

Statistiques bayésiennes.

Master 1 Mathématiques et applications parcours ISN- Ingénierie statistique numérique

par

Henry Watiotienne et Othmane Ettaqi

Encadré par
Mme Ophélie Guin

mémoire soutenue devant le jury

Mme Ophélie Guin
Mr Nicolas Wicker

Remerciement

Nous tenons à exprimer notre gratitude envers Madame Ophélie Guin qui a su encadrer l'écriture de ce TER durant ces derniers mois.

Nous vous remercions de votre temps, de votre disponibilité, de votre soutien constant, et plus globalement de l'aide précieuse que vous nous avez apporté.

Résumé

Ce TER est une exploration approfondie des statistiques bayésiennes, avec un accent particulier sur les calculs de loi a posteriori. Cette exploration se fait en plusieurs étapes, avec une première focalisation sur les lois normales, puis s'étend au modèle de régression linéaire simple et enfin, au modèle de régression semi-paramétrique.

La statistique bayésienne est une sous-branche de la statistique qui a pour caractéristique principale de s'appuyer sur la loi de Bayes pour actualiser les connaissances a priori au regard des données. Il s'agit d'une méthode d'analyse statistique qui a démontré son efficacité dans de nombreux domaines d'application, notamment pour la modélisation de phénomènes complexes.

La première étape de notre recherche se concentre sur les lois normales. Cette partie se veut être une mise en lumière des fondements et des mécanismes essentiels des statistiques bayésiennes. Nous examinons ensuite le modèle de régression linéaire simple, qui est un outil statistique utilisé pour comprendre la relation entre deux variables continues.

La suite de notre exploration se tourne vers le modèle de régression semi-paramétrique, qui combine les caractéristiques des modèles paramétriques et non-paramétriques, permettant ainsi une flexibilité accrue dans la modélisation. Cette partie est plus technique, mais offre une perspective intéressante sur les possibilités offertes par l'approche bayésienne.

Finalement, nous appliquons notre modèle de régression semi-paramétrique à des données climatiques, dans le but d'analyser la tendance des températures dans différentes villes européennes. L'objectif ici est double : d'une part, démontrer l'application concrète de notre modèle en environnement réel, et d'autre part, contribuer à l'analyse du changement climatique, sujet d'une importance capitale à notre époque.

Ce TER est donc une exploration de la puissance et de la polyvalence des statistiques bayésiennes, tout en mettant en lumière son potentiel d'application dans des domaines aussi divers que le climat.

Tables des matières

1.Introduction générale sur la statistique bayésienne. P.5

- 1.1 L'approche bayésienne une approche intuitive.....P.5
- 1.2 Une opposition à la vision fréquentiste.....P.6

2. Prérequis théorique.P.7

- 2.1 Quelques définitions.....P.7
- 2.2 Choix de la loi a priori.....P.8
 - 2.2.1 Loi a priori impropre.....P.8
 - 2.2.2 Loi a priori non informative.....P.8
 - 2.2.3 Loi a priori conjugués.....P.9

3.Modèle gaussien à moyenne et variance inconnus P.11

- 3.1 Modèle bayésien et vraisemblanceP.11
- 3.2Lois a posteriori.....P.11
 - 3.2.1 Illustration avec un jeu de données simulées.....P.13

4.Modèle de régression linéaire simple P.15

- 4.1 Modèle bayésien et vraisemblanceP.15
- 4.2 Loi a priori non informatives.....P.16
- 4.3 Loi a priori informatives.....P.18
 - 4.3.1 Calculs de la loi a posteriori.....P.18
 - 4.3.2 Illustration avec un jeu de données simulées.....P.20

5.Modèle de régression semi-paramétrique P.24

- 5.1 Définition du modèle.....P.24
- 5.2 Modèle bayésien et vraisemblance.....P.25
- 5.3 Illustration avec un jeu de données simulées.....P.27

6. Application à des données réelles P.31

- 6.1 Présentation du jeu de données.....P.31
- 6.2 Analyse des sorties graphiques.....P.34

7.Conclusion et perspectives P.35

8. Annexe P.36

9. Références P.43

Chapitre 1 Introduction générale sur la statistique bayésienne

1.1 L'approche bayésienne une approche intuitive

Dans le domaine de la statistique, le praticien bayésien se confronte à l'idée selon laquelle le futur, toujours fluide et incertain, ne peut être prédit avec une exactitude absolue. En revanche, il peut s'appuyer sur les connaissances déjà acquises, qu'il met à jour en fonction des nouvelles données collectées. Cette démarche dynamique, constamment en mouvement, reflète l'essence même de l'approche bayésienne.

Pour illustrer cette approche, prenons l'exemple d'un automobiliste effectuant un trajet quotidien entre son domicile et l'Université Lille 1 à Villeneuve d'Ascq. Au fil du temps, et après avoir effectué ce trajet à de nombreuses reprises, l'automobiliste développe une intuition a priori : il s'attend à rencontrer quatre feux rouges parmi les dix feux de signalisation potentiels sur son parcours.

Durant une semaine, il décide de mettre à l'épreuve cette intuition en recueillant des données concrètes : il compte le nombre de feux rouges rencontrés lors de ses déplacements quotidiens. Au terme de cette semaine, le total s'élève à six feux rouges, soit deux de plus que son estimation a priori.

L'automobiliste fait donc face à une divergence entre ses connaissances a priori et les données empiriques recueillies. En bon statisticien bayésien, il effectue alors un compromis entre son intuition de départ et les nouvelles informations collectées.

La semaine suivante, l'automobiliste continue de compter les feux rouges rencontrés lors de ses trajets. À sa grande surprise, il en compte cette fois une dizaine. Cette observation le conduit à remettre en question son intuition a priori, qui semble bien éloignée de la réalité. C'est à ce moment qu'il prend une décision cruciale : donner davantage de poids aux données empiriques plutôt qu'à son intuition a priori.

Cette expérience illustre l'approche bayésienne en statistique, où les connaissances préalables ne sont pas gravées dans le marbre, mais sont sujettes à révision à la lumière des nouvelles données.

L'importance relative accordée aux connaissances a priori et aux données est flexible et dépend de la confiance que l'on accorde à ces sources d'information. En fin de compte, c'est la confrontation constante entre nos croyances préalables et la réalité empirique qui permet d'affiner notre compréhension du monde et de nous adapter aux circonstances changeantes.

1.2 Une opposition à la vision fréquentiste

La statistique, en tant que discipline, abrite deux courants de pensée majeurs : la vision fréquentiste et la vision bayésienne. Ces deux écoles se distinguent principalement par leur interprétation respective de la notion de probabilité, mais aussi par leur traitement des paramètres et des données, créant ainsi deux approches distinctes pour la résolution de problèmes statistiques.

Du côté fréquentiste, la probabilité est perçue comme une limite de fréquence à long terme. Autrement dit, si une expérience est répétée un grand nombre de fois, la probabilité d'un événement est définie par la fréquence relative de son occurrence. Par ailleurs, dans cette vision, les paramètres de l'expérience sont considérés comme des constantes, tandis que les données collectées sont envisagées comme des variables aléatoires. Ainsi, chaque expérience est traitée comme une réalisation indépendante d'un processus aléatoire, sans prendre en compte les connaissances préalablement acquises.

En revanche, la perspective bayésienne offre une interprétation radicalement différente. Dans cette approche, la probabilité est vue comme une mesure de l'incertitude ou du degré de croyance. Les paramètres de l'expérience, bien qu'inconnus, ne sont pas traités comme des constantes, mais plutôt comme des variables aléatoires. À la différence de l'approche fréquentiste, les bayésiens disposent de croyances a priori sur ces paramètres, croyances qui sont constamment mises à jour avec l'arrivée de nouvelles données, pour donner naissance à des croyances a posteriori. De ce fait, les statisticiens bayésiens ont la capacité d'intégrer des connaissances préalablement acquises et de réajuster leurs croyances en fonction des nouvelles données collectées.

Ces différences de philosophie ne sont pas sans conséquences. Elles se traduisent par de nombreuses divergences dans la manière dont les fréquentistes et les bayésiens abordent les problèmes de statistique. Les fréquentistes, ancrés dans le concret de la répétition d'expériences, et les bayésiens, plus flexibles et intégrant un degré de subjectivité, proposent ainsi deux prismes distincts pour comprendre et interpréter les phénomènes statistiques. Ces deux approches, malgré leurs divergences, restent complémentaires et contribuent chacune à leur manière à la richesse et à la complexité du domaine de la statistique.

Chapitre 2 Prérequis théoriques

2.1 Quelques définitions

On suppose que l'on dispose d'un échantillon $X = (X_1, \dots, X_n)$ de variables aléatoires i.i.d de loi P_θ avec $\theta \in \Theta \subset \mathbb{R}^d$ et $X_i \in \mathbb{R}^p$. On supposera que P_θ admet une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^p ou par rapport à la mesure de comptage.

Definition 1 (*Loi à priori et hyper paramètres*) On appelle loi a priori toute mesure de probabilité π sur l'ensemble Θ , et les hyperparamètres sont les paramètres de la loi a priori, ils sont considérés fixes et connus.

Definition 2 (*Densité jointe*) La loi jointe des observations de $X = (X_1, \dots, X_n)$ est conditionnelle à θ et est noté

$$f(x|\theta) = f(x_1, x_2, x_3, \dots, x_n|\theta) \text{ dans le cas continue et}$$
$$P(X = x|\theta) = P(X_1 = x_1, \dots, X_n = x_n|\theta).$$

Dans le cas continue on note, $f(X|\theta)$ la densité jointe de la variable aléatoire X .

Definition 3 (*Modèle Bayésien*) Un modèle bayésien est la donnée pour une variable aléatoire, ou pour une suite de variable aléatoire d'une loi conditionnelle et d'une loi à priori:

$$X \sim f(X|\theta)$$
$$\theta \sim \pi$$

A partir de ce modèle on peut alors calculer une loi a posteriori sur θ

Definition 4 (*Loi a posteriori*) on définit la loi a posteriori de manière suivante:

$$p(\theta|x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n) \times \pi(\theta)}{f_X(x_1, \dots, x_n)} \quad \text{Avec } L(x_1, \dots, x_n) \text{ la fonction de vraisemblance}$$

On appelle le dénominateur la loi marginale de X

En pratique il ne sera pas nécessaire de toujours déterminer le dénominateur, le loi marginale de X ne dépendant pas du paramètre θ , on peut la considérer comme une constante de normalisation, et on dit alors que la loi a posteriori est proportionnelle à la loi marginale de X , on adopte alors la notation suivante:

$$p(\theta|x_1, \dots, x_n) \propto L(x_1, \dots, x_n) \times \pi(\theta)$$

Enfin on évoque le théorème de Bayes, qui sera essentiel pour les calculs de loi a posteriori dans les prochaines sections.

