

République Algérienne Démocratique et Populaire.
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.
Université de Mouloud Mammeri, Tizi-Ouzou

Faculté des Sciences
Département de Mathématiques

Spécialité : Mathématiques

Option : Processus Aléatoires et Statistique de la décision

Mémoire de Master intitulé :

Approche Bayésienne dans les modèles de régression.

Réalisé par :

Bekda Lilia

Sous la direction de :

M^{lle} **Atil Lynda**

Soutenu devant le jury d'examen composé de :

M^r **Fellag Hocine,**

Professeur, UMMTO,

Président.

M^{lle}. **Atil Lynda ,**

Maître de conférence A, UMMTO,

Rapporteur.

M^{me} **Belkacem Cherifa,**

Maitre de conférence B, UMMTO,

Examinatrice.

12 octobre 2017

Table des matières

Introduction	3
1 Inférence classique et Bayésienne	5
1.1 Inférence classique	5
1.2 Inférence Bayésienne	11
1.3 Entre école classique et école Bayésienne	21
2 Modèles de Régression	22
2.1 Présentation du modèle de Régression	22
2.2 Régression linéaire	23
2.3 Régression logistique	27
2.4 Aperçu sur les modèles linéaires généralisés	29
2.5 Régression linéaire Bayésienne	31
2.6 Régression logistique Bayésienne	34
3 Estimation Bayésienne d'un modèle de régression linéaire à deux phases	35
Introduction	35
3.1 Modèle de régression linéaire à deux phases (TPLR)	36
3.2 Estimation Bayésienne des paramètres	36
3.3 Estimation Bayésienne sous les fonctions de coûts symétriques	43
3.4 Estimation Bayésienne sous les fonctions de coûts asymétriques	45
Conclusion	53
Bibliographie	54

Introduction

La statistique est une discipline mathématique qui consiste à recueillir, traiter et interpréter des données issues d'expériences et d'études de un ou plusieurs phénomènes. Son objectif principal est de mener, grâce à l'observation d'un phénomène aléatoire, une inférence sur la distribution de probabilité qui est à l'origine de ce phénomène.

L'inférence se divise en deux parties, nous avons d'une part l'estimation et d'autre part les tests. Nous nous intéressons en particulier à la partie estimation.

La théorie de l'estimation dont l'origine vient en principe de la méthode des moindres carrés qui fut développée par Gauss a donné deux arguments en faveur de cette méthode. Le premier argument est que les valeurs éventuelles des paramètres inconnus sont celles qui donnent la plus grande probabilité possible des résultats observés. Le second que Gauss lui-même préférait est de trouver des estimateurs qui rendent les erreurs quadratiques moyennes les plus petites possibles. Fisher a développé cette méthode en étudiant des problèmes d'estimations plus larges et plus généraux, ce développement a mené à la méthode du maximum de vraisemblance.

L'approche Bayésienne est également appliquée à la théorie de l'estimation ce qui permet de déduire la probabilité d'un événement à partir de celles d'autres événements préalablement évalués. Le paradigme Bayésien 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 décisions dans le contexte de l'incertitude. Dans l'analyse des données statistiques, on s'intéresse à l'estimation de plusieurs paramètres inconnus dans des modèles comme les modèles de régression. La régression s'intéresse à l'évolution en moyenne d'une variable dépendante à expliquer conditionnellement à des variables explicatives.

Habituellement, les paramètres inconnus ou les coefficients de régression du modèle sont estimés par la méthode des moindres carrés ou par la méthode du maximum de vraisemblance, néanmoins, on pourrait désirer vouloir plus d'informations à propos de la distribution conditionnelle de la variable à expliquer.

Un modèle statistique peut être également perturbé, ainsi on obtient des modèles à plusieurs phases, en d'autres termes, plusieurs points de perturbations sont observés et causent un changement des paramètres. Ces points de perturbations sont appelés points de rupture. Dans le cas présent, nous nous intéressons aux modèles de régression à deux-phases où un seul point de rupture est remarqué.

L'objectif du présent travail est d'étudier un modèle de régression linéaire à deux-phases, travail basé sur l'article de **Pandya et al.** (2011), une lecture critique de cet article est faite. Le premier chapitre donne des généralités et les notions fondamentales pour l'approche classique et l'approche Bayésienne nécessaires pour l'estimation. Le deuxième chapitre introduit quelques modèles de régression, on citera le modèle de régression linéaire et le modèle de régression logis-

tique bien qu'il existe beaucoup d'autre types de régression. Le troisième chapitre est la partie principale de ce mémoire. On définira en premier lieu qu'est ce qu'un modèle de régression linéaire à deux phases. Les estimateurs des différents paramètres inconnus de ce modèle, en particulier le point de rupture, seront également obtenus via l'approche Bayésienne. À cette fin, les différentes lois a posteriori sont requises, elles représentent l'actualisation des informations a priori des paramètres du modèle. Une étude sur l'estimation Bayésienne est considérée en utilisant des fonctions de coût, appelées aussi fonctions de pertes. Notre intérêt se portera sur la fonction perte symétrique connue qui est la fonction quadratique et les fonctions pertes asymétriques usuelles comme la fonction coût Linex et la fonction de coût générale entropie.

D'autres travaux ont été réalisés récemment, on cite l'article de **A. Chaturvedi et A. Shrivastava** publié en 2016, dans la revue *Communications in statistics-Theory and methods*, L'article est intitulé : **Bayesian analysis of linear model involving structural changes in either regression parameters or disturbance precision**, où ils étudient un modèle de régression linéaire avec des ruptures au niveau des paramètres de régression ou au niveau de la précision de la perturbation.

Nous terminons par une conclusion où on citera quelques perspectives de recherche.

Chapitre 1

Inférence classique et Bayésienne

Un aspect important de l'inférence statistique consiste à obtenir des estimations fiables des caractéristiques d'une population. L'école classique (fréquentiste) et Bayésienne trouvent dans ce domaine un terrain de désaccord, chacune a ses méthodes et arguments spécifiques qui défendent leurs cause.

Les statisticiens qui utilisent l'approche fréquentiste supposent que les données sont issues d'un échantillon de variables aléatoires, DeGroot (1988) exprime que cette hypothèse est logiquement non atteinte. Plusieurs critiques se sont faites à ce sujet, mais des expériences suggèrent que cette hypothèse est une excellente approximation de la réalité.

Cependant ceci n'est pas la principale cause du désaccord entre les deux écoles, les méthodes classiques tendent à être jugées sur leurs performances théoriques moyennes, par exemple comparaison entre les pertes quadratiques moyennes (*i.e* MSE de l'anglais "mean squart error"), les Bayésiens eux, disent que le "moyen" n'est pas approprié pour comparer et que les observations *i.i.d* sont impossibles dans quelques cas.

Dans cette section, nous allons présenter l'inférence faite par les deux écoles, on exposera les différentes méthodes de recherche des estimateurs ainsi que les critères d'un bon estimateur et puis quelques définitions et notions basiques pour les test qui nous donnera une idée plus au moins générale sur les tests d'hypothèses.

1.1 Inférence classique

On considère (Ω, A, P) un espace probabilisé. Soit X une variable aléatoire de loi de probabilité connue dépendant d'un paramètre $\theta \in \Theta$ inconnu, et soit (x_1, x_2, \dots, x_n) une réalisation d'un n-échantillon aléatoire issu de la variable aléatoire X .

1.1.1 Définitions élémentaires

Définition 1.1.1. .

On appelle **statistique** toute variable aléatoire fonction du n-échantillon

$$T : (\Omega, A) \longrightarrow f(X_1(\omega), X_2(\omega), \dots, X_n(\omega)),$$

indépendante du paramètre θ .

Définition 1.1.2. .

On appelle **estimateur** T de θ toute statistique prenant ses valeurs dans l'ensemble Θ après expérience.

Définition 1.1.3. .

On appelle **biais d'un estimateur** $\hat{\theta}$ de θ la valeur :

$$b_{\theta}(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

Si $b_{\theta}(\hat{\theta}) = 0$ l'estimateur est dit sans biais pour θ .

Définition 1.1.4. .

On définit l'**erreur quadratique moyenne** de $\hat{\theta}$ par rapport à θ par la quantité :

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

La variance d'un estimateur désigne la dispersion autour de $E[\hat{\theta}]$ des valeurs de $\hat{\theta}$ dans l'univers des échantillons possibles.

Propriété 1. .

On a :

$$MSE(\hat{\theta}) = b_{\theta}(\hat{\theta})^2 + var_{\theta}(\hat{\theta})$$

Si $\hat{\theta}$ est un estimateur sans biais pour θ alors on a :

$$MSE(\hat{\theta}) = var_{\theta}(\hat{\theta})$$

Définition 1.1.5. .

On dit qu'un estimateur $\hat{\theta}_1$ **domine** $\hat{\theta}_2$ si $MSE(\hat{\theta}_1) \leq MSE(\hat{\theta}_2)$

Définition 1.1.6. .

On appelle estimateur **admissible** un estimateur tel qu'il n'existe aucun autre le dominant.

1.1.2 Estimateur sans biais à variance minimale

L'un des critères essentiel pour un meilleur estimateur est qu'il soit sans biais et à variance la plus petite possible, on verra dans cette section trois théorèmes fondamentaux pour l'obtention de cette estimateur

Définition 1.1.7. .

On dit qu'un estimateur $\hat{\theta}_*$ est **UMVUE** (de l'anglais "uniformly minimum variance unbiased estimator") (i.e. un estimateur sans biais à variance minimale) s'il est sans biais pour θ et si pour tout autre estimateur sans biais $\hat{\theta}$ on a :

$$var(\hat{\theta}_*) \leq var(\hat{\theta})$$

Définition 1.1.8. .

Une statistique T est dite **exhaustive** (ou suffisante) si la distribution de $X = (X_1, X_2, \dots, X_n)$ sachant T ne dépend pas du paramètre inconnu θ , c'est à dire :

$$P[X = x|T = t] = h(x, t)$$

La statistique exhaustive porte toute l'information sur le paramètre inconnu, plus que la variable X elle-même, c'est la raison pour laquelle l'inférence est basée sur cette statistique.

Définition 1.1.9.

Une statistique exhaustive est dite **complète** pour θ si, pour une fonction donnée g , $E[g(T)] = 0$ alors $g(T) = 0$ avec probabilité 1.

La notion de complétude assure l'existence d'une unique fonction de la statistique exhaustive qui est non biaisée pour θ , en d'autre terme, assure l'unicité de l'estimateur sans biais de θ .

Théorème 1.1.1 (Rao-Blackwell).

On suppose que les variables aléatoires X_1, X_2, \dots, X_n sont i.i.d de loi de probabilité f , et que T , fonction de X_1, X_2, \dots, X_n , est une statistique exhaustive pour θ .

Soit $\hat{\theta}$ un estimateur sans biais pour θ . On définit l'estimateur $\hat{\theta}_*$ par :

$$\hat{\theta}_* = \hat{\theta}_*(T) = E[\hat{\theta}|T]$$

$\hat{\theta}_*$ ainsi défini est un UMVUE de θ .

Théorème 1.1.2 (Lehmann-scheffe).

On suppose que les variables aléatoires X_1, X_2, \dots, X_n sont i.i.d de loi de probabilité f , et que T , fonction de X_1, X_2, \dots, X_n , est une statistique exhaustive et complète pour θ .

Soit $\hat{\theta} = h(T)$ un estimateur sans biais pour θ . Alors pour un θ arbitraire dans Θ , $\hat{\theta}$ a la variance la plus minimale possible, c'est donc un UMVUE de θ (estimateur sans biais à variance minimale).

Ce théorème donne un bon algorithme pour trouver le meilleur estimateur :

- Identifier une statistique exhaustive T .
- Vérifier que T est complète.
- Trouver un estimateur sans biais qui est une fonction de T .

Le résultat sera un UMVUE de θ (estimateur sans biais à variance minimale).

Théorème 1.1.3 (Cramér-Rao).

Soit $\hat{\theta}$ un estimateur sans biais pour θ , sous certaines conditions de régularités, on a nécessairement pour tout $\theta \in \Theta$:

$$\text{var}_\theta(\hat{\theta}) \geq \frac{1}{nI(\theta)}$$

où $I(\theta)$ est l'information de Fisher et vaut :

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \right]$$

Remarque : Si le domaine de définition de la va X ne dépend pas de θ alors :

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \right] = -E \left[\frac{\partial^2}{\partial \theta^2} \ln f(X, \theta) \right].$$

L'inégalité présentée dans le théorème de Cramér-Rao signifie que la variance de l'estimateur ne peut pas descendre au-dessous d'un certain seuil (appelé borne de Cramer-Rao) qui est fonction de θ

Si un estimateur sans biais atteint cette borne, alors cet estimateur est un UMVUE pour θ (estimateur sans biais à variance minimale).

1.1.3 Méthodes d'estimation du paramètre θ

Estimation par la méthode du maximum de vraisemblance (MV)

Définition 1.1.10.

Soit un n -échantillon (X_1, X_2, \dots, X_n) issu d'une variable aléatoire X de loi de probabilité $f(x, \theta), \theta \in \Theta$. On appelle fonction de **vraisemblance** de θ pour une réalisation de l'échantillon (x_1, x_2, \dots, x_n) la fonction :

$$\mathcal{L}(\theta, x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta)$$

Définition 1.1.11.

La statistique

$$T \mapsto \operatorname{argmax}_{\theta \in \Theta} \left(\prod_{i=1}^n f(x_i, \theta) \right)$$

s'appelle **estimateur du maximum de vraisemblance** noté $\hat{\theta}_{MV}$.

Souvent, la fonction de vraisemblance a une écriture complexe, le logarithme lui est appliqué pour la simplifier, on obtient ainsi la fonction dite *log-vraisemblance* définie par :

$$\ell = \ell(\theta, x_1, x_2, \dots, x_n) = \ln \mathcal{L}(\theta, x_1, x_2, \dots, x_n)$$

L'estimateur du MV est obtenu en résolvant le système suivant :

$$\begin{cases} \frac{\partial \ell}{\partial \theta} = 0 \\ \frac{\partial^2 \ell}{\partial \theta^2} < 0 \end{cases}$$

Estimation par la méthode des moments

La méthode des moments consiste à trouver une fonction m , inversible et continue, et une fonction h telle que $E[|h(X)|] < \infty$ et $m(\theta) = E_\theta[h(X_i)]$ pour $\theta \in \Theta$.

Définition 1.1.12.

L'estimateur des moments de θ , noté $\hat{\theta}_M$ est défini par :

$$\hat{\theta}_M = m^{-1} \left(\frac{1}{n} \sum_{i=1}^n h(X_i) \right)$$

Souvent, on prend $h(X_i) = x_i^r$, m correspond au moment d'ordre r de la variable X_i dont x_i est une réalisation. Autrement dit, cette méthode consiste à égaliser les moments théoriques aux moments empiriques de l'échantillon.

1.1.4 Tests d'hypothèses

Les tests d'hypothèses constituent un autre aspect important d'inférence statistique permettant de contrôler à partir d'un ou plusieurs échantillons la validité d'une hypothèse relative à une ou plusieurs populations.

Les méthodes d'inférences statistiques permettent de déterminer si les différences entre les valeurs estimées et les valeurs considérées dans les hypothèses sont importantes ce qui signifie que les échantillons proviennent de populations différentes.

Il existe deux classes de tests :

- Les *test paramétriques* exigeant la normalité des distributions des échantillons ou une approximation gaussienne pour les échantillons de grandes tailles.
- Les *tests non-paramétriques* dont le modèle ne précise pas les conditions que doivent suivre les paramètres de la population.

On distingue plusieurs types de tests, il y'a :

- Le *test de conformité* qui consiste à confronter un paramètre calculé sur un échantillon à une valeur pré-établie, les plus connus de ce type sont les tests portant sur la moyenne, la variance ou sur la proportion.
- Le *test d'ajustement* qui consiste à vérifier la compatibilité des données avec une distribution choisie a priori.
- Le *test de comparaison*, il consiste à vérifier que K échantillons proviennent de la même population ($K \geq 2$).
- Le *test d'indépendance*, il consiste à prouver l'existence d'une liaison entre deux variables.

Principe des tests

Le principe des tests d'hypothèses est de poser une hypothèse de travail et de prédire les conséquences de cette dernière pour la population ou l'échantillon, puis on compare les valeurs estimées avec les observations, de ce fait, on accepte ou on rejette l'hypothèse posée.

Concepts utiles aux tests d'hypothèses

Définition 1.1.13. .

On appelle hypothèse statistique, un énoncé concernant les caractéristiques d'une population (valeur des paramètres, distributions des observations,...).

Définition 1.1.14. .

On appelle test d'hypothèses, la démarche qui a pour but de donner une règle de décision qui permet de faire un choix entre deux hypothèses.

Définition 1.1.15. .

L'hypothèse nulle est l'hypothèse que l'on désire contrôler, notée H_0 , elle consiste à dire qu'il n'y a pas de différence entre les paramètres comparés ou que la différence existante est négligeable.

Définition 1.1.16. .

L'hypothèse alternative est toute autre hypothèse qui diffère de H_0 , notée H_1 , elle équivaut à dire que " H_0 est fausse".

L'hypothèse nulle est celle soumise au test dont toutes les démarches s'effectuent en considérant l'hypothèse " H_0 est vraie".

Définition 1.1.17. .

