W}_t$  à partir des résultats obtenus à l'instant  $t - 1$ , et puis appliquer les récursions du filtre de Kalman avec les paramètres donnés par les formules (2.37) et (2.39) avec :

$$\alpha_t = \alpha_0 + \frac{t}{2}$$

$$\beta_t = \beta_0 + \frac{1}{2} \sum_{i=1}^t (y_i - f_i)^2 \tilde{Q}_i^{-1} = \beta_0 + \frac{1}{2} \sum_{i=1}^t \tilde{e}_i^2$$

avec :

$$Q_t = Var(Y_t | y_{1:t-1}) = \tilde{Q}_t \frac{\beta_{t-1}}{\alpha_{t-1} - 1}$$

$$C_t = Var(\theta_t | y_{1:t}) = \tilde{C}_t \frac{\beta_t}{\alpha_t - 1}$$

Ces étapes peuvent être facilement mises en application sous le  $R$ , avec  $\delta = 0.95$  d'où on obtient les graphes suivants qui représentent les prévisions en une étape en avant pour le nombre de compteurs ainsi que les bandes d'incertitude calculées avec une probabilité de 0.95 ce qui est équivalent à un intervalle de crédibilité prédictifs à 95%.

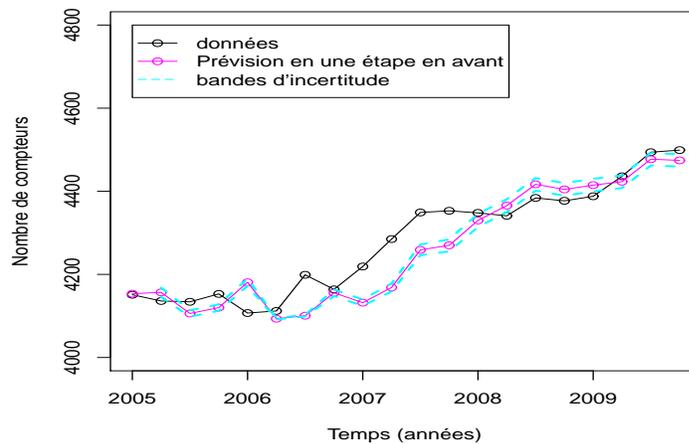


FIGURE 3.11 – Prévision en une étape en avant à Beni-Douala

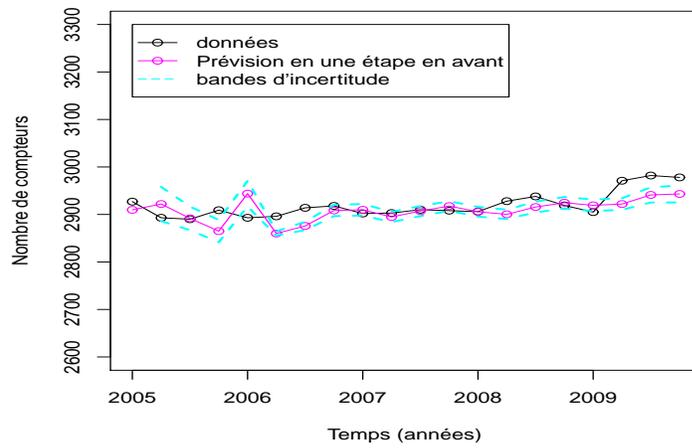


FIGURE 3.12 – Prévion en une étape en avant à Beni-Zmenzer

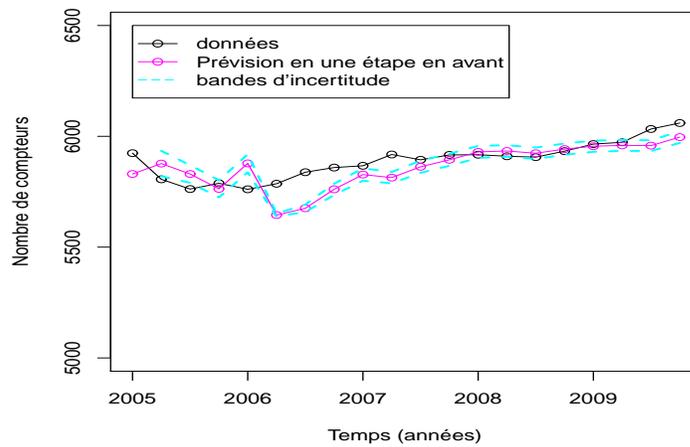


FIGURE 3.13 – Prévion en une étape en avant à Maatkas

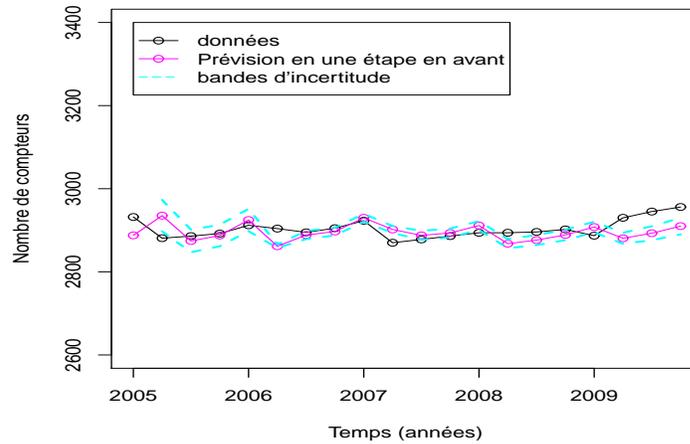


FIGURE 3.14 – Prédiction en une étape en avant à Bouhinoune

La prédiction en une étape en avant pour le nombre de compteurs est bien visible dans les différents villages, bien que ce ne soit pas vraiment le cas pour le premier bloc qui représente la tendance obtenue avec la fonction de prédiction  $f_t(k) = E(Y_{t+k}|y_{1:t})$ , ou on constate que la prédiction ne suit pas vraiment le comportement des données en particulier au niveau des villages Maatkas et Bouhinoune ce qui s'explique par les récursions du filtre de Kalman, par contre la précision de la prédiction augmente dans le bloc qui représente la saisonnalité avec les bandes d'incertitude calculées avec une probabilité de 0.95.

### 3.4.2 Représentation du modèle sous forme d'un DLM + une composante *ARMA*

Une deuxième approche intéressante pour la modélisation de nos séries c'est de les représenter sous forme d'un modèle dynamique linéaire plus une composante *ARMA* pour notre cas sa sera un *AR(1)*. Les résultats de cette étude sont donnés sous forme de graphes