Theorem 1 (*Théorème de Bayes*)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2.2 Choix de la loi a priori

2.2.1 Loi a priori impropre

Les lois a priori impropres représentent une catégorie particulière de distributions de probabilité. Contrairement aux lois a priori propres, où l'intégrale de la densité de probabilité sur l'ensemble de l'espace est égale à un, les lois a priori impropres ne satisfont pas cette condition et on a :

$$\int \pi(\theta) d\theta = +\infty.$$

Bien qu'à première vue elles semblent contre-intuitives, ces distributions jouent un rôle important en statistique bayésienne. Par exemple, la distribution uniforme sur tout l'espace réel est une loi a priori impropre qui est souvent utilisée en pratique, car elle exprime une absence totale d'information préalable. Cependant, leur utilisation nécessite une attention particulière. Il est crucial de vérifier que la procédure d'inférence donne des résultats cohérents et bien définis, car ce n'est pas toujours le cas lors de l'utilisation de lois a priori impropres. Dans certains cas, le résultat peut être une loi a posteriori impropre, indiquant généralement un problème avec les données ou le modèle, ou même si cela peut s'avérer contre-intuitif, il est tout à fait possible que la loi a posteriori associée à une loi a priori impropre soit une mesure de probabilité.

2.2.2 Loi a priori non informative

Dans la méthodologie bayésienne, la loi a priori joue un rôle crucial, permettant d'intégrer des connaissances antérieures avant l'arrivée des données. Toutefois, dans certains contextes, nous pouvons nous trouver dans une situation où peu ou pas d'informations préalables sont disponibles. Dans de tels cas, nous faisons appel aux lois a priori non informatives, aussi appelées lois a priori dites "impartiales" ou "vagues".

Ces lois a priori non informatives ont pour but de minimiser l'influence de l'information a priori sur les conclusions tirées des données. Elles sont formulées de manière à ne privilégier aucune valeur

spécifique du paramètre en question. Par conséquent, elles ont une large application, en particulier dans les situations où nous voulons que nos conclusions soient principalement influencées par les données observées plutôt que par nos croyances a priori.

Cependant, il est important de noter que le concept de loi a priori non informative peut être trompeur. En réalité, toute loi a priori introduit une certaine forme d'information, même si cette information est conçue pour être aussi "impartiale" que possible. Par ailleurs, la notion de non informativité peut dépendre du paramétrage choisi. Par conséquent, le choix d'une loi a priori non informative doit toujours être effectué avec prudence, en tenant compte du contexte spécifique de l'étude.

Malgré ces mises en garde, les lois a priori non informatives constituent un outil précieux dans l'arsenal du statisticien bayésien. Elles offrent un moyen de traiter les problèmes de manière bayésienne même en l'absence de connaissances a priori substantielles, permettant ainsi d'exploiter pleinement les avantages de l'approche bayésienne en termes de flexibilité et d'intégration des informations.

2.2.3 Loi a priori conjugués

Dans le cadre de l'analyse bayésienne, les lois a priori conjuguées occupent une place prépondérante. Une loi a priori est dite conjuguée à la loi de vraisemblance d'un modèle si la loi a posteriori appartient à la même famille que la loi a priori. Cela signifie que, lorsque l'on met à jour notre connaissance a priori avec les données observées pour obtenir une loi a posteriori, cette dernière demeure dans la même classe de distributions que la loi a priori.

L'usage des lois conjuguées facilite grandement les calculs dans l'analyse bayésienne. En raison de leur structure mathématique, ces lois permettent d'obtenir des formes analytiques simplifiées pour les lois a posteriori, ce qui facilite l'inférence bayésienne. C'est là l'un des avantages majeurs de l'emploi des lois a priori conjuguées : elles rendent l'analyse bayésienne plus maniable et évitent la nécessité de recourir à des méthodes numériques plus complexes.

Definition 5 Une famille de loi \mathcal{F} sur Θ est dite conjuguées si pour tout π appartenant à cette famille, la loi $p(\theta|x)$ appartient également à celle ci

On donne ci dessous un tableau de quelques familles de loi conjuguées usuelles.

$L(x \theta)$	$\pi(\theta)$	$p(\theta x)$
$N(\theta, \sigma^2) \sigma^2 \text{ connu}$	$N(\mu, \tau^2)$	$N\left(\frac{\sigma^2\mu + \tau^2\sum x_i}{\sigma^2 + n\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right)$

$\mathcal{G}(k, \theta)$ <i>k connu</i>	$\mathcal{G}(a, b)$	$\mathcal{G}\left(a + nk, b + \sum x_i\right)$
$\mathcal{B}(\theta)$	$Beta(a, b)$	$Beta\left(a + \sum x_i, b + n - \sum x_i\right)$

Chapitre 3 Modèle gaussien à moyenne et variance inconnus

3.1 Modèle bayésien et vraisemblance

On considère un modèle gaussien à moyenne et précision inconnus. Soit $\theta = (\mu, \tau)$ les paramètres de ce modèle, tel que $\tau = \frac{1}{\sigma^2}$.

Le modèle bayésien sera le suivant:

$$x_i | \mu, \tau \sim \mathcal{N}(\mu, \tau) \text{ iid.} \quad i = 1, \dots, n$$

$$\mu | \tau \sim \mathcal{N}(\mu_0, n_0 \tau)$$

$$\tau \sim \mathcal{G}(\alpha, \beta)$$

On considère la vraisemblance des données:

$$L(x_1, \dots, x_n | \mu, \tau) \propto \tau^{\frac{n}{2}} \exp \left(-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2 \right)$$

3.2 Lois a posteriori

Définissons un lemme que nous utiliserons directement pour avoir la loi a posteriori du paramètre μ .

Lemme 1 On considère un modèle d'échantillonnage gaussien à variance σ^2 connue et une vraisemblance gaussienne avec $x | \mu \sim \mathcal{N}(\mu, \sigma^2)$ avec une loi a priori sur la moyenne de $\mathcal{N}(\mu_0, \sigma_0^2)$.

La loi a posteriori sera :

$$\mathcal{N} \left(p\bar{x} + q\mu_0, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1} \right) \text{ avec } p = 1 - q = \frac{\sigma_0^2}{\frac{\sigma^2}{n} + \sigma_0^2}$$

D'après le lemme ci-dessus, si la variance est fixe, $\pi(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2)$ et $x | \mu \sim \mathcal{N}(\mu, \sigma^2)$ alors

$$\mu | x \sim \mathcal{N} \left(\frac{\frac{\sigma_0^2}{n} \bar{x} + \frac{\sigma^2}{\sigma_0^2 + \frac{\sigma^2}{n}} \mu_0, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1} \right)$$

Avec $\sigma^2 = \frac{1}{\tau}$ et $\sigma_0^2 = \frac{1}{n_0\tau}$ on obtient alors après simplification

$$\mu|x, \tau \sim N\left(\frac{n}{n+n_0}\bar{x} + \frac{n_0}{n+n_0}\mu_0, \frac{1}{(n+n_0)\tau}\right)$$

Maintenant intéressons-nous à $p(\tau|x)$, on l'exprimera en fonction de la densité jointe puis en intégrant par rapport à μ nous aurons la loi a postériori voulu.

$$p(\tau, \mu|x) \propto p(\tau) \cdot p(\mu|\tau) \cdot p(x|\tau, \mu)$$

$$\propto \tau^{\alpha-1} \cdot \exp(-\beta\tau) \cdot \tau^{\frac{1}{2}} \cdot \exp\left(-\frac{n_0\tau}{2}(\mu - \mu_0)^2\right) \cdot \tau^{\frac{n}{2}} \cdot \exp\left(-\frac{\tau}{2}\sum (x_i - \mu)^2\right)$$

En utilisant $x_i - \mu = x_i - \bar{x} + \bar{x} - \mu$

$$\propto \tau^{\alpha+\frac{n}{2}-1} \cdot \exp\left(-\tau\left(\beta + \frac{1}{2}\sum (x_i - \bar{x})^2\right)\right) \cdot \tau^{\frac{1}{2}} \cdot \exp\left(\frac{\tau}{2}(n_0(\mu - \mu_0)^2 + n(\bar{x} - \mu)^2)\right)$$

On calcule :

$$\begin{aligned} n_0(\mu - \mu_0)^2 + n(\bar{x} - \mu)^2 &= (n+n_0)\left(\mu^2 - \frac{2\mu}{n+n_0}(n_0\mu_0 + n\bar{x})\right) + n_0\mu_0^2 + n\bar{x}^2 \\ &= (n+n_0)\left(\mu - \frac{1}{n+n_0}(n_0\mu_0 + n\bar{x})\right)^2 - \frac{1}{n+n_0}(n_0\mu_0 + n\bar{x})^2 + n_0\mu_0^2 + n\bar{x}^2 \end{aligned}$$

De plus

$$-\frac{1}{n+n_0}(n_0\mu_0 + n\bar{x})^2 + n_0\mu_0^2 + n\bar{x}^2 = \frac{nn_0}{n+n_0}(\mu_0^2 + \bar{x}^2 - 2\mu_0\bar{x}) = \frac{nn_0}{n+n_0}(\mu_0 - \bar{x})^2$$

Et donc

$$\begin{aligned} \int \exp\left(-\frac{\tau}{2}(n_0(\mu - \mu_0)^2 + n(\bar{x} - \mu)^2)\right) d\mu &\propto \exp\left(-\frac{\tau}{2}\frac{nn_0}{n+n_0}(\mu_0 - \bar{x})^2\right) \int \exp\left(-\frac{\tau(n+n_0)\mu^2}{2}\right) d\mu \\ &\propto \tau^{-\frac{1}{2}} \cdot \exp\left(-\frac{\tau}{2}\frac{nn_0}{n+n_0}(\mu_0 - \bar{x})^2\right) \end{aligned}$$

on trouve finalement

$$p(\tau|x) \propto \tau^{\frac{n}{2}+\alpha-1} \exp\left(-\beta\tau - \frac{\tau}{2} \sum (x_i - \bar{x})^2 - \frac{\tau}{2} \frac{nn_0}{n+n_0} (\mu_0 - \bar{x})^2\right)$$

$$\text{Et donc } \tau|x_1, \dots, x_n \sim \mathcal{G}\left(\alpha + \frac{n}{2}, \frac{\sum (x_i - \bar{x})^2}{2} + \frac{nn_0}{n+n_0} (\mu_0 - \bar{x})^2\right)$$