On définit le seuil de signification du test, noté α , comme étant le risque consenti à l'avance de rejeter à tort l'hypothèse nulle alors qu'elle est vraie, il est donné en probabilité comme suit :

$$\alpha = P[\text{rejeter } H_0 | H_0 \text{ vraie}]$$

À ce seuil, on associe une zone dite région critique

Définition 1.1.18. .

On appelle région critique, notée R_c , d'air α et de probabilité α , sous l'hypothèse H_0 , l'ensemble des valeurs de la variable aléatoire, une statistique notée S , de décision qui conduisent à écarter H_0 au profit de l'hypothèse H_1 .

La région complémentaire de R_c , notée $\overline{R_c}$, d'air $(1 - \alpha)$ et de probabilité $(1 - \alpha)$ représente la région d'acceptation de l'hypothèse H_0 .

$$P[S \in R_c] = \alpha$$

La règle de décision peut se formuler de la façon suivante :

Si la valeur de la statistique considérée appartient à la région d'acceptation $\overline{R_c}$, on favorise l'hypothèse nulle H_0 , si elle appartient à la région critique R_c , on favorise l'hypothèse alternative H_1 .

Définition 1.1.19. .

On appelle risque d'erreur de première espèce la probabilité de rejeter H_0 et d'accepter H_1 alors que H_0 est vraie.

Définition 1.1.20. .

On appelle risque d'erreur de seconde espèce, noté β , la probabilité de rejeter H_1 et d'accepter H_0 alors que H_1 est vraie.

(Pour quantifier β , il faut connaître la distribution de la statistique sous l'hypothèse H_1).

Définition 1.1.21. .

On appelle puissance d'un test la probabilité de rejeter H_0 et d'accepter H_1 alors que H_1 est vraie, elle est égale à $1 - \beta$

1.2 Inférence Bayésienne

L'analyse Bayésienne, créée par Pierre-Simon de Laplace et Thomas Bayes (1774), a pour première étape l'étude d'une situation et d'identifier une incertitude portée sur un paramètre inconnu θ puis de quantifier celle-ci à travers une répartition probabiliste en utilisant les éléments du calcul de probabilités.

L'incertitude sur θ est modélisée sous la forme d'une distribution, dite *a priori*, qui apporte une information sur θ pris comme étant une variable aléatoire contrairement à l'analyse fréquentiste qui la considère comme une constante.

Cette *a priori* est actualisée en extrayant de l'information contenu dans les observations de X , pour obtenir une autre distribution, maitresse du Bayésien, dite distribution *a posteriori*.

1.2.1 Théorème de Bayes

Soit $(\Omega, \mathcal{A}, \mathcal{P}_\theta)$ un espace probabilisé et A, B deux événements tel que $P[B] \neq 0$. La théorie de Bayes donne l'expression définie par :

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

qui donne la probabilité conditionnelle de A sachant la réalisation de B .

Cette formule peut être réécrite sous la forme :

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

exprimant une relation entre les probabilités conditionnelles $P[A|B]$ et $P[B|A]$.

Le théorème de Bayes est un principe d'actualisation, car il décrit la mise à jour de l'observation de A , de $P[A]$ vers $P[A|B]$ une fois B observé.

Bayes donne une version de ce résultat pour le cas continu, à savoir pour deux variables aléatoires X et Y , de réalisations x et y respectivement, de distribution conditionnelle $f(x|y)$ et de marginale pour Y , $g(y)$.

La distribution conditionnelle de Y sachant X est donc donnée par :

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}$$

Les détails de cette formule seront donnés correctement dans ce qui suit.

1.2.2 Présentation du modèle Bayésien

Définition 1.2.1. .

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

Définissons les espaces intervenant dans un modèle Bayésien :

* **Espace des observations**

Noté \mathcal{X} , il représente l'ensemble des résultats suite à une étude d'un phénomène.

* **Espace des actions**

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

*** Espace des états de la nature**

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

*** Espace des règles de décisions**

Noté \mathcal{D} , il représente l'ensemble des règles de décisions qu'on définit comme une application de \mathcal{X} dans \mathcal{A} notée par δ :

$$\begin{aligned}\delta : \mathcal{X} &\longrightarrow \mathcal{A} \\ x_i &\longrightarrow \delta(x_i) = a_i\end{aligned}$$

a_i étant la *i*^{ème} action.

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

*** La vraisemblance**

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

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

*** Loi a priori**

Notée par $\pi(\theta)$, désigne la distribution du paramètre inconnu θ , elle porte l'information sur ce dernier. L'appellation *a priori* exprime le fait qu'elle a été établie préalablement à l'observation des données x .

*** Loi jointe du couple (θ, x)**

Généralement notée par $f(\theta, x)$, sa formule est donnée par :

$$f(\theta, x) = f(x|\theta)\pi(\theta).$$

*** Loi marginale de x**

On la note par $f(x)$, elle est calculée comme suit :

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

*** Loi a posteriori :**

C'est la loi conditionnelle de θ sachant X , notée par :

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

1.2.3 Les bases de la décision et approche Bayésienne

Définition 1.2.2.

On appelle fonction de coût (ou fonction de perte) une fonction mesurable de (Θ, \mathcal{D}) à valeur dans \mathbb{R}^+

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

La fonction de coût est censée évaluer l'erreur associée à la décision δ quand le paramètre prend la valeur θ . Cependant, dans la pratique, cette fonction est souvent difficile à déterminer, par exemple la difficulté est présente quand les espaces Θ et \mathcal{D} sont grands, voir de dimensions infinies, il est alors impossible de déterminer chaque action $a = \delta(x)$ pour chaque valeur de θ .

Définition 1.2.3.

Soit $\mathcal{L}(\theta, \delta)$ une fonction de coût et soit π une loi a priori pour θ .
On appelle estimateur de bayes, la règle de décision $\delta^\pi(x)$ qui vérifie :

$$\delta^\pi(x) = \arg \min_{\delta \in \mathcal{D}} \mathbb{E}^\pi [\mathcal{L}(\theta, \delta) | x].$$

Cet estimateur sera déterminé analytiquement ou numériquement selon la complexité de la fonction de coût \mathcal{L} et de la loi a posteriori $\pi(\theta|x)$.

L'intérêt de l'approche décisionnelle est de trouver la meilleure évaluation possible pour la fonction de θ , ce qui peut entraîner au mieux un coût nul lorsque ce paramètre est connu. Dans le cas contraire, la fonction de coût perdrait sa continuité, ce qui pourrait empêcher le choix d'une procédure de décision.

Pour obtenir un critère de comparaison à partir d'une fonction de coût, on considère la moyenne de celui-ci, c'est ce qu'on appelle le *risque fréquentiste*.

Définition 1.2.4.

On appelle *risque fréquentiste* le coût moyen d'une règle de décision noté $R(\theta, \delta(x))$ et défini par :

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

Ce risque a été proposé par l'approche fréquentiste, qui repose sur cette notion pour comparer les estimateurs, sauf que cette approche restreint l'espace des estimateurs en préférant la classe des estimateurs sans biais.

L'école Bayésienne définit un autre risque qui intègre l'espace des paramètres pour remédier aux difficultés présentes dans la première approche.

Définition 1.2.5.

On appelle *risque a posteriori* la moyenne du coût par rapport à la loi a posteriori, communément noté par $\rho(\pi, \delta(x))$, il est défini par :

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

Il est aussi possible de définir le risque intégré en se donnant une distribution a priori π .

Définition 1.2.6.

On définit le *risque intégré* comme étant le *risque fréquentiste moyenné* sur les valeurs de θ selon sa distribution a priori, il est noté $r(\pi, \delta(x))$;

$$r(\pi, \delta(x)) = \mathbb{E}^\pi [R(\theta, \delta(x))] = \int_{\Theta} \int_{\mathcal{X}} \mathcal{L}(\theta, \delta(x)) f(x|\theta) \pi(\theta) dx d\theta.$$

L'intérêt de ce risque est qu'il associe un nombre réel à chaque estimateur, ce qui permet une comparaison directe entre ces estimateurs. La valeur $r(\pi, \delta^\pi) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta) < \infty$ est appelé *risque Bayésien*

Une autre formulation de l'estimateur Bayésien est proposé dans ce qui suit

Définition 1.2.7.

Un estimateur de Bayes associé à une distribution a priori π et une fonction coût \mathcal{L} est un estimateur δ^π vérifiant, pour chaque $x \in \mathcal{X}$,

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

Définition 1.2.8.

On appelle risque minimax associé à la fonction coût \mathcal{L} , la valeur \bar{R} :

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

Le critère de minimaxité vise à minimiser le coût moyen dans le cas le moins favorable, c'est une sorte d'assurance contre le pire.

Proposition 1.2.1.

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

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

\underline{R} est appelé le risque maximin, et \bar{R} désigne le risque minimax.

Définition 1.2.9.

Un estimateur δ_0 est dit inadmissible s'il existe un estimateur δ_1 tel que pour tout θ :

$$R(\theta, \delta_0) \geq R(\theta, \delta_1).$$

En d'autre terme, δ_0 est admissible s'il n'existe aucun estimateur δ tel que :

$$R(\theta, \delta) \leq R(\theta, \delta_0).$$

Proposition 1.2.2.

Si l'estimateur de Bayes associé à une loi a priori π est unique, il est admissible.

Proposition 1.2.3.

Si δ_0 est admissible et de risque constant, alors δ_0 est l'unique estimateur minimax.

1.2.4 Les fonctions de coût usuelles

La fonction de coût quadratique

Définition 1.2.10.

La fonction de coût quadratique est la fonction définie par :

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

Proposition 1.2.4.

L'estimateur Bayésien $\delta^\pi(x)$ associé à la loi a priori π et au coût quadratique est la moyenne de la loi a posteriori de θ :

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta|x].$$

Une autre variante du coût quadratique est présentée, c'est le *coût quadratique pondéré* défini par :

$$\mathcal{L}(\theta, \delta(x)) = \omega(\theta)(\theta - \delta(x))^2$$

où $\omega(\theta)$ est une fonction positive.

Corollaire 1.2.1. .

L'estimateur Bayésien associé à la loi a priori π et au coût quadratique pondéré est :

$$\delta^\pi(x) = \frac{\mathbb{E}^\pi[\omega(\theta)\theta|x]}{\mathbb{E}^\pi[\omega(\theta)|x]}$$

La fonction de coût absolu

Définition 1.2.11. .

La fonction de coût absolu est définie comme suit :

$$\mathcal{L}(\theta, \delta(x)) = |\theta - \delta(x)|$$

Proposition 1.2.5. .

L'estimateur Bayésien associé à la loi a priori π et à la fonction de coût absolu est la médiane a posteriori, c'est-à-dire la médiane de $\pi(\theta|x)$.

Dans un cas plus général, cette fonction est pondérée, ce qui donne la *fonction de coût linéaire par morceaux* :

$$\mathcal{L}(\theta, \delta(x)) = \begin{cases} k_2(\theta - \delta(x)) & ; \theta > \delta(x) \\ k_1(\delta(x) - \theta) & ; \theta < \delta(x) \end{cases}$$

Proposition 1.2.6. .

L'estimateur Bayésien associé à la loi à priori π et au coût linéaire par morceaux est le fractile d'ordre $\frac{k_2}{k_1+k_2}$ de la loi a posteriori $\pi(\theta|x)$.

La fonction de coût 0-1

Ce coût est utilisé pour les tests d'hypothèses, c'est un exemple de coût non-quantitatif.

Définition 1.2.12. *Soit la fonction de coût \mathcal{L} définie par :*

$$\mathcal{L}(\theta, \delta(x)) = \begin{cases} 0 & ; (\theta - \delta(x)) < \varepsilon \\ 1 & ; (\theta - \delta(x)) > \varepsilon \end{cases}$$

où ε est très petit.

Proposition 1.2.7. .

L'estimateur Bayésien associé à la loi a priori π et à la fonction de coût 0-1 est le mode de la distribution a posteriori $\pi(\theta|x)$

La fonction de coût Linex

Définition 1.2.13. On définit la fonction de coût Linex comme suit :

$$\mathcal{L}(\theta, \delta(x)) = \exp \{c(\delta(x) - \theta)\} - c(\delta(x) - \theta) - 1$$

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

Proposition 1.2.8. L'estimateur Bayésien associé à la loi a priori π et à la fonction de coût Linex est :

$$\delta^\pi(x) = -\frac{1}{c} \ln [\mathbb{E}_\theta(e^{-c\theta})]$$

où \mathbb{E}_θ est l'espérance a posteriori.

1.2.5 Le choix des lois a priori

Dans l'approche Bayésienne, il est nécessaire et crucial de déterminer la loi a priori $\pi(\theta)$ du paramètre inconnu θ .

Dans la pratique, l'information a priori est généralement insuffisante, ce qui rend le choix de la loi a priori difficile et non précis. Faute de ressources ou de temps, le chercheur ne peut construire un a priori exacte, il doit alors compter sur l'information partielle portée sur les données du modèle.

Des critiques sur cette approche Bayésienne s'élèvent, qui ont tout de même une certaine validité, par exemple sur le fait qu'il n'y a pas de façon unique de choisir une loi a priori du fait que plusieurs lois de probabilités peuvent être compatibles avec ces données sur le paramètre.

Approche partiellement informative

Maximum d'entropie

Supposons que certaines caractéristiques de la loi a priori π peuvent s'écrire comme des espérances : $\mathbb{E}^\pi[g_k(\theta)] = \omega_k$, $k=1, \dots, n$, pour une fonction donnée g_k .

La méthode d'entropie maximale permet de choisir un a priori qui satisfait certaines contraintes, cette méthode a été développée par Jaynes (1968, 1983), ce concept est mieux adapté pour des distributions discrètes, c'est-à-dire où $\theta \in \{1, 2, \dots, n\}$.

Définition 1.2.14.

Soit θ une variable aléatoire discrète et soit π sa loi de probabilités telle que $\pi(\theta) = (\pi_1, \pi_2, \dots, \pi_n)$. L'entropie de π , notée $Ent(\pi)$ est définie comme suit :

$$Ent(\pi) = -\sum_{i=1}^n \pi_i \ln \pi_i.$$

Par convention, si $\pi_i = 0$, la quantité $\pi_i \log \pi_i = 0$.

L'entropie a une relation directe avec la théorie de l'information, en un sens, elle mesure l'incertitude présente dans la distribution a priori. Une entropie petite s'interprète comme une loi concentrée et informative.

La maximisation de l'entropie sous des contraintes, permet de chercher la loi la moins informative, le principe de la méthode est de calculer $\arg \min Ent(\pi)$, avec contrainte $\mathbb{E}^\pi[g_k(\theta)] = \omega_k$.

La solution de ceci donne :

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

où les λ_k sont les multiplicateurs de Lagrange.

Un exemple bien connu est celui de la distribution de Dirac, c'est la loi la plus désordonnée et la plus incertaine, où $\pi_i = \frac{1}{n}$ pour $i = 1, \dots, n$, son entropie est égale à :

$$Ent(\pi) = - \sum_{k=1}^n \frac{1}{n} \ln \frac{1}{n} = \ln n.$$

Notons comme remarque que pour toute loi a priori π on a :

$$Ent(\pi) \leq \ln n.$$

Dans le cas où θ est continu, il devient plus compliqué d'employer la méthode du maximum d'entropie. Une première difficulté étant qu'il n'y a pas de définition formelle de l'entropie, Jaynes (1968) propose alors la formule suivante :

$$Ent(\pi) = -\mathbb{E} \left[\ln \frac{\pi(\theta)}{\pi_0(\theta)} \right] = - \int_{\Theta} \ln \frac{\pi(\theta)}{\pi_0(\theta)} \pi(\theta) d\theta$$

avec $\pi_0(\theta)$ est une loi non-informative invariante pour ce problème appelée *mesure de référence*. L'autre difficulté présente est que la détermination de cette mesure rend la définition donnée par Jaynes ambiguë.

La loi qui maximise cette entropie, solution de ce problème, est donnée par :

$$\pi^* = \frac{\pi_0(\theta) \exp \{ \sum_{k=1}^n \lambda_k g_k(\theta) \}}{\int_{\Theta} \pi_0(\theta) \exp \{ \sum_{k=1}^n \lambda_k g_k(\theta) \}}$$

Les lois a priori conjuguées

En inférence Bayésienne, le terme "conjuguée" décrit un lien particulier qui relie le modèle que suit les données et les distributions a priori des paramètres inconnus. L'approche a priori conjuguée, introduite par Raiffa et Schlaifer (1961), peut être considérée comme un point de départ pour l'élaboration de distributions a priori fondées sur des informations a priori limitées.

Définition 1.2.15. .

Une famille F de distributions de probabilités sur Θ est dite **conjuguée** par rapport à la vraisemblance $f(x|\theta)$ si pour tout $\pi \in F$, la distribution a posteriori $\pi(\theta|x)$ appartient également à F .

Les lois a priori conjuguées sont généralement associées à un type particulier de lois d'échantillonnage qui permet toujours leur obtention, ces lois constituent ce qu'on appelle les *familles exponentielles*.

Définition 1.2.16. .

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

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

Où C et h sont des fonctions respectivement de \mathcal{X} et Θ dans \mathbb{R}^+ , et R et T des fonctions de Θ et \mathcal{X} dans \mathbb{R}^k (T est une statistique).

Dans le cas particulier où $\Theta \subset \mathbb{R}^k$, $\mathcal{X} \subset \mathbb{R}^k$ et

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

la famille est dite naturelle.

Proposition 1.2.9. .