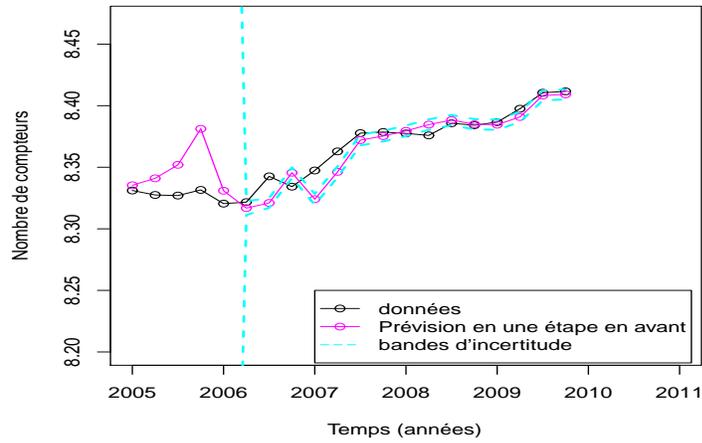


FIGURE 3.15 – Préviation en une étape en avant à Beni-Douala

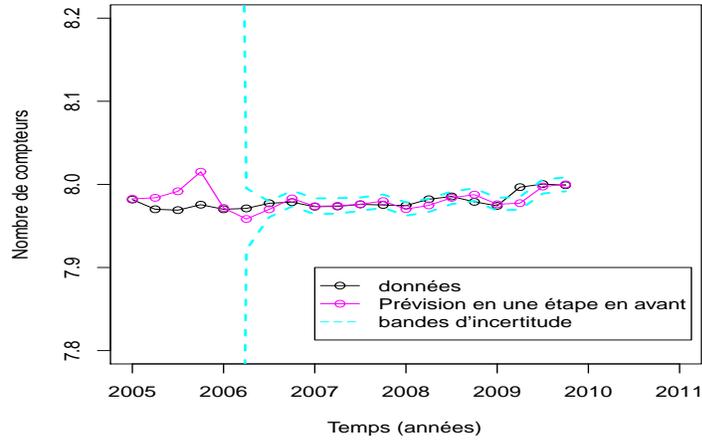


FIGURE 3.16 – Préviation en une étape en avant à Beni-Zmenzer

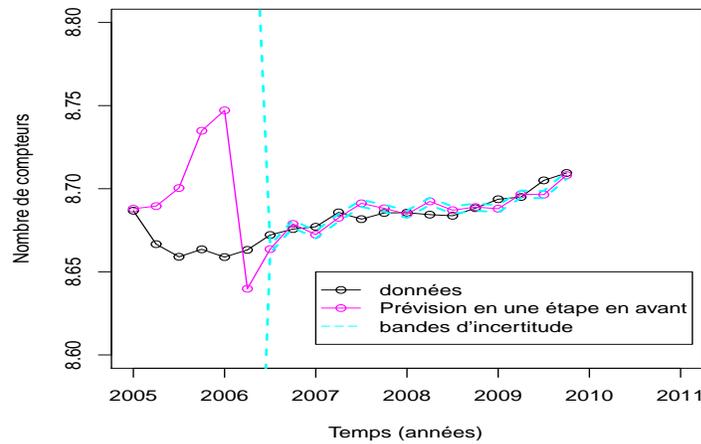


FIGURE 3.17 – Prévision en une étape en avant à Maatkas

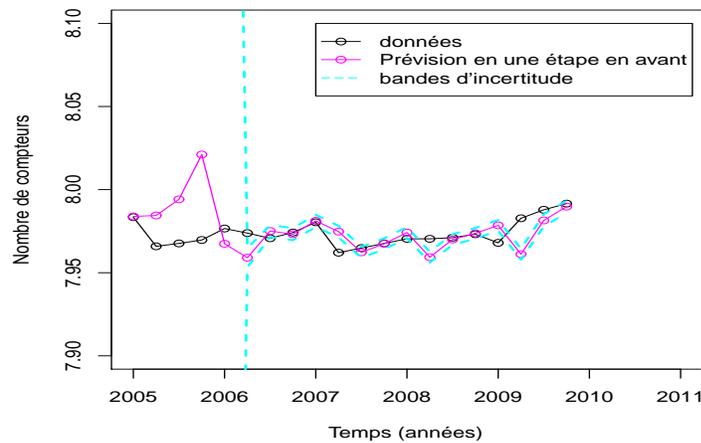


FIGURE 3.18 – Prévision en une étape en avant à Bouhinoune

C'est la même chose que dans la première approche la prévision pour la composante tendancielle n'est pas bien visible ce qui s'explique par les récursions de l'algorithme et spécialement le filtre de Kalman, par contre on a obtenu une bonne prévision pour la composante saisonnière.

Comme on peut bien le constater les prévisions sont meilleurs avec la première approche où la matrice de covariance  $W_t$  est spécifiée par le facteur d'escompte.

### 3.5 Modèle pour la demande d'eau

Dans cette partie on s'intéresse à la quantité d'eau demandée au niveau de chaque village. En supposant que c'est la même, comme l'a affirmé l'algérienne des eaux, elle est de 150l par jour et par personne.

En utilisant les prévisions des compteurs d'eau obtenus dans la section précédente c'est facile de représenter cette quantité sous forme de graphe. Chaque compteur représente une famille et en moyenne cette dernière est composée de quatre personnes, donc la quantité d'eau demandée ( $d$ ) sera calculée de la façon suivante :

$$d = 150 * \text{le nombre de compteurs} * \text{le nombre de personnes} * \text{le nombre de jours.}$$

Les résultats sont représentés sous forme de graphe :