Finalement, si

$$x_i|\mu, \tau \sim \mathcal{N}(\mu, \tau) \text{ iid.}$$

$$\mu|\tau \sim \mathcal{N}(\mu_0, n_0\tau)$$

$$\tau \sim \mathcal{G}(\alpha, \beta)$$

Alors les lois a posteriori seront :

$$\mu|x, \tau \sim \mathcal{N}\left(\frac{n}{n+n_0}\bar{x} + \frac{n_0}{n+n_0}\mu_0, \frac{1}{(n+n_0)\tau}\right)$$

$$\tau|x_1, \dots, x_n \sim \mathcal{G}\left(\alpha + \frac{n}{2}, \frac{\sum (x_i - \bar{x})^2}{2} + \frac{nn_0}{n+n_0} (\mu_0 - \bar{x})^2\right)$$

3.2.1 Illustration avec un jeu de données simulées

L'objectif de cet exemple est de montrer l'impact de la loi a priori sur la a posteriori en fonction du nombres de données. Le code est donné en Annexe (1)

On commence par simuler n échantillons de loi $\mathcal{N}(9, 2.25)$

On choisit les a priori suivantes:

$$\mu|x, \sigma^2 \sim \mathcal{N}\left(8, \frac{\sigma^2}{0.5}\right)$$

$$\sigma^2 \sim \text{IG}(1, 1)$$

Avec un algorithme de Gibbs, on peut simuler selon la loi a posteriori de μ et σ^2 . On obtient les graphiques suivants:

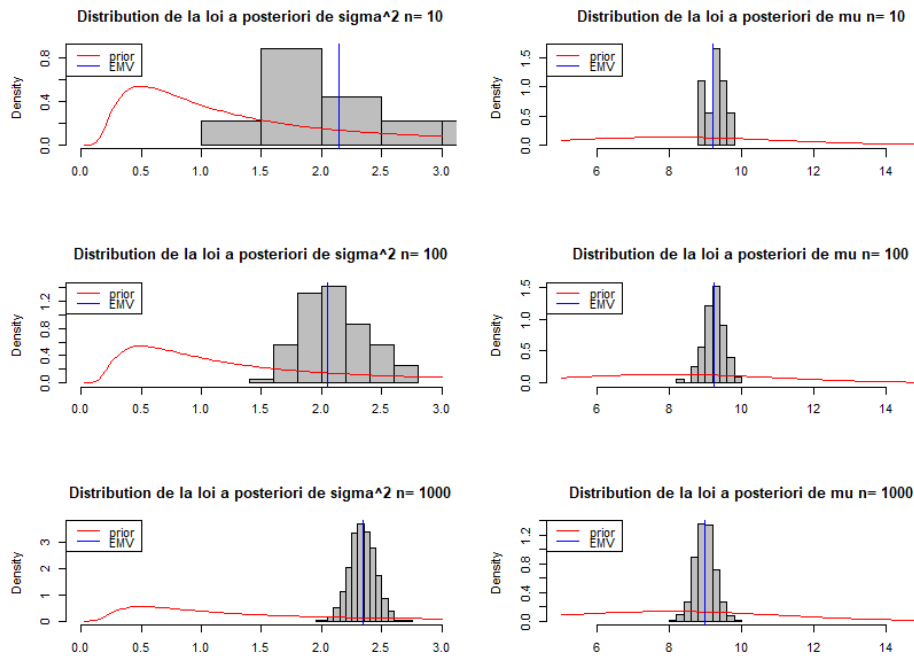


Figure 1: Impact de la loi a priori et du nombre de données sur le posterior.

Un aspect qui ressort de cette analyse est l'influence du nombre de données sur l'interaction entre la loi a priori et la loi a posteriori. Plus précisément, la taille de notre jeu de données (n) joue un rôle crucial dans la détermination de l'importance relative de la loi a priori dans l'influence de la loi a posteriori. En effet, plus la taille de l'échantillon est grande (c'est-à-dire plus n est grand), moins la loi a priori influence la loi a posteriori. Inversement, lorsque la taille de l'échantillon est plus petite, la loi a priori a une influence plus importante sur la loi a posteriori.

Ce phénomène est clairement visible lorsque nous comparons les résultats pour $n=10$ et $n=1000$. Avec un grand échantillon ($n=1000$), la distribution a posteriori s'éloigne du prior, indiquant que les données exercent une plus grande influence sur les résultats.

Concernant l'estimation de μ , notre loi a priori était centrée sur 8, alors que la vraie valeur de μ est 9. Nous avons observé que, à mesure que n augmente, l'estimateur du maximum de vraisemblance (EMV) se déplace vers la droite et se rapproche de plus en plus de la vraie valeur de 9. De même pour l'estimation de σ^2 , nous constatons qu'avec l'augmentation de n , l'EMV se rapproche progressivement de la vraie valeur de 2,25.

Ces observations confirment une notion fondamentale de l'inférence bayésienne: plus notre jeu de données est grand, moins la loi a priori a d'impact sur la loi a posteriori.

Chapitre 4 Modèle de régression linéaire simple

4.1 Modèle bayésien et vraisemblance

Dans cette partie nous nous intéressons à l'approche bayésienne dans le cadre de modèle de régression linéaire.

Commençons par définir le modèle de régression que nous allons étudier :

Soit Y la variable à expliquer, X une variable explicative et ϵ le terme d'erreur.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \dots, n$$

Où y_i est la i^{eme} observation de Y , x_i est la i^{eme} observation de X et ϵ_i est la i^{eme} observation du terme d'erreur. β_1 et β_0 sont les paramètres de régression.

On suppose que les $\epsilon_i \sim N(0, \sigma^2)$, les x_i sont connus, et les paramètres β_1 et β_0 sont vus comme des variables aléatoires.

Ainsi la variable conditionnelle $y_i | x_i, \beta_0, \beta_1, \sigma^2$ est une loi normale de moyenne $\beta_0 + \beta_1 x_i$ et de variance σ^2 .

L'objectif est de s'intéresser à l'estimation bayésienne des paramètres β_0 (l'intercepte) et β_1 (la pente) dans le cas de lois a priori non-informatives et le cas de lois a priori informatives.

Connaissant la loi de y nous pouvons écrire la vraisemblance du modèle:

On pose $Y = (y_1, \dots, y_n)$ et $X = (x_1, \dots, x_n)$

$$\mathcal{L}(Y|X, \beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right\}$$

$$\mathcal{L}(Y|X, \beta_0, \beta_1, \sigma^2) \propto \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right\}$$

Dans la suite nous allons détailler les calculs de la loi a posteriori uniquement pour un prior non

informatif.

4.2 Loi a priori non informatives

En utilisant la méthode de Jeffreys nous pouvons obtenir une loi a priori non informative. Les détails des calculs pour cette loi sont aux pages 139, 141-142 dans Le choix bayésien-Principe et pratique de Robert, C.P. (2006)

$$\text{Loi a priori jointe : } \pi(\beta_0, \beta_1, \sigma) \propto \frac{1}{\sigma}$$

En combinant la fonction de vraisemblance avec la loi a priori jointe on obtient la loi a posteriori suivante :

$$p(\beta_0, \beta_1, \sigma | x, y) \propto \pi(\beta_0, \beta_1, \sigma) * \mathcal{L}(Y|X, \beta_0, \beta_1, \sigma^2)$$
$$p(\beta_0, \beta_1, \sigma | x, y) \propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right\}$$

L'objectif maintenant est d'extraire les lois a posteriori marginales de nos trois paramètres β_0, β_1, σ

On pose:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1} \bar{x}$$

$$\widehat{\beta_1} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n \left(y_i - (\widehat{\beta_0} + \widehat{\beta_1} x_i) \right)^2$$

$\widehat{\beta_0}, \widehat{\beta_1}$ et S^2 sont les estimateurs obtenus par maximum de vraisemblance des paramètres $\beta_0, \beta_1, \sigma^2$

Calcul préliminaire:

Remarquons que :

$$\sum_{i=1}^n (\beta_0 - \widehat{\beta}_0) (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) (\beta_1 - \widehat{\beta}_1) x_i = 0$$

Considérons alors les équations algébriques suivantes :

$$\begin{aligned} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 &= \sum_{i=1}^n \left[(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) - (\beta_0 - \widehat{\beta}_0) - (\beta_1 - \widehat{\beta}_1) x_i \right]^2 \\ &= (n-2)S^2 + n(\beta_0 - \widehat{\beta}_0)^2 + (\beta_1 - \widehat{\beta}_1)^2 \sum_{i=1}^n x_i^2 + 2(\beta_0 - \widehat{\beta}_0)(\beta_1 - \widehat{\beta}_1) \sum_{i=1}^n x_i \end{aligned}$$

En insérant cette égalité dans la formule de densité a posteriori jointe de β_0, β_1, σ on obtient :

$$\begin{aligned} p(\beta_0, \beta_1, \sigma | x, y) &\propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} [(n-2)S^2 + n(\beta_0 - \widehat{\beta}_0)^2 + (\beta_1 - \widehat{\beta}_1)^2 \sum_{i=1}^n x_i^2 \right. \\ &\quad \left. + 2(\beta_0 - \widehat{\beta}_0)(\beta_1 - \widehat{\beta}_1) \sum_{i=1}^n x_i] \right\} \end{aligned}$$

On reconnaît une loi normale bivariée de moyenne $(\widehat{\beta}_0, \widehat{\beta}_1)$ et de matrice de variance-covariance notée M

$$M = \sigma^2 \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} = \sigma^2 \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} & \frac{-\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})} \\ \frac{-\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})} & \frac{1}{\sum_{i=1}^n (x_i - \bar{x})} \end{bmatrix}$$

En intégrant par rapport à σ , on obtient la loi a posteriori jointe de β_0 et de β_1 :

$$p(\beta_0, \beta_1 | x, y) = \int_0^\infty p(\beta_0, \beta_1, \sigma | x, y) d\sigma$$

$$\propto \left[(n-2)S^2 + n(\beta_0 - \widehat{\beta}_0)^2 + (\beta_1 - \widehat{\beta}_1)^2 \sum_{i=1}^n x_i^2 + 2(\beta_0 - \widehat{\beta}_0)(\beta_1 - \widehat{\beta}_1) \sum_{i=1}^n x_i \right]^{\frac{-n}{2}}$$

En intégrant une nouvelle fois par rapport à β_0 et β_1 , on obtient les lois a posteriori marginales de β_0 et de β_1 respectivement :

$$p(\beta_0|y, x) \propto \left[(n-2) + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2 \sum_{i=1}^n x_i^2 / n} (\beta_1 - \widehat{\beta}_1)^2 \right]^{\frac{-((n-2)+1)}{2}}$$

et

$$p(\beta_1|y, x) \propto \left[(n-2) + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2} (\beta_0 - \widehat{\beta}_0)^2 \right]^{\frac{-((n-2)+1)}{2}}$$

On remarque que :

$$T_1 = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2 \sum_{i=1}^n x_i^2 / n} \right]^{1/2} (\beta_0 - \widehat{\beta}_0) \quad \text{et} \quad T_2 = \frac{(\beta_1 - \widehat{\beta}_1)}{\left[S / \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}}$$

suivent une distribution de Student à $(n-2)$ degrés de libertés.