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

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

Où $K(\mu, \lambda)$ est une constante de normalisation. La loi a posteriori est de la forme :

$$\pi(\theta|\mu + x, \lambda + 1).$$

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

Exemples de lois conjuguées .

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

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\varrho(\sigma^2\mu + \tau^2x), \varrho^2\sigma\tau^2)$
Poisson $\mathcal{P}(\theta)$	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 $\mathcal{B}(n, \theta)$	Bêta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + x, \beta + n - x)$
Binomiale négative $\mathcal{N}eg(m, \theta)$	Bêta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + m, \beta + x)$
Normale $\mathcal{N}(\mu, \frac{1}{\theta})$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + 0.5, \beta + (\mu - x)^2/2)$

Approche non-informative

L'analyse Bayésienne est aussi appliquée même quand nous disposons d'aucune information a priori, ce qui ajoute un nouveau point dans les critiques contre cette approche, mais ceci fait intervenir un type particulier de lois a priori construites à partir de la distribution d'échantillonnage, seule information disponible. Ces lois sont dites *non-informatives*.

Définition 1.2.17. .

Une loi non-informative est une loi qui ne porte aucune information sur les paramètres à estimer, elle ne donne pas d'avantage de poids à une telle ou telle valeur du paramètre.

Définition 1.2.18. .

Une loi a priori $\pi(\theta)$ est dite impropre si elle est une mesure σ -finie et qui vérifie :

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

les lois a priori de Laplace

Laplace fut le premier à utiliser des techniques non-informatives dans le cas d'absence d'informations, il utilise alors la loi uniforme qui est l'une des lois les plus simples et les plus utilisées parmi les lois a priori pour l'approche non-informative. En effet, ce choix repose sur l'équiprobabilité des valeurs du paramètre θ dans son domaine de définition. Supposons que Θ est un ensemble de taille k alors :

$$\pi(\theta) = \frac{1}{k}.$$

Bien que cette loi peut mener à des lois a priori impropres, il reste possible de travailler avec, du moment que nous n'essayons pas de les interpréter comme des lois de probabilités.

La loi a priori de Jeffreys

Les lois a priori de Jeffreys sont fondées sur l'information de Fisher $I_x(\theta)$ qui mesure la quantité d'information sur θ contenue dans l'observation.

Définition 1.2.19.

Soit θ un paramètre inconnu réel, on appelle loi a priori non-informative de Jeffreys, la loi de la forme :

$$\pi_j(\theta) = c\sqrt{I_x(\theta)}$$

Où c est une constante de normalisation et l'information de Fisher est définie par :

$$I_x(\theta) = \mathbb{E} \left[\left(\frac{\partial \ln f(x|\theta)}{\partial \theta} \right)^2 \right].$$

Exemple 1. Soit $X \sim \text{Exp}(\theta)$ de fonction de densité

$$f(x|\theta) = \theta e^{-\theta x} \mathbb{I}_{[0, +\infty[}(x).$$

L'information de Fisher est égale à :

$$I_x(\theta) = \mathbb{E} \left[\frac{1}{\theta^2} \right] = \frac{1}{\theta^2}$$

La loi non-informative de Jeffreys est :

$$\pi_J(\theta) \propto \sqrt{\frac{1}{\theta^2}} = \frac{1}{\theta^2}$$

qui est une loi a priori impropre qu'on peut normaliser en considérant que $\theta \in [a, b]$, $a > 0$, on détermine alors la constante c telle que

$$c \int_a^b \pi_J(\theta) d\theta = 1$$

On obtien au final la loi a priori de Jeffreys suivante :

$$\pi_J(\theta) = \frac{1}{\ln \left(\frac{a}{b} \right) \theta} = \left(\ln \left(\frac{a}{b} \right) \theta \right)^{-1}.$$

1.2.6 Tests et intervalles de crédibilité

Les tests d'hypothèses classiques opposent deux hypothèses spécifiées, l'hypothèse nulle $H_0 : \theta \in \Theta_0$, contre l'hypothèse alternative $H_1 : \theta \in \Theta_1$, ces tests évaluent la probabilité d'avoir une erreur de type 1 ou de type 2. Ces erreurs représentent la chance qu'un échantillon soit observé, ce qui fait qu'une hypothèse est acceptée ou rejetée.

Intervalle de crédibilité

Une autre approche usuelle en inférence est la donnée d'un intervalle de confiance pour un paramètre inconnu θ . L'analogie Bayésien d'un intervalle de confiance classique s'appelle région ou intervalle de crédibilité défini comme suit

Définition 1.2.20. .

Une région α -crédible de $100(1 - \alpha)\%$ pour θ ($0 < \alpha < 1$) est un sous-ensemble C de Θ tel que

$$1 - \alpha \leq P[\theta \in C|x] = \begin{cases} \int_C \pi(\theta|x) d\theta; & (\text{cas continu}) \\ \sum_{\theta \in C} \pi(\theta|x); & (\text{cas discret}). \end{cases}$$

Il existe une infinité de régions α -crédibles pour θ , pour choisir le bon ensemble, on s'intéresse à la région qui a le volume minimal, le volume V étant défini par :

$$V(C) = \int_C d\nu(\theta),$$

si $\pi(\theta|X)$ est absolument continue par rapport à une mesure ν .

Pour cela, nous allons introduire la notion de région HPD (en anglais "Highest Posterior Density").

Définition 1.2.21. .

Une région HPD α -crédible pour θ ("Highest Posterior Density" soit : densité a posteriori la plus forte), le sous ensemble C_x de Θ de la forme

$$C_x = \{\theta \in \Theta : \pi(\theta|x) \geq k(\alpha)\}$$

où $k(\alpha)$ est la plus grande borne telle que

$$P^\pi[\theta \in C_x|x] \geq 1 - \alpha.$$

Les régions HPD peuvent être calculées numériquement ou approximativement ou calculées par des méthodes de simulations comme par les méthodes de Monte Carlo ou par intégration numérique.

En analyse Bayésienne, décider entre les hypothèses nulle et alternative, respectivement H_0 et H_1 , n'est pas aussi direct et simple, des probabilités a posteriori sont calculées, $\alpha_0 = P[\theta \in \Theta_0|x]$ et $\alpha_1 = P[\theta \in \Theta_1|x]$, puis H_0 ou H_1 est acceptée.

L'avantage de ce concept est que les probabilités α_0 et α_1 sont les vraies probabilités des hypothèses, en considérant les données et les informations a priori disponibles.

Bien que ces probabilités a posteriori soit des mesures essentielles dans les tests Bayésiens, le concept qui suit est aussi d'un intérêt fondamental.

Le facteur Bayésien

Définition 1.2.22. .

On appelle *facteur Bayésien*, le rapport des probabilités a posteriori des hypothèses nulle et alternative ($\frac{\alpha_0}{\alpha_1}$ en anglais "posterior odds ratio"), sur le rapport des probabilités a priori des mêmes hypothèses ($\frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)} = \frac{\pi_0}{\pi_1}$ en anglais "prior odds ratio"), soit :

$$B = \frac{\text{posterior odds ratio}}{\text{prior odds ratio}} = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\alpha_0\pi_1}{\alpha_1\pi_0}$$

Ce rapport permet d'évaluer la modification de la vraisemblance de l'ensemble Θ_0 par rapport à celle de l'ensemble Θ_1 due à l'observation.

Une interprétation claire dans le cas particulier où les hypothèses sont simples, c'est-à-dire quand $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$, le facteur de Bayes s'écrit sous la forme

$$B = \frac{f(x|\theta_0)}{f(x|\theta_1)}$$

et devient simplement le rapport de vraisemblance classique de H_0 et H_1 .

En général, ce facteur dépend de l'information a priori, il peut être perçu comme un rapport de vraisemblance Bayésien. B est alors écrit comme suit :

$$B = \frac{\int_{\Theta_0} f(x|\theta_0)\pi_0(\theta)d\theta}{\int_{\Theta_1} f(x|\theta_1)\pi_1(\theta)d\theta},$$

en rappelant que π_0 et π_1 sont les densités a priori de θ_0 et θ_1 respectivement.

1.3 Entre école classique et école Bayésienne

Il y a en effet deux conceptions des probabilités, toutes deux présentes dès les débuts de la mathématisation du probable. On parle ainsi, d'une part des probabilités fréquentistes, il s'agit de probabilités dites objectives et a posteriori, visant à dégager les lois stochastiques de processus aléatoires tendanciels dans des statistiques de fréquence à long terme, et d'autre part des probabilités Bayésiennes, où il s'agit de probabilité subjective, de degré de certitude a priori. On doit notamment à Condorcet d'avoir présenté les probabilités subjectives comme étant des « raisons de croire », c'est à dire de présenter le calcul des probabilités comme relevant d'une théorie de la connaissance plus que d'une théorie de la nature.

Il y a ainsi d'un côté des probabilités dites « fréquentistes » qui s'intéressent à l'occurrence d'un événement parmi un nombre total et significatif d'observations, et de l'autre les probabilités bayésiennes qui sont des mesures du degré de connaissances subjectives qui représentent la traduction chiffrée d'un état de connaissance.

Chapitre 2

Modèles de Régression

L'origine du mot régression revient à Sir Francis Galton, en 1885, qui travaillait sur l'hérédité (variation de la taille d'un homme en fonction de celle de son père), les résultats obtenus l'ont conduit à sa théorie dite théorie de "régression toward mediocrity".

La régression est l'un des outils mathématiques les plus utilisés pour l'analyse des données, elle fournit des méthodes simples pour l'étude des relations entre variables.

L'approche standard de la régression utilise un échantillon de données pour le calcul et l'estimation des paramètres inconnus. Un modèle explicatif est un modèle exprimant une variable Y , appelée variable à *expliquer* (ou réponse), sous forme d'une fonction à une ou plusieurs variables dites variables *explicatives* notées X_i .

Dans un modèle, on cherche à déterminer la variation de l'espérance mathématique de Y en fonction des X_i , autrement dit, on étudie comment Y évolue en moyenne en fonction des variables explicatives.

Il existe plusieurs types de régression, paramétriques et non-paramétriques, chaque type ayant son importance et ses conditions d'application. Dans ce chapitre nous allons présenter les deux types de modèles paramétriques les plus communément utilisés : la régression linéaire et la régression logistique.

2.1 Présentation du modèle de Régression

Pour simplifier, on utilisera uniquement une seule variable explicative notée X , cette variable sera quantitative à valeurs dans un intervalle I de \mathbb{R} ($X(\omega) = x \in I \subset \mathbb{R}$).

Par hypothèse, une famille de variables aléatoires (v.a) distinctes $\{Y(x)|x \in I\}$ est associée aux différentes valeurs de X dans I .

On admettra que l'espérance mathématique $E[Y(x)]$ existe pour tout $x \in I$ (x étant une réalisation de la variable aléatoire X), c'est une fonction à rechercher qui met en évidence l'évolution moyenne de l'entité Y à expliquer en fonction des valeurs de X , elle est appelée **fonction de régression**.

Le modèle simple de la régression s'écrit généralement sous la forme :

$$Y(x) = g(x) + \varepsilon(x),$$

où : $E[Y(x)] = g(x)$ et $\varepsilon(x)$ sont des variable aléatoire d'espérances nulles et qui sont de même loi quelque soit x , ceci permet de simplifier l'écriture et d'avoir : $Y(x) = g(x) + \varepsilon$.

2.2 Régression linéaire

2.2.1 Le modèle

La fonction de régression est supposée être sous la forme :

$$g(x) = E[Y|X = x] = \beta_0 + \beta_1 x$$

pour $x \in I$. On suppose que la variance de la loi conditionnelle $Y|X = x$ (Y sachant $X = x$) ne dépend pas de x et est égale à σ^2 , nous faisons l'hypothèse que cette loi est gaussienne quelque soit x .

La linéarité du modèle se fait par rapport aux paramètres β_0 et β_1 , de ce fait, des transformées de X comme $\ln X$, \sqrt{X} , X^2 , ...etc, peuvent être relayer à X , éventuellement pour avoir la linéarité du modèle qui contient donc trois paramètres inconnus β_0 , β_1 et σ^2 à estimer.

L'estimation se fait en considérant une série d'observations Y_1, Y_2, \dots, Y_n situées aux niveaux x_1, x_2, \dots, x_n de la variable explicative fixé. Ainsi pour tout $i = 1, \dots, n$, Y_i est de loi normale $N(g(x), \sigma^2)$.

La notation $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ sera adoptée pour $i = 1, \dots, n$, en notant que ε_i sont des variables aléatoires *i.i.d* de loi normale $N(0, \sigma^2)$.

2.2.2 Estimation des paramètres du modèle de régression

Les estimateurs du maximum de vraisemblance (MV)

Étant donné que les variables aléatoires Y_i sont gaussiennes pour tout i , on dispose alors de leurs fonction de densité :

$$f_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - g(x_i))^2\right\}$$

La fonction de vraisemblance des trois paramètres associée aux observations $(y_1, x_1), \dots, (y_n, x_n)$ et dépendant des paramètres β_0 , β_1 et σ^2 est comme suit :

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2, (y_1, x_1), \dots, (y_n, x_n)) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - g(x_i))^2\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2\right\}. \end{aligned}$$

La log-vraisemblance est :

$$l = \ln \{L(\beta_0, \beta_1, \sigma^2, (y_1, x_1), \dots, (y_n, x_n))\} = \frac{-n}{2}(\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2.$$

Le système des équations de vraisemblance est alors construit en annulant les dérivées par rapport aux paramètres inconnus :

$$\begin{cases} \frac{\partial l}{\partial \beta_0} = 0, \\ \frac{\partial l}{\partial \beta_1} = 0, \\ \frac{\partial l}{\partial \sigma^2} = 0. \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) = 0, \\ \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0, \\ \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 = 0. \end{cases}$$

On obtient de la première équation :

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}.$$

En remplaçant dans la deuxième équation on obtient :

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i}$$

Or on sait que :

$$\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i = \sum_{i=1}^n (x_i - \bar{x})^2$$

et

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

En substituant \bar{y} par \bar{Y} et y_i par Y_i , on obtient :

$$\begin{cases} \widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x} \\ \widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{cases}$$

$\widehat{\beta}_0$ et $\widehat{\beta}_1$ sont les estimateurs du maximum de vraisemblance de β_0 et β_1 .

Remarque : Les estimateurs du maximum de vraisemblance $\widehat{\beta}_0$ et $\widehat{\beta}_1$ sont aussi appelés estimateurs des moindres carrés, du fait que le couple $(\widehat{\beta}_0, \widehat{\beta}_1)$ est la solution qui minimise la somme des carrés des résidus définie par $\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$.

Pour une solution quelconque (β_0^*, β_1^*) , la différence $y_i - (\beta_0^* + \beta_1^* x_i)$ est appelée résidu (souvent noté par ε_i), elle correspond à l'écart entre la valeur observée et la valeur ainsi estimée donné par le modèle.

L'existence des résidus ε_i fait que les points de coordonnées (x_i, y_i) ne sont pas portés par une même droite.

Par le même raisonnement de calcul, l'estimateur du maximum de vraisemblance de σ^2 est trouvé, donné par la formule :

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n [y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)]^2,$$

en rappelant que $\widehat{\beta}_0 + \widehat{\beta}_1 x_i$ est un estimateur sans biais pour $E(Y|X = x)$, l'espérance de la variable aléatoire à expliquer Y , pour la valeur x de la variable explicative.

Propriétés de $\widehat{\beta}_0, \widehat{\beta}_1$ et $\widehat{\sigma}^2$

$\widehat{\beta}_1$ est une fonction linéaire de variable aléatoire normales, ce qui entraîne qu'elle soit aussi normale, son espérance est égale à β_1 , c'est donc un estimateur sans biais pour β_1

la variance de $\widehat{\beta}_1$ est :

$$var(\widehat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Idem pour l'estimateur $\widehat{\beta}_0$ qui est aussi de loi normale, d'espérance β_0 , ce qui en fait un estimateur sans biais pour le paramètre β_0 , il est de variance :

$$\text{var}(\widehat{\beta}_0) = \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}$$

Sachant que la variable aléatoire $Z = \frac{1}{\sigma^2} \sum_{i=1}^n [Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)]^2$ suit une loi de khi-deux à $n - 2$ degrés de liberté, on déduit que $\mathbb{E}[\sigma^2 Z] = (n - 2)\sigma^2$.

L'espérance mathématique de $\widehat{\sigma}^2$ est donc égale à $\frac{n-2}{n}\sigma^2$, ce qui fait de $\widehat{\sigma}^2$ un estimateur biaisé pour σ^2 , c'est la raison pour laquelle un autre estimateur lui est préféré, noté par S^2 , il est donné par la formule :

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

S^2 est un estimateur sans biais pour σ^2 , sa variance est : $\text{var}(S^2) = \frac{2\sigma^4}{n-2}$.

Convergence de $\widehat{\beta}_0, \widehat{\beta}_1$ et $\widehat{\sigma}^2$

Si les données x_i sont sélectionnées de façon à ce que pour tout n , leur variance $\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$ admette une borne inférieure strictement positive indépendante de n , et que la moyenne \bar{x} admette une borne supérieure indépendante de n , alors $\text{var}(\widehat{\beta}_0)$ et $\text{var}(\widehat{\beta}_1)$ tendent vers 0 quand n tend vers l'infini, on dit alors que $\widehat{\beta}_0$ et $\widehat{\beta}_1$ sont convergents en moyenne quadratique. On a également S^2 qui converge en moyenne quadratique vers σ^2 sachant que la variance d'une loi de khi deux à $n - 2$ degrés de liberté est $2(n - 2)$.