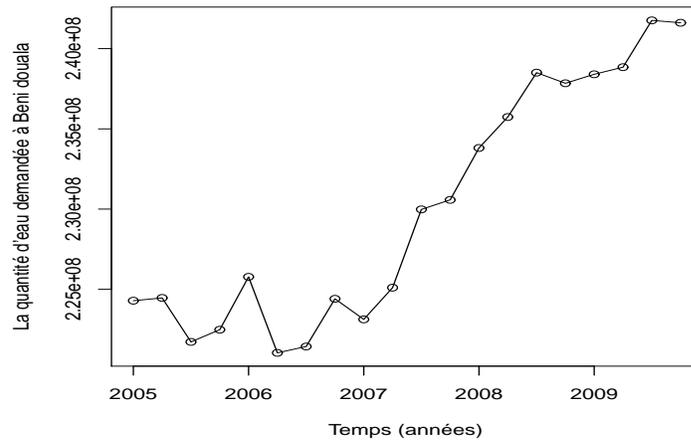


FIGURE 3.19 – La quantité d'eau demandée à Beni Douala

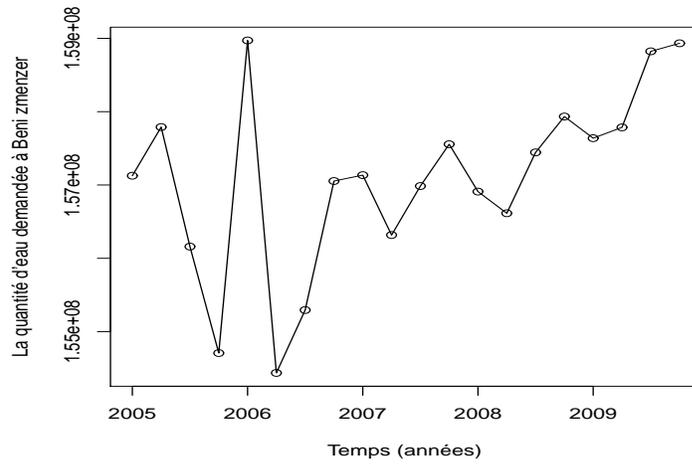


FIGURE 3.20 – La quantité d'eau demandée à Beni Zmenzer

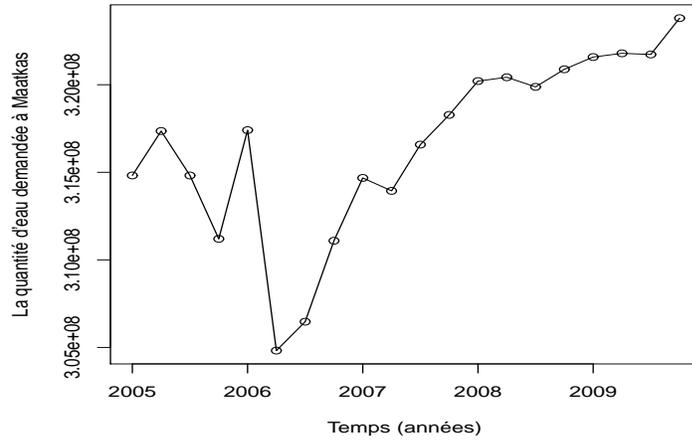


FIGURE 3.21 – La quantité d'eau demandée à Maatkas

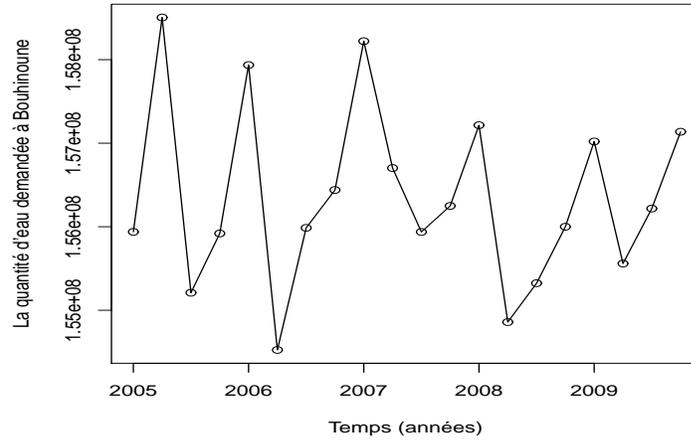


FIGURE 3.22 – La quantité d'eau demandée à Bouhinoune

Comme on le constate la quantité d'eau demandée est régulière au niveau du premier village Beni Douala. On remarque que cette quantité augmente au fur et à mesure, ce qui n'est pas le cas pour les trois autres villages surtout le dernier où la quantité est loin d'être stable.

# Conclusion générale

Ce travail présente une méthode originale de traitement des séries temporelles : les modèles dynamiques linéaires. Les modèles ayant été appliqués à des séries temporelles représentant le nombre de compteurs d'eau dans les différents villages de la Kabylie, les résultats se sont avérés être une approche particulièrement intéressante du traitement des séries temporelles hydrauliques.

En perspectives, ça sera intéressant de faire une modélisation pour la quantité d'eau consommée dans les différents village sous forme d'un modèle dynamique linéaire avec une incertitude sur la demande, puis construire un modèle hiérarchique composé du modèle de nombre de compteurs et la quantité d'eau consommée dans les différents villages de la kabylie.

# Annexes

## A. Quelques lois de probabilités d'usage courant

### A.1. La loi binomiale

La loi de probabilité la plus simple qu'on puisse imaginer est la loi dite de Bernoulli. Elle est définie sur l'ensemble discret  $\{0, 1\}$  et elle affecte la valeur  $p \in [0, 1]$  