La densité à posteriori de σ est obtenue en intégrant la fonction de densité a posteriori jointe par rapport aux deux paramètres β_0 et β_1 , il en résulte la fonction suivante:

$$p(\sigma|y, x) \propto \frac{1}{\sigma^{(n-2)+1}} \exp \left\{ -\frac{(n-2)S^2}{2\sigma^2} \right\} \quad \text{qui est une gamma inverse } IG \left(\left(\frac{n}{2} - 1 \right), \frac{(n-2)S^2}{2} \right)$$

4.3 Loi a priori informatives

4.3.1 Calculs de la loi a postérieure

Commençons par réécrire le modèle.

$$Y = X\beta + \epsilon \text{ avec } Y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix}; X = \begin{pmatrix} 1 & x_1 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}; \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}; \epsilon = \begin{pmatrix} \epsilon_1 \\ \dots \\ \epsilon_n \end{pmatrix}$$

et $\epsilon_i \sim N(0, \sigma^2)$. On suppose le X connu.

Vraisemblance:

On peut donc écrire la vraisemblance avec cette nouvelle écriture.

$$\mathcal{L}(Y|X, \beta, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\right)$$

Choix des priors:

Pour faciliter les calculs nous choisissons des priors conjugués.

Nous avons :

$$\pi(\beta, \sigma^2) = \pi(\sigma^2) \pi(\beta|\sigma^2) \text{ avec}$$

$\pi(\sigma^2)$ suit une Gamma inverse et $\pi(\beta|\sigma^2)$ suit une loi normal multivariée

$$\pi(\sigma^2) \sim IG\left(\frac{v_0}{2}, \frac{v_0 s_0}{2}\right) \quad \pi(\beta|\sigma^2) \sim N(\mu_0, \sigma^2 \Lambda_0^{-1})$$

Ce qui nous donne donc pour notre prior:

$$\pi(\beta, \sigma^2) \propto (\sigma^2)^{-\frac{v_0}{2}-1} \exp\left(-\frac{v_0 s_0^2}{2\sigma^2}\right) (\sigma^2)^{-k/2} \exp\left(-\frac{1}{2\sigma^2}(\beta - \mu_0)^T \Lambda_0 (\beta - \mu_0)\right)$$

Loi a posteriori :

$$p(\beta, \sigma^2|y, X) \propto \mathcal{L}(Y|X, \beta, \sigma^2) \pi(\beta, \sigma^2)$$

$$\begin{aligned} & (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\right) (\sigma^2)^{-\frac{v_0}{2}-1} \exp\left(-\frac{v_0 s_0^2}{2\sigma^2}\right) (\sigma^2)^{-k/2} \\ & \propto \exp\left(-\frac{1}{2\sigma^2}(\beta - \mu_0)^T \Lambda_0 (\beta - \mu_0)\right) \end{aligned}$$

Avec quelques réarrangements on obtient $p(\beta, \sigma^2|y, X)$ les détails et étapes intermédiaires des calculs se trouvent dans " O'Hagan, Anthony (1994) Inférence bayésienne" au début du chapitre sur les modèles linéaires.

On trouve la loi a posteriori $p(\sigma^2 | Y, X)$ en intégrant par rapport à β et inversement en intégrant par rapport à σ^2 on obtient la loi a posteriori $p(\beta | \sigma^2, Y, X)$.

$$p(\beta | \sigma^2, Y, X) \sim \mathcal{N}(\mu_n, \sigma^2 \Lambda_n^{-1})$$

$$\Lambda_n = (X^T X + \Lambda_0) \quad \mu_n = (\Lambda_n)^{-1} (X^T X \hat{\beta} + \Lambda_0 \mu_0) \quad \hat{\beta} \text{ l'estimateur des moindres carrés}$$

$$p(\sigma^2 | Y, X) \sim IG(a_n, b_n)$$

$$a_n = \frac{v_0 + n}{2} \quad b_n = \frac{v_0 s_0^2 (Y^T Y + \mu_0^T \Lambda_0 \mu_0 - \mu_n^T \Lambda_n \mu_n)}{2}$$

4.3.2 Illustration avec un jeu de donnée simulé

Dans cet exemple, nous allons réaliser une régression linéaire dans le cadre bayésien et observer théoriquement l'importance des lois à priori.

Nous allons étudier la régression suivante :

$$Y = \alpha + \beta X + \epsilon$$

Pour cette régression nous cherchons à expliquer les malaises cardiaque (les Y) en fonction de la consommation d'alcool (les X). Nous utilisons les données présentent dans la library wooldridge. Commençons par une régression linéaire classique .

Code en Annexe (2)

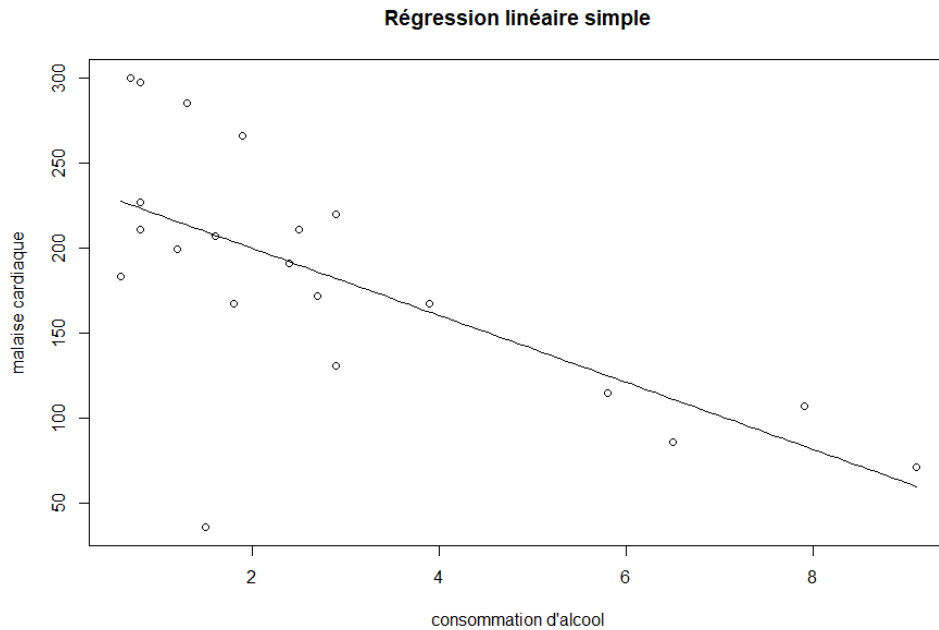


Figure 2: Régression linéaire simple.

Coefficients	Estimateurs	IC
α	239.147	[199.31 , 278.98]
β	-19.683	[-30.4,-8.96]

Regardons maintenant ce modèle du point de vue bayésien. En utilisant la fonction `bayes.lin.reg()` on peut choisir nos lois à priori et cette fonction renvoie les graphiques des lois a priori, posteriori et vraisemblance pour la pente (slope) et l'intercept.

On choisit comme lois a priori:

$$\alpha \sim N(180, 10^2)$$

$$\beta \sim N(-15, 5^2)$$

et un a priori non informatif pour σ^2 .

```
bayesien_modele=bayes.lin.reg(
  data$heart,
  data$alcohol,
  slope.prior = "normal",
  intcpt.prior = "flat",
  mb0=-15,
  sb0=5,
  ma0=0,
  sa0=0,
  sigma=NULL,
  alpha=0.05,
  plot.data=TRUE,
  pred.x = NULL,
)
```

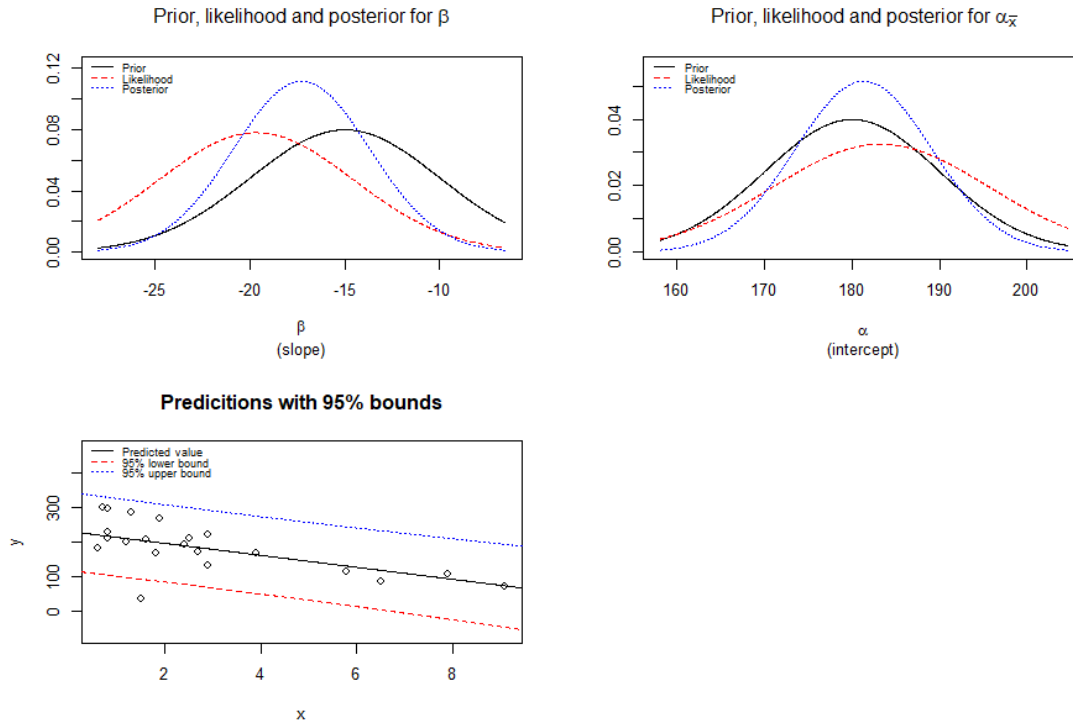


Figure 3: Régression linéaire bayésienne.