2.2.3 Intervalles de confiance

Proposition 2.2.1. .

Lorsque la variance σ^2 est estimée par S^2 qui est un estimateur sans biais, nous avons :

$$i) \frac{\widehat{\beta}_0 - \beta_0}{S} \sim T_{n-2}$$

$$ii) \frac{\widehat{\beta}_1 - \beta_1}{S} \sim T_{n-2}$$

où T_{n-2} est une loi de Student à $n-2$ degrés de liberté.

Ces propriétés nous permettent de donner des intervalles de confiance (IC) des paramètres inconnus. En effet, la valeur ponctuelle d'un estimateur étant insuffisante, c'est pourquoi un intervalle de confiance lui est adjoit.

Proposition 2.2.2. .

i) Un intervalle de confiance du paramètre de régression β_i ($i \in \{0, 1\}$) est donné par :

$$\left[\widehat{\beta}_i - t_{n-2}\left(\frac{1-\alpha}{2}\right)S_{\widehat{\beta}_i}, \widehat{\beta}_i + t_{n-2}\left(\frac{1-\alpha}{2}\right)S_{\widehat{\beta}_i} \right],$$

où $t_{n-2}\left(\frac{1-\alpha}{2}\right)$ est le fractile de niveau $\frac{1-\alpha}{2}$ d'une loi T_{n-2} et $S_{\widehat{\beta}_i}$ l'écart-type de $\widehat{\beta}_i$ avec la variance σ^2 estimée par S^2 .

ii) un intervalle de confiance de la variance σ^2 est donné par :

$$\left[\frac{(n-2)S^2}{C_{n-2}\left(\frac{1-\alpha}{2}\right)}, \frac{(n-2)S^2}{C_{n-2}\left(\frac{\alpha}{2}\right)} \right],$$

où $C_{n-2}\left(\frac{1-\alpha}{2}\right)$ et $C_{n-2}\left(\frac{\alpha}{2}\right)$ sont les fractiles de niveaux $\frac{1-\alpha}{2}$ et $\frac{\alpha}{2}$ respectivement d'une loi du khi-deux à $(n-2)$ degrés de liberté.

2.2.4 Test de signification des paramètres

Ce test permet de déterminer l'intérêt du modèle et la significativité de la variable explicative. Si on pose $H_0 : \beta_1 = 0$ comme hypothèse nulle, alors sous H_0 , $\frac{\widehat{\beta}_1}{S_{\widehat{\beta}_1}}$ suit une loi de Student T_{n-2} à $n-2$ degrés de liberté. H_0 est rejetée au niveau α si :

$$\frac{\widehat{\beta}_1}{S_{\widehat{\beta}_1}} \notin \left[-t_{n-2}\left(\frac{1-\alpha}{2}\right), t_{n-2}\left(\frac{1-\alpha}{2}\right) \right].$$

Notons que si l'hypothèse nulle est acceptée, la loi conditionnelle $Y|X = x$ ne dépend pas de x , les variables Y_i sont alors gaussiennes *i.i.d* de moyenne β_0 et de variance σ^2 .

Ce test est mieux abordé par l'utilisation de la relation de décomposition de la somme des carrés totale, cette relation est valable pour tout type de modèle linéaire.

Par simplification, on pose $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, la relation de décomposition est alors définie par :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\widehat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \widehat{y}_i)^2,$$

où

$\sum_{i=1}^n (y_i - \bar{y})^2 = SST$ est appelé *somme des carrés totale*, il exprime la variabilité des y_i indépendamment du modèle explicatif.

$\sum_{i=1}^n (\widehat{y}_i - \bar{y})^2 = SSR$ est appelé *somme des carrés dû à la régression*, il prend en compte que les valeurs modélisées.

$\sum_{i=1}^n (y_i - \widehat{y}_i)^2 = SSE$ est la *somme des carrés des résidus*.

La formule est réécrite sous la forme : $SST = SSR + SSE$.

Le rapport $R^2 = \frac{SSR}{SST}$ est appelé "*coefficient de détermination*" qui est une mesure de la qualité de la prévision d'une régression linéaire.

2.3 Régression logistique

2.3.1 Le modèle

La régression logistique est l'un des modèles statistiques les plus couramment utilisés en analyse des données. Ce modèle est utilisé de plus en plus souvent dans de larges domaines d'applications notamment en épidémiologie et biomédecine, récemment, des études en sciences sociales et marketing ont été faites.

La régression logistique s'utilise lorsque la variable à expliquer Y est qualitative, plus souvent binaire. Le modèle cherche à estimer la probabilité de succès de cette variable, notée $p(x)$, par la linéarisation des variables explicatives.

L'apparition des logiciels informatiques et langages de programmation statistiques rend cette étude plus facile, plusieurs méthodes sont employées comme la méthodes MCMC d'échantillonnage (Markov chains Monte-Carlo).

La fonction de régression à estimer est alors :

$$E[Y|X = x] = p(x).$$

En pratique, on souhaite déterminer comment la probabilité de "succès" évolue en fonction de X .

L'application du modèle linéaire pose problème et n'est pas approprié dans ce cas vu que la fonction $p(x)$ n'est pas forcément contenue dans l'intervalle $[0, 1]$, des estimations négatives ou supérieures à 1 peuvent être obtenues ce qui est contradictoire avec le fait que $p(x)$ soit une probabilité.

2.3.2 Présentation du modèle

Pour une variable aléatoire binaire Y et une variable explicative X soit :

$$p(x) = P[Y = 1|X = x] = 1 - P[Y = 0|X = x].$$

Le modèle de régression logistique est :

$$p(x) = \frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}$$

Cette probabilité est aussi appelé fonction logistique.

On appelle *fonction logit* la fonction définie par la formule suivante :

$$\begin{aligned} \text{logit}(p(x)) &= \ln\left(\frac{p(x)}{1 - p(x)}\right) \\ &= \beta_0 + \beta_1 x. \end{aligned}$$

On obtient ainsi une relation linéaire avec des paramètres inconnus β_0 et β_1 à estimer.

Le rapport entre la probabilité de succès $p(x)$ et la probabilité d'échec $(1 - p(x))$ est appelé *cote* ou chance, plus souvent l'appellation anglaise "*odds ratio*" est utilisée.

La cote n'est jamais négative, mais n'a pas de borne supérieure, si sa valeur est égale à 1, dans ce cas la variable explicative n'a pas d'effet sur la cote et le coefficient de régression est alors nul.

Lorsque la variable explicative X est continue, l'odds ratio est parfois proche de 1, puisque une différence d'une unité de x , réalisation de X , est insuffisante pour modifier de manière significative les rapports de cotes.

2.3.3 Interprétation des paramètres

Ayant toujours la fonction logit écrite comme définie précédemment, le signe du paramètre β_1 détermine si la fonction logistique (autrement dit la probabilité) $p(x)$ est croissante ou décroissante pour des données x de la variable explicative X ordonnées et croissantes.

Si β_1 tend vers 0, alors la courbe représentative des données tend à s'aplatir horizontalement, et si $\beta_1 = 0$, dans ce cas la variable Y est indépendante de X .

Le paramètre β_0 n'a pas d'intérêt particulier, cependant, en centrant les données autour de 0 (*i.e.* remplacer x par $(x - \bar{x})$).

Remarque : Dans les modèles de régression ordinaires, le centrage des données est pratique pour les cas de modèles complexes contenant des termes quadratiques ou liés entre eux pour réduire les corrélations entre ces derniers intervenant dans l'estimation des paramètres.

2.3.4 Estimation de la fonction $p(x)$

La méthode la plus commune pour estimer les valeurs des paramètres β_0 et β_1 lorsque nous disposons de n observations indépendantes, soit $(y_1, x_1), \dots, (y_n, x_n)$ est la méthode du maximum de vraisemblance qui consiste d'abord à définir la fonction de vraisemblance de Y_1, \dots, Y_n , la loi de probabilité de $Y_i|X_i = x_i$ étant modélisée par une loi de Bernoulli de paramètre $p(x)$ (cette modélisation est due au fait que la variable à expliquer est binaire), on a alors sa fonction de densité pour l'observation i définie par :

$$f(y_i, x_i) = p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}$$

La fonction de vraisemblance associée aux observations $(y_1, x_1), \dots, (y_n, x_n)$ est :

$$L(\beta_0, \beta_1, (y_1, x_1), \dots, (y_n, x_n)) = \prod_{i=1}^n p(x_i)^{y_i}(1 - p(x_i))^{1-y_i}$$

Il s'en suit que la fonction log-vraisemblance est :

$$\begin{aligned} \ell(\beta_0, \beta_1) = \ln L(\beta_0, \beta_1, (y_1, x_1), \dots, (y_n, x_n)) &= \sum_{i=1}^n [y_i \ln(p(x_i)) + (1 - y_i) \ln(1 - p(x_i))] \\ &= \sum_{i=1}^n \ln(1 - p(x_i)) + \sum_{i=1}^n y_i \ln \left(\frac{p(x_i)}{1 - p(x_i)} \right). \end{aligned}$$

La deuxième étape de la méthode est la résolution du système d'équation de vraisemblance suivant :

$$\begin{cases} \frac{\partial \ell}{\partial \beta_0} = \sum_{i=1}^n \{(y_i - p(x_i))\} = 0 \\ \frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^n \{x_i(y_i - p(x_i))\} = 0 \end{cases}$$

autrement écrit :

$$\begin{cases} \sum_{i=1}^n \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} = \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} = \sum_{i=1}^n x_i y_i \end{cases}$$

Sauf que la résolution de ce système n'est pas possible analytiquement, c'est pourquoi des méthodes itératives, telle que la méthode de Newton, permet de résoudre ce problème pour avoir les estimateurs du maximum de vraisemblance $\hat{\beta}_0$ et $\hat{\beta}_1$ de β_0 et β_1 respectivement. L'estimateur de la fonction $p(x)$ est alors écrit comme suit :

$$\hat{p}(x) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x)}$$

2.3.5 Matrice des variances-covariances

Les estimateurs du maximum de vraisemblance n'ayant pas de formules analytiques, on utilise les propriétés asymptotiques pour l'approximation des estimations. La variance de $\hat{\beta}_0$ et de $\hat{\beta}_1$ est obtenue en considérant les secondes dérivées de la fonction log-vraisemblance, on forme ainsi la matrice d'information de Fisher.

$$\mathbb{I} = \begin{pmatrix} \sum_{i=1}^n p(x_i)[1 - p(x_i)] & \sum_{i=1}^n x_i p(x_i)[1 - p(x_i)] \\ \sum_{i=1}^n x_i p(x_i)[1 - p(x_i)] & \sum_{i=1}^n x_i^2 p(x_i)[1 - p(x_i)] \end{pmatrix}.$$

On pose l'inverse de cette matrice égale à $\widehat{var}(\hat{\beta})$ qui est un estimateur convergent de $var(\hat{\beta})$ la matrice de variance-covariance de $\hat{\beta}$ avec $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$.

2.3.6 Intervalle de confiance

Un intervalle de confiance pour $\hat{\beta}_0 + \hat{\beta}_1 x$ est proposé, asymptotiquement, son estimateur du maximum de vraisemblance est gaussien et de variance :

$$V(\hat{\beta}_0 + \hat{\beta}_1 x) = v(\hat{\beta}_0) + x cov(\hat{\beta}_0, \hat{\beta}_1) + x^2 v(\hat{\beta}_1)$$

estimé par :

$$S^2(\hat{\beta}_0 + \hat{\beta}_1 x) = S_{\hat{\beta}_0}^2 + 2x S_{\hat{\beta}_0, \hat{\beta}_1}^2 + x^2 S_{\hat{\beta}_1}^2$$

L'intervalle de confiance est donc

$$\left[\hat{\beta}_0 + \hat{\beta}_1 x - t_{n-2}(1 - \alpha) S_{\hat{\beta}_0}, \hat{\beta}_0 + \hat{\beta}_1 x + t_{n-2}(1 - \alpha) S_{\hat{\beta}_0} \right]$$

Où $t_{n-2}(1 - \alpha)$ est le fractile d'ordre $(1 - \alpha)$ d'une loi de student à $(n - 2)$ degrés de liberté. En appliquant la fonction $g(u) = \frac{e^u}{1 + e^u}$ aux bornes de l'intervalle de confiance, on déduit l'intervalle de confiance de $p(x)$.

2.4 Aperçu sur les modèles linéaires généralisés

Pour l'analyse des données, plusieurs modèles d'études existent tels les modèles de régression. Nous allons ici présenter brièvement les modèles linéaires généralisés.

Soit Y une variable à expliquer et X une variable explicative, on définit le modèle de régression linéaire comme suit : Une observation y_i de Y peut s'écrire sous la forme

$$y_i = \beta_0 + \beta_1 x_i = \mu_i + \varepsilon_i$$

où μ_i est appelé *prédicteur linéaire* et ε_i des variables aléatoires représentant l'erreur qui est supposée être gaussienne (*i.e.* $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ avec σ^2 la variance).

Une généralisation de ce modèle donne une classe de modèles dits "linéaires généralisés", cette généralisation peut se faire de deux manières, la première est de supposer que les ε_i sont de distribution différente de la loi normale, appartenant à la famille exponentielle, autrement dit les ε_i seront de loi binomiale, de loi de Poisson, de loi Gamma ou de loi normale inverse.

La seconde manière est que la fonction principale μ_i n'est pas forcément le prédicteur linéaire mais une fonction de celui-ci, cette fonction que l'on notera par h doit être monotone, dérivable et qui sera $h(\mu_i) = \beta_0 + \beta_1 x_i$.

$h(\mu)$ désigne le lien reliant μ au prédicteur linéaire, cette fonction est dite "fonction de lien".

De ces généralisations, les modèles linéaires généralisés qui en résultent peuvent être utilisés dans des situations et phénomènes où la régression linéaire simple n'est plus appropriée, les paramètres inconnus de ces modèles sont estimés et calculés par plusieurs méthodes dont le maximum de vraisemblance.

Les modèles linéaires généralisés ont été introduits par Nelder et Wedderburn (1972), une version plus approfondie a été développée par McCullagh et Nelder (1989).

La régression logistique vue dans la section précédente est un exemple de modèle linéaire généralisé, la distribution des variables ε_i étant binomiale induit à ce que la variable à expliquer Y soit aussi binomiale, en d'autres termes, le modèle de régression logistique peut être présenté tel que :

$\mu_i = p(x_i)$ avec $p(x_i)$ la probabilité de succès, et $h(x_i) = h(p(x_i)) = \ln\left(\frac{p(x_i)}{1-p(x_i)}\right)$ est le prédicteur linéaire.

Un autre modèle de cette classe est connu sous le nom de "modèle de régression de Poisson", qui est adapté aux cas où la variable Y est de comptage, par exemple le nombre de personnes hospitalisées atteintes d'une certaine maladie ou selon des critères spécifiques pour une recherche dans le domaine de la santé publique, le nombre de morts ou de dégâts après le passage d'une tornade en tenant compte du mois et de l'année, ou encore de la sévérité de la tornade. Chatterjee et Hadi (2006) présentent un exemple illustratif pour ce modèle qui étudie le nombre de médicaments achetés au marché américain entre 1992 et 2005 pour 16 maladies.

2.5 Régression linéaire Bayésienne

Dans cette partie, nous allons appliquer l'approche Bayésienne aux modèles de régression. L'approche Bayésienne diffère de l'approche statistique classique, en effet elle prend en considération des données et informations obtenues préalablement d'études précédentes, la formule fondamentale nommée formule de Bayes permet de combiner ces informations avec les données observées récentes, elle a été introduite par le révérend Thomas Bayes en 1760, la première application de cette formule aux modèles de régression est dû à Harold Jeffreys en 1939.

2.5.1 Le modèle linéaire et la fonction de vraisemblance

Soit Y une variable dépendante à expliquer, et soit X une variable explicative, le modèle de régression linéaire reliant ces deux variables est représenté comme suit :

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

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 d'une variable aléatoire désignant l'erreur. β_0 et β_1 sont les paramètres de régression.

Notons que la linéarité de la relation est en β_0 , β_1 et ε_i et on suppose que les variables ε_i sont de loi normale de moyenne nulle et de variance σ^2 , on suppose aussi que les x_i sont des données fixées ou bien qu'elles sont des variables aléatoires indépendantes des ε_i dont la fonction de densité ne dépend pas de β_0 , β_1 et de σ^2 .

La variable conditionnelle $y|x_i, \beta_0, \beta_1, \sigma^2$, $y = (y_1, \dots, y_n)$, est de loi normale de moyenne $\mathbb{E}[y|x_i, \beta_0, \beta_1, \sigma^2] = \beta_0 + \beta_1 x_i$ et de variance $\text{var}(y|x_i, \beta_0, \beta_1, \sigma^2) = \sigma^2$.

La fonction de vraisemblance des paramètres inconnus est :

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

Dans ce modèle, on s'intéresse de près à l'estimation Bayésienne de β_1 qui représente la pente. Pour ceci, il nous faudrait obtenir la loi a posteriori de ce paramètre, on distingue alors deux cas : absence d'informations sur les paramètres, ce qui implique l'utilisation de lois a priori non-informatives, et le cas où l'on dispose d'informations issues d'études et d'expérience similaires précédentes, ce qui implique alors l'utilisation de lois a priori informatives.