au nombre 1 et  $1-p$  au nombre 0. Cette loi est normalement associée au modèle d'urne : si on considère une urne remplie de boules de deux couleurs (blanches et noires, par exemple) aléatoirement mélangées, telle que le rapport entre le nombre de boules blanches et le nombre total de boules vaille  $p$ , alors la variable aléatoire (v.a.) qui associe à un tirage dans l'urne, le chiffre 1 si la boule extraite est blanche et 0 si elle est noire, est une v.a. de Bernoulli de paramètre  $p$ .

On peut facilement vérifier que l'espérance et la variance de cette variable valent  $p$  et  $p(1-p)$  respectivement.

Imaginons maintenant de réaliser  $m$  tirages avec la précaution de remettre dans l'urne la boule extraite après chaque tirage, de manière que la proportion entre boules blanches et noires soit toujours constante, et de remélanger aléatoirement l'urne à chaque fois. Le nombre de fois qu'on a extrait une boule blanche (appelons le  $x$ ) est alors la somme de  $m$  v.a. de Bernoulli de même paramètre  $p$  et indépendantes. On exprime cette circonstance en disant que les  $m$  v.a. sont indépendantes et identiquement distribuées (i.i.d.).

La loi binomiale de paramètres  $p$  et  $m$  est celle suivie par la somme de  $m$  v.a. de Bernoulli i.i.d.

C'est une loi discrète et son support est l'ensemble  $0, 1, \dots, m$ . Elle s'écrit :

$$P[x|p, m] = \binom{m}{x} p^x (1-p)^{m-x}$$

$$\text{où } \binom{m}{x} = \frac{m!}{x!(m-x)!}$$

L'espérance et la variance s'obtiennent facilement à partir de l'espérance et de la variance de la loi de Bernoulli :

$$E(X) = mp$$

$$\text{Var}(X) = mp(1-p)$$

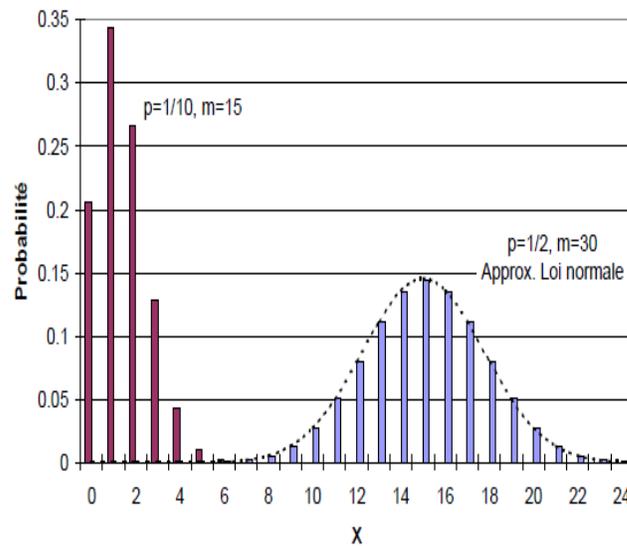


FIGURE 3.23 – Exemples de lois binomiales.