On constate que la loi à posteriori est une combinaison entre la vraisemblance et la loi a priori.

Les moyennes a posteriori:

$$E(\alpha|x, y) = 181.3 \quad \text{et} \quad E(\beta|x, y) = -17.29.$$

Dans le contexte de notre étude, nous avons mené deux types de régression linéaire : une régression linéaire simple classique et une régression linéaire simple basée sur l'approche bayésienne. Après avoir réalisé ces analyses, nous avons comparé les estimations des paramètres obtenues à partir de ces deux méthodes.

Nous constatons que les moyennes a posteriori obtenues à partir de la régression bayésienne sont assez éloignées de celles obtenues à partir de la régression linéaire simple classique.

L'explication de cela réside dans l'influence significative des priors dans l'approche bayésienne. Comme notre ensemble de données est relativement petit, ne contenant que 21 points, les priors ont une incidence importantes sur l'estimation de nos paramètres. Autrement dit, dans le contexte bayésien, les priors contribuent à orienter, voire à contraindre, les résultats de l'analyse, particulièrement lorsque la taille de l'échantillon est limitée.

Dans un jeu de données plus volumineux, cette divergence entre les résultats de la régression linéaire

simple et de la régression bayésienne serait probablement moins prononcée. Cela est dû au fait qu'avec un grand nombre de points de données, l'influence des priors se dilue, et les données elles-mêmes commencent à exercer une influence plus dominante sur les estimations des paramètres. En d'autres termes, l'impact des données augmente tandis que celui du prior diminue.

Une nouvelle fois cela met en évidence un aspect important de l'analyse bayésienne, qui est la manière dont les priors peuvent influencer les résultats, surtout lorsque la taille de l'échantillon est faible.

Chapitre 5 Modèle de régression semi-paramétrique

5.1 Définition du modèle

Cette partie se base sur l'article "Bayesian Backfitting, Trevor Hastie and Robert Tibshirani".

Supposons que nous avons un jeu de données que l'on note (x_i, y_i) . On considère le modèle suivant :

$$y_i = f(x_i) + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Avec une approche classique, on peut obtenir f comme le minimiseur de la somme des carrés pénalisée suivante:

$$J(f) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int (f''(x))^2 dx$$

λ est un paramètre de lissage utilisé pour contrôler l'allure de la courbe et ainsi obtenir des courbes plus ou moins lisses.

Pour comprendre ce modèle et l'importance du paramètre λ nous allons l'appliquer sur un exemple avec des données simulées.

Dans notre exemple : $f(x) = \sin(2\pi x)$ et 0.04

Nous allons utiliser la fonction `smooth.spline` du package `splines` pour visualiser la fonction f^* qui est la solution de minimisation du critère de somme des carrés pénalisée pour un certain paramètre de lissage.

Code en Annexe (3).

On obtient les sorties graphiques suivantes :

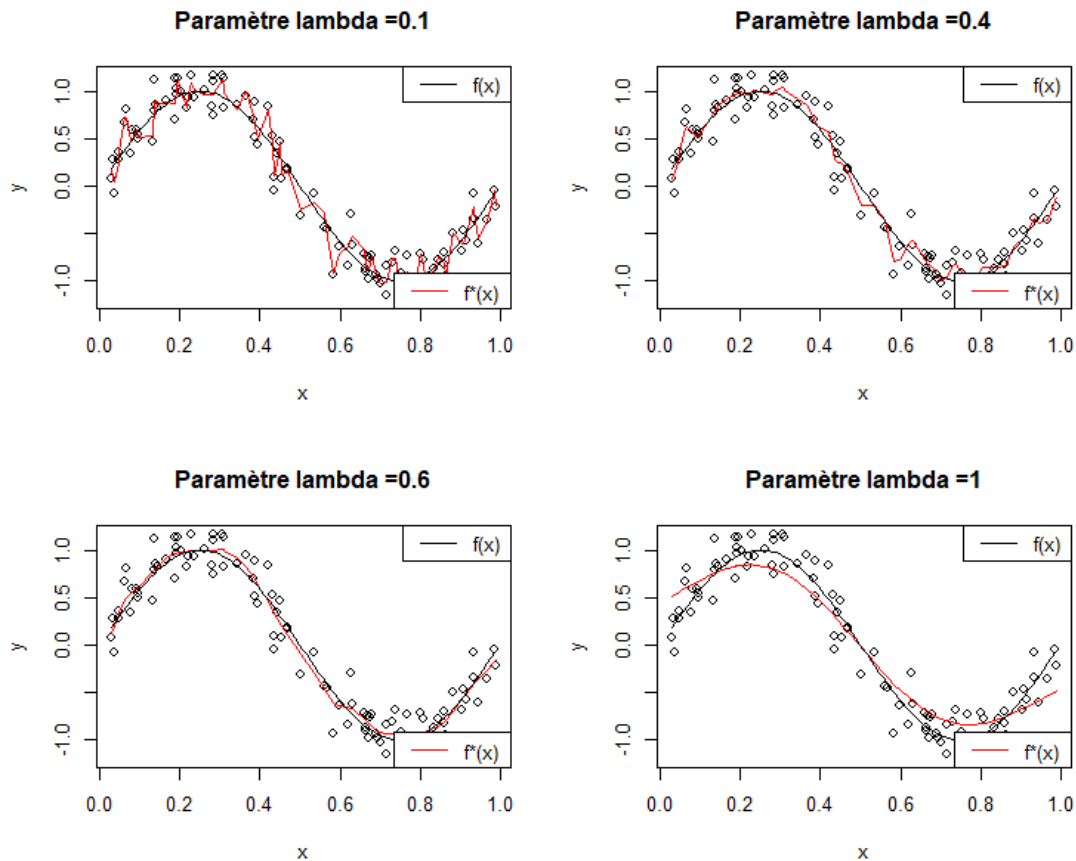


Figure 4: Régressions semi-paramétrique.

Les résultats montrent comment le paramètre λ affecte le lissage de la courbe. Un λ plus petit (proche de zéro) conduit à une courbe qui s'ajuste de près aux données. À mesure que λ augmente, la courbe s'adoucit et suit davantage la tendance générale des données, ce qui correspond à un lissage plus important.

5.2 Modèle bayésien et vraisemblance

Approche bayésienne:

Nous allons regarder ce modèle avec une approche bayésienne. f n'est plus vu comme un paramètre mais comme une variable aléatoire. Pour l'instant on suppose σ^2 connu.

On note:

$y = (y_1, \dots, y_n)$ $f = (f(x_1), \dots, f(x_n))$ N_m pour la loi normale multidimensionnelle et I pour la matrice identité de taille $n \times n$.

Ainsi pour notre modèle:

$$y|f, \sigma^2 \sim N_m(f, \sigma^2 I)$$

L'article suggère de choisir comme loi a priori pour f une gaussienne partiellement impropre :

$$f \sim N_m(0; K^- \tau^2)$$

K^- est l'inverse généralisé de K

Ainsi d'après l'article la loi a posteriori de f est:

$$f|y, \sigma^2 \sim N_m(Sy, \sigma^2 S)$$

$$\text{Avec } S = B(B^T B + \lambda \Omega)^{-1} B^T$$

B correspond à la matrice B-spline où la i -ème ligne de cette matrice est $(b_1(x_i), \dots, b_M(x_i))$

$$\text{avec } f(x) = \sum_{j=1}^M b_j(x) \theta_j \quad \text{et } \theta_j \text{ des poids.}$$

On peut voir cette matrice B comme la combinaison de plusieurs fonctions de base appelées B-spline de base qui sont définies sur un intervalle spécifique et capturent une partie de la forme générale de la courbe.

$$\Omega \text{ et la matrice de pénalité : } \Omega_{ij} = \int b_i''(t) b_j''(t) dt.$$

$$\text{et d'après l'article la matrice de covariance du prior } K^- = B \Omega^- B^T$$

Nous allons vérifier par le calcul que l'on retombe bien sur la loi a posteriori de l'article:

Vraisemblance du modèle :

$$\mathcal{L}(y|X, f, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}(y-f)^T(y-f)\right)$$

La loi a priori:

$$\pi(f) \propto \exp\left(\frac{1}{\tau^2} f^T K f\right)$$

La loi a posteriori:

$$p(f|y, \sigma^2) \propto \mathcal{L}(y|X, f, \sigma^2) \pi(f)$$

$$\propto \exp\left(\frac{-1}{2\sigma^2}(y-f)^T(y-f) - \frac{1}{2\tau^2}f^TKf\right)$$

$$\propto \exp\left(\frac{-1}{2\sigma^2}(y^Ty - 2f^Ty + f^Tf) - \frac{1}{2\tau^2}f^TKf\right)$$

y^Ty est une constante

$$\propto \exp\left(\frac{-1}{2}\left(f^T\left(\frac{I}{\sigma^2} + \frac{K}{\tau^2}\right)f - \frac{2f^Ty}{\sigma^2}\right)\right)$$

$$\propto \exp((f-\mu)^T\Sigma^{-1}(f-\mu))$$

$$\text{avec } \mu = \frac{\Sigma y}{\sigma^2} \quad \Sigma^{-1} = \left(\frac{I}{\sigma^2} + \frac{K}{\tau^2}\right)$$

$$\text{On pose } S = \frac{\Sigma}{\sigma^2} \quad \text{Ainsi } f|y, \sigma^2 \sim \mathcal{N}_m(Sy, \sigma^2 S)$$

Dans l'article $S = B(B^TB + \lambda\Omega)^{-1}B^T$ vérifions que si $K^- = B\Omega^-B^T$ l'on retombe sur le même S.

$$B(B^TB + \lambda\Omega)^{-1}B^T = \frac{1}{\sigma^2}\left(\frac{I}{\sigma^2} + \frac{K}{\tau^2}\right)^{-1}$$

$$\Leftrightarrow (B^TB + \lambda\Omega)^{-1} = B^{-1}\frac{1}{\sigma^2}\left(\frac{I}{\sigma^2} + \frac{K}{\tau^2}\right)^{-1}(B^T)^{-1}$$

$$\Leftrightarrow (B^TB + \lambda\Omega)^{-1} = \left(B^T\sigma^2\left(\frac{I}{\sigma^2} + \frac{K}{\tau^2}\right)B\right)^{-1}$$

$$\Leftrightarrow B^TB + \lambda\Omega = B^T(I + \lambda K)B$$

$$\Leftrightarrow \lambda\Omega = \lambda B^TKB$$

$$\Leftrightarrow K^- = B\Omega^-B^T$$

$$\lambda = \sigma^2 / \tau^2$$

5.3 Illustration avec un jeu de données simulées

On reprend le même exemple que précédemment.

$$f(x) = \sin(2\pi x) \text{ et } \sigma^2 = 0.04$$

On génère des x auxquels on applique la fonction f et l'on rajoute un bruit.