2.5.2 Loi a priori non-informative

L'une des lois a priori généralement utilisée est la loi a priori dite standard, bien qu'elle soit impropre, la loi a posteriori qui en résulte est une densité de probabilité propre.

Pour cette loi a priori standard, nous disposons d'aucune information sur les paramètres autre que σ est positif, on suppose alors que les paramètres β_0 , β_1 , et $\log \sigma$ sont uniformément distribués et indépendants, ce qui donne la loi a priori jointe de ces paramètres qui suit :

$$\pi(\beta_0, \beta_1, \sigma) \propto \frac{1}{\sigma},$$

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

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

Cette loi a posteriori jointe est la base de l'inférence Bayésienne.
 Considérons alors les équations algébriques suivantes :

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 = (n-2)S^2 + n(\beta_0 - \hat{\beta}_0)^2 + (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n x_i^2 + 2(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) \sum_{i=1}^n x_i.$$

Avec

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ S^2 &= \frac{1}{n-2} \sum_{i=1}^n \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right)^2 \end{aligned}$$

On aura alors :

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

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

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

On peut voir qu'il s'agit d'une loi normale bivariée de moyenne $(\hat{\beta}_0, \hat{\beta}_1)$ et de matrice de variance-covariance notée M

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

(Vu que la variance σ^2 n'est pas toujours connue, ce résultat n'est pas très pratique).

La loi a posteriori jointe de β_0 et de β_1 est obtenue en intégrant par rapport à σ , ce qui donne :

$$\begin{aligned} \pi(\beta_0, \beta_1 | y, x) &= \int_0^\infty \pi(\beta_0, \beta_1, \sigma | y, x) d\sigma \\ &\propto \left[(n-2)S^2 + n(\beta_0 - \hat{\beta}_0)^2 + (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n x_i^2 + 2(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) \sum_{i=1}^n x_i \right]^{-n/2} \end{aligned}$$

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

$$\pi(\beta_0|y, x) \propto \left[(n-2) + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2 \sum_{i=1}^n x_i^2/n} (\beta_0 - \hat{\beta}_0)^2 \right]^{-((n-2)+1)/2}, \quad -\infty < \beta_0 < \infty.$$

et

$$\pi(\beta_1|y, x) \propto \left[(n-2) + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2} (\beta_1 - \hat{\beta}_1)^2 \right]^{-((n-2)+1)/2}, \quad -\infty < \beta_1 < \infty.$$

La table de la loi de Student à $(n-2)$ degrés de libertés nous servira pour l'estimation des paramètres de régression, en effet les variables aléatoires

$$T_1 = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S^2 \sum_{i=1}^n x_i^2/n} \right]^{1/2} (\beta_0 - \hat{\beta}_0)$$

et

$$T_2 = \frac{(\beta_1 - \hat{\beta}_1)}{S / [\sum_{i=1}^n (x_i - \bar{x})^2]^{1/2}}$$

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

La fonction de densité 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 :

$$\pi(\sigma|y, x) \propto \frac{1}{\sigma^{(n-2)+1}} \exp \left\{ -\frac{(n-2)S^2}{2\sigma^2} \right\}, \quad 0 < \sigma < \infty.$$

qui est une gamma inverse $IG \left(\left(\frac{n}{2} - 1 \right), \frac{(n-2)S^2}{2} \right)$.

2.5.3 Loi a priori informative

On suppose qu'on dispose d'informations a priori sur les paramètres inconnus β_0 , β_1 et σ^2 , il est pratique, pour des raisons de calculs, d'exprimer ces informations sous forme de distributions a priori conjuguées, l'usage de ces lois n'est pas basé sur l'information en elle même mais plutôt par commodité, l'information intervient lors de la sélection de la loi a priori conjuguée qui convient, cette information peut être quantifier en spécifiant son espérance.

La manière la plus simple et la plus communément adoptée est d'admettre qu'il existe une indépendance a priori des paramètres β_0 , β_1 et σ^2 , on suppose par suite que β_0 et β_1 sont de loi de probabilités gaussienne et la variance σ^2 suit une gamma inverse (ou de façon équivalente $\frac{1}{\sigma^2}$ suit une loi gamma), cependant, le résultat de ces trois fonctions de densités a priori n'est pas une densité conjuguée puisque la densité a posteriori obtenue n'est pas un produit de trois densités indépendantes de la même famille.

2.6 Régression logistique Bayésienne

En considérant le modèle de régression logistique présenté dans la section précédente

$$p(x) = \frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}$$

l'étude Bayésienne sur les coefficients de régression est faite en supposant que les coefficients, β_0 et β_1 , ne sont plus fixés mais plutôt des variables aléatoires suivant une certaine loi de probabilité. Par conséquent, on laisse dorénavant les coefficients de la régression logistique provenir de lois de probabilités et non simplement être de nature déterministe, c'est ce qu'on appelle la régression logistique Bayésienne.

Ainsi, l'approche classique est modifiée, nous appelons alors densités a priori, les lois de probabilités de β_0 et de β_1 , notées respectivement $\pi(\beta_0)$ et $\pi(\beta_1)$. Ces lois modélisent toutes les informations connues sur β_0 et β_1 indépendamment des observations de la variable à expliquer Y .

Le principe de l'approche Bayésienne est d'actualiser ces informations a priori en introduisant d'autres informations contenues dans les observations, la loi qui résulte de cette actualisation est la loi a posteriori des paramètres β_0 et β_1 , notées respectivement $\pi(\beta_0|Y)$ et $\pi(\beta_1|Y)$.

On considère généralement β_0 et β_1 comme des variables aléatoires gaussiennes, on obtient alors la loi a priori jointe de β_0 et β_1 qui est une loi normale bivariée $\pi(\beta_0, \beta_1)$. En vertu du théorème de Bayes, la loi a posteriori β_0 et β_1 est

$$\pi(\beta_0, \beta_1|Y) = \frac{L(\beta_0, \beta_1, (y_1, x_1), \dots, (y_n, x_n))\pi(\beta_0, \beta_1)}{\int \int L(\beta_0, \beta_1, (y_1, x_1), \dots, (y_n, x_n))\pi(\beta_0, \beta_1)d\beta_0d\beta_1},$$

Rappelons que

$$L(\beta_0, \beta_1, (y_1, x_1), \dots, (y_n, x_n)) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

est la fonction de vraisemblance associée aux observations $(y_1, x_1), \dots, (y_n, x_n)$.

Cependant, le calcul de l'intégrale au dénominateur s'avère difficile, voir impossible sans passer par des estimations numériques, c'est pourquoi des méthodes d'approximations sont utilisées pour pallier à ce problème. On peut citer les méthodes MCMC (Markov Chain Monte Carlo) qui donnent des approximations pour estimer la densité a posteriori.

Chapitre 3

Estimation Bayésienne d'un modèle de régression linéaire à deux phases

Soit le modèle de régression $Y_t = \beta_1 X_t + \varepsilon_t$, où les erreurs ε_t , sont des variables aléatoires *i.i.d* de loi $\mathcal{N}(0, \sigma^2)$ avec $\sigma^2 > 0$. Cependant, une rupture est observée dans le système en un certain point m , il est reflété par un changement dans la pente, c'est-à-dire un changement dans le paramètre de régression β_1 qui devient β_2 .

On s'intéresse à savoir où et quand ce changement se produit, cette étude s'appelle problème d'inférence du point de rupture. Les estimateurs de m , β_1 et β_2 sont obtenus sous les fonctions de perte symétriques usuelles ainsi que sous les fonctions de pertes asymétriques dont on cite la Linex et la générale d'entropie.

Introduction

La régression est un outil statistique important pour l'analyse des données dans différents domaines tels qu'en science sociale et médecine.

Souvent, en pratique, les coefficients de régression sont supposés être constants. Cependant, en théorie, on suggère des modèles avec un ou plusieurs paramètres qui varient (c'est-à-dire non constants). Le paramètre d'intérêt dans cette analyse est le point de rupture qui indique où et quand le changement apparaît.

Plusieurs situations ont été étudiées dans les deux dernières décennies. Holbert (1982), en re-voyant le développement Bayésien dans le changement structurel donne une variété d'exemples intéressants en économie et en biologie. La monographie de Broemeling et Tsurumi (1987) fournit une étude complète du changement structurel dans un modèle linéaire avec une approche Bayésienne.

L'inférence Bayésienne du point de rupture suppose la disponibilité des distributions a priori des paramètres inconnus du modèle en étude.

Dans ce qui suit, nous allons étudier un modèle de régression linéaire à deux phases, noté TPLR (en anglais "two-phases linear regression model").

Dans ce chapitre, nous allons présenter une lecture critique de l'article fait par **Pandya et al.** publié en 2011, l'article étudie un modèle de régression linéaire à deux phases. L'approche Bayésienne est adoptée pour cette étude et utilise différentes fonctions de pertes.

La stabilité d'un modèle est l'un des critères important étudié par les statisticiens, toutefois, cette stabilité peut-être rompue, il est donc impératif d'étudier la rupture du modèle et de la déceler, dans un cas contraire, les résultats peuvent être erronés ce qui conduirait probablement à une interprétation fautive.

Pour cette étude, le statisticien fait appel aux tests pour détecter des éventuels changements des paramètres du modèle et estimer où et quand la rupture a eu lieu.

On dira qu'il y'a rupture dans une suite de variables aléatoires X_1, X_2, \dots, X_n s'il existe un entier m ($1 \leq m \leq n - 1$), tel que les variables aléatoires X_1, X_2, \dots, X_m et X_{m+1}, \dots, X_n suivent deux lois de probabilités différentes de fonctions de densités $f_1(x|\theta_1)$ et $f_2(x|\theta_2)$ respectivement, où θ_1 et θ_2 sont des paramètres inconnus.

Le point m est appelé *point de rupture* ou point de changement, c'est un paramètre aléatoire prenant les valeurs $\{1, 2, \dots, n - 1\}$.

La nature du changement du modèle diffère d'une étude à une autre, par exemple, en analyse des échantillons aléatoires, la rupture peut être observée sur la moyenne, la variance ou sur d'autres paramètres. En régression, le changement concerne les paramètres de régression ou bien la distribution des erreurs.

Dans notre travail basé sur l'article de **M.Pandya et al.** (2011), nous allons étudier un modèle de régression linéaire à deux phases, où l'on a uniquement un seul point de rupture, cette étude est fait via l'approche Bayésienne pour estimer les différents paramètres du modèle présenté ci-dessous.

3.1 Modèle de régression linéaire à deux phases (TPLR)

Le modèle TPLR est l'un des modèles qui exhibe un changement structurel, c'est-à-dire un modèle perturbé, il est défini comme suit :

$$y_t = \begin{cases} \alpha_1 + \beta_1 x_t + \varepsilon_t & t=1,2,\dots,m, \\ \alpha_2 + \beta_2 x_t + \varepsilon_t & t=m+1,\dots,n, \end{cases} \quad (3.1)$$

$\varepsilon_t, t = 1, \dots, n$, sont les erreurs du modèle, elles représentent des variables aléatoires *i.i.d* suivant une loi normale $\mathcal{N}(0, \sigma^2)$, les paramètres de régression sont tels que $(\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)$.

m est le point de rupture, c'est un paramètre aléatoire prenant la valeur n si aucune perturbation ne se remarque. Si m prend les valeurs de l'ensemble $\{1, \dots, n - 1\}$ alors une seule perturbation a lieu.

3.2 Estimation Bayésienne des paramètres

La méthode du maximum de vraisemblance (ML : maximum likelihood) comme d'autres approches classiques sont basées uniquement sur l'information observée après l'étude du phénomène. Toutefois, quand on dispose de données techniques et d'informations données par un expert sur les paramètres inconnus, l'approche Bayésienne est alors mieux adaptée. Cette approche est basée sur la densité a posteriori, notée $g(\beta_1, \beta_2, \sigma^{-2}, m|z)$, où $z = (y, x)$, qui est proportionnelle au produit de la fonction de vraisemblance $\mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z)$ et de la densité

jointe a priori, noté $g(\beta_1, \beta_2, \sigma^{-2}, m)$ représentant l'incertitude sur les valeurs des paramètres.

Les formules des différentes densités citées ci-dessus sont données dans ce qui suit :

$$g(\beta_1, \beta_2, \sigma^{-2}, m|z) = \frac{\mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) \cdot g(\beta_1, \beta_2, \sigma^{-2}, m)}{\sum_{m=1}^{n-1} \int_{\beta_1} \int_{\beta_2} \int_{\sigma^{-2}} \mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) \cdot \mathfrak{z}}. \quad (3.2)$$

Où \mathfrak{z} désigne $g(\beta_1, \beta_2, \sigma^{-2}, m)d\sigma^{-2}d\beta_2d\beta_1$.

Calcul de la fonction de vraisemblance

La fonction de vraisemblance étant donné un échantillon d'informations $z_t = (x_t, y_t)$, $t = 1, 2, \dots, m, m+1, \dots, n$ est :

$$\begin{aligned} \mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) &= \frac{1}{(2\pi)^{n/2}} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 s_{m1} + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \\ &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1}) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \\ &\times \exp \left\{ -A/2\sigma^2 \right\} \sigma^{-n}. \end{aligned} \quad (3.3)$$

Où :

$$\begin{aligned} s_{k1} &= \sum_{t=1}^k x_t^2; & s_{k2} &= \sum_{t=1}^k x_t y_t; \\ s_{m3} &= s_{m2} - \alpha B_m; & s_{m4} &= s_{n2} + 2\alpha(B_n - B_m); \\ A &= \sum_{t=1}^n y_t^2 + n\alpha^2 - 2\alpha \sum_{t=1}^n y_t; & B_k &= \sum_{t=1}^k x_t, \\ \text{et} \quad \alpha &= \alpha_1 + \alpha_2. \end{aligned} \quad (3.4)$$

Démontrons cela.

En considérant que $Y_t \sim \mathcal{N}(\alpha_i + \beta_i x_t, \sigma^2)$, $i = 1, 2$, soit $f(z_t)$ la densité de Y_t pour $t = 1, \dots, n$, on a :

$$\mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) = \prod_{t=1}^n f(z_t) = \prod_{t=1}^m f_1(z_t) \times \prod_{t=m+1}^n f_2(z_t).$$

où

$$\begin{aligned} f(z_t) &= f_1(z_t), & \text{pour } t &= 1, \dots, m \\ f(z_t) &= f_2(z_t), & \text{pour } t &= m+1, \dots, n. \end{aligned}$$

Posons $\mathcal{L}_1 = \prod_{t=1}^m f_1(z_t)$ et $\mathcal{L}_2 = \prod_{t=m+1}^n f_2(z_t)$, calculons \mathcal{L}_1 et puis \mathcal{L}_2 :

$$\begin{aligned} \mathcal{L}_1 &= \prod_{t=1}^m f(z_t) = \prod_{t=1}^m \left[\frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ \frac{-1}{2\sigma^2} (y_t - (\alpha_1 + \beta_1 x_t))^2 \right\} \right] \\ &= \frac{1}{\sigma^m (2\pi)^{m/2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{t=1}^m (y_t - (\alpha_1 + \beta_1 x_t))^2 \right\} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{\sigma^m (2\pi)^{m/2}} \exp \left\{ \frac{-1}{2\sigma^2} \left[\sum_{t=1}^m y_t^2 - 2\alpha_1 \sum_{t=1}^m y_t + m\alpha_1^2 + \beta_1^2 \sum_{t=1}^m x_t^2 - 2\beta_1 \left(\sum_{t=1}^m x_t y_t - \alpha_1 \sum_{t=1}^m x_t \right) \right] \right\} \\
 &= \frac{1}{\sigma^m (2\pi)^{m/2}} \exp \left\{ \frac{-1}{2\sigma^2} \left[\sum_{t=1}^m y_t^2 - 2\alpha_1 \sum_{t=1}^m y_t + m\alpha_1^2 \right] \right\} \times \exp \left\{ -\frac{1}{2\sigma^2} \beta_1^2 s_{m1} + \beta_1 \frac{s_{m3}}{\sigma^2} \right\}.
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{L}_2 &= \prod_{t=m+1}^n f(z_t) = \prod_{t=m+1}^n \left[\frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ \frac{-1}{2\sigma^2} (y_t - (\alpha_2 + \beta_2 x_t))^2 \right\} \right] \\
 &= \frac{1}{\sigma^{n-m} (2\pi)^{(n-m)/2}} \exp \left\{ \frac{-1}{2\sigma^2} \left[\sum_{t=m+1}^n y_t^2 - 2\alpha_2 \sum_{t=m+1}^n y_t + (n-m)\alpha_2^2 \right] \right\} \\
 &\times \exp \left\{ \left[\beta_2^2 \sum_{t=m+1}^n x_t^2 - 2\beta_2 \left(\sum_{t=m+1}^n x_t y_t - \alpha_2 \sum_{t=m+1}^n x_t \right) \right] \right\} \\
 &= \frac{1}{\sigma^{n-m} (2\pi)^{(n-m)/2}} \exp \left\{ \frac{-1}{2\sigma^2} \left[\sum_{t=m+1}^n y_t^2 - 2\alpha_2 \sum_{t=m+1}^n y_t + (n-m)\alpha_2^2 \right] \right\} \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \left[\beta_2^2 (s_{n1} - s_{m1}) - 2\beta_2 (s_{n2} - s_{m2} - \alpha_2 (B_n - B_m)) \right] \right\} \\
 &= \frac{1}{\sigma^{n-m} (2\pi)^{(n-m)/2}} \exp \left\{ \frac{-1}{2\sigma^2} \left[\sum_{t=m+1}^n y_t^2 - 2\alpha_2 \sum_{t=m+1}^n y_t + (n-m)\alpha_2^2 \right] \right\} \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1}) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\}.
 \end{aligned}$$

En multipliant \mathcal{L}_1 et \mathcal{L}_2 , on obtient la fonction de vraisemblance qui suit :

$$\begin{aligned}
 \mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) &= \frac{1}{(2\pi)^{n/2}} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 s_{m1} + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1}) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \\
 &\times \exp \left\{ -A/2\sigma^2 \right\} \sigma^{-n}.
 \end{aligned}$$

Deux cas se présentent pour le calcul de la densité a posteriori, le cas où σ^2 est inconnue et le cas où σ^2 est connue. Dans la suite de cette section, nous allons calculer les différentes densités a priori et a posteriori, jointes et marginales, des paramètres du modèles.