Pour des valeurs de  $m$  très grandes (en pratique supérieures ou égales à 30) la loi binomiale peut être approximée par une loi normale de même espérance et même variance. Cette approximation trouve sa justification théorique dans le théorème centrale limite.

## A.2. La loi normale

La loi normale, dite aussi gaussienne ou (parfois) de Laplace-Gauss est une loi continue définie sur l'ensemble des nombres réels. Son expression, paramétrée par  $\mu$  et  $\sigma^2$  qui s'écrit :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

donne lieu à la célèbre courbe en cloche de Gauss.

L'espérance et la variance d'une loi normale sont  $\mu$  et  $\sigma^2$  respectivement. En particulier pour  $\mu = 0$  et  $\sigma^2 = 1$  on parle de loi normale centrée-réduite ou standard.

La loi normale a un rôle de premier plan en statistique en vertu du théorème centrale limite qui affirme que si  $(X_1, X_2, \dots, X_n, \dots)$  est une suite de v.a réelles (non nécessairement normales) i.i.d. d'espérance  $\mu$  et variance  $\sigma^2$ , alors pour  $n \rightarrow +\infty$  la v.a. :

$$\frac{\sqrt{n}}{\sigma} \left( \sum_{i=1}^n \frac{X_i}{n} - \mu \right)$$

tend, en loi, vers la loi normale centrée-réduite  $\mathcal{N}(0, 1)$ .

En vertu du théorème centrale limite (Bernier et al. 2000) la loi normale peut être utilisée pour décrire un phénomène dont la variabilité résulte de la combinaison d'un grand nombre de causes dont les effets s'additionnent mais restent individuellement petits par rapport à leur somme, comme dans le cas des erreurs de mesure. La loi normale est ainsi utilisée pour décrire la variabilité naturelle. Dans un livre fameux de Science-Fiction (Jurassic Park de M. Chrichton), la preuve définitive que des animaux théoriquement incapables de se reproduire arrivent à procréer naturellement est que la taille des jeunes individus est distribuée selon la courbe en cloche de Gauss.

D'autres propriétés mathématiques de la loi normale (symétrie, additivité) et des échantillons gaussiens (théorème de Cochran), qui simplifient énormément les calculs dans le cadre des techniques statistiques le plus couramment utilisées (modèles linéaires, analyse de la variance), ont contribué à consolider le statut prédominant de la loi normale en

statistique.

### A.3. La loi gamma

La densité de probabilité d'une loi Gamma, définie sur l'intervalle  $]0, +\infty[$  est exprimée par la relation :

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

où  $\Gamma(\alpha)$  est la fonction Gamma ou intégrale d'Euler de deuxième espèce :

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} \exp(-t) dt$$

et on écrit  $X \sim \mathcal{G}(\alpha, \beta)$ .

La loi Gamma est paramétrée par les deux réels positifs  $\alpha$  et  $\beta$  et son espérance et variance valent respectivement :

$$E(X) = \frac{\alpha}{\beta}$$

$$Var(X) = \frac{\alpha}{\beta^2}$$

Cette loi est habituellement utilisée pour décrire des v.a. strictement positives et notamment en statistique Bayésienne comme loi a priori du paramètre d'une loi de Poisson ou de l'inverse de la variance d'une loi normale.

En fonction des valeurs des paramètres  $\alpha$  et  $\beta$  cette loi peut prendre des formes très différentes :

Si  $X \sim \mathcal{G}(\alpha, \beta)$ , la densité de  $Y = 1/X$  est une loi inverse gamma de paramètres  $(\alpha, \beta)$ , et on a

$$E(Y) = \beta/(\alpha - 1) \quad \text{si } \alpha > 1$$

$$Var(Y) = \beta^2/((\alpha - 1)^2(\alpha - 2)) \quad \text{si } \alpha > 2$$

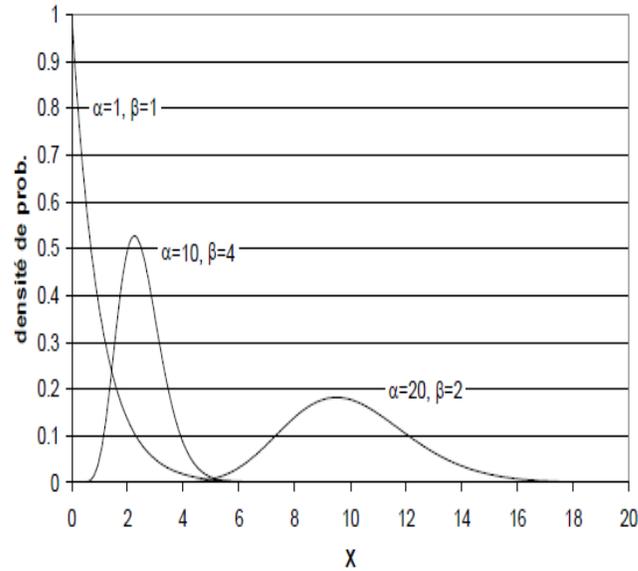


FIGURE 3.24 – Exemples de lois Gamma.

