

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira de Béjaia
Faculté des Sciences Exactes
Département Mathématiques



Mémoire de Master
Option : Probabilité Statistique et
Applications

Thème

RÉGRESSION QUANTILE
BAYÉSIENNE

Présenté par :

M^r Barkat khaled

Devant le Jury :

Présidente : *Pr* Timeridjine Karima Université de Béjaia.

Examinatrice : *Dr* Saadi Nora Université de Béjaia.

Encadreur : *Dr* Lagha Karima Université de Béjaia.

Promotion 2020/2021

Remerciements

Tout d'abord, je souhaite remercier ma promotrice Mme. LAGHA KARIMA pour sa disponibilité, ses remarques constructives ainsi que remercie aussi les membres du jury d'avoir accepté d'examiner ton mémoire. pour l'intérêt porté à mon travail tout au long de cette année. Enfin, et surtout, merci à mes parents, ma famille, mes amis (ceux qui ne comptent que sur les doigts d'une seule main comme on dit) d'avoir été et d'être toujours présent , et ce quelque soit les circonstances. Je sais que cela n'a pas dû être facile pour vous de devoir supporter la maladie, les incertitudes, mes sautes d'humeur, mes angoisses. Merci pour tous...

Dédicaces

A mes chers parents, pour tous leurs sacrifices, leur amour, leur soutien

et leurs prières tout au long de mes études,

A toute ma famille pour leur soutien tout au long de mon parcours

universitaire,

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le

fruit de votre soutien infailible,

Merci d'être toujours là pour moi.

Table des matières

Introduction générale	3
1 Inférence bayésienne	5
1.1 Introduction	5
1.2 Principales lois de probabilités utiles dans l'approche bayésienne	5
1.3 Bases décisionnelle de l'analyse bayésienne	6
1.4 Estimation bayésienne	8
1.5 Approche bayésienne des tests	10
1.6 Modélisation de l'information a priori	12
1.7 Méthodes de simulation de Monte-Carlo	15
2 Les modèles de régression	17
2.1 Introduction	17
2.2 Modèle de régression linéaire	18
2.2.1 Régression linéaire simple	18
2.2.2 La régression linéaire multiple	21
2.2.3 Limites du modèle linéaire	24
2.3 Régression quantile (RQ)	24
2.3.1 Définition de Quantile	26
2.3.2 Exemple de modèle	28
2.4 Le modèle de régression quantile standard	33
2.4.1 Estimation des paramètres du modèle	34
2.4.2 Propriétés des estimateurs	36
2.4.3 Tests sur les quantiles	40
2.4.4 Cas particuliers	44
3 Régression quantile bayésienne	49
3.1 Introduction	49
3.2 Le modèle de régression quantile bayésienne	49
3.3 APPLICATION :	55

TABLE DES MATIÈRES

3.3.1	Présentation des données	55
3.3.2	Étude du quantile conditionnel d'ordre $\tau = 0.75$	55
3.3.3	Étude du quantile conditionnel d'ordre $\tau = 0,95$	59
3.3.4	Étude du quantile conditionnel d'ordre $\tau = 0,5$	61
	Conclusion	64
	Annexe	67
	Bibliographie	77

Introduction générale

La régression quantile est un type de régression utilisé en statistiques. Elle prend de plus en plus d'importance, comme le montre le nombre croissant d'articles publiés dans les revues scientifiques[11].

Dans la régression linéaire le coefficient de régression représente le changement opéré dans la variable réponse produit par une unité de changement dans la variable prédictive associée à ce coefficient. Le paramètre de régression quantile procure une estimation du changement dans un quantile spécifique de la variable réponse produit par une unité de changement de la variable prédictive.

En ce qui concerne la régression quantile, on peut dire qu'elle étend la notion de régression ordinaire aux quantiles de la variable expliquée [13]. Cela nous donne plus d'informations sur la distribution de cette variable, car les quantiles sont des points particuliers pris à intervalles réguliers de la fonction de distribution d'une variable aléatoire. Nous rappelons que la régression ordinaire est un modèle de moyenne conditionnelle, où le conditionnement porte sur les variables explicatives. De manière similaire, la régression quantile est un modèle pour les quantiles conditionnels.

Le principal avantage de la régression quantile par rapport à la régression linéaire est sa flexibilité dans la modélisation de données avec des distributions conditionnelles hétérogènes. De nombreux chercheurs ont travaillé sur ce thème, notamment Roger Koenker[35], Yue et al. et plusieurs statisticiens poursuivent leurs recherches et leur développement.

L'objectif de la régression quantile diffère de celui de la régression linéaire classique. Un modèle de régression linéaire exprime la variable dépendante

comme une combinaison linéaire de certaines covariables fixes plus un terme d'erreur aléatoire. Ainsi, la régression linéaire s'ajuste la moyenne de la réponse, car il est supposé que le terme d'erreur a une espérance égale à zéro. La régression quantile étudie les quantiles plutôt que l'espérance de la variable de réponse. Les méthodes d'estimation de ces modèles peuvent être classées en paramétriques et non paramétriques. La méthode d'estimation dépend de la généralisation du critère d'erreur absolue pondérée symétriquement (voir Koenker et Zhao (1994)[38]. Dans le cas non-paramétrique, les méthodes les plus utilisées dépendent des β -splines et de l'approche localement polynomiale Yu et Jones [54].

Ce mémoire se veut être une présentation de deux approches de la régression quantile, à savoir l'approche classique et l'approche bayésienne dans le cas paramétriques. Dans le première chapitre nous présentons le principe de l'inférence bayésienne, la méthode d'estimation bayésienne et les algorithmes de simulation de Monte Carlo.

Dans le deuxième chapitre, nous abordons les modèles de régression linéaire simple et multiple, ensuite nous donnons quelques définitions du quantile et quelques modèles de régression quantile (RQ). Ensuite, nous discutons des méthodes d'estimation paramétriques classiques et nous terminons le chapitre par quelques cas particuliers avec leur interprétation générale.

Dans le troisième chapitre, nous développons la régression quantile bayésienne (RQB) en utilisant une fonction de vraisemblance basée sur la distribution de Laplace asymétrique (voir l'annexe pour la définition et les propriétés). Il est démontré que, quelle que soit la distribution originale des données, l'utilisation de la distribution asymétrique de Laplace est une manière très naturelle et efficace de modéliser la RQB par rapport à d'autres distributions [57]. Le chapitre se termine par une application dans laquelle nous modélisons le problème par la RQB et nous présentons les résultats de simulation par des graphiques et des tableaux pour les quantiles ($q=0.5; 0.75$ et 0.95). Le document se termine par une conclusion et quelques perspectives.

Chapitre 1

Inférence bayésienne

1.1 Introduction

Considérons un modèle statistique $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$ où θ est un paramètre inconnu à estimer. La démarche bayésienne consiste à traiter le paramètre inconnu θ comme une variable aléatoire (v.a.), en lui associant une loi de probabilité sur l'espace des paramètres Θ , dite loi a priori, notée $\pi(\theta)$. Cette loi reflète la connaissance a priori (éventuelle) du paramètre [23]. Le modèle bayésien est alors $(\mathcal{X}, \mathcal{A}, P_\theta, \pi(\theta), \theta \in \Theta)$.

Soit X une v.a. de probabilité P_θ dépendant du paramètre inconnu θ à estimer de densité f_θ avec $f_\theta(x) = f(x | \theta)$. Dans le modèle bayésien on interprète la densité f_θ comme la loi conditionnelle par rapport à θ .

Le théorème de Bayes est fondamental dans l'inférence statistique bayésienne, permettant de mettre à jour un état de connaissance a priori (i.e., avant d'observer certaines données) en un état de connaissance a posteriori (i.e., après avoir observé ces données).

1.2 Principales lois de probabilités utiles dans l'approche bayésienne

a- Loi a priori

La loi a priori de θ notée par $\pi(\theta)$ est une loi de probabilité qui modélise toute l'information disponible sur le paramètre d'intérêt θ , avant le recueil des observations, sa détermination est l'essence de la statistique

bayésienne.

b- La loi a posteriori

La loi a posteriori décrit le comportement de la variable aléatoire θ sachant l'observation x . Sa densité est donnée par :

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

Dans ce contexte bayésien on a la notion de proportionnalité, c'est-à-dire :

$$\pi(\theta | x) \propto f(x | \theta)\pi(\theta) = g(\theta, x)$$

où $g(\theta, x) = f(x | \theta)\pi(\theta)$ est la densité de probabilité du couple (θ, x) .

c- La loi marginale de X

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

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

Remarque 1.2.1 Lorsque'on manipule un échantillon $\underline{X} = (X_1, \dots, X_n)$, on remplace dans les expressions précédentes la v.a. X par \underline{X} et la réalisation x par $\underline{x} = (x_1, \dots, x_n)$ et $f_{\theta}(x)$ par la vraisemblance $f_{\theta}(\underline{x}) = L(\theta, \underline{x})$.

1.3 Bases décisionnelle de l'analyse bayésienne

En pratique l'inférence statistique conduit à une décision finale prise par le décideur et il est important de pouvoir comparer les différentes décisions au moyen d'un critère d'évaluation, qui va apparaitre sous forme de fonction de coût (perte) qui est une fonction mesurable de $(\Theta \times \mathcal{D})$ à valeurs dans \mathbb{R}_+ notée par $\ell(\theta, \delta(x))$.