3.2.1 Cas où σ^2 est inconnue

On considère le modèle TPLR présenté en (3.1) avec la variance σ^2 inconnue. Nous supposons que le point de rupture m a une loi a priori uniformément distribuée sur l'ensemble $\{1, 2, \dots, n-1\}$ notée $g_{1m}(m)$ indépendant de β_1 et β_2 . Nous supposons aussi que des informations sur β_1 et β_2 sont disponibles sous forme de distributions conditionnelles a priori. Les densités a priori, sachant σ^2 , de β_1 et de β_2 sont respectivement :

$$g_{1\beta_1}(\beta_1|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[\frac{-1}{2\sigma^2} \beta_1^2 \right], \quad -\infty < \beta_1 < \infty \quad (3.5)$$

$$g_{1\beta_2}(\beta_2|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[\frac{-1}{2\sigma^2} \beta_2^2 \right], \quad -\infty < \beta_2 < \infty. \quad (3.6)$$

Nous remarquons aisément qu'elles sont Gaussiennes, $\mathcal{N}(0, \sigma^2)$. On suppose que des informations sur $\frac{1}{\sigma^2}$ sont disponibles, elles sont données via la moyenne a priori μ et le coefficient de variation ϕ , on utilise donc la loi gamma de paramètres c et d (notée $\mathcal{G}(c, d)$) comme distribution a priori marginale de $\frac{1}{\sigma^2}$ de moyenne μ dont la formule est comme suit

$$g_{1\sigma^2}\left(\frac{1}{\sigma^2}\right) = \frac{c^d}{\Gamma(d)} \left(\frac{1}{\sigma^2}\right)^{d-1} e^{-c/\sigma^2}, \quad \sigma^2 > 0, \quad c, d > 0, \quad (3.7)$$

où $\Gamma(d)$ désigne la fonction gamma définie par

$$\Gamma(d) = \int_0^\infty e^{-t} t^{d-1} dt.$$

Les paramètres de la loi gamma, c et d sont obtenus par

$$d = \frac{1}{\phi^2}, \quad c = \frac{1}{\mu\phi^2}.$$

Par conséquent, la densité a priori jointe de $\beta_1, \beta_2, \sigma^{-2}$ et m , qu'on note $g_1(\beta_1, \beta_2, \sigma^{-2}, m)$ est :

$$g_1(\beta_1, \beta_2, \sigma^{-2}, m) = k_1 \exp\left\{\frac{-1}{2\sigma^2}(\beta_1^2 + \beta_2^2)\right\} \left(\frac{1}{\sigma^2}\right)^d e^{-c/\sigma^2}, \quad (3.8)$$

où

$$k_1 = \frac{c^d}{2\pi\Gamma(d)(n-1)} \quad (3.9)$$

Cette loi a priori jointe est obtenue comme suit

Calcul de la densité a priori jointe .

En vertu de l'indépendance des paramètres $\beta_1, \beta_2, \sigma^{-2}$, et m on a :

$$\begin{aligned} g_1(\beta_1, \beta_2, \sigma^{-2}, m) &= g_{1\beta_1}(\beta_1|\sigma^2) \cdot g_{1\beta_2}(\beta_2|\sigma^2) \cdot g_{1\sigma^2}\left(\frac{1}{\sigma^2}\right) \cdot g_{1m}(m) \\ &= \frac{1}{2\pi\sigma^2} \exp\left\{\frac{-1}{2\sigma^2}(\beta_1^2 + \beta_2^2)\right\} \cdot \frac{c^d}{\Gamma(d)} \left(\frac{1}{\sigma^2}\right)^{d-1} e^{-c/\sigma^2} \cdot \frac{1}{n-1} \\ &= \frac{c^d}{2\pi\Gamma(d)(n-1)} \exp\left\{\frac{-1}{2\sigma^2}(\beta_1^2 + \beta_2^2)\right\} \left(\frac{1}{\sigma^2}\right)^d e^{-c/\sigma^2} \\ &= k_1 \exp\left\{\frac{-1}{2\sigma^2}(\beta_1^2 + \beta_2^2)\right\} \left(\frac{1}{\sigma^2}\right)^d e^{-c/\sigma^2}. \end{aligned}$$

Calcul de la densité a posteriori jointe .

La densité a posteriori jointe de $\beta_1, \beta_2, \sigma^{-2}$ et m , notée $g_1(\beta_1, \beta_2, \sigma^{-2}, m|z)$, est

$$g_1(\beta_1, \beta_2, \sigma^{-2}, m|z) = \frac{\mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) \cdot g_1(\beta_1, \beta_2, \sigma^{-2}, m)}{h_1(z)}$$

$$\begin{aligned}
 &= \frac{k_2}{h_1(z)} \cdot \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 [(s_{n1} - s_{m1}) + 1] + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \\
 &\times e^{-((A/2)+c(1/\sigma^2))} \left(\frac{1}{\sigma^2} \right)^d (\sigma^2)^{-n/2},
 \end{aligned} \tag{3.10}$$

où

$$h_1(z) = \sum_{m=1}^{n-1} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) \cdot g_1(\beta_1, \beta_2, \sigma^{-2}, m) d\beta_1 d\beta_2 d\sigma^{-2} \tag{3.11}$$

$$k_2 = k_1 \frac{1}{(2\pi)^{n/2}}. \tag{3.12}$$

Pour calculer cette densité, procédons par étapes. Tout d'abord, le calcul du produit de la fonction de vraisemblance et de la densité a priori jointe ;

$$\begin{aligned}
 \mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) \cdot g_1(\beta_1, \beta_2, \sigma^{-2}, m) &= \frac{1}{(2\pi)^{n/2}} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 s_{m1} + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1}) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \times \exp [-A/2\sigma^2] \sigma^{-n} \\
 &\times k_1 \exp \left\{ \frac{-1}{2\sigma^2} (\beta_1^2 + \beta_2^2) \right\} \left(\frac{1}{\sigma^2} \right)^d e^{-c/\sigma^2} \\
 &= k_2 \left(\frac{1}{\sigma^2} \right)^{d+n/2} \times \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \times \exp \left\{ -(A/2 + c)/\sigma^2 \right\}.
 \end{aligned}$$

où s_{n1} , s_{m1} , s_{m3} , s_{m4} , et A sont donnés dans les formules de (3.4).

Nous calculons $h_1(z)$ comme suit

$$\begin{aligned}
 h_1(z) &= \sum_{m=1}^{n-1} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \mathcal{L}(\beta_1, \beta_2, \sigma^{-2}, m|z) \cdot g_1(\beta_1, \beta_2, \sigma^{-2}, m) d\beta_1 d\beta_2 d\sigma^{-2} \\
 &= \sum_{m=1}^{n-1} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \left[k_2 \left(\frac{1}{\sigma^2} \right)^{d+n/2} \times \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \right. \\
 &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \times \exp \left\{ -(A/2 + c)/\sigma^2 \right\} \left. \right] d\beta_1 d\beta_2 d\sigma^{-2} \\
 &= k_2 \sum_{m=1}^{n-1} \int_0^\infty \left[\left(\frac{1}{\sigma^2} \right)^{d+n/2} \exp \left\{ -(A/2 + c)/\sigma^2 \right\} \times \int_{-\infty}^\infty \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} d\beta_1 \right. \\
 &\times \left. \int_{-\infty}^\infty \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} d\beta_2 \right] d\sigma^{-2} \\
 &= k_2 \sum_{m=1}^{n-1} \int_0^\infty \left[\left(\frac{1}{\sigma^2} \right)^{d+n/2} \exp \left\{ -(A/2 + c)/\sigma^2 \right\} \times \frac{\sqrt{2\pi} \cdot e^{s_{m3}^2/2\sigma^2(s_{m1}+1)}}{\sqrt{(s_{m1} + 1)}/\sigma} \right.
 \end{aligned}$$

$$\begin{aligned}
 & \times \left. \frac{\sqrt{2\pi} \cdot e^{s_{m4}^2/2\sigma^2(s_{n1}-s_{m1}+1)}}{\sqrt{(s_{n1}-s_{m1}+1)/\sigma}} \right] d\sigma^{-2} \\
 & = k_2 \sum_{m=1}^{n-1} \int_0^\infty \frac{1}{(\sigma^2)^{n/2+d-1}} \frac{2\pi}{\sqrt{s_{m1}+1}\sqrt{s_{n1}-s_{m1}+1}} \\
 & \times \exp \left\{ \frac{-1}{\sigma^2} \left(\frac{A}{2} + c - \frac{s_{m3}^2}{2(s_{m1}+1)} - \frac{s_{m4}^2}{2(s_{n1}-s_{m1}+1)} \right) \right\} d\sigma^{-2} \\
 & = k_2 \sum_{m=1}^{n-1} 2\pi \frac{\Gamma(n/2+d)}{\left(\frac{A}{2} + c - h_{m1} - h_{m2} \right)^{n/2+d}} \cdot \mathbf{x} \\
 & \times \int_0^\infty \left(\frac{1}{\sigma^2} \right)^{n/2+d-1} \exp \left\{ \frac{-1}{\sigma^2} \left(\frac{A}{2} + c - h_{m1} - h_{m2} \right) \right\} \cdot \frac{\left(\frac{A}{2} + c - h_{m1} - h_{m2} \right)^{n/2+d}}{\Gamma(n/2+d)} d\sigma^{-2} \\
 & = k_2 \sum_{m=1}^{n-1} 2\pi \frac{\Gamma(n/2+d)}{\left(\frac{A}{2} + c - h_{m1} - h_{m2} \right)^{n/2+d}} \cdot \mathbf{x} \\
 & = k_2 \sum_{m=1}^{n-1} T_1(m)
 \end{aligned}$$

$$T_1(m) = 2\pi \frac{\Gamma(n/2+d)}{\left(\frac{A}{2} + c - h_{m1} - h_{m2} \right)^{n/2+d}} \cdot \mathbf{x} \quad (3.13)$$

En notant que :

$$h_{m1} = \frac{s_{m3}^2}{2(s_{m1}+1)}, \quad h_{m2} = \frac{s_{m4}^2}{2(s_{n1}-s_{m1}+1)}, \quad \text{et} \quad \mathbf{x} = \sqrt{s_{m1}+1}\sqrt{s_{n1}-s_{m1}+1}.$$

(Quelques passages de cette démonstration seront vus plus en détails dans les sections qui suivent).

On obtient la fonction de densité a posteriori jointe donnée précédemment par la formule (3.10).

En intégrant cette dernière par rapport à β_1 , β_2 et σ^{-2} , on obtient aisément la densité a posteriori marginale de m , qui est donnée par l'équation

$$g_1(m|z) = \frac{T_1(m)}{\sum_{m=1}^{n-1} T_1(m)} \quad (3.14)$$

3.2.2 Cas où σ^2 est connue

On considère maintenant le modèle présenté en (3.1) avec une variance σ^2 connue. Les distributions a priori des paramètres β_1 , β_2 et m restent inchangées, c'est-à-dire, m est de loi uniforme sur $\{1, 2, \dots, n-1\}$, la fonction de densité de β_1 est celle donnée en (3.5) et la fonction de densité de β_2 est donnée par (3.6).

Par conséquent, la loi a priori jointe de β_1 , β_2 et m , notée $g_2(\beta_1, \beta_2, m)$ est

$$g_2(\beta_1, \beta_2, m) = k_3 \exp \left\{ \frac{-1}{2\sigma^2} (\beta_1^2 + \beta_2^2) \right\}, \quad (3.15)$$

où

$$k_3 = \frac{1}{2\pi(n-1)\sigma^2} \quad (3.16)$$

$$(3.17)$$

La densité a posteriori jointe des paramètres, notée $g_2(\beta_1, \beta_2, m|z)$ est

$$\begin{aligned} g_2(\beta_1, \beta_2, m|z) &= \frac{k_4}{h_2(z)} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \\ &\times \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\}, \end{aligned} \quad (3.18)$$

où

$$k_4 = k_3 \frac{1}{(2\pi)^{n/2}} e^{-A/2\sigma^2} \sigma^{-n}, \quad (3.19)$$

et

$$\begin{aligned} h_2(z) &= k_4 \sum_{m=1}^{n-1} \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} d\beta_1 \\ &\times \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} d\beta_2, \end{aligned} \quad (3.20)$$

On peut réécrire $h_2(z)$ comme suit :

$$\begin{aligned} h_2(z) &= k_4 \sum_{m=1}^{n-1} T_2(m) \\ T_2(m) &= G_{1m} \cdot G_{2m}. \end{aligned} \quad (3.21)$$

avec

$$\begin{aligned} G_{1m} &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} d\beta_2 \\ &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \left(\beta_2 \sqrt{s_{n1} - s_{m1} + 1} - \frac{s_{m4}}{\sqrt{s_{n1} - s_{m1} + 1}} \right)^2 - \frac{s_{m4}^2}{(s_{n1} - s_{m1} + 1)} \right\} d\beta_2 \\ &= \exp \left\{ \frac{1}{2\sigma^2} \frac{s_{m4}^2}{(s_{n1} - s_{m1} + 1)} \right\} \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} (s_{n1} - s_{m1} + 1) \left(\beta_2 - \frac{s_{m4}}{\sqrt{s_{n1} - s_{m1} + 1}} \right)^2 \right\} d\beta_2 \\ &= \frac{\sqrt{2\pi} \cdot e^{s_{m4}^2/2\sigma^2(s_{n1}-s_{m1}+1)}}{\sqrt{s_{n1} - s_{m1} + 1}/\sigma}. \end{aligned} \quad (3.22)$$

Et

$$\begin{aligned} G_{2m} &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} d\beta_1 \\ &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \left(\beta_1 \sqrt{s_{m1} + 1} - \frac{s_{m3}}{\sqrt{s_{m1} + 1}} \right)^2 - \frac{s_{m3}^2}{(s_{m1} + 1)} \right\} d\beta_1 \end{aligned} \quad (3.23)$$

$$\begin{aligned}
 &= \exp \left\{ \frac{1}{2\sigma^2} \frac{s_{m3}^2}{(s_{m1} + 1)} \right\} \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} (s_{m1} + 1) \left(\beta_1 - \frac{s_{m3}}{\sqrt{s_{m1} + 1}} \right)^2 \right\} d\beta_1 \\
 &= \frac{\sqrt{2\pi} \cdot e^{s_{m3}^2/2\sigma^2(s_{m1}+1)}}{\sqrt{s_{m1} + 1}/\sigma}.
 \end{aligned}$$

En intégrant par rapport à β_1 et β_2 , on obtient la densité marginale a posteriori de m , donnée par

$$g_2(m|z) = \frac{T_2(m)}{\sum_{m=1}^{n-1} T_2(m)}. \quad (3.24)$$

On calcule aussi les marginales a posteriori de β_1 et de β_2 , données respectivement par

$$g_2(\beta_1|z) = \frac{k_4}{h_2(z)} \left[\sum_{m=1}^{n-1} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \right] \cdot G_{1m} \quad (3.25)$$

$$g_2(\beta_2|z) = \frac{k_4}{h_2(z)} \left[\sum_{m=1}^{n-1} \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \right] \cdot G_{2m}. \quad (3.26)$$

3.3 Estimation Bayésienne sous les fonctions de coûts symétriques

L'estimateur Bayésien d'un paramètre α (ou une fonction de ce paramètre) associé à la fonction de coût quadratique (en littérature anglaise "squared error loss"), définie par

$$L_1(\alpha, d) = (\alpha - d)^2, \quad (3.27)$$

est la moyenne a posteriori de α , où d désigne une règle de décision pour estimer le paramètre α . Par suite, la fonction coût quadratique relative à l'estimation d'un paramètre à valeur entière est

$$L_1(m, v) = (m - v)^2, \quad m, v = 1, 2, \dots \quad (3.28)$$

L'estimateur Bayésien dans ce cas n'est plus la moyenne a posteriori, il peut être obtenu numériquement en minimisant le risque a posteriori associé. Généralement un tel estimateur est égale à la valeur entière la plus proche de la moyenne a posteriori.