A l'aide des fonctions du package "fda" on peut construire les matrices B et Ω qui interviennent et

ainsi calculer la matrice S ($S = B(B^T B + \lambda \Omega)^{-1} B^T$).

Connaissant la moyenne et la matrice de variance-covariance on peut simuler $f|y$.

On simule 2000 échantillons puis nous prenons la médiane pour chaque composant du vecteur f pour construire notre courbe de régression a posteriori.

On construit également un intervalle de crédibilité en utilisant la fonction quantile.

Le paramètre lambda est déterminé empiriquement et dépend du lissage que l'on souhaite.

Code en Annexe (4)

Nous obtenons la régression suivante:

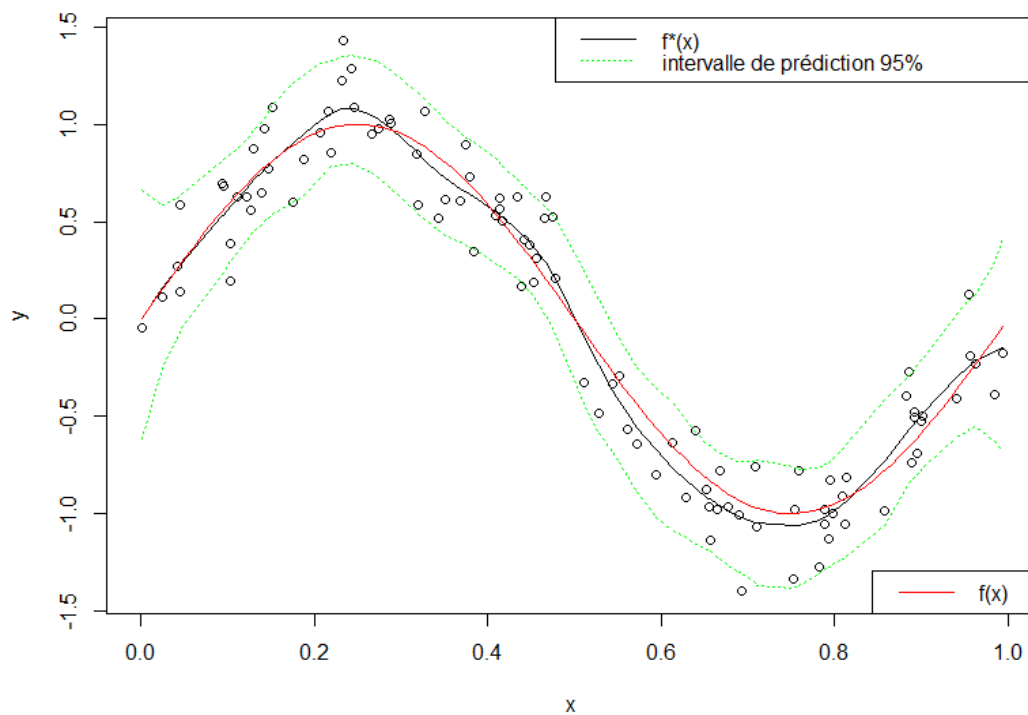


Figure 5: Régression bayésienne semi-paramétrique avec σ^2 connu.

La courbe de régression suit bien la tendance générale de notre fonction f .

σ^2 inconnu:

On considère le même modèle mais maintenant σ^2 est inconnu, nous allons donc le considérer comme une variable aléatoire. On va donc mettre une loi a priori pour σ^2 et calculer sa loi a posteriori.

On choisit un a priori conjugué :

$$\pi(\sigma^2) \sim IG(a, b)$$

On calcul la loi a posteriori :

$$p(\sigma^2|f, y) \propto \mathcal{L}(y|X, f, \sigma^2)\pi(\sigma^2).$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+a+1} \exp\left(-\frac{1}{2\sigma^2}(y-f)^T(y-f) - \frac{b}{\sigma^2}\right)$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+a+1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{1}{2}(y-f)^T(y-f) + b\right)\right)$$

On remarque que $\sigma^2 \sim IG\left(\frac{n}{2} + a, \frac{1}{2}(y-f)^T(y-f) + b\right)$

L'objectif est de vérifier que l'algorithme que nous allons utiliser dans notre application à des données climatiques fonctionne correctement.

Pour cela on reprend le même jeu de données simulées et on réalise un algorithme de Gibbs pour simuler la loi a posteriori de f . L'algorithme de Gibbs consiste à simuler $f|\sigma^2$ en utilisant le σ^2 précédemment obtenu et simuler $\sigma^2|f$ en utilisant le f que l'on vient de simuler puis mettre à jour le λ et à chaque nouvelle itération calculer la nouvelle matrice S .

Pour tracer la fonction on prend la médiane de f sur les 300 dernières itérations.

L'algorithme est donné en Annexe.

code en Annexe (5)

Nous obtenons le graphique suivant:

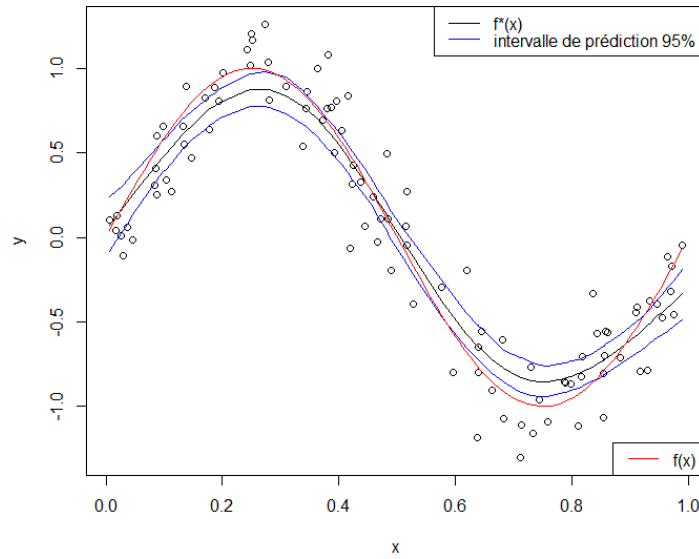


Figure 6: Régression bayésienne semi-paramétrique avec σ^2 inconnu.

La courbe de régression obtenu suit une nouvelle fois la tendance générale de la courbe. On vérifie que σ^2 converge vers sa vraie valeur qui est 0.04.

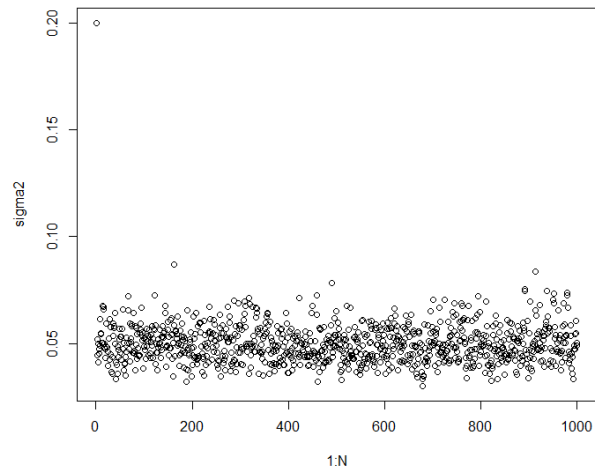


Figure 7: Chaîne de Markov pour σ^2 .

Chapitre 6 Application à des données réelles

6.1 Présentation du jeu de données

Nous avons les relevés des températures maximales par jour depuis 1977 dans les villes suivantes: Lyon (France), Brindisi (Italie), Akureyri (Islande),Helsinki (Finlande) et Cape Sterlegov (Russie)



Figure 8: Localisation des villes étudiées.

Avant d'appliquer notre code précédent nous devons réarranger nos données. En effet par exemple pour la ville de Lyon nous avons 14380 relevés de température, pour des raisons de temps de calcul nous ne pouvons pas toute les prendre en compte, c'est pourquoi nous réalisons une moyenne des températures maximales par années.

Nous pouvons alors appliquer le même algorithme de Gibbs que précédemment. En déterminant empiriquement le paramètre τ^2 en fonction du lissage souhaité.

Les codes sont donnés en Annexe (6).

On obtient les courbes suivantes:

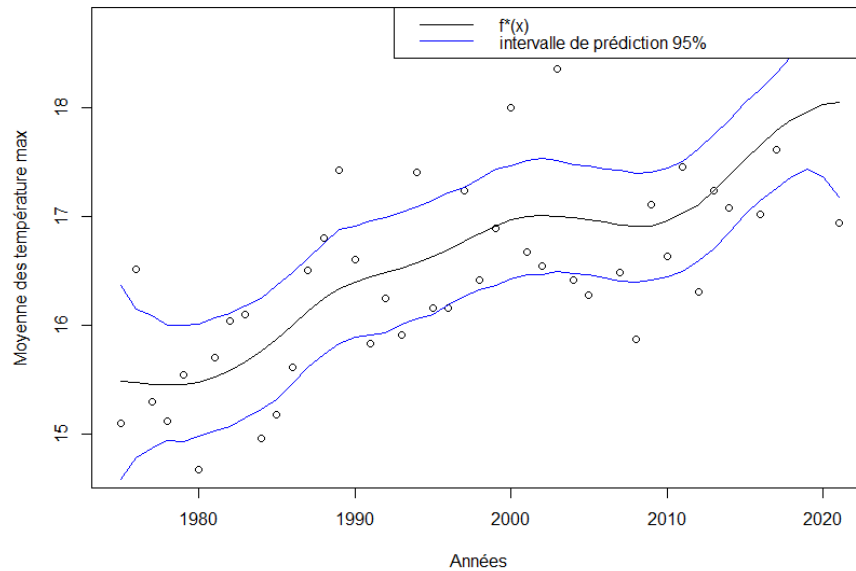


Figure 9: Régression bayésienne semi-paramétrique pour les températures maximales en fonction des années à Lyon.