#### A.4. La loi Bêta

Une v.a.  $X$ , définie dans l'intervalle  $[0, 1]$ , distribuée selon une loi Bêta, a une densité de probabilité exprimée par la relation :

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

ou  $B(\alpha, \beta)$  est la fonction Bêta ou intégrale d'Euler de première espèce :

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

La moyenne et la variance d'une v.a. de type Bêta valent respectivement :

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

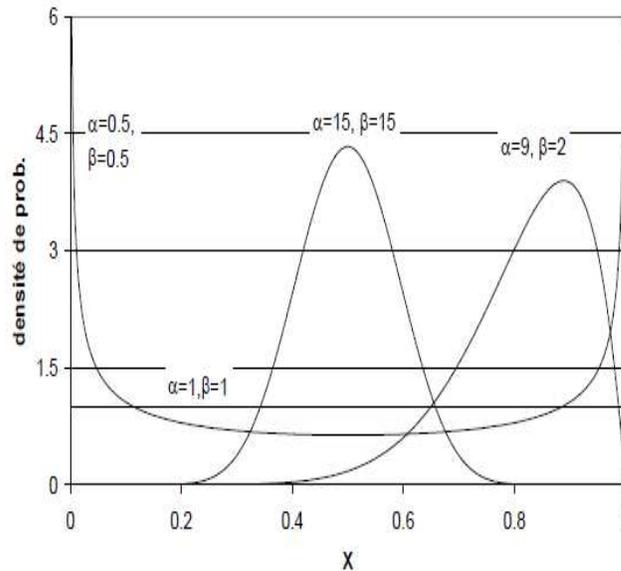


FIGURE 3.25 – Exemples de lois bêta.

Les lois de la famille Bêta sont utilisées pour décrire des variables bornées et notamment en statistique Bayésienne comme lois a priori conjuguées du paramètre  $p$  des lois binomiales.

### A.5. La loi multinomiale

La loi multinomiale s'obtient en généralisant le modèle d'urne décrit précédemment, dans le cas où le nombre de résultats aléatoires possibles est supérieur à 2 (par exemple on peut imaginer une urne avec des boules de 3 ou 4 couleurs différentes). Si on associe le nombre 1 au résultat 1, le nombre 2 au résultat 2 etc. la v.a. ainsi construite, qui généralise la v.a. de Bernoulli, est appelée *catégorielle*.

Imaginons de répéter  $m$  fois une expérience qui peut résulter en  $k$  valeurs possibles. Le phénomène est régi par le vecteur de  $\mathbb{R}^k$  :

$$\underline{P} = (P_1, P_2, \dots, P_k)$$

dont les composantes expriment les probabilités de réalisation de chacun des  $k$  événements. Les  $P_i$  vérifient les conditions :

$$0 \leq P_i \leq 1 \quad \sum_{i=1}^k P_i = 1$$

Le résultat des  $m$  répétitions peut être exprimé sous forme d'un vecteur  $\underline{x} = (x_1, x_2, \dots, x_k)$ , dont la composante  $i$ ème est le nombre de fois que l'expérience a donné le résultat  $i$ . Evidemment on a :

$$0 \leq x_i \leq m \quad \sum_{i=1}^k x_i = m$$

Si les expériences sont répétées dans les mêmes conditions, de manière que tous les tirages puissent être décrits par des v.a. catégorielles *iid* de paramètre  $\underline{P}$ , alors le vecteur  $\underline{x}$  est la réalisation d'une variable  $\underline{X}$  dite multinomiale de paramètres  $\underline{P}$  et  $m$  :

$$\underline{X} \sim M(\underline{P}, m)$$

La loi de probabilité d'une v.a. multinomiale est donnée par :

$$P[\underline{x}|\underline{P}, m] = \frac{m!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k P_i^{x_i}$$

### A.6. La loi gamma normale

Soit  $(X, Y)$  un vecteur aléatoire, si  $X|Y = y \sim \mathcal{N}(\mu, (n_0 y)^{-1})$ , et  $Y \sim \mathcal{G}(a, b)$  alors  $(X, Y)$  suit une loi gamma normale de paramètres  $(\mu, n_0^{-1}, a, b)$ , avec  $\mu \in \mathbb{R}$ ,  $a, b \in \mathbb{R}^+$ , on écrit  $(X, Y) \sim \mathcal{NG}(\mu, n_0^{-1}, a, b)$ . La loi marginale de  $X$  est  $t$ -student,  $X \sim \mathcal{T}(\mu, (n_0 \frac{a}{b})^{-1}, 2a)$ .