Estimateur Bayésien de m associé à la fonction de perte quadratique où la variance est inconnue

L'estimateur Bayésien de m associé à la fonction de coût quadratique dans le cas où la variance σ^2 est inconnue est

$$m^* = \frac{\sum_{m=1}^{n-1} m T_1(m)}{\sum_{m=1}^{n-1} T_1(m)}. \quad (3.29)$$

L'estimateur Bayésien de m sous la fonction de coût quadratique est calculé comme suit

$$m^* = \mathbb{E}_{g_{1m}(m|z)} [m|z]$$

$$\begin{aligned}
 &= \sum_{m=1}^{n-1} m g_{1m}(m|z) \\
 &= \sum_{m=1}^{n-1} m \frac{T_1(m)}{\sum_{m=1}^{n-1} T_1(m)} \\
 &= \frac{\sum_{m=1}^{n-1} m T_1(m)}{\sum_{m=1}^{n-1} T_1(m)}.
 \end{aligned}$$

Estimateur Bayésien de m associé à la fonction de perte quadratique où la variance est connue

L'estimateur Bayésien de m associé à la fonction de coût quadratique dans le cas où la variance σ^2 est connue est égale à

$$m^{**} = \frac{\sum_{m=1}^{n-1} m T_2(m)}{\sum_{m=1}^{n-1} T_2(m)}. \quad (3.30)$$

m^{**} se calcule exactement de la même façon que m^* .

Estimateur Bayésien de β_1 et de β_2 associé à la fonction de perte quadratique où la variance est connue

L'estimateur Bayésien de β_1 , noté β_1^{**} sous la fonction coût quadratique est l'espérance de la distribution a posteriori de β_1 donnée par la formule (3.25). Par conséquent, l'estimateur β_1^{**} est donné par le calcul de l'intégrale suivante

$$\beta_1^{**} = \int_{-\infty}^{\infty} \beta_1 g_2(\beta_1|z) d\beta_1 = \int_{-\infty}^{\infty} \beta_1 \frac{k_4}{h_2(z)} \left[\sum_{m=1}^{n-1} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \right] \cdot G_{1m} d\beta_1$$

L'estimateur Bayésien de β_2 , noté β_2^{**} sous la fonction coût quadratique est l'espérance de la distribution a posteriori de β_2 donnée par la formule (3.26). Par conséquent, l'estimateur β_2^{**} est donné par le calcul de l'intégrale suivante

$$\beta_2^{**} = \int_{-\infty}^{\infty} \beta_2 g_2(\beta_2|z) d\beta_2 = \int_{-\infty}^{\infty} \beta_2 \frac{k_4}{h_2(z)} \left[\sum_{m=1}^{n-1} \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} \right] \cdot G_{2m} d\beta_2$$

La formule analytique de ces deux estimateurs ne peut être qu'obtenue numériquement étant donné la complexité des intégrales.

D'autres fonctions de coûts symétriques existent, on cite la fonction de coût absolue, définie par

$$L_2(\alpha, d) = |\alpha - d|, \quad (3.31)$$

et la fonction de coût 0-1, définie par

$$L_3(\alpha, d) = \begin{cases} 0, & \text{si } |\alpha - d| < \varepsilon, \quad \varepsilon > 0 \\ 1, & \text{sinon,} \end{cases} \quad (3.32)$$

dont les estimateurs Bayésiens associés sont respectivement, la médiane et le mode a posteriori.

Remarque : la démonstration de ces résultats est présente dans la plupart des ouvrages classiques de la statistique Bayésienne.

3.4 Estimation Bayésienne sous les fonctions de coûts asymétriques

Une fonction de perte $L(\alpha, d)$ mesure, par exemple, les conséquences d'une chute financière dû à un mauvais choix d'une règle de décision d pour l'estimation d'une quantité inconnue α (un paramètre général ou une fonction de ce paramètre). Le choix de la fonction de perte adéquate dépend de la situation financière uniquement et est indépendant de la méthode d'estimation utilisée.

On présente alors une fonction de perte asymétrique usuellement pratique, c'est la fonction de perte Linex, qui a été introduite par Varian (1975), en supposant que la perte minimale est obtenue au point α . La fonction de perte Linex est donnée par la formule suivante

$$L_4(\alpha, d) = \exp\{q_1(d - \alpha)\} - q_1(d - \alpha) - 1, \quad q_1 \neq 0. \quad (3.33)$$

où q_1 désigne le paramètre de forme dont le signe indique la direction de l'asymétrie, autrement dit, si $q_1 > 0$ alors la sur-estimation est plus grande que la sous-estimation, et vice versa.

L'espérance a posteriori, ou risque a posteriori, sous la fonction de perte Linex est

$$\mathbb{E}_\alpha [L_4(\alpha, d)] = \exp(q_1 d) \mathbb{E}_\alpha [\exp(-q_1 \alpha)] - q_1(d - \mathbb{E}_\alpha(\alpha)) - 1, \quad (3.34)$$

où $\mathbb{E}_\alpha [f(\alpha)]$ représente l'espérance de $f(\alpha)$ par rapport à la densité a posteriori $g(\alpha|z)$.

$$\begin{aligned} \mathbb{E}_\alpha [L_4(\alpha, d)] &= \int L_4(\alpha, d)g(\alpha|z)d\alpha \\ &= \int [\exp\{q_1(d - \alpha)\} - q_1(d - \alpha) - 1]g(\alpha|z)d\alpha \\ &= \exp(q_1 d) \int \exp(-q_1 \alpha)g(\alpha|z)d\alpha - q_1 \left(d \int g(\alpha|z)d\alpha - \int \alpha g(\alpha|z)d\alpha \right) - 1 \\ &= \exp(q_1 d) \mathbb{E}_\alpha [\exp(-q_1 \alpha)] - q_1(d - \mathbb{E}_\alpha(\alpha)) - 1. \end{aligned}$$

L'estimateur Bayésien du paramètre α , noté α_L^* est la valeur de d qui minimise le risque a posteriori, *i.e.* $\mathbb{E}_\alpha [L_4(\alpha, d)]$, cette estimateur est donné par la formule

$$\alpha_L^* = \arg \min_{d \in \mathcal{D}} \mathbb{E}_\alpha [L_4(\alpha, d)] = -\frac{1}{q_1} \ln [\mathbb{E}_\alpha [\exp(-q_1 \alpha)]], \quad (3.35)$$

en supposant que l'espérance $\mathbb{E}_\alpha [\exp(-q_1 \alpha)]$ existe et qui doit être finie.

Cet estimateur est obtenu en résolvant le système

$$\begin{cases} \frac{\partial}{\partial d} \mathbb{E}_\alpha [L_4(\alpha, d)] = 0 \\ \frac{\partial^2}{\partial d^2} \mathbb{E}_\alpha [L_4(\alpha, d)] > 0 \end{cases}$$

Ce qui donne

$$\begin{aligned}
 \frac{\partial}{\partial d} \mathbb{E}_\alpha [L_4(\alpha, d)] &= q_1 e^{q_1 d} \mathbb{E}_\alpha [e^{-q_1 \alpha}] - q_1 \\
 \Rightarrow q_1 (e^{q_1 d} \mathbb{E}_\alpha [e^{-q_1 \alpha}] - 1) &= 0 \\
 \Rightarrow e^{q_1 d} \mathbb{E}_\alpha [e^{-q_1 \alpha}] &= 1 \\
 \Rightarrow \ln \{ \mathbb{E}_\alpha [e^{-q_1 \alpha}] \} &= -q_1 d \\
 \Rightarrow \hat{d}^* = \alpha_L^* &= -\frac{1}{q_1} \ln \{ \mathbb{E}_\alpha [e^{-q_1 \alpha}] \}.
 \end{aligned}$$

$$\frac{\partial^2}{\partial d^2} \mathbb{E}_\alpha [L_4(\alpha, d)] = q_1^2 e^{q_1 d} \mathbb{E}_\alpha [e^{-q_1 \alpha}] > 0.$$

On conclut alors que α_L^* est bien l'estimateur Bayésien de α associé à la fonction de perte Linex.

Une autre fonction de coût asymétrique, appelée entropie générale, est aussi utilisée, elle fut introduite par Calabria et Pulcini (1994), elle est donnée par la formule qui suit

$$L_5(\alpha, d) = \left(\frac{d}{\alpha} \right)^{q_3} - q_3 \ln \left(\frac{d}{\alpha} \right) - 1, \quad (3.36)$$

où q_3 est aussi un paramètre de forme.

Le risque a posteriori $\mathbb{E}_\alpha [L_5(\alpha, d)]$ associé à la fonction de perte entropie générale est

$$\mathbb{E}_\alpha [L_5(\alpha, d)] = d^{q_3} \mathbb{E}_\alpha [\alpha^{-q_3}] - q_3 \mathbb{E}_\alpha \left[\ln \left(\frac{d}{\alpha} \right) \right]. \quad (3.37)$$

Nous le calculons comme suit

$$\begin{aligned}
 \mathbb{E}_\alpha [L_5(\alpha, d)] &= \int L_5(\alpha, d) g(\alpha|z) d\alpha \\
 &= \int \left(\left(\frac{d}{\alpha} \right)^{q_3} - q_3 \ln \left(\frac{d}{\alpha} \right) - 1 \right) g(\alpha|z) d\alpha \\
 &= d^{q_3} \int \alpha^{-q_3} g(\alpha|z) d\alpha - q_3 \int \ln \left(\frac{d}{\alpha} \right) g(\alpha|z) d\alpha - 1 \\
 &= d^{q_3} \mathbb{E}_\alpha [\alpha^{-q_3}] - q_3 \mathbb{E}_\alpha \left[\ln \left(\frac{d}{\alpha} \right) \right].
 \end{aligned}$$

L'estimateur Bayésien d'un paramètre général α est la valeur d qui minimise le risque a posteriori $\mathbb{E}_\alpha [L_5(\alpha, d)]$, ce estimateur est donné par

$$\alpha_E^* = \arg \min_d (\mathbb{E}_\alpha [L_5(\alpha, d)]) = [\mathbb{E}_\alpha [\alpha^{-q_3}]]^{-1/q_3} \quad (3.38)$$

avec $\mathbb{E}_\alpha [\alpha^{-q_3}]$ est supposée finie. On l'obtient en résolvant le système suivant

$$\begin{cases} \frac{\partial}{\partial d} \mathbb{E}_\alpha [L_5(\alpha, d)] = 0 \\ \frac{\partial^2}{\partial d^2} \mathbb{E}_\alpha [L_5(\alpha, d)] > 0 \end{cases}$$

$$\begin{aligned}
 \frac{\partial}{\partial d} \mathbb{E}_\alpha [L_5(\alpha, d)] &= q_3 d^{q_3-1} \mathbb{E}_\alpha [\alpha^{-q_3}] - \frac{q_3}{d} \\
 \Rightarrow q_3 d^{-1} (d^{q_3} \mathbb{E}_\alpha [\alpha^{-q_3}] - 1) &= 0 \\
 \Rightarrow d^{q_3} \mathbb{E}_\alpha [\alpha^{-q_3}] &= 1 \\
 \Rightarrow \widehat{d}^* = \alpha_E^* &= [\mathbb{E}_\alpha [\alpha^{-q_3}]]^{-1/q_3}.
 \end{aligned}$$

La condition de la positivité de la deuxième dérivée est facile à vérifier.

3.4.1 Cas où σ^2 est inconnue

Estimateur Bayésien de m associé à la fonction de perte Linex

En minimisant le risque a posteriori $\mathbb{E}_\alpha [L_4(\alpha, d)]$ associé à la fonction de perte Linex $L_4(\alpha, d)$ et à la densité a posteriori $g_1(m|z)$ donnée par (3.14), on obtient l'estimateur Bayésien de m associé à la fonction de perte Linex, noté m_L^* que l'on donne par

$$\begin{aligned}
 m_L^* &= -\frac{1}{q_1} \ln [\mathbb{E}_m [\exp \{-q_1 m\}]] \\
 &= -\frac{1}{q_1} \ln \left[\frac{\sum_{m=1}^{n-1} e^{-q_1 m} T_1(m)}{\sum_{m=1}^{n-1} T_1(m)} \right].
 \end{aligned} \tag{3.39}$$

Estimateur Bayésien de m associé à la fonction de coût entropie générale

L'estimateur Bayésien de m associé à la fonction de coût entropie générale et à la densité a posteriori $g_1(m|z)$ donnée par la formule (3.14) est explicité comme suit

$$\begin{aligned}
 m_E^* &= [\mathbb{E}_m [m^{-q_3}]]^{-1/q_3} \\
 &= \left[\frac{\sum_{m=1}^{n-1} m^{-q_3} T_1(m)}{\sum_{m=1}^{n-1} T_1(m)} \right]^{-1/q_3}.
 \end{aligned} \tag{3.40}$$

Où $T_1(m)$ est donnée précédemment par (3.12).

3.4.2 Cas où σ^2 est connue

Estimateur Bayésien de m associé à la fonction de perte Linex

Combinant la fonction de coût Linex avec la densité a posteriori $g_2(m|z)$ explicitée par (3.24), on obtient l'estimateur de Bayes du paramètre m , noté m_L^{**} , donné par la formule suivante

$$\begin{aligned}
 m_L^{**} &= -\frac{1}{q_1} \ln (\mathbb{E}_m [e^{-q_1 m}]) \\
 &= -\frac{1}{q_1} \ln \left(\frac{\sum_{m=1}^{n-1} e^{-q_1 m} T_2(m)}{\sum_{m=1}^{n-1} T_2(m)} \right)
 \end{aligned} \tag{3.41}$$

Estimateurs Bayésiens de β_1 et de β_2 associé à la fonction de perte Linex

Les estimateurs Bayésiens associés à la fonction de perte Linex et aux densités a posteriori données par les formules (3.25) et (3.26) des paramètres de régression β_1 et β_2 respectivement, notés β_1^{**} et β_2^{**} sont comme suit

$$\begin{aligned}
 \beta_{1L}^{**} &= -\frac{1}{q_1} \ln \left(\mathbb{E}_{\beta_1} \left[e^{-q_1 \beta_1} \right] \right) \\
 &= -\frac{1}{q_1} \ln \left\{ \frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} \frac{\sqrt{2\pi} e^{(s_{m3}-q_1\sigma^2)^2/2\sigma^2(s_{m1}+1)}}{\sqrt{s_{m1}+1}/\sigma} \cdot G_{1m} \right\}
 \end{aligned} \tag{3.42}$$

Détaillons ce calcul

$$\begin{aligned}
 \beta_{1L}^{**} &= -\frac{1}{q_1} \ln \left(\mathbb{E}_{\beta_1} \left[e^{-q_1 \beta_1} \right] \right) \\
 &= -\frac{1}{q_1} \ln \left(\int_{-\infty}^{\infty} e^{-q_1 \beta_1} g_2(\beta_1|z) d\beta_1 \right) \\
 &= -\frac{1}{q_1} \ln \left(\int_{-\infty}^{\infty} e^{-q_1 \beta_1} \frac{k_4}{h_2(z)} \left[\sum_{m=1}^{n-1} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1}+1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} \right] \cdot G_{1m} d\beta_1 \right) \\
 &= -\frac{1}{q_1} \ln \left(\frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} G_{1m} \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1}+1) + \beta_1 \left(\frac{s_{m3}}{\sigma^2} - q_1 \right) \right\} d\beta_1 \right)
 \end{aligned}$$

Pour alléger l'écriture, on pose I égale à l'intégrale suivante

$$\begin{aligned}
 I &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1}+1) + \beta_1 \left(\frac{s_{m3}}{\sigma^2} - q_1 \right) \right\} d\beta_1 \\
 &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-1}{2\sigma^2} \left[(s_{m1}+1) \left(\beta_1 - \left(\frac{s_{m3} - \sigma^2 q_1}{s_{m1}+1} \right) \right)^2 - \frac{(s_{m3} - \sigma^2 q_1)^2}{s_{m1}+1} \right] \right\} d\beta_1 \\
 &= \int_{-\infty}^{\infty} \exp \left\{ \frac{1}{2\sigma^2} \frac{(s_{m3} - \sigma^2 q_1)^2}{s_{m1}+1} \right\} \cdot \exp \left\{ \frac{-1}{2\sigma^2} (s_{m1}+1) \left(\beta_1 - \left(\frac{s_{m3} - \sigma^2 q_1}{s_{m1}+1} \right) \right)^2 \right\} d\beta_1 \\
 &= \exp \left\{ \frac{1}{2\sigma^2} \frac{(s_{m3} - \sigma^2 q_1)^2}{s_{m1}+1} \right\} \cdot \frac{\sqrt{2\pi}}{\sqrt{s_{m1}+1}/\sigma},
 \end{aligned}$$

d'où

$$\begin{aligned}
 \beta_{1L}^{**} &= -\frac{1}{q_1} \ln \left(\frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} \exp \left\{ \frac{1}{2\sigma^2} \frac{(s_{m3} - \sigma^2 q_1)^2}{s_{m1}+1} \right\} \cdot \frac{\sqrt{2\pi}}{\sqrt{s_{m1}+1}/\sigma} \cdot G_{1m} \right) \\
 &= -\frac{1}{q_1} \ln \left\{ \frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} \frac{\sqrt{2\pi} e^{(s_{m3}-q_1\sigma^2)^2/2\sigma^2(s_{m1}+1)}}{\sqrt{s_{m1}+1}/\sigma} \cdot G_{1m} \right\}.
 \end{aligned}$$

L'estimateur Bayésien du paramètre β_2 associé à la fonction de perte Linex est

$$\begin{aligned}
 \beta_{2L}^{**} &= -\frac{1}{q_1} \ln \left(\mathbb{E}_{\beta_2} \left[e^{-q_1 \beta_2} \right] \right) \\
 &= -\frac{1}{q_1} \ln \left\{ \frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} \frac{\sqrt{2\pi} e^{(s_{m4}-q_1\sigma^2)^2/2\sigma^2(s_{n1}-s_{m1}+1)}}{\sqrt{s_{n1}-s_{m1}+1}/\sigma} \cdot G_{2m} \right\},
 \end{aligned} \tag{3.43}$$

qui se calcule exactement de la même manière que β_{1L}^{**} .

k_4 , $h_2(z)$, G_{1m} et G_{2m} sont les mêmes donnés précédemment par les formules (3.19), (3.20), (3.22) et (3.23) respectivement, s_{m1} , s_{m2} , s_{m3} , s_{m4} , sont donnés par les formules en (3.4).