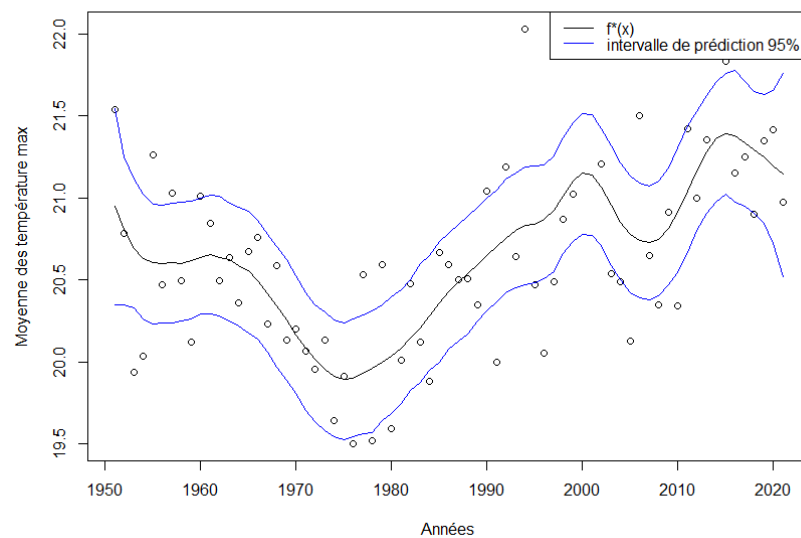


Figure 10: Régression bayésienne semi-paramétrique pour les températures maximales en fonction des années à Brindisi.

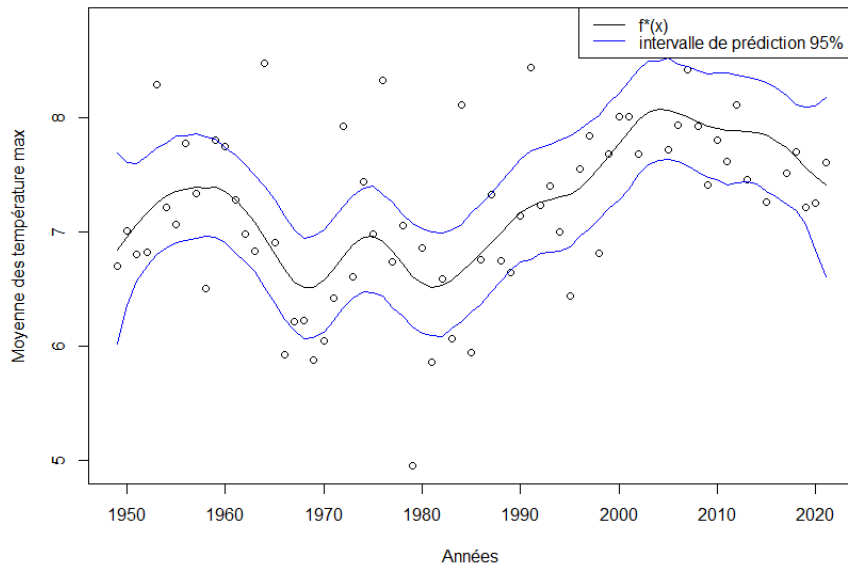


Figure 11: Régression bayésienne semi-paramétrique pour les températures maximales en fonction des années à Akureyri.

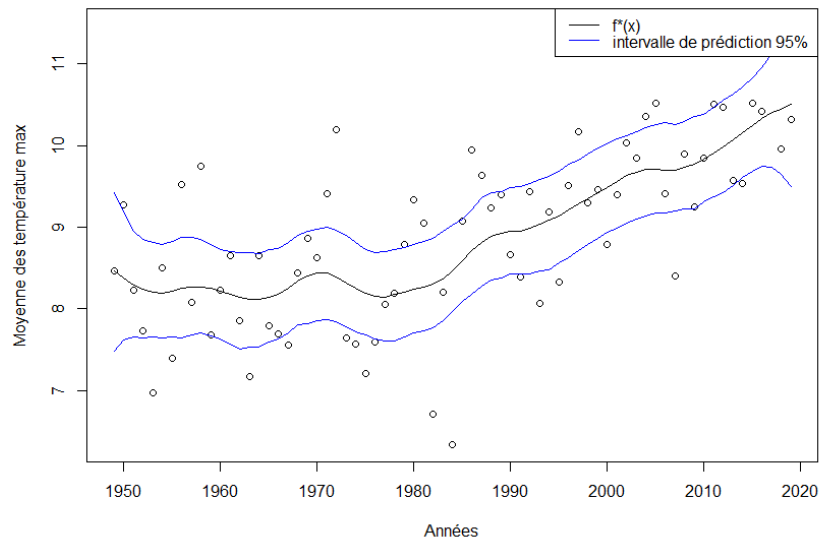


Figure 12: Régression bayésienne semi-paramétrique pour les températures maximales en fonction des années à Helsinki.

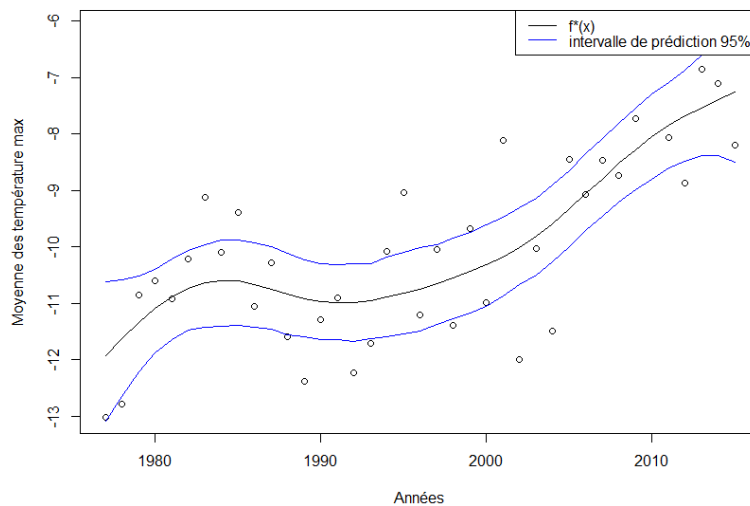


Figure 13: Régression bayésienne semi-paramétrique pour les températures maximales en fonction des années à Cape Sterlegov.

6.2 Analyse des sorties graphiques

Cet ensemble de graphiques dépeint l'évolution des températures dans cinq villes européennes au cours des deux dernières décennies.

Ces graphiques mettent en évidence une tendance commune qui traverse toutes les cinq villes. En calculant la moyenne annuelle des températures maximales, nous obtenons une image claire de l'évolution des températures d'une année à l'autre. Cette approche nous donne une vision globale de l'évolution du climat et met en évidence des tendances importantes qui ne sont pas immédiatement visibles dans les variations quotidiennes ou mensuelles.

Il est clairement observable à partir de ces graphiques que la température moyenne annuelle a augmenté dans chaque ville au cours des 70 dernières années. Il est intéressant de noter que bien qu'il y ait eu des fluctuations - des baisses temporaires de température - la tendance générale montre une hausse constante des températures. Cela indique que, malgré les variations saisonnières et annuelles naturelles, les températures aujourd'hui sont généralement plus élevées qu'elles ne l'étaient il y a 70 ans.

Cette observation met en évidence le phénomène de réchauffement climatique en cours. Le fait que cette tendance soit visible dans un éventail de villes à travers l'Europe renforce l'idée que c'est un problème global, pas seulement localisé dans certaines régions.

Conclusion et perspectives

Dans le cadre de ce TER, nous avons exploré en profondeur l'approche bayésienne en statistique, une méthode qui nous permet d'intégrer des connaissances a priori et de les mettre à jour de manière dynamique avec l'arrivée de nouvelles données. Nous avons mis en lumière les différentes composantes de cette méthode, allant des lois a priori non informatives aux lois a priori conjuguées, et leur application dans divers contextes, notamment la régression linéaire simple et la régression semi-paramétrique.

La principale contribution de ce travail a été d'appliquer la régression semi-paramétrique bayésienne aux données climatiques afin d'étudier les tendances des températures dans différentes villes. Nos résultats ont montré le potentiel de cette approche pour tirer des informations précieuses à partir de données complexes, et en offrant la possibilité de modéliser des relations complexes et non linéaires.

Cependant, il faut noter que notre analyse n'a pas pris en compte l'aspect spatial des données. C'est un domaine d'étude potentiellement riche qui mérite d'être exploré davantage. L'aspect spatial, qui se réfère à la manière dont les variables interagissent et changent en fonction de leur position ou de leur localisation, pourrait révéler des informations supplémentaires qui ne sont pas immédiatement apparentes lorsque l'on considère uniquement l'aspect temporel.

7. Annexe

1. Code de la partie 3.2.1: Impact de la loi a priori et du nombre de données sur la loi a posteriori.

```
library(invgamma)
library(ggplot2)
# Modele bayesien

#  $\xi \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ 
#  $\mu \mid \sigma^2 \sim N(a, \sigma^2 / b)$ 
#  $\sigma^2 \sim IG(c, d)$ 

# les hyperparamètres
a <- 8
b <- 0.5
c <- 1
d <- 1

par(mfrow=c(3,2))
chaine <- matrix(c(1,1),nc=2,nr=1) #initialisation
for(j in 1:3){
  N <- c(10,100,1000)
  n=N[j]
  X=rnorm(n,9,1.5)
  Xn=mean(X)
  sum_X=sum(X)
  SEM=sum((X -mean(X))^2)
  chaine <- matrix(c(1,1),nc=2,nr=1)

  for (i in 2:n){
    sigma_2 <- rinvgamma(1,c+n/2,d+0.5*(SEM+(n*b)*(Xn-a)^2/(2*(n+b))))
    mu <- rnorm(1,(sum_X+a*b)/(n+b),sigma_2/(a+b))
    chaine <- rbind(chaine,c(sigma_2,mu))
  }
}
```

```

print(n[j])
print(n)
hist(chaine[2:n,1],main=paste("Distribution de la loi a
posteriori de sigma^2
n=",N[j]),freq=F,col="grey",xlim=c(0,3),xlab=" ")
curve(dinvgamma(x,c,d),col="red",add=T)
abline(v=SEM/n,col="blue")
legend("topleft",legend = c("prior","EMV"), col =
c("red","blue"), lty = c(1,1))
hist(chaine[2:n,2],freq = F,main=paste("Distribution de la loi
a posteriori de sigma^2 n=",N[j]),col="grey",xlim=c(5,15))
curve(dnorm(x,a,1.5/b),col="red",add=T)
abline(v=Xn,col="blue")
legend("topleft",legend = c("prior","EMV"), col =
c("red","blue"), lty = c(1,1))
}

```

2. Code de la partie 4.2.1 : Régression linéaire simple.

```

#loi a priori informative

library(wooldridge)
data = wooldridge::wine
data;
modele=lm(data$heart~data$alcohol)
summary(modele)
plot(data$alcohol,data$heart)
curve(239.147-19.683*x,add=T)

```

3. Code de la partie 5.1: Régression semi-paramétrique et étude du paramètre λ .

```

library(splines)
# Générer des données
n <- 100 # Nombre de points
x <- runif(n,0,1)
f <- sin(2 * pi * x)
epsilon <- rnorm(n, mean = 0, sd = 0.2) # Erreurs aléatoires ε
y <- f + epsilon # Valeurs de y avec bruit

```

```

lambdas <- c(0.1, 0.4, 0.6, 1) # Différentes valeurs de lambda

par(mfrow = c(2, 2)) # Organiser les graphiques

for (i in 1:4) {
  lambda <- lambdas[i]
  smooth_spline <- smooth.spline(x, y, spar = lambda)

  # Visualiser la courbe ajustée
  plot(x, y, main = paste0("Paramètre lambda =", lambda), xlab =
"x", ylab = "y")
  lines(smooth_spline, col = "red") # Ajouter la courbe ajustée
en rouge
  curve(sin(2*pi*x), add=TRUE, col='black')
  legend("topright", legend = "f(x)", col = "black", lty = 1)
  legend("bottomright", legend = "f*(x)", col = "red", lty = 1)
}

```

4. Code de la partie 5.3 Régression semi-paramétrique, approche bayésienne avec σ^2 connu.

```

#chargement des bibliothèques
library(splines)
library(fda)
library(MASS)

#On génère un jeu de données
n <- 100
x <- runif(n, 0, 1)
y <- sin(2 * pi * x) + rnorm(n, 0, 0.2) # On ajoute un bruit
plot(x,y) #on visualise nos données
curve(sin(2 * pi * x), add=T, col="green") #la vraie fonction

#Création de la matrice B et Omega
r=4 #spline cubique
M=(n-2)+r #nombre de fonctions de bases
z=x[order(x)]

```

```

bspl=create.bspline.basis(range(z),nbasis=M,norder=r,z) #création
d'une base BSpline - range(z) = intervalle sur lequel les
fonctions doivent être évaluées
Omega=bsplinepen(bspl,Lfdobj=2,rng=bspl$rangeval) #création de
la matrice de pénalisation (pénalité de lissage pour fonctions
exprimées en base B-spline)
B=bsplineS(x,z,nord=r)

#paramètre de lissage
lambda=0.0001 #déterminer empiriquement
S=B%*%solve((t(B)%*%B+lambda*Omega))%*%t(B) #la matrice S

sigma2=0.2 #On suppose sigma2 connu

f=mvrnorm(2000,S%*%y,sigma2*S) #On génère 2000 échantillons de f|y
f_med=apply(f,2,median) #on récupère la médiane
f_lower=apply(f,2,quantile,probs=0.025)
f_upper=apply(f,2,quantile,probs=0.975)
plot(x,y)
lines(z,f_med[order(x)])
lines(z,f_lower[order(x)],lty=3,col='green')
lines(z,f_upper[order(x)],lty=3,col='green')
curve(sin(2 * pi * x),add=T,col="red")
legend("topright", legend = c("f*(x)","intervalle de prédiction
95%"), col = c("black","green"), lty = c(1,3))
legend("bottomright", legend = "f(x)", col = "red", lty = 1)

```

5. Code de la partie 5.3 .Régression semi-paramétrique, approche bayésienne avec σ^2 inconnu.

```

#chargement des bibliothèques
library(splines)
library(fda)
library(MASS)
library(invgamma)

#On génère un jeu de données

```

```

n <- 100
x <- runif(n, 0, 1)
y <- sin(2 * pi * x) + rnorm(n, 0, 0.2) # On ajoute un bruit
plot(x,y) #on visualise nos données
curve(sin(2 * pi * x),add=T,col="green") #la vraie fonction

#Construction de B et Omega
r=4 #spline cubique
M=(n-2)+r #nombre de fonctions de bases
z=x[order(x)]
bspl=create.bspline.basis(range(z),nbasis=M,norder=r,z) #création
d'une base BSpline
Omega=bsplinepen(bspl,Lfdobj=2,rng=bspl$rangeval) #création de
la matrice de pénalisation (pénalité de lissage pour fonctions
exprimées en base B-spline)
B=bsplineS(x,z,nord=r)

#hyperparamètre de la loi a priori sur sigma^2 suit une
invgamma(a,b)
a=0.01
b=0.01
curve(dinvgamma(x,a,b),xlim = c(0,10),main="Loi a priori de
sigma^2")

tau2=8 #déterminer empiriquement

N=1000 # Nombre d'itération
#initialisation
f_iter=matrix(y,N,n) #on stocke les vecteurs f dans une matrice
sigma2=rep(0.2,N)
lambda=rep(0.0015,N)

# Algorithme Gibbs
for(i in 2:N){
S=B%%solve((t(B)%%B+lambda[i-1]*Omega))%%t(B)#on calcul
S(lambda)

```



```

f=mvrnorm(1,S%%y,sigma2[i-1]*S) #On simule  $f|\sigma^2, Y$  en
utilisant le précédent  $\sigma^2$ 
f_iter[i,]=f #On stocke le f dans la matrice
sigma2_value=rinvgamma(1,n/2+a,(1/2)*norm(y-
f_iter[i,],type="2")^2+b) #on simule  $\sigma^2|Y, f$  en utilisant le f
que l'on vient de simuler
sigma2[i]=sigma2_value #on stocke la valeur mediane
lambda[i]=sigma2[i]/tau2
}

f_final=apply(f_iter[floor(0.8*N):N,],2,median)
f_lower=apply(f_iter[floor(0.8*N):N,],2,quantile,probs=0.025)
f_upper=apply(f_iter[floor(0.8*N):N,],2,quantile,probs=0.975)
plot(x,y)
lines(z,f_final[order(x)])
lines(z,f_lower[order(x)],lty=1,col='blue')
lines(z,f_upper[order(x)],lty=1,col='blue')
curve(sin(2 * pi * x),add=T,col="red")
legend("topright", legend = c("f*(x)", "intervalle de prédiction
95%"), col = c("black", "blue"), lty = c(1,1))
legend("bottomright", legend = "f(x)", col = "red", lty = 1)

#représentation de la chaine
plot(1:N,sigma2)

```

6. Code de la partie 6.1 Application à notre jeu de données.

```

library(splines)
library(fda)
library(MASS)
library(invgamma)
#Importation des données
chemin_Lyon <-
"C:/Users/hrywa/Desktop/Donnees_temperature/TX_STAID000037.txt"
chemin_ROME<-
"C:/Users/hrywa/Desktop/Donnees_temperature/TX_STAID000176.txt"
chemin_AKUREYRI<-
"C:/Users/hrywa/Desktop/Donnees_temperature/TX_STAID002943.txt"

```

```

chemin_HELSINKI<-
"C:/Users/hrywa/Desktop/Donnees_temperature/TX_STAID007015.txt"
chemin_STERLEGOVA<-
"C:/Users/hrywa/Desktop/Donnees_temperature/TX_STAID008044.txt"

donnees <- read.table(chemin_Lyon, sep = ",", header=TRUE , skip =
20)
#on retire les valeurs manquantes
donnees_filtrees <- subset(donnees, TX != -9999)
# Nombre de groupes de 30 valeurs
nb_groupes <- length(donnees_filtrees$TX) %/% 365

# Vecteur pour stocker les moyennes
vecteur_moyennes <- numeric(nb_groupes)

# Calculer les moyennes de chaque groupe de 30 valeurs
for (i in 1:nb_groupes) {
  debut <- (i - 1) * 365 + 1
  fin <- i * 365
  vecteur_groupe <- donnees_filtrees$TX[debut:fin]
  moyenne_groupe <- mean(vecteur_groupe)
  vecteur_moyennes[i] <- moyenne_groupe
}

y=0.1*vecteur_moyennes #Pour avoir les température en °C

n=length(y)
x=1975:(n-1+1975) # à adapter en fonction de la ville

tau2=0.009 #déterminer empiriquement
#Puis on applique l'algorithme de gibbs du code précédent avec
nos nouveaux vecteurs x et y

```

8. Références

- [1] Robert, C. (2005). Le Choix Bayésien. Springer-Verlag, Paris.
- [2] O'Hagan, Anthony (1994) Inférence bayésienne, chapitre Modèles linéaires
- [3] Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. Journal of the Royal Statistical Society, 40(3):364–372.
- [4] Hastie, T. and Tibshirani, R. (2000). Bayesian backfitting. Statistical Science, 15(3):196– 223.
- [5] Bekda Lilia, (2017) Approche Bayésienne dans les modèles de régression.
lien: <https://www.ummo.dz/dspace/bitstream/handle/ummo/2866/Bekda%2C%20Lilia%20.pdf?sequence=1&isAllowed=y>
- [6] Wikipedia, Bayesian linear regression
lien: https://en.wikipedia.org/wiki/Bayesian_linear_regression
- [7] The Conjugate Prior for the Normal Distribution
lien: <https://people.eecs.berkeley.edu/~jordan/courses/260-spring10/lectures/lecture5.pdf>