Soit \mathcal{D} l'ensemble des règles de décisions δ , qui sont des applications de \mathcal{X} dans \mathcal{A} , on note $a = \delta(x)$ est une estimation. Le but est de trouver une règle de décision $\delta \in \mathcal{X}$.

1- Fonctions de coût

Les fonctions de coût les plus utilisées sont

1-1 Le coût quadratique

La fonction de coût quadratique est la fonction définie par :

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

1-2 Le coût absolu

La fonction de coût absolu ou le coût L^1 est la fonction définie par :

$$\ell(\theta, \delta(x)) = |\theta - \delta(x)|$$

1-3 Le coût 0 – 1

La fonction de perte 0-1 est l'application ℓ définie par :

$$\ell(\theta, \delta(x)) = \begin{cases} 0 & \theta \in \Theta_0 \\ 1 & \theta \in \Theta_1 \end{cases}$$

où

$$\Theta_0 \cup \Theta_1 = \Theta \quad \text{et} \quad \Theta_0 \cap \Theta_1 = \emptyset$$

Fonctions de risque

– Risque fréquentiste

Pour une fonction de perte donnée $\ell(\delta, \theta)$, le risque fréquentiste est le coût moyen (l'espérance mathématique) du coût d'une règle de décision qui est défini par :

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

– Risque intégré

Pour une fonction de perte donnée, on définit le risque intégré comme étant le risque fréquentiste moyenné sur les valeurs de θ selon leurs distributions a priori π noté $R^*(\pi, \delta)$, par :

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

– **Risque a posteriori**

Le risque a posteriori noté par $R^*(\pi, \delta | x)$ est défini comme étant la moyenne du coût par rapport à la loi a posteriori :

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

– **Risque de bayes**

Le risque de bayes pour la décision $\delta \in \mathcal{D}$ est la quantité donnée par :

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

Dans le cas ou $r(\pi, \delta) < \infty$, la décision

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

où $R^*(\pi, \delta | x)$ est le risque a posteriori, est appelée estimateur bayésien.

– **Risque minimax**

On appelle risque minimax (minimum du risque maximum) associé à la fonction de coût ℓ , la valeur :

$$\bar{R} = \inf_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta) = \inf_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} \mathbb{E}_\theta[\ell(\theta, \delta(x))]$$

1.4 Estimation bayésienne

On appelle estimateur de bays associé à une fonction de coût $\ell(\theta, \delta(x))$, la décision $\delta^\pi(x)$ qui minimise le risque a posteriori de θ , c'est-à-dire : :

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

Parmi les fonction de coût usuelles, on peut citer :

1. **L'estimateur MMSE** (Minimum Mean Square Error)[46]

L'estimateur MMSE de θ , noté $\hat{\theta}_{MMSE}(x)$ est l'estimateur qui mini-

mise le coût a posteriori relativement à la perte quadratique moyenne

$$R(\theta, \hat{\theta}) = \mathbb{E} \left[(\hat{\theta} - \theta)^2 \right],$$

avec $\hat{\theta} = \hat{\theta}(x)$ un estimateur de θ .

Il est défini par :

$$\hat{\theta}_{MMSE}(x) = \mathbb{E}(\theta | x), \quad x \in \mathbb{R}$$

L'estimateur *MMSE* est donc la moyenne a posteriori de θ .

2. Estimateur MAP (Maximum A Posteriori)

L'estimateur MAP de θ est obtenu par maximisation de la loi a posteriori, c'est-à-dire :

$$\hat{\theta} = \arg \max_{\theta} \pi(\theta | x) = \arg \max_{\theta} \ln \pi(\theta | x).$$

Cet estimateur est associé au coût $0 - 1$.

3. La médiane a posteriori

L'estimateur de bays associe á la loi a priori π et á la fonction de coût absolu est le fractile d'ordre $\frac{1}{2}$ de la loi a posteriori .

Alors le médiane a posteriori donnée par

$$p(\theta/\delta) = p(\theta < \delta | x) = \frac{1}{2}$$

L'estimateur de bays associe á la loi a priori π est le fractile d'ordre $\frac{k_2}{k_1 + k_2}$ de $\pi(\theta | x)$. c-á-d :

$$\frac{k_2}{k_1 + k_2} = p(\theta | \delta)$$

Propriétés de l'estimateur de Bayes

1. L'estimateur de Bayes est admissible.

2. L'estimateur de Bayes est biaisé.

Sous certaines hypothèses de régularité le plus souvent satisfaites en pratique, on a les deux propriétés :

a L'estimateur de Bayes est convergent en probabilité (quand la taille de l'échantillon $n \rightarrow +\infty$).

b La loi a posteriori peut être asymptotiquement (c.a.d. pour de grandes valeurs de n) approximée par une loi normale $\mathcal{N}(E[\theta | \mathbf{x}], \text{Var}[\theta | \mathbf{x}])$. Cette dernière propriété est particulièrement utile pour construire des intervalles de confiance a posteriori [15].

1.5 Approche bayésienne des tests

On veut tester l'hypothèse suivant :

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

où

$$\Theta_0 \cup \Theta_1 = \Theta \quad \text{et} \quad \Theta_0 \cap \Theta_1 = \emptyset$$

Par définition, les décisions bayésiennes sont celles qui minimisent le coût a posteriori $R^*(\pi, \delta | x)$

On a deux décisions possibles :

d_0 : On accepte $H_0 : \theta \in \Theta_0$ et d_1 : On rejette H_0 .

En pratique, on accepte l'hypothèse H_0 ou H_1 dès que sa probabilité a posteriori $\varrho_0 = \pi(\theta \in \Theta_0)$ ou $\varrho_1 = \pi(\theta \in \Theta_1) = 1 - \varrho_0$, respectivement, est forte c'est à dire supérieure à 0,9 ou 0,95 .

1. Le facteur de bayes

Le facteur de Bayes est le rapport des probabilités a posteriori des hypothèses nulle et alternative sur le rapport des probabilités a priori de ces même hypothèses, soit

$$B(x) = \frac{P(\theta \in \Theta_0 | x)}{P(\theta \in \Theta_1 | x)} / \frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)} .$$

Ce rapport évalue la modification de la vraisemblance de l'ensemble Θ_0 par rapport à celle de l'ensemble Θ_1 dûe à l'observation.

Cas particulier

Soit $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$. Le facteur de Bayes se simplifie et devient le rapport de vraisemblance classique

$$B(x) = \frac{\int_{\Theta_0} f(x | \theta_0) \pi_0(\theta) d\theta}{\int_{\Theta_1} f(x | \theta_1) \pi_1(\theta) d\theta} = \frac{m_0(x)}{m_1(x)}$$

Comme indiqué précédemment le facteur de Bayes est un point de vue décisionnel, complètement équivalent à la probabilité a posteriori de l'hypothèse nulle puisque

$$B(x) = \frac{a_1 / \varrho_0}{a_0 / \varrho_1} = \frac{a_1 \varrho_1}{a_0 \varrho_0}$$

où

$$a_0 = \int_{\Theta_0} \pi(\theta \in \Theta_0) \quad \text{et} \quad a_1 = \int_{\Theta_1} \pi(\theta \in \Theta_1) .$$

Echelle de Jeffreys :

2. Intervalles de confiances bayésiens

L'approche bayésienne présente l'avantage de permettre une construction directe d'une région de confiance, deux types d'intervalles seront définis par la suite, qui sont présentés comme suit :

– Intervalle de confiance a priori

Un intervalle de confiance a priori J de niveau $1 - \alpha$, l'intervalle donné par :

$$\mathbb{P}(\theta \in J) = \int_J \pi(\theta) = 1 - \alpha,$$

– **Intervalle de confiance a posteriori**

Un intervalle de confiance a posteriori I , de niveau $1 - \alpha$, l'intervalle donné par :

$$\mathbb{P}(\theta \in I | X) = \int_I \pi(\theta | X) = 1 - \alpha,$$

1.6 Modélisation de l'information a priori

Le choix de la loi a priori est une étape fondamentale dans l'analyse bayésienne pour ce la il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori, dans le sens où plusieurs lois de probabilité peuvent être compatibles avec cette information.

En pratique, parfois on utilise comme loi a priori les lois usuelles (lois normales, lois gamma, etc) ou des lois dites conjuguées, l'information a priori étant alors utilisée pour déterminer les paramètres de la loi a priori, appelés hyperparamètres. Le choix de la loi a priori étant cruciale, il existe principalement deux méthodes : subjective et informative, nous ne parlerons que du cas informatif.

– **Lois conjuguées naturelles**

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. Ce calcul est facilité lorsque la loi a priori et la loi a posteriori ont la même forme. On parle dans ce cas de la loi a priori conjuguée.

Définition 1.6.1 *Supposant que la loi des observations est connue [49], on se donne une famille \mathcal{F} de lois de probabilité sur Θ . On suppose que la loi a priori $\pi(\theta)$ appartient à \mathcal{F} . Si dans ces conditions, la loi a posteriori $\pi(\theta | X)$ appartient toujours à \mathcal{F} , on dit que la loi a priori est conjuguée naturelles.*

Il y a plusieurs des exemples sur les lois conjuguée naturelles comme la famille des lois Beta est conjuguée pour des vraisemblances binomiales ou la famille des lois de Dirichlet est conjuguée pour des vraisemblances multinomiales.