Estimateur Bayésien de m associé à la fonction de coût entropie générale

L'estimateur Bayésien de m associé à la fonction coût entropie générale est obtenu en minimisant le risque a posteriori $\mathbb{E}[L_5(m, d)]$, noté m_E^{**} , est

$$\begin{aligned} m_E^{**} &= [\mathbb{E}_m [m^{-q_3}]]^{-1/q_3} \\ &= \left[\frac{\sum_{m=1}^{n-1} m^{-q_3} T_2(m)}{\sum_{m=1}^{n-1} T_2(m)} \right]^{-1/q_3}. \end{aligned} \quad (3.44)$$

$T_2(m)$ est donné par la formule (3.21).

Nous allons définir les fonctions hypergéométrique qui vont nous être utiles pour les estimateurs Bayésiens des paramètres β_1 , et β_2 sous la fonction de coût générale entropie.

Fonction hypergéométrique

Définition 3.4.1. *La fonction hypergéométrique confluyente du premier type, notée ${}_1F_1(a, b; z)$ est une forme dégénérée de la fonction hypergéométrique ${}_2F_1(a, b, c; z)$ qui est une solution de l'équation différentielle hypergéométrique confluyente $zy''(z) + (c - z)y'(z) - ay(z) = 0$ pour $c \notin \mathbb{Z}$.*

$${}_1F_1(a, b; z) = \lim_{b \rightarrow \infty} {}_2F_1(a, b, c; \frac{z}{b}).$$

La fonction hypergéométrique confluyente est aussi connue sous le nom de la fonction de Kummer du premier type, elle est définie comme suit

$${}_1F_1(a, b; z) = \sum_{m=0}^{\infty} \frac{(a, m)z^m}{(b, m)m!}, \quad \text{pour } |z| < 1, \quad (3.45)$$

avec les coefficients dit de Pochhammer défini par $(a, m) = \frac{\Gamma(a + m)}{\Gamma(a)}$, $m \geq 1$.

Cette fonction a une autre écriture sous forme d'intégrale

$${}_1F_1(a, b; z) = \int_0^1 \frac{e^{zu}u^{a-1}(1-u)^{b-a-1}}{\mathcal{B}(a, b-a)} du. \quad (3.46)$$

Γ et \mathcal{B} désignent la fonction Gamma et la fonction Bêta respectivement.

Pour a et b entiers, on obtient des résultats particuliers, c'est-à-dire, si $a < 0$ et $b > 0$ ou $b < a$, la série devient un polynôme avec un nombre fini de termes, et si $b \leq 0$, la fonction hypergéométrique confluyente n'est pas définie.

Définition 3.4.2. *On appelle une série hypergéométrique généralisée, notée par ${}_pF_q[\{(a_1), \dots, (a_p)\}, \{(b_1), \dots, (b_q)\}, z]$ la série définie comme suit*

$${}_pF_q[\{(a_1), \dots, (a_p)\}, \{(b_1), \dots, (b_q)\}, z] = \sum_{m=0}^{\infty} \frac{(a_1)_m \dots (a_p)_m z^m}{(b_1)_m \dots (b_q)_m m!} \quad (3.47)$$

où $(a_i)_m$ est le coefficient de Pochhammer défini par $(a_i)_m = \frac{\Gamma(a_i + m)}{\Gamma(a_i)}$, $i = 1, \dots, p$.

La série est divergente pour tout $|z| > 1$ ou si $p > q + 1$ dans ce cas la série diverge

pour toute valeur de $z \neq 0$.

Si $p = q + 1$ la série converge si l'une de ces conditions est vérifiée

- $|z| < 1$.
- $z = 1$ et $\left[\sum_{j=1}^q b_j - \sum_{j=1}^{q+1} a_j \right] > 0$.
- $z = -1$ et $\left[\sum_{j=1}^q b_j - \sum_{j=1}^{q+1} a_j \right] + 1 > 0$.

La notation hypergéométrique permet de regrouper un nombre infini de fonctions spéciales. En effet, en substituant les arguments de ${}_pF_q [\{(a_1), \dots, (a_p)\}, \{(b_1), \dots, (b_q)\}, z]$ par des valeurs particulières, on reconnaît des fonctions usuelles, plusieurs exemples de ces fonctions sont données dans Abramowitz et Stegun (1972).

Définition 3.4.3. *La fonction Bêta $\mathcal{B}(x, y)$, dite également intégrale d'Euler du premier type est définie par*

$$\mathcal{B}(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(xy)}.$$

Estimateurs Bayésiens de β_1 et de β_2 associé à la fonction de coût entropie générale

En minimisant l'espérance a posteriori $\mathbb{E}[L_5(\beta_i, d)]$, $i = 1, 2$, associée à la fonction de coût entropie générale et aux distributions a posteriori (3.25) et (3.26) des paramètres β_1 et β_2 , on obtient leurs estimateurs Bayésiens donnés respectivement comme suit

$$\beta_{iE}^{**} = [\mathbb{E}[\beta_i^{-q_3}]]^{-1/q_3} \quad (3.48)$$

$$\begin{aligned} \beta_{1E}^{**} &= [\mathbb{E}[\beta_1^{-q_3}]]^{-1/q_3} \\ &= \left[\frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} J_{1m} G_{1m} \right]^{-1/q_3}, \end{aligned} \quad (3.49)$$

où

$$\begin{aligned} J_{1m} &= \int_{-\infty}^{\infty} \beta_1^{-q_3} \exp \left\{ \frac{-1}{2\sigma^2} \beta_1^2 (s_{m1} + 1) + \beta_1 \frac{s_{m3}}{\sigma^2} \right\} d\beta_1 \\ &= \frac{1}{((s_{m1} + 1)/\sigma^2)(-1 + q_3)} \times \left\{ k_{5m} \Gamma \left(\frac{1 - q_3}{2} \right) {}_1F_1 \left(\frac{q_3}{2}, \frac{1}{2}, -\frac{s_{m3}^2/\sigma^2}{2(s_{m1} + 1)} \right) \right\} \\ &\quad + \left\{ G_{3m} \Gamma \left(1 - \frac{q_3}{2} \right) {}_1F_1 \left(\frac{1 + q_3}{2}, \frac{3}{2}, -\frac{s_{m3}^2/\sigma^2}{2(s_{m1} + 1)\sigma^2} \right) \right\} \\ &\quad + \left\{ G_{4m} {}_pF_q \left(\left\{ \frac{1}{2}, 1 \right\}, \left\{ 1 - \frac{q_3}{2}, \frac{3}{2} - \frac{q_3}{2} \right\}, -\frac{s_{m3}^2/\sigma^2}{2(s_{m1} + 1)\sigma^2} \right) \right\}, \end{aligned} \quad (3.50)$$

avec

$$\begin{aligned} k_{5m} &= (-1)^{-q_3} 2^{(1/2)(-1-q_3)} \left\{ -\frac{(s_{m1} + 1)^2}{s_{m3}^2} \right\}^{-q_3} \left\{ -\frac{s_{m3}^2}{(s_{m1} + 1)^2} \right\}^{-q_3} \exp \left\{ \frac{s_{m3}^2/\sigma^2}{2(s_{m1} + 1)} \right\} \\ &\quad \times \left[\left(\frac{s_{m1} + 1}{\sigma^2} \right)^{q_3/2} (-1 + q_3) \left\{ \sqrt{\frac{s_{m1} + 1}{\sigma^2}} \left((-1)^{q_3} \left(-\frac{s_{m1} + 1}{s_{m3}} \right)^{q_3} \left(-\frac{s_{m3}}{s_{m1} + 1} \right)^{q_3} + 1 \right) \right\} \right], \end{aligned} \quad (3.51)$$

$$G_{3m} = \sqrt{2} \frac{s_{m3}}{\sigma^2} \left[(-1)^{q_3} \left(-\frac{s_{m1} + 1}{s_{m3}} \right)^{q_3} \left(-\frac{s_{m3}}{s_{m1} + 1} \right)^{q_3} - 1 \right], \quad (3.52)$$

$$G_{4m} = 2^{(1+q_3)/2} \left[\left\{ -\frac{(s_{m1} + 1)^2}{s_{m3}^2} \right\}^{q_3} \cdot \frac{s_{m3}}{\sigma^2} \left\{ (-1)^{q_3} \left(-\frac{s_{m3}}{s_{m1} + 1} \right)^{q_3} - \left(\frac{s_{m3}}{s_{m1} + 1} \right)^{q_3} \right\} \right]. \quad (3.53)$$

Toutes les notations utilisées sont vus dans les sections précédentes.

$$\begin{aligned} \beta_{2E}^{**} &= [\mathbb{E} [\beta_2^{-q_3}]]^{-1/q_3} \\ &= \left[\frac{k_4}{h_2(z)} \sum_{m=1}^{n-1} J_{2m} G_{2m} \right]^{-1/q_3}, \end{aligned} \quad (3.54)$$

où

$$\begin{aligned} J_{2m} &= \int_{-\infty}^{\infty} \beta_2^{-q_3} \exp \left\{ \frac{-1}{2\sigma^2} \beta_2^2 (s_{n1} - s_{m1} + 1) + \beta_2 \frac{s_{m4}}{\sigma^2} \right\} d\beta_2 \\ &= \frac{1}{((s_{n1} - s_{m1} + 1)/\sigma^2) (-1 + q_3)} \left\{ k_{6m} \cdot \Gamma \left(\frac{1 - q_3}{2} \right) {}_1F_1 \left(\frac{q_3}{2}, \frac{1}{2}, -\frac{s_{m4}^2/\sigma^2}{2(s_{n1} - s_{m1} + 1)} \right) \right\} \\ &\quad + \left\{ G_{5m} \Gamma \left(1 \frac{q_3}{2} \right) {}_1F_1 \left(\frac{1 + q_3}{2}, \frac{3}{2}, -\frac{s_{m4}^2/\sigma^2}{2(s_{n1} - s_{m1} + 1)} \right) \right\} \\ &\quad + G_{6m} {}_pF_q \left(\left\{ \frac{1}{2}, 1 \right\}, \left\{ 1 - \frac{q_3}{2}, \frac{3}{2} - \frac{q_3}{2} \right\}, \left\{ -\frac{s_{m4}^2/\sigma^2}{2(s_{n1} - s_{m1} + 1)} \right\} \right), \end{aligned} \quad (3.55)$$

avec

$$\begin{aligned} k_{6m} &= (-1)^{-q_3} 2^{(1/2)(-1-q_3)} \cdot \left\{ -\frac{(s_{n1} - s_{m1} + 1)^2}{s_{m4}^2} \right\}^{-q_3} \left\{ -\frac{s_{m4}^2}{(s_{n1} - s_{m1} + 1)^2} \right\}^{-q_3} \\ &\quad \times \exp \left\{ \frac{s_{m4}^2/\sigma^2}{2(s_{n1} - s_{m1} + 1)} \right\} \left[\left(\frac{(s_{n1} - s_{m1} + 1)}{\sigma^2} \right)^{q_3/2} (-1 + q_3) \right. \\ &\quad \left. \times \left\{ \sqrt{\frac{(s_{n1} - s_{m1} + 1)}{\sigma^2}} \left((-1)^{q_3} \left(-\frac{(s_{n1} - s_{m1} + 1)}{s_{m4}} \right)^{q_3} \left(-\frac{s_{m4}}{(s_{n1} - s_{m1} + 1)} \right)^{q_3} + 1 \right) \right\} \right] \end{aligned} \quad (3.56)$$

$$G_{5m} = \sqrt{2} \frac{s_{m4}}{\sigma^2} \left[(-1)^{q_3} \left(-\frac{(s_{n1} - s_{m1} + 1)}{s_{m4}} \right)^{q_3} \left(-\frac{s_{m4}}{(s_{n1} - s_{m1} + 1)} \right)^{q_3} - 1 \right] \quad (3.57)$$

$$G_{6m} = 2^{(1+q_3)/2} \left[\left\{ -\frac{(s_{n1} - s_{m1} + 1)^2}{s_{m4}^2} \right\}^{q_3} \frac{s_{m4}}{\sigma^2} \left\{ (-1)^{q_3} \left(\frac{s_{m4}}{(s_{n1} - s_{m1} + 1)} \right)^{q_3} - \left(\frac{s_{m4}}{(s_{n1} - s_{m1} + 1)} \right)^{q_3} \right\} \right]. \quad (3.58)$$

Remarque : Si on pose $q_3 = -1$ dans la formule (3.40) et dans (3.44), l'estimateur Bayésien de m coïncide avec la moyenne a posteriori associée à la fonction de coût quadratique donnée précédemment, on note également que pour $q_3 = -1$ la fonction coût entropie générale est égale à la fonction de coût quadratique.

La partie numérique et la comparaison des estimateurs Bayésiens obtenus sous différentes fonctions de coût de cette étude n'ont pas été faites vu que nous disposons de données fausses dans l'article prises d'un tableau de données d'un modèle de régression linéaire où les erreurs sont autocorrélées (table 4.1 de Zellner(1996)), c'est à dire, les données du tableau ne sont pas compatibles avec le modèle de régression linéaire à deux-phases. Également, les formules des différentes fonctions vues dans l'article sont aussi fausses, des corrections ont été apportées sur ces dernières dans ce mémoire.

Conclusion

Dans ce travail, nous avons vu les principaux outils nécessaires pour l'inférence classique et l'inférence Bayésienne entre autre l'utilisation des lois a priori et le choix des fonctions coût qui représentent des étapes cruciales pour l'estimation Bayésienne. Aussi, nous avons abordé les différents types de modèles de régression les plus connus, la régression linéaire et la régression logistique, ainsi que les méthodes d'estimation de leurs paramètres inconnus.

Dans ce mémoire, le modèle utilisé est le modèle de régression linéaire à deux-phases , c'est-à-dire où l'on constate un seul point de rupture. L'estimation bayésienne de ce point de rupture ainsi que les paramètres de régression sont étudiés sous les différentes fonctions coûts usuelles.

Ce travail est basé sur l'article de **Pandya et al.** (2011) qui peut mener à d'autres travaux plus poussés. Comme perspectives, nous pouvons citer une étude sur les différents choix de lois a priori pour une meilleure estimation. On peut également considérer l'utilisation d'autres fonctions de pertes comme la fonction de perte quadratique pondérée et la quadratique généralisée qui nous permettra de faire une comparaison directe avec les travaux de cet article. Un autre aspect peut être abordé qui est l'existence de plusieurs points de ruptures.

Bibliographie

- [1] Abramowitz and Irene A.Stegun, Handbook of mathematical functions with formulas, graphs, and mathematical tables, *Nationa Bureau of Standards Applied Mathematics Series.55* (1964).
- [2] Alan Agresti, Categorical data analysis, 2nd edition *Wiley Interscience* (2002).
- [3] Annette J.Dobson, An introduction to statistical modelling, *Springer-science+Business Media, B.V. (Originally published by Chapman and Hall in 1983* 2006.
- [4] Berger J.O, Statistical decision theory and Bayesian analysis, 2nd edition, *Springer-Verlag* (1985).
- [5] Birkes D. and Dodge Y., Alternative methods of régression, *Wiley* (1993).
- [6] Broemeling L.D. and Tsurumi H., Economics an structural change, *Marcel Dekker, New York,NY,USA* (1987).
- [7] Calabria R. and Pulcini G., An engineering approach to Bayes estimation for the Weibull diqtribution, *Microelectronics Reliability, vol. 34* (1994).
- [8] Chatterjee S. and Hadi A.S., regression analysis by exemple, 4th edition, *Wiley* (2006).
- [9] Chaturvedi A. et Shrivastava A. Bayesian analysis of a linear model involving structural changes in either regression parameters or disturbances precision
Communications in Statistics - Theory and Methods (2016).
- [10] Christensen R., Log-linear model and logistic regression, 2nd edition, *Springer* (1997).
- [11] Christian P.Robert, Le choix bayésien-Principe et pratique, *Springer* (2006).
- [12] Cornillon P.A. et Metzner-Lober E., Régression-Théorie et applications, *Springer* (2006).
- [13] Cowles M.K., Applied Bayesian statistics, *Springer-Verlag* (2013).
- [14] Holber D., A Bayesian analysis of a switchinglinear model : a known number of regimes, *Journal of the American statistical association, vol. 70* (1982).
- [15] McCullagh P. and Nelder J.A, Generelized linear model, 2nd edition, *Chapman and Hall* (1983).
- [16] Michel Lejeune, Statistique-Théorie et ses applications, 2ème édition, *Springer* (2010).
- [17] Pandya M., Bhatt K. and Anharia P., Bayes estimation of two-phases linear regression model, *International journal of quality, statistics and reliability* (2011).
- [18] Samaniego F.J., A Comparison of the Bayesian and frequentist approaches to estimation, *Springer series in statistics* (2010).
- [19] Van der Waerden B.L., Mathematical statistics, *Springer-Verlag* (1969).
- [20] Wikefield J., Bayesian and frequentist regression models, *Springer series in statistics* (2013).
- [21] Zellner A., An introduction to Bayesian inference in econometrics, *Wiley* (1996).