# Bibliographie

- [1 ] Abramowitz M. et Stegun I., *Handbook of Mathematical Functions*, Dover, New York, (1964).
- [2 ] Alexander C. *The Kalai-Smorodinsky Bargaining Solution in Wage Negotiations*, Journal of the Operational Research Society 43(8), (779-786), (1992).
- [3 ] Alexander C. and W. Ledermann, *The Constrained Nash Bargaining Solution*, Journal of the Operational Research Society 45(8), (954-958), (1994).
- [4 ] Bayes T., *Essay Towards Solving a Problem in the Doctrine of Chances*. Philosophical Transactions of the Royal Society of London n. 53, pp. (370-418) et n. 54, pp. (296-325), (1763). Re-imprime dans Biometrika, vol. 45, pp. (293-315), (1958).
- [5 ] Berger J.O., *Statistical Decision Theory and Bayesian Analysis*, 2ème édition Springer-Verlag, (1985).
- [6 ] Bernardo J.M., Smith A.F.M, *Bayesian Theory*, John Wiley & Sons, (1994).
- [7 ] Bernier J., Parent E., Boreux J.J., *Statistiques pour l'Environnement. Traitement Bayésien des Incertitudes*, Tec & Doc, (2000).
- [8 ] Berry D.A., *Statistics. A Bayesian perspective*, Duxbury Press, (1996).
- [9 ] Berry D.A., Thor A., Cirrincione C., Edgerton S., Muss H., Marks J., Liu E., Wood W., Budman D., Perloff M., Peters W., Henderson I.C., *Scientific Inference and Predictions. Multiplicities and Convincing Stories : A Case Study in Breast Cancer Therapy (avec discussion)*, Dans Bayesian Statistics 5, Ed. Bernardo J.M, Berger J.O., Dawid A.P., Smith A.F.M., pp. 45-67. Clarendon Press, (1996).
- [10 ] Birge J. and F. Louveaux, *Introduction to Stochastic Programming* New York : Springer Verlag, (1997)
- [11 ] Bouache M., Tabani I., *Gestion optimale et équitable de l'eau dans une région de Kabylie*, Memoire d'ingeniorat en recherche opérationnelle, U.M.M.T.O, (2010).

- 
- [12 ] Box G.E.P., Tiao G.C., *Bayesian inference in statistical analysis*, Addison-Wesley, (1973).
- [13 ] Brown L., *Notes on statistical decision theory*, Technical report, Ithaca, New York, (1976)
- [14 ] Caines P., *Linear Stochastic Systems*, Wiley, New York (1988).
- [15 ] Christian P. Robert, *Le Choix Bayésien - Principes et pratique*, Springer, (2006).
- [16 ] Cullen A.C., Frey H.C., *Probabilistic Techniques in Exposure Assessment*, Plenum Press, (1999).
- [17 ] De Finetti B., *La prévision : ses lois logiques, ses sources subjectives*, Annales de l'Institut Henri Poincaré, vol.7, pp 86-133, (1937).
- [18 ] De Finetti B., *Theory of probability, a critical introductory treatment*, vol. 1. John Wiley & Sons, (1974).
- [19 ] Durbin J. and Koopman S. *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, (2001).
- [20 ] Garthwaite P., O'Hagan A., *Quantifying expert opinion in the water industry : an experimental study*, The Statistician, vol. 49, pp. 455-477, (2000).
- [21 ] Gourieroux C. and Monfort A., *Time Series and Dynamic Models*, Cambridge University Press, Cambridge, (1997).
- [22 ] Hannan E. and Deistler M., *The Statistical Theory of Linear Systems*, Wiley, New York, (1988).
- [23 ] Hartigan J.A. *Bayes Theory*. Springer-Verlag, (1983).
- [24 ] Harvey A., *Forecasting, Structural Time Series Models and the Kalman filter*, Cambridge University Press, Cambridge, (1989).
- [25 ] Hills S. et Smith A., *Parametrization issues in Bayesian inference*. In Bernardo J., Berger J., Dawid A., et Smith A., éditeurs, Bayesian Statistics 4, pages 641–649. Oxford University Press, Oxford (1992).
- [26 ] Hutchinson C., *The Kalman filter applied to aerospace and electronic systems*, Aerospace and Electronic Systems, IEEE Transactions on Aerospace and Electronic Systems AES-20 : 500–504, (1984).
- [27 ] Kadane J.B., Wolfson L.J. *Experiences in elicitation*, The Statistician, vol. 47 pp. 1-20, (1998).