– **Lois a priori non informatives**

Une loi non informative est une loi qui porte une information sur le paramètre à estimer dont le poids dans l'inférence est réduit. Aussi est une démarche à mettre en place en l'absence d'information a priori.

Définition 1.6.2 Soit θ un paramètre réel. On appelle loi a priori non informative de Jeffreys est invariant par transformation bijective, la loi (éventuellement impropre) de densité

$$\pi_J(\theta) \propto [I_X(\theta)]^{\frac{1}{2}} \mathbb{I}_{\Theta}(\theta)$$

où $I_X(\theta)$ désigne l'information de Fisher apportée par X sur θ .

Ou encore, $\pi_J(\theta) = k\sqrt{I_X(\theta)}$, k est une constante, alors le paramètre θ est définie par

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

si le domaine de X est indépendant de θ alors

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

.

On peut justifier que l'a priori de Jeffreys offre une méthode automatisée pour obtenir un a priori non informatif pour n'importe quel modèle paramétrique.

– **Lois invariantes et a priori de Laplace**

Laplace a été le premier à utiliser des techniques non-informatives, puisque, bien que ne disposant d'aucune information a priori pour les paramètres qu'il étudiait, il munit ces paramètres d'une loi qui prend en compte son ignorance en donnant la même vraisemblance à chaque

valeur possible, soit donc en utilisant une loi uniforme. Son raisonnement, appelé plus tard principe de la raison insuffisante, se fondait sur l'équiprobabilité des événements élémentaires. Il y a plusieurs cas dans le critère de Laplace qui sont définies dans [1].

Il s'agit de trouver une distribution a priori invariante par certaines opérations. Si la famille de distributions est stable par une classe de transformations, on pourra chercher une distribution a priori invariante par cette famille de transformations. Il y a deux cas de la loi a priori invariante, les cas la loi a priori invariante est invariance par translation ou invariance par changement d'échelle [60].

L'approche invariante n'est que partiellement satisfaisante car elle implique la référence à une structure d'invariance qui peut parfois être choisie de plusieurs manières, ne pas exister, ou être sans intérêt pour le décideur.

– Maximum d'entropie

Si certaines caractéristiques de la loi a priori sont connues (moments, quantiles, etc), on suppose qu'elles peuvent s'écrire comme des espérances a priori,

$$E^\pi[g_k(\theta)] = w_k, \text{ avec } k = 1, \dots, K$$

où g_k est une fonction de θ , $\theta \in \mathbb{R}^n$.

La définition de l'entropie dans un cadre fini et discret donnée par

$$\text{Ent}(\pi) = - \sum_{i=1}^n \pi(\theta_i) \log(\pi(\theta_i))$$

La maximisation de l'entropie permet de chercher la loi qui apporte le maximum d'information. Le principe à la base de cette méthode est donc de chercher à calculer,

$$\arg \min_{\pi} \text{Ent}(\pi)$$

Plus d'information sera présenté sur [43].

1.7 Méthodes de simulation de Monte-Carlo

Le terme méthode de Monte-Carlo, ou méthode Monte-Carlo, désigne une famille de méthodes algorithmiques visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes.

Les méthodes de Monte Carlo permettent d'estimer des quantités en utilisant la simulation de variables aléatoires [27]. Les problèmes pouvant être rencontrés comprennent le calcul d'intégrales, les problèmes d'optimisation et la résolution de systèmes linéaires. La simplicité, la flexibilité et l'efficacité pour les problèmes en grande dimension de la méthode en font un outil intéressant, pouvant servir d'alternative ou de référence pour d'autres méthodes numériques[4].

– Algorithme de Metropolis-Hastings

L'algorithme de Metropolis-Hastings, proposé par METROPOLIS et collab. [1953] et généralisé par HASTINGS [1970] pour engendrer une chaîne de Markov qui satisfasse le principe du bilan détaillé, est un des plus simples algorithmes Monte Carlo. Il est en principe applicable à tout système. Il est extrêmement facile à implémenter pour échantillonner une densité cible, Pour une revue complète concernant cette méthode consulter [6] et [53] pour une densité $\pi(\theta | x)$ connue a une constante prés et une loi conditionnelle $q(./.),$ appelée loi de proposition choisie symétrique au sens ou $q(\ddot{\theta}/\theta) = q(\theta/\ddot{\theta}).$

Pour $\theta^{(0)}$ est une valeur initiale, on définit par récurrence les valeurs de $\theta^{(k)}$

A l'étape $k,$ à partir de $\theta^{(k-1)}, \theta^{(k)}$ est construit en tirant un θ' à l'aide d'une distribution de probabilité instrumentale : $\theta' \sim q(| \theta^{(k-1)}) . \theta^{(k)}$ est donné par :

$$\theta^{(k)} = \begin{cases} \theta' & \text{avec une probabilité } \alpha(\theta', \theta^{(k-1)}) \\ \theta^{(k-1)} & \text{avec une probabilité } 1 - \alpha(\theta', \theta^{(k-1)}) \end{cases}$$

avec

$$\alpha(\theta', \theta^{(k-1)}) = \min \left(\frac{\pi(\theta')}{\pi(\theta^{(k-1)})} \frac{q(\theta^{(k-1)} | \theta')}{q(\theta' | \theta^{(k-1)})}, 1 \right)$$

– l'Algorithme de Gibbs

Dans le cadre bayésien, [49] l'algorithme de Gibbs va permettre d'obtenir une réalisation du paramètre $\theta = (\theta_1, \dots, \theta_P)$ suivant la loi a posteriori $\pi(\theta/x)$ dès que l'on est capable d'exprimer les lois conditionnelles $\pi(\theta_i/\theta_j; x), j \neq i$.

L'échantillonnage de Gibbs consiste à : Partant d'un vecteur initial $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_P^{(0)})$ où $P \geq 2$ est la dimension de $\theta^{(0)}$. A la $(k+1)^{\text{ième}}$ étape, disposant du vecteur $\theta^{(k)} = (\theta_1^{(k)}, \dots, \theta_P^{(k)})$, on écrit :

- simuler $\theta_1^{(k+1)} \sim \pi(\theta_1/\theta_2^{(k)}, \dots, \theta_P^{(k)}; x)$
- simuler $\theta_2^{(k+1)} \sim \pi(\theta_2/\theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_P^{(k)}; x)$
- ⋮
- simuler $\theta_P^{(k+1)} \sim \pi(\theta_P/\theta_1^{(k+1)}, \dots, \theta_{P-1}^{(k+1)}; x)$

Cette chaîne admet une mesure invariante qui est la loi a posteriori. Pour un nombre d'itérations suffisamment grand, le vecteur θ obtenu peut donc être considéré comme étant une réalisation de la loi a posteriori.

Chapitre 2

Les modèles de régression

2.1 Introduction

En statistique, le mot régression désigne un type de modèles bien précis. Or, la notion de modèle est absolument essentielle dans toute étude statistique. La régression est l'un des outils, les plus utilisés en statistique. Elle est très pratique lorsqu'on s'intéresse à la relation entre une variable réponse Y et une ou plusieurs covariables X (quantitative ou qualitative). La régression peut aussi être utilisée pour prédire la valeur de la variable réponse, à partir de valeurs connues d'une ou plusieurs covariables. Les applications de la régression, couvrent la plupart des domaines, par exemple dans le domaine de la santé, économie et éducation, ... ect.

Le modèle de régression s'en sert de plus en plus comme d'un simple point de départ de l'analyse complète, il demeure l'outil de référence pour commencer toute recherche empirique. Il est aussi une sorte de lentille à travers laquelle on voit des relations entre les variables, en général ces modèles sont construits dans le but d'expliquer la variance d'un phénomène (variable dépendante) à l'aide d'une combinaison de facteurs explicatifs (variables indépendantes).

Si on s'intéresse à la relation entre deux variables, on parlera de la régression simple en exprimant une variable en fonction de l'autre. Si la relation porte entre une variable et plusieurs autres variables, on parlera de la régression multiple. La mise en oeuvre d'une régression impose l'existence d'une relation de cause à effet entre les variables prises en compte dans le

modèle.

Il existe plusieurs types de régressions paramétriques et non paramétriques chaque type ayant son importance et ses conditions d'application, au sein de ce chapitre nous présenterons en détail deux types de modèles paramétriques : la régression linéaire et la régression quantile.

2.2 Modèle de régression linéaire

2.2.1 Régression linéaire simple

La régression linéaire est un modèle simple et facile d'utilisation d'une variable expliquée Y sur une ou plusieurs variable X explicatives dans lequel on fait l'hypothèse que la fonction de lien. On suppose un échantillon de n points (x_i, y_i) du plan. $f(x)$ est linéaire dans ses paramètres [9] c'est à dire :

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

- les coefficients β_1 et β_0 sont les paramètres inconnus (mais non aléatoires!) du modèle,
- Les quantités ϵ_i viennent du fait que les points ne sont jamais parfaitement alignés sur une droite. On les appelle les erreurs (ou bruits) et elles sont supposées aléatoires. Il y a deux hypothèses principales sur les erreurs :

$$\begin{cases} (\mathcal{H}_1) : \mathbb{E}[\epsilon_i] = 0 \text{ pour tout indice } i = 1, \dots, n \\ (\mathcal{H}_2) : \text{Cov}(\epsilon_i, \epsilon_j) = \delta_{ij} \sigma^2 \text{ pour tout couple } (i, j), \sigma^2 = \text{Var}(\epsilon) < \infty \end{cases}$$

Les erreurs $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ sont supposées centrées et de même variance (homoscédasticité) et non corrélées entre elles (δ_{ij} est le symbole de Kronecker, i.e. $\delta_{ij} = 1$ si $i = j$, $\delta_{ij} = 0$ si $i \neq j$).

Notons que le modèle de régression linéaire simple peut encore écrire sous la forme matricielle suivante :

$$Y = X\beta + \epsilon \quad (1)$$

où :

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\text{et } \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

avec

- Y désigne le vecteur à expliquer de taille $n \times 1$,
- X la matrice explicative de taille $n \times 2$
- ϵ le vecteur d'erreurs de taille $n \times 1$.
- β le vecteur des paramètres inconnus du modèle

Estimation de paramètres

Définition 2.2.1 On appelle estimateurs des Moindres Carrés Ordinaires (en abrégé MCO) des paramètres β_0 et β_1 les valeurs minimisant la quantité :

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^n \epsilon_i^2$$

Autrement dit, la droite des moindres carrés minimise la somme des carrés des distances verticales des points (x_i, y_i) du nuage à la droite ajustée $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ [26].

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

obtenu en résolvant le système des deux équations à deux inconnues suivant :

$$\nabla S(\hat{\beta}_0, \hat{\beta}_1) = 0 \Leftrightarrow \begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{S_{xy}}{S_{xx}} \end{cases}$$

avec

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

Les estimateurs des MCO vérifient les propriétés suivantes :

1. $\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs linéaires en $(Y_i)_{i=1\dots n}$, sans biais de β_0 et β_1 respectivement ;
2. $\text{Var}(\hat{\beta}_0) = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2$
3. $\text{Var}(\hat{\beta}_1) = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2$;
4. $\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2$.

Estimation de variance σ^2

Un estimateur sans biais de σ^2 est donnée par

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} = \frac{\sum_{i=1}^n (\varepsilon_i)^2}{n - 2} = \frac{SCR}{n - 2}$$

Où $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$.

Coefficient de détermination

On peut juger la qualité de l'ajustement linéaire par le coefficient de détermination R^2 qui mesure la proportion de la variance totale de Y expliquée par la régression qui peut être prise en compte par les variables, Y qui est donné par :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{i=1}^n \varepsilon_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (0 \leq R^2 \leq 1)$$

On retrouve la propriété fondamentale $SCT = SCE + SCR$ qui permet de mesurer l'ajustement du modèle par le coefficient de détermination.

$$- SCE = \sum_{j=1}^n (\hat{y}_i - \bar{y})^2$$

$$\begin{aligned}
- \text{SCR} &= \sum_{i=1}^h \epsilon_i^2 \\
- \text{SCT} &= \sum_{i=1}^n (y_i - \bar{y})^2
\end{aligned}$$

De façon générale, l'interprétation est la suivante : le modèle de régression linéaire permet d'expliquer $(100 \times R^2)$ % de la variance totale des données.

2.2.2 La régression linéaire multiple

Le modèle de régression linéaire multiple est une généralisation du modèle de régression linéaire simple lorsque les variables explicatives sont en nombre fini (≥ 2). Nous supposons donc que les données collectées suivent le modèle suivant :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n$$

où

- ϵ_i est une variable aléatoire, non observée, $i = 1, \dots, n$
- x_{ij} est observé et non aléatoire, $i = 1, \dots, n, j = 1, \dots, p$
- les paramètres β_j du modèle sont inconnus, mais non aléatoires ;

Il est convenable d'écrire le modèle de régression linéaire sous sa forme matricielle :

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

avec

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} \quad b = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{et } \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

La matrice X est appelée matrice du plan d'expérience. On suppose que cette matrice est de rang p . Ses vecteurs colonnes sont linéairement indépendants. Cela implique en particulier que la matrice symétrique $X'X$ est

définie positive.

Hypothèses du modèle :

- H_1 : Les variables explicatives sont observées sans erreurs (donc non aléatoires).
- H_2 : $\mathbb{E}(\epsilon_i) = 0$, pour $i = \overline{1..n}$, erreurs centrées.
- H_3 : $\mathbb{E}(\epsilon_i^2) = \sigma_e^2$, la variance de l'erreur est constante (hypothèse d'homoscédasticité).
- H_4 : $\mathbb{E}(\epsilon_i \epsilon_j) = 0 \quad \forall i \neq j$, les erreurs sont non corrélées.
- H_5 : $\text{Cov}(\epsilon_i, x) = 0$, pour $i = \overline{1..n}$, l'erreur est indépendante des variables explicatives.
- H_6 : $\epsilon_i \sim \mathcal{N}(0, \sigma_e^2)$, $i = \overline{1..n}$

Si l'une de ces hypothèses n'est pas respectées, donc on ne peut pas appliquer le modèle.

Le modèle de régression linéaire est estimé par plusieurs méthodes telles que : la méthode des moindres carrés, la méthode robuste (robust fit) et l'ajustement par étape, la méthode classique utilisée dans cette partie sera celle des Moindres Carrés Ordinaires (MCO).

Estimation des paramètres du modèle

A partir de l'échantillon (aléatoire) de n observations

$$\{(x_{i1}, \dots, x_{ip}, y_i), i = 1, \dots, n\}$$

on veut estimer les paramètres

$$\beta_0, \beta_1, \dots, \beta_p \text{ et } \sigma^2$$

Pour estimer $\beta = (\beta_0, \beta_1, \dots, \beta_p)$, on peut utiliser la méthode des moindres carrés ordinaires qui ne nécessite pas d'hypothèse supplémentaire sur la distribution de ϵ_i , L'estimateur des MCO des coefficients du modèle s'écrit :

$$\hat{\beta} = (X'X)^{-1} X'Y \quad .$$

l'estimateur sans biais des moindres carrés de la variance des résidus $\hat{\sigma}_\epsilon^2$ est défini comme suit[8] :

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n - (p + 1)} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{SCR}{n - p - 1} .$$

Par conséquent, l'estimateur de la matrice de variances-covariances des $\hat{\beta}$ est donnée par :

$$Var(\hat{\beta}) = \hat{\sigma}_\epsilon^2 (X'X)^{-1} .$$

Coefficient de détermination

La propriété fondamentale d'analyse de la variance $SCT = SCE + SCR$ permet de mesurer l'ajustement du modèle par le coefficient de détermination. Le coefficient de détermination R^2 qui mesure la proportion de la variance totale de Y expliquée, donné par :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (0 \leq R^2 \leq 1)$$

avec $\hat{\epsilon}_i = y_i - \hat{y}_i$.

Tests statistiques

Le test de Fisher

$$\mathcal{H}_0 : " \beta_1 = \dots = \beta_p = 0 " \text{ contre } \mathcal{H}_1 : " \exists j \in \{1, \dots, p\}, \beta_j \neq 0 "$$

est fondé sur la statistique suivante :

$$F^* = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 / p}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - p - 1)} = \frac{SCE/p}{SCR/(n - p - 1)} .$$

Il permet de juger la qualité globale de l'ajustement . Sous \mathcal{H}_0 , F_n suit une loi de Fisher à p et $n - p - 1$ degrés de libertés. \mathcal{H}_0 avec un risque $0 \leq \alpha \leq 1$ si

$$F^* \geq f_{(p, n-p-1)},$$

- Si $F^* > F_{(p, n-p-1, \alpha)}$ alors on rejette l'hypothèse nulle H_0 donc on accepte l'hypothèse alternative H_1 , le modèle est globalement significatif. D'où il ya au moins une variable explicative significative de Y .
- Si $F^* \leq F_{(p, n-p-1, \alpha)}$ on accepte H_0 , alors le modèle est rejeté. D'où , il n' y a aucune variable explicative significative de Y .

Intervalle de confiance

Au seuil de confiance α , l'intervalle de confiance pour chaque coefficient de régression est donné par :

$$IC_{\beta_j} = \left[\hat{\beta}_j - \hat{\sigma}_{\beta_j} t_{\left(\frac{\alpha}{2}, n-p-1\right)}; \hat{\beta}_j + \hat{\sigma}_{\beta_j} t_{\left(\frac{\alpha}{2}, n-p-1\right)} \right],$$

avec $t_{\left(\frac{\alpha}{2}, n-p-1\right)}$ représente le quantile d'ordre $\frac{\alpha}{2}$ de la loi de student à $(n - p - 1)$ ddl. et $\hat{\sigma}_{\beta_j}$ est l'écart type estimé de $\hat{\beta}_j$, $j = \overline{1, \dots, p}$.

2.2.3 Limites du modèle linéaire

La principale limite de la régression linéaire est l'hypothèse de linéarité entre la variable dépendante et les variables indépendantes. Dans le monde réel, les données sont rarement séparables linéairement. Elle suppose qu'il existe une relation linéaire entre les variables dépendantes et indépendantes, ce qui est souvent incorrect.

Sujet au bruit et au sur-ajustement : Si le nombre d'observations est inférieur au nombre de caractéristiques, la régression linéaire ne doit pas être utilisée, sinon elle peut conduire à un ajustement excessif car elle commence à prendre en compte le bruit dans ce scénario lors de la construction du modèle.

Sujet à la multicollinéarité : Avant d'appliquer la régression linéaire, la multicollinéarité doit être éliminée (à l'aide de techniques de réduction de la dimensionnalité) car elle suppose qu'il n'y a pas de relation entre les variables indépendantes. Le modèle linéaire ne permet pas de modéliser des variables de réponse discrètes. En résumé, la régression linéaire est un excellent outil pour analyser les relations entre les variables, mais elle n'est pas recommandée pour la plupart des applications pratiques car elle simplifie trop les problèmes du monde réel en supposant une relation linéaire entre les variables.

2.3 Régression quantile (RQ)

La régression quantile s'impose progressivement comme une approche globale de l'analyse statistique des modèles de réponse linéaires et non li-

néaires. La régression quantile permet d'étudier les effets des covariables, non seulement sur la moyenne d'une variable de réponse, mais aussi sur la distribution conditionnelle complète de la réponse compte tenu des covariables.

La régression quantile évite certaines hypothèses restrictives du modèle linéaire car elle ne nécessite pas d'homoscédasticité ou un type spécifique de distribution pour les réponses (ou de manière équivalente les termes d'erreur).

La régression sur les quantiles peut se faire sous plusieurs angles et nécessite beaucoup de notions statistiques c'est un outil statistique dont l'objet est de décrire l'impact de variables explicatives sur une variable d'intérêt. Elle permet une description plus riche que les régressions linéaires classiques. L'usage des régressions quantiles s'est beaucoup répandu au cours de la dernière décennie. Elles peuvent être aujourd'hui effectuées aisément avec de nombreux logiciels statistiques (R, SAS, STATA ... ect).

Dans les applications, les quantiles de régression qui décrivent les observations "extrêmes" en termes de covariables présentent souvent un intérêt réel. En complétant la concentration exclusive des méthodes basées sur les moindres carrés sur l'estimation des fonctions moyennes conditionnelles par une technique générale d'estimation des familles de fonctions quantiles conditionnelles, la régression quantile est capable d'étendre considérablement la flexibilité des méthodes de régression paramétriques et non paramétriques [33]. Dans cette partie, avant de présenter les modèles de régression quantile, nous rappelons la définition des quantiles et régression, nous donnerons le modèle de régression quantile et les méthodes d'estimation. Nous donnerons également les cas particuliers, de modèles régression quantile et quelques extensions.

2.3.1 Définition de Quantile

- **Définition 2.3.1** *Le quantile théorique d'ordre τ et $0 < \tau < 1$ Soit une variable aléatoire Y est définie par les inégalités suivantes :*

$$P(Y \geq Q_Y(\tau)) \geq \tau \quad \text{et} \quad P(Y \leq Q_Y(\tau)) \geq 1 - \tau$$

on peut en déduire que la probabilité d'observer une valeur inférieure (ou égale) à $Q_Y(\tau)$ devrait être supérieur à $1 - \tau$ tandis que la probabilité d'observer une valeur supérieure (ou égale) à $Q_Y(\tau)$ devrait être supérieur à τ de. Pour la régression quantile, il est utile de reformuler cette définition implicite comme solution de problème d'optimisation

$$Q_Y(\tau) = \arg \min_q (E(\rho_\tau(Y, q) | Y - q)) \quad (2, 1)$$

où la pondération

$$\rho_\tau(y, q) = \begin{cases} 1 - \tau & y < q \\ 0 & y = q \\ \tau & y > q \end{cases} \quad (2, 2)$$

De tels poids définissent $Q_Y(\tau)$ comme le minimisateur d'un écart absolu à pondération asymétrique. On notera que les poids $\rho_\tau(y, q)$ sont définis différemment pour les valeurs supérieures et inférieures à q et donc déplacer effectivement la solution vers le haut ou vers le bas en fonction du choix de q .

Il existe une autre définition du quantile :

Définition 2.3.2 *Supposons que Y est une variable aléatoire dont la fonction de distribution de probabilité*

$$F_Y(x) = P(Y \leq x)$$

Pour tout $0 < \tau < 1$, on définit le quantile d'ordre τ de Y par

$$Q_Y(\tau) = \inf\{y : F_Y(y) \geq \tau\} \quad (2, 3)$$

Si Y est une variable aléatoire continue, le quantile d'ordre τ est dès lors

$$Q_Y(\tau) = F^{-1}(\tau) \quad \text{et} \quad F(Q_Y(\tau)) = \tau.$$

– **La fonction quantile empirique**

La fonction quantile empirique est un estimateur de la fonction quantile $Q_Y(\tau)$ lorsque la fonction de distribution de probabilité F_Y de la variable Y est inconnue. On considère y_1, y_2, \dots, y_n un échantillon de n observations dont la distribution de probabilité est inconnue. Au moins une fraction τ des observations devrait être inférieure ou égale à $\hat{Q}_Y(\tau)$ et au moins une fraction de $1 - \tau$ observations devrait être supérieure ou égale à $\hat{Q}_Y(\tau)$. On a alors la définition suivante :

Définition 2.3.3 *Le quantile empirique d'ordre τ écrit comme suit :*

$$\frac{1}{n} \sum_{i=1}^n I(y_i \leq \hat{Q}_Y(\tau)) \geq \tau \text{ et } \frac{1}{n} \sum_{i=1}^n I(y_i \geq \hat{Q}_Y(\tau)) \geq 1 - \tau$$

où $I(\cdot)$ désigne la fonction indicatrice. À des fins d'estimation et pour généraliser le contexte de la régression, il est préférable d'utiliser l'équivalent comme la solution d'un critère d'optimisation, [18] : où

$$\hat{Q}_Y(\tau) = \arg \min_q \sum_{i=1}^n \rho_\tau(Y_i, q) |Y_i - q|$$

est l'analogie empirique de (2.1) où ρ_τ définie dans (2,2) est appelée aussi la fonction d'influence .

Nous pouvons aussi définir le quantile empirique par :

Définition 2.3.4 *La fonction de distribution empirique, notée \hat{F}_Y , se détermine par le quotient entre le nombre des observations inférieures ou égales à une quantité spécifiée y et le nombre total des observations :*

$$\hat{F}_Y(y) = \frac{\sum_{i=1}^n 1_{\{Y_i \leq y\}}}{n}$$

De manière analogue à l'équation (2,3), on peut en déduire la fonction quantile empirique :

$$\hat{Q}_Y(\tau) = \hat{F}_Y^{-1}(\tau) = \inf \left\{ y : \frac{\sum_{i=1}^n 1_{\{Y_i \leq y\}}}{n} \geq \tau \right\}, 0 < \tau < 1.$$

– **Quantile conditionnel [47]**

Dans le cas d'un problème de régression on s'intéresse au quantile

conditionnel. Considérons deux variables quantitatives continues : une variable Y , appelée variable d'intérêt, et une variable X , appelée co-variable, et $\tau \in]0, 1[$. Le quantile conditionnel d'ordre τ de la variable Y sachant que $X = x$ est défini de la manière suivante :

$$Q_Y(\tau, x) = F_Y^{-1}(\tau/x) = \inf\{y : F(y/x) \geq \tau\} \quad (2.4)$$

Exemple 2.3.1 *Considérons l'échantillon de "Engel Data" (Koenker et Basset, 1982) [10] de 235 observations avec comme variable dépendante la dépense alimentaire annuelle du ménage en euros et comme covariable le salaire annuel du ménage en euros. Sur le graphique (2.1) sont représentées les droites de régressions relatives aux valeurs de $\tau = 0.05, 0.25, 0.75, 0.95$ ainsi que celle correspondant à la médiane et la droite de régression linéaire obtenue par la méthode des moindres carrés. On peut constater que, par exemple, pour un revenu annuel de 20,25 % percent des dépenses alimentaires seraient inférieures à 11.849 et 75% inférieures à 14.427. Par contre, pour un revenu annuel de 80, 25% des dépenses alimentaires seraient inférieures à 40.295 et 75% percent inférieures à 53.078percent. En d'autres termes, les personnes ayant un revenu de 20 ont 25 de chance de dépenser 28.446 en moins et 75% percent de chances de dépenser 35.651 de moins que ceux ayant un revenu de 80 %. Ceci tend à confirmer que plus le salaire est élevée, plus les dépenses le sont également et ce d'autant plus qu'on s'intéresse aux quantiles conditionnels plus élevés.*

2.3.2 Exemple de modèle

1. Régression quantile additive

Depuis l'introduction des modèles additifs, ils ont reçu une attention considérable de Breiman et Friedman (1985) et Hastie et Tibshirani (1986,1990)[2] .

Définition 2.3.5 [55] *Soit L_2 les normes au carré comme mesures de la rugosité des fonctions ajustées. Sont remplacées par des normes L_1 correspondantes mesurant la variation totale, Nous considérerons*

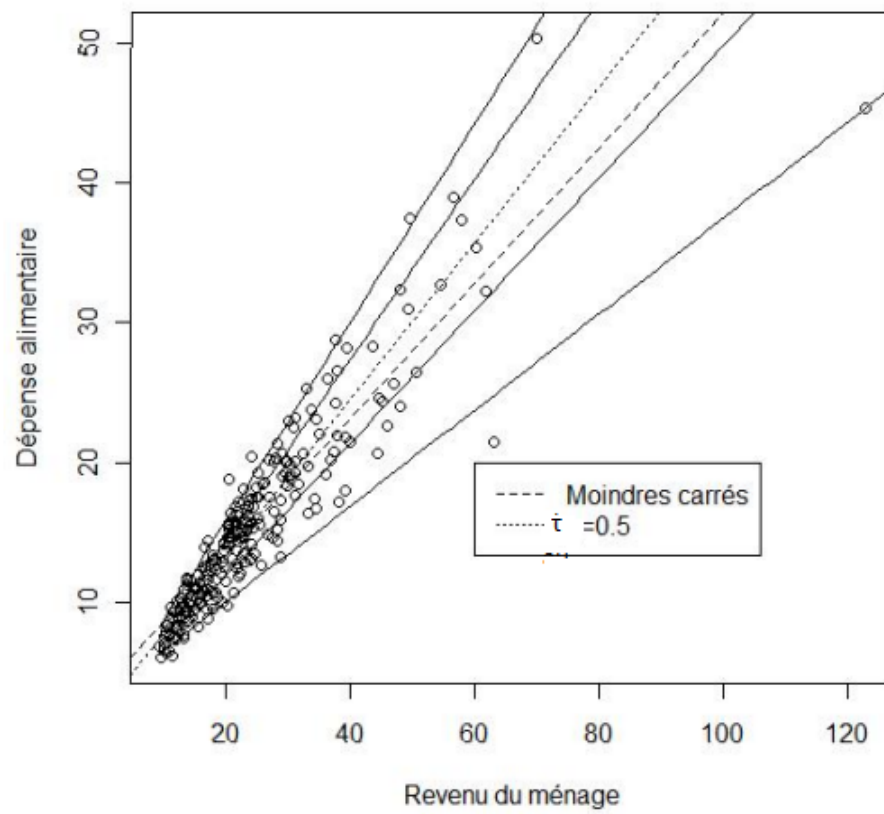


FIGURE 2.1 – Représentation des droites de régression pour $\tau = (0.05, 0.25, 0.5, 0.75, 0.95)$ et droite obtenue par la méthode des moindres carrés

des modèles pour les quantiles conditionnels d'ordre $\tau \in (1, 0)$ et $j = 1, \dots, J$ indexés par de la forme générale :

$$Q_{Y_i|x_i, z_i}(\tau | x_i, z_i) = x_i \theta'_1 + \sum_{j=1}^J g_j(z_{ij})$$

Les composantes non paramétriques g_j seront supposés être des fonctions continues, soit univariées, $\mathcal{R} \rightarrow \mathcal{R}$, ou bivarié, $\mathcal{R}^2 \rightarrow \mathcal{R}$. Nous désignerons le vecteur de ces fonctions comme étant $g = (g_1, \dots, g_J)$. Notre tâche consiste à estimer ces fonctions ainsi que le paramètre euclidien $\theta' \in \mathcal{R}^{p_1}$, en résolvant

$$\min_{(\theta'_1, g)} \sum \rho_\tau \left(y_i - x_i^\top \theta'_1 - \sum g_j(z_{ij}) \right) + \lambda_1 \|\theta'_1\|_1 + \sum_{j=1}^J \lambda_j \mathcal{V}(\nabla g_j)$$

avec $\rho_\tau(u) = u(\tau - I(u < 0))$ est la fonction perte quantile habituelle, $\|\theta'_1\|_1 = \sum_{k=1}^K |\theta'_{0k}|$ et $\mathcal{V}(\nabla g_j)$ désigne la variation totale de la dérivée ou du gradient de la fonction g . exprimer la variation totale de g' : $\mathcal{R} \rightarrow \mathcal{R}$ Rappelons que pour g avec une dérivée absolument continue g' nous pouvons comme

$$\mathcal{V}(g'(z)) = \int |g''(z)| dz$$

tandis que pour $g : \mathcal{R}^2 \rightarrow \mathcal{R}$ avec un gradient absolument continu,

$$\mathcal{V}(\nabla g) = \int \|\nabla^2 g(z)\| dz$$

où $\nabla^2 g(z)$ désigne le hessien de g , et $\|\cdot\|$ désignera la norme habituelle de Hilbert-Schmidt pour les matrices.

2. **Foret de régression quantile**[40] Le principe des forêts de régression quantile est la méthode des arbres de régression. Lorsque l'explication d'un quantile de la variable de réponse, les covariables explicatives confirment une structure arborescente. Cela signifie qu'une covariable est utilisée pour former deux groupes, qui correspondent à

deux branches. Dans chaque branche, une autre covariable est utilisée pour diviser à nouveau l'échantillon. La séquence d'enjambement de l'arbre ne doit pas nécessairement être la même dans chacune des branches. Une forêt est simplement une collection d'arbres, où chaque arbre a été obtenu avec un sous-échantillon bootstrapped, obtenu avec un sous-échantillon bootstrap de l'échantillon de formation original. Plusieurs décisions doivent être prises en compte lors de la génération d'arbres. L'une d'entre elles est la taille du sous-échantillon qui correspond à chaque arbre; cette taille est généralement constante et égale à la taille de l'échantillon d'entraînement initial, où le bootstrap est mis en œuvre avec remplacement. Lorsqu'un arbre comporte moins d'un petit nombre d'observations, par exemple dix, le n'ud est éliminé ou n'est plus portion-né. La distribution conditionnelle de la variable réponse est estimée par la distribution pondérée.

Supposons que nous étudions Y étant donné $X = x$ sur la base de l'arbre. Supposons que la feuille qui contient x est le poids $\xi(x, \omega)$ de l'observation pour cette feuille est soit 0 si elle n'est pas contenue dans la feuille ou 1 divisé par le nombre d'observations dans cette feuille. Soit, les T arbres des forêts aléatoires sont $1, \dots, T$ et $\xi(x)$ est la moyenne de $\xi(x, \omega)$ sur tous les arbres. Alors, le quantile conditionnel est l'estimation empirique pondérée.[48] Les forêts de régression quantile sont faciles à mettre en œuvre via les paquets R existants, mais elles sont beaucoup moins interprétables que les forêts classiques. mais elles sont beaucoup moins interprétables que les régressions quantiles classiques. C'est la raison pour laquelle sont utilisées principalement à des fins prédictive.

3. Régression quantile pondérée

Lorsque les densités conditionnelles de la réponse sont hétérogènes, il est naturel de se demander si la régression par quantile pondéré peut améliorer l'efficacité. naturel de se demander si la régression quantile pondérée peut conduire à des améliorations de l'efficacité. Plutôt que de pondérer par les réciproques des écarts types des observations, les

pondérations de la régression quantile devraient être proportionnelles à la densité locale évaluée au quantile d'intérêt. Les asymptotiques de la régression quantile pondérée [52] peuvent sembler quelque peu chimériques jusqu'à ce que des méthodes raisonnables d'estimation des pondérations souhaitées soient spécifiées. L'analyse de tels schémas d'estimation nécessite certaines hypothèses sur la variabilité de la densité conditionnelle. On considère l'estimateur suivant :

$$\tilde{\beta}_0(\tau) = \operatorname{argmin}_{b \in \mathbb{R}^p} \sum_{i=1}^n f_i(\xi_i) \rho_\tau(y_i - x_i' b)$$

et

$$\sqrt{n}(\tilde{\beta}(\tau) - \beta(\tau)) \sim \mathcal{N}(0, \tau(1 - \tau)D_2^{-1}(\tau))$$

fourni par $D_2(\tau) = \lim_{n \rightarrow \infty} n^{-1} \sum f_i^2(\xi_i) x_i x_i'$ est définie positive.

Définition 2.3.6 *La manière informelle ce qui est nécessaire pour estimer les poids $w_i = f_i(\xi_i)$ avec une précision suffisante pour que l'estimateur résultant soit efficace. On considère l'estimateur*

$$\hat{\beta}_{\hat{w}}(\tau) = \operatorname{argmin} \sum_{i=1}^n \hat{w}_i \rho_\tau(y_i - x_i^\top b)$$

avec des estimations $\{\hat{w}_i\}$ en remplaçant les poids optimaux $\{w_i\}$. Il n'est généralement pas difficile de montrer qu'une représentation de Bahadur est toujours valable pour une valeur τ fixe [5] :

$$n^{1/2} \left(\hat{\beta}_{\hat{w}} - \beta \right) = n^{-1/2} \sum_{i=1}^n \hat{v}_i R_i + o_p(1)$$

où $R_i = x_i (\tau - I(U_i \leq 0))$. Il s'agit essentiellement de montrer que la représentation pour des poids fixes est en fait uniforme dans les poids ; pour rendre cela rigoureux, il faudrait évidemment imposer certaines restrictions sur les poids.

Considérons l'estimation des poids comme une fonction lisse des résidus de a \sqrt{n} estimateur préliminaire cohérent de $\beta(\tau)$. Puisque l'estimation préliminaire donne \sqrt{n} cohérente des quantiles de la population, les résidus sont \sqrt{n} des estimations cohérentes des erreurs. Cela

suggère une expansion pour les poids de la forme

$$\hat{v}_i = w_i + m_n^{-1} \sum_{j \neq i}^n W_i(u_j) + o_p(m_n^{-1})$$

ou v_i est une fonction bornée et où la contribution de l'erreur $W_i(u_i)$ to \hat{v}_i est absorbé dans le terme d'erreur. Lorsque $\sigma(x)$ est linéaire x , on peut s'attendre à obtenir des taux de convergence paramétriques tels que $m_n = n$. dans le cas non paramétriques, nous pouvons nous attendre à pouvoir prendre. Maintenant, considérons l'estimation des poids comme une fonction lisse des résidus de $m_n = n^{4/5}$ pour les estimateurs à noyau et spline, Dans les deux cas, nous avons

$$\frac{1}{m_n} = o(n^{-1/2}) \quad \text{et} \quad \frac{n}{m_n^2} = o(n^{-1/2})$$

$$n^{-1/2} \sum_{i=1}^n \hat{v}_i R_i = n^{-1/2} \sum_{i=1}^n v_i R_i + n^{-1/2} \sum_{i=1}^n S_i R_i + o_P(1)$$

Avec

$$S_i = m_n^{-1} \sum_{j \neq i}^n v_i(u_j).$$

et la variance :

$$Q \equiv \text{Var} \left(n^{-1/2} \sum_{i=1}^n S_i R_i \right) = n^{-1} \sum_{i=1}^n \sum_{j=1}^n E S_i S_j R_i R_j$$

notons que $S_i = T_{ij} + v_i(u_j)/m_n$, est définie comme

$$T_{ij} = m_n^{-1} \sum_{k \notin \{i,j\}} W_i(u_k)$$

2.4 Le modèle de régression quantile standard

Pour présenter le modèle de régression quantile standard, nous allons transférer la régression de la moyenne à la régression des quantiles. Rappelons que le modèle de régression linéaire standard prend la forme suivante :

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

et en plus d'autres propriétés pour le terme d'erreur nous supposons que $E[\epsilon] = 0$. Ceci implique que les coefficients de régression ont un impact sur l'espérance de la réponse puisque

$$E[Y] = X\beta + E[\epsilon] = X\beta$$

l'hypothèse sur l'espérance du terme d'erreur implique une interprétation spécifique des coefficients de régression [18]. Pour la régression quantile, nous supposons également que le modèle s'écrit sous la forme

$$Y = X\beta_\tau + \epsilon_\tau \quad (2, 5)$$

où le terme d'erreur et les coefficients dépendent du quantile d'ordre τ de la variable d'intérêt Y . On suppose que le quantile d'ordre τ de l'erreur est nul c'est à dire $F_{\epsilon_\tau}(0) = \tau$. Alors

$$\tau = F_{\epsilon_\tau}(0) = P(\epsilon_\tau \leq 0) = P(X'_i\beta_\tau + \epsilon_\tau \leq X'_i\beta_\tau) = P(Y \leq X'_I\beta_\tau) = F_y(X'_i\beta_\tau)$$

donc le τ -quantile de la réponse Y est donné par le prédicteur $X'_i\beta_\tau$.

2.4.1 Estimation des paramètres du modèle

L'estimation des paramètres de ce modèle peut être effectuée en utilisant la fonction de perte absolue pondérée asymétriquement.

L'équation d'estimation est donnée par :

$$\hat{Q}_Y(\tau) = \arg \min_{\beta} \sum_{i=1}^n \rho_\tau(y_i, \eta_{i\tau}) |y_i - \eta_{i\tau}| \quad (2, 6)$$

avec $\eta_{i\tau} = X'_i\beta_\tau$ et

$$\rho_\tau(y_i, \eta_{i\tau}) = \begin{cases} 1 - \tau & y_i < \eta_{i\tau} \\ 0 & y_i = \eta_{i\tau} \\ \tau & y_i > \eta_{i\tau} \end{cases}$$

Elle est obtenue en utilisant la formule (2,4) à partir du critère d'erreur absolue à pondération asymétrique. Au lieu de supposer que l'erreur moyenne dans le modèle de régression est nulle, nous supposons que la "moyenne" pondérée des erreurs est égale à 0. La caractéristique la plus importante de la régression quantile à noter ici est généralement l'utilisation de la distribution des erreurs, sans oublier l'indépendance des termes d'erreur avec les observations. Pour minimiser le critère d'erreur absolue à pondération asymétrique nous pouvons procéder comme dans le cas d'une régression moyenne, c'est-à-dire en prenant des dérivées partielles par rapport aux coefficients de régression et en mettant les dérivées à zéro.

Malheureusement, cela n'est plus possible puisque (2,6) n'est pas différentiable à l'origine. Nous élargissons donc le problème d'estimation en introduisant $2n$ variables auxiliaires.

$$u_{i\tau} = (y_i - X'_i\beta_\tau)_+ \quad \text{et} \quad v_{i\tau} = (X'_i\beta_\tau - y_i)_+$$

où $(x)_+ = \min(x, 0)$ et par conséquent $\varepsilon_{i\tau} = u_{i\tau} - v_{i\tau}$. Le problème de minimisation (2.6) peut alors être réécrit comme suit

$$\sum_{i=1}^n \rho_\tau(y_i, \eta_{i\tau}) |y_i - \eta_{i\tau}| = \sum_{i=1}^n \tau u_{i\tau} + \sum_{i=1}^n (1 - \tau) v_{i\tau} = \tau \mathbf{1}' u_\tau + (1 - \tau) \mathbf{1}' v_\tau,$$

avec

$$y_i = \eta_{i\tau} + \varepsilon_{i\tau} = X'_i\beta_\tau + u_{i\tau} - v_{i\tau}, i = \overline{1, n}$$

ou en notation matricielle

$$Y = X\beta_\tau + u_\tau + v_\tau$$

avec $u_\tau = (u_{1\tau}, \dots, u_{n\tau})'$, $v_\tau = (v_{1\tau}, \dots, v_{n\tau})'$ et $\mathbf{1} = (1, \dots, 1)'$. En résumé, le problème d'optimisation après avoir introduit les variables auxiliaires est maintenant donné par

$$\min_{\beta_\tau, u_\tau, v_\tau} \{ \tau \mathbf{1}' u_\tau + [(1 - \tau) \mathbf{1}' v_\tau \mid X\beta_\tau + u_\tau - v_\tau = y] \}$$

Il s'agit d'un problème de minimisation avec des contraintes polyédriques, c'est-à-dire que la contrainte définit un objet géométrique avec des faces plates et des bords droits, ce qui signifie que nous pouvons contraindre un problème qui correspond à la restriction quantile originale, mais écrit en termes de variables auxiliaires. Le problème de la minimisation est maintenant linéaire dans les paramètres et peut donc être abordé avec des techniques de programmation linéaire qui permettent d'incorporer les contraintes polyédriques. Se reporter à Lange (2000) pour une brève introduction à la programmation linéaire en mettant l'accent sur les applications liées aux statistiques et à Lange (2004) pour une étude détaillée. Une autre approche pour minimiser l'écart absolu pondéré, asymétrique est l'utilisation de l'approche boosting [39], proposé par Fenske et al (2011) [19], dont le principe est de substituer les résidus en utilisant la fonction de perte

$$\rho(y, \eta) = \rho_\tau(y, \eta)|y - \eta|$$

Remarque 2.4.1 Nous pouvons écrire [57]

$$\rho_\tau(u) = \tau u 1_{[0, +\infty)}(u) - (1 - \tau) u 1_{(-\infty, 0)}(u) = (\tau - 1 \{u < 0\})u = \frac{|u| + (2\tau - 1)u}{2}$$

Elle est également Lipschitz continue de constante de Lipschitz

$$|L_\tau| = \max\{\tau, 1 - \tau\}.$$

De plus,

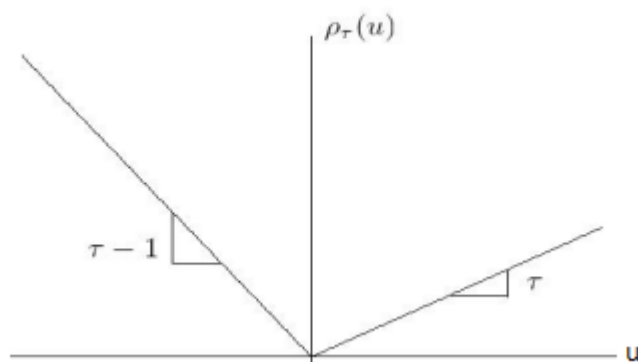
$$\min\{\tau, 1 - \tau\}|u| \leq \rho_\tau(u) \leq |L_\tau||u|.$$

pour tout $u \in] - \infty, +\infty[$.

Le cas de $y = \eta$ se produit avec une probabilité nulle .

2.4.2 Propriétés des estimateurs

L'une des propriétés les plus importantes pour l'estimation de la régression quantile, héritée des quantiles estimés (déterminés à partir d'un échantillon i.i.d.), est son invariance sous les transformations monotones.

FIGURE 2.2 – Fonction de perte $\rho_\tau(u)$

- **Les Propriétés asymptotiques** La méthode RQ offre une démarche plus riche pour estimer les paramètres des modèles en utilisant l'ensemble des quantiles conditionnels. Celle-ci consiste à estimer les coefficients de régression en minimisant la somme pondérée des valeurs des termes d'erreurs positifs ou négatifs [28], respectivement par le τ d'intérêt et son complémentaire $(1 - \tau)$. Ainsi, le coefficient de régression est alors en fonction de τ .

Soit $\beta_\tau = \beta(\tau)$, le vecteur des coefficients de régression à estimer et $\hat{\beta}(\tau)$ un vecteur d'estimation de $\beta(\tau)$ tel que :

$$\beta(\tau) = (\beta_0(\tau), \beta_1(\tau), \dots, \beta_p(\tau))$$

et

$$\hat{\beta}(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}_1(\tau), \dots, \hat{\beta}_p(\tau))$$

Les paramètres à estimer correspondent à la meilleure estimation $\hat{\beta}(\tau)$ du vecteur $\beta(\tau)$, déterminée en minimisant la somme pondérée des valeurs positives et négatives des différences entre la valeur observée y_i de Y_i et la valeur prédite $\hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)x_{1i} + \dots + \hat{\beta}_p(\tau)x_{pi}$ de y_i , notée \hat{y}_i avec $(i = 1, \dots, n)$, respectivement par le quantile d'ordre τ de la variable d'intérêt et son complémentaire $(1 - \tau)$. La somme pondérée est donnée par :

$$\sum_{y_i \geq x_i \hat{\beta}(\tau)} \tau \left| y_i - x_i' \hat{\beta}(\tau) \right| + \sum_{y_i < x_i \hat{\beta}(\tau)} (1 - \tau) \left| y_i - x_i' \hat{\beta}(\tau) \right|$$

avec $x_i' = (1, x_{1i}, x_{2i}, \dots, x_{pi})'$, $i = 1, \dots, n$ en posant

$$\rho_\tau(\varepsilon_i) = \varepsilon_i [\tau - \mathbf{I}\{\varepsilon_i < 0\}]$$

avec $\rho_\tau(\varepsilon_i)$ la fonction de pondération dépendant des termes d'erreurs pour $0 < \tau < 1$, la somme pondérée devient :

$$\sum_{i=1}^n \rho_\tau(\varepsilon_i) = \sum_{i=1}^n \varepsilon_i [\tau - \mathbf{I}\{\varepsilon_i < 0\}]$$

La meilleure estimation de $\hat{\beta}(\tau)$ de $\beta(\tau)$ est donnée par

$$\hat{\beta}(\tau) = \underset{\beta(\tau) \in \mathbb{R}^{p+1}}{\operatorname{argmin}} V(\beta(\tau); \tau)$$

où

$$V(\beta(\tau); \tau) = \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i' \beta(\tau))$$

est la fonction objective $V(\beta(\tau); \tau)$ du modèle avec $p + 1$ paramètres à estimer.

Dans le cas de la régression quantile standard $Y = X' \beta_\tau + \varepsilon_\tau$, la propriété asymptotique de l'estimateur et l'estimation de la précision sont dans le théorème suivant :

Théorème 2.4.1 [34] *Supposons que $\varepsilon_\tau = Y - X\beta_\tau$ admette, conditionnellement en X , une densité en 0 $f_{\varepsilon_\tau/X}(0/X)$ et que $J_\tau = E[f_{\varepsilon_\tau/X}(0/X) \cdot X]$ soit inversible. Alors :*

$$\sqrt{n} \left(\hat{\beta}_\tau - \beta_\tau \right) \xrightarrow{d} N \left(0, \tau(1 - \tau) J_\tau^{-1} E[XX'] J_\tau^{-1} \right)$$

Pour plus de détails consulter les références [17] et [13].

– L'intervalle de confiance et le p-value

A partir de la méthode directe proposé par selon Pourele (1991), l'intervalle de confiance est donnée par :

$$IC_\alpha = \left[\hat{\beta}_\tau - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_{as}}; \hat{\beta}_\tau + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_{as}} \right]$$

et

$$\hat{V}_{as} = \frac{\tau(1-\tau)}{f_\varepsilon(q_\tau(\varepsilon))^2} \left[\frac{1}{n} \sum_{i=1}^n X_i X_i' \right]^{-1}$$

où $z_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{N}(0, 1)$.

– Coefficient de corrélation

Par définition, le coefficient de corrélation mesure la qualité du modèle avec la régression quantile, pour le quantile spécifié en termes de somme pondérée des valeurs absolues des résidus. La notation utilisée pour décrire le pseudo R^2 dans la régression quantile est R^2 . La condition est la même pour la MCO ($0 < R^2 < 1$), R^2 est donc en mesure d'explorer plus d'informations à différents niveaux de la distribution conditionnelle de la variable d'intérêt, mais R^1 mesure une qualité locale pour chaque quantile ce qui n'est pas le cas pour la régression linéaire simple car dans la méthode de moindre carré ordinaire, R^2 mesure une qualité globale sur l'ensemble de la distribution conditionnelle.

Koenker et Machado(1999)[36] ont proposé une alternative pour calculer R^1 . Pour la MCO, R^2 est la somme des carrés des résidus alors qu'en régression quantile, elle est remplacée par la fonction objective du modèle non-contraint, avec p variables indépendantes, notée $\tilde{V}(\tau)$, et la somme des carrés totaux par la fonction objective du modèle contraint, avec uniquement la constante, notée $\hat{V}(\tau)$. Cela peut se traduire par les hypothèses ci-dessous, sur les coefficients excepté la constante $\beta_0(\tau)$:

$$H_0 : \beta_1(\tau) = \dots = \beta_p(\tau) = 0.$$

contre

$$H_1 : \text{Au moins un des } \beta_j(\tau) \neq 0.$$

Le rapport de la différence entre les fonctions objectives R^1 issues de chacune des hypothèses H_0 et H_1 , la fonction objective sous l'hypothèse H_0 :

$$R^1 = \frac{\tilde{V}(\tau) - \hat{V}(\tau)}{\tilde{V}(\tau)}$$

2.4.3 Tests sur les quantiles

Le principe de test de la régression quantile est le même que celui de la régression linéaire standard, mais la différence réside dans le calcul des statistiques de test. En général, il existe deux types de tests dans la régression quantile : le test d'égalité des quantiles et les tests de régression quantile (test de qualité de la prédiction et test de stabilité des paramètres).

– Test d'égalité des quantiles

Li et al (2012) décrit un test d'égalité des quantiles. La méthode est identique au test ANOVA dans la régression linéaire simple. Mais d'autre part au lieu de comparer les moyennes de $k \geq 2$ populations indépendantes et de distribution normale $\mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, k$, il compare les quantiles de ces k populations, en trouve les hypothèses suivants :

$$H_0 \quad : \quad \tau_{l1} = \dots = \tau_{lk}$$

$$H_1 \quad : \quad \tau_{li} \neq \tau_{lj}, i \neq j, i, j \in \{1, \dots, k\}$$

On a la fonction $\tau_{li} = \mu_i + \sigma_i z_l$ est le l -ème quantile de la i -ème population et z_l le l -ème quantile de la distribution normale standard. On rencontre souvent la résolution matricielle suivante pour ce type de test, Sous les hypothèses suivants :

$$H_0 \quad : \quad H\theta = 0,$$

$$H_1 \quad : \quad H\theta \neq 0,$$

la statistique du test T est donnée par :

$$T = \tilde{n}(H\hat{\theta} - H\theta)'(HAH')^{-1}(H\hat{\theta} - H\theta) \sim \chi_{k-1}^2$$

avec \tilde{n} est la plus grande taille d'échantillons tirés parmi les k populations

$$H_{(k-1) \times k} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix}, \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \vdots \\ \theta_k \end{pmatrix}, \theta_i = \tau_{li}, i = 1, \dots, k$$

$$A = \text{diag} \left(\frac{\sigma_1^2}{\lambda_1} \left(1 + \frac{z_l^2}{2} \right), \dots, \frac{\sigma_k^2}{\lambda_k} \left(1 + \frac{z_l^2}{2} \right) \right)$$

$$\lambda_i = \lim_{n \rightarrow +\infty} \left(\frac{n_i}{\tilde{n}} \right), 0 < \lambda_i \leq 1, i = 1, \dots, k$$

où n_i est la taille de l'échantillon i et χ_{k-1}^2 la loi de Chi-deux à $(k-1)$ degré de liberté. Concernant la règle de décision, le p -value est obtenu par :

$$\begin{aligned} p\text{-value} &= P(T \geq t_{vfd}) \\ &= 1 - F_{\chi_{k-1}^2}(t_{vfd}) \end{aligned}$$

avec t_{vfd} la valeur calculée de T et $F_{\chi_{k-1}^2}(\cdot)$ la fonction de distribution cumulative de la distribution de Khi-deux à $k-1$ degré de liberté. Si p -value est inférieur au niveau de signification α , on rejette l'hypothèse H_0 .

– Tests en régression quantile

Les tests présentés dans cette section ont été développés par Furno (2011). Ils se concentrent globalement sur l'hypothèse d'invariance des coefficients de régression telle que décrite par Johnston et Di Nardo (1997)[30], en comparant un modèle de régression avec contraintes à un ou plusieurs autres modèles de régression sans contraintes. Chow (1960)[7] utilise la même démarche lorsqu'il analyse des données temporelles. Nous présentons ci-dessous ces tests.

1. Test de qualité de prédiction

Soit k' la taille de l