- [28] Kalai E., *Proportional Solutions to Bargaining Situations : Interpersonal Utility Comparisons*. *Econometrica : Journal of the Econometric Society* 45(7), (1623-1630), (1977).
- [29] Kalman R., *A new approach to linear filtering and prediction problems*, *Trans. of the AMSE - Journal of Basic Engineering (Series D)* 82 : 35–45, (1960).
- [30] Koopman B., *On distributions admitting a sufficient statistic*, *Trans. Amer. Math. Soc.*, 39, 399–409, (1936).
- [31] Krzysztofowicz R. *Why should a forecaster and a decision maker use Bayes Theorem*, *Water Resources Research*, vol. 19, n. 2, pp 327-336, (1983).
- [32] Kunsch, H., *State space and hidden Markov models*, in O. Barndorff-Nielsen, D. Cox and C. Klüppelberg (eds), *Complex stochastic systems*, Chapman & Hall/CRC, Boca Raton, pp. 109–173, (2001).
- [33] Larbi L., *Sur la décision statistique dans le contexte Bayésien*, *Memoire de master en statistique*, U.M.M.T.O, (2011).
- [34] Le Cam L., *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York, (1986).
- [35] Lecoutre B., Poitevineau J. *Traitement statistique des données expérimentales : des pratiques traditionnelles aux pratiques bayésiennes*, CISIA-CERESTA, (1996).
- [36] Lee P.M. *Bayesian Statistics, an Introduction*, 2ème édition, Arnold, (1997).
- [37] Lindley D.V. *Introduction to Probability and Statistics from a Bayesian Viewpoint*, University Press, (1965).
- [38] Lindley D., *Making Decisions*, John Wiley, New York, (1985).
- [39] Lipster R. and Shiriyayev A., *Statistics of conditionally Gaussian random sequences, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Univ. California Press, Berkeley, (1972).
- [40] Mariotti M., *Fair Bargains, Distributive Justice and Nash Bargaining Theory*, *Review of Economic Studies* 66(3), (733-741), (1999).
- [41] Metropolis N. et Ulam S., *The Monte Carlo method*. *J. American Statist. Assoc.*, 44, (335–341), (1949).
- [42] Migon H., Gamerman D., Lopez H. and Ferreira M., *Bayesian dynamic models*, in D. Day and C. Rao (eds), *Handbook of Statistics*, Vol. 25, Elsevier B.V., chapter 19, pp. 553–588, (2005).

- [43 ] Naylor J. et Smith A., *Application of a method for the efficient computation of posterior distributions*, Applied Statistics, 31, 214–225, (1982).
- [44 ] Nydegger R. and G. Owen, *Two-Person Bargaining : An Experimental Test of the Nash Axioms*, International Journal of Game Theory 3(4), (239-249), (1974).
- [45 ] O’Hagan A., *Eliciting expert beliefs in substantial practical applications*, The Statistician, vol. 47, pp. 21-35, (1998).
- [46 ] Parent E., Prevost E., *Inférence bayésienne de la taille d’une population de saumons par utilisation de sources multiples d’informations*, Revue de Statistique Appliquée, n. 3, pp. 5-38, (2003).
- [47 ] Perreault L., *Analyse bayésienne rétrospective d’une rupture dans les séquences de variables aléatoires hydrologiques*, Thèse de doctorat, ENGREF, (2000).
- [48 ] Petris G., Petrone S., Campagnoli P., *Dynamic Linear Models with R*, édition : Springer, (2009)
- [49 ] Pitman E., *Sufficient statistics and intrinsic accuracy*, Proc. Cambridge Philos. Soc., 32, 567–579, (1936).
- [50 ] Ponsati C. and J. Watson ., *Multiple-Issue Bargaining and Axiomatic Solutions*, International Journal of Game Theory 26(4), (501-524), (1997).
- [51 ] Press S.J., Tanur J.M., *The Subjectivity of Scientists and the Bayesian Approach* John Wiley & Sons, (2000).
- [52 ] Robert C.P., *L’Analyse Statistique Bayésienne*, Economica, (1992).
- [53 ] Rousseau J., *Statistique Bayésienne*, notes de cours, (2010).
- [54 ] Roustant O., *Introduction aux séries chronologiques*, Ecole Nationale Supérieure des Mines de Saint-Etienne, (2008).
- [55 ] Savage L. J., *The Foundations of Statistics*, New York : Dover Publications, (1954).
- [56 ] Savage L.J., *The Foundations of Statistics*, 2ème édition. Dover Publications, (1972).
- [57 ] Shumway R. and Stoffer D., *Time Series Analysis and its Applications*, Springer-Verlag, New York, (2000).
- [58 ] Sivia D.S., *Data Analysis, a Bayesian Tutorial*, Oxford University Press, (1996).
- [59 ] Soncini-Sessa R., A. Castelletti, and E. Weber . *Integrated and Participatory Water Resources Management : Theory*, Amsterdam : Elsevier, (2007).
- [60 ] Strasser H., *Mathematical Theory of Statistics*, W. de Gruyter, Berlin, (1985).

- 
- [61 ] Thiele T., *Om anvendelse af mindste kvadraters metode i nogle tilfælde, hvor en komplikation af visse slags uensartede tilfældige fejlkilder giver fejlene en "systematisk" karakter*, Det Kongelige Danske Videnskabernes Selskabs Skrifter – Naturvidenskabelig og Matematisk Afdeling pp :381–408 (1880). English Translation in : Thiele : Pioneer in Statistics, S. L. Lauritzen, Oxford University Press (2002).
- [62 ] Udias A., D. Rios Insua, J. Cano, and H. Fellag, *Reliability and optimization of the operational cost of water distribution in kabylia*, In The 4th International Symposium on Water Resources and Sustainable Development (CIRED4), (2011).
- [63 ] Von Mises R., Geiringer H., *The Mathematical Theory of Probability and Statistics*, Academic Press, (1964).
- [64 ] Von Neumann, J., *Various techniques used in connection with random digits*, J. Resources of the National Bureau of Standards–Applied Mathematics Series, 12, (36–38), (1951).
- [65 ] West M. and Harrison J., *Bayesian Forecasting and Dynamic Models*, 2nd edn, Springer, New York, (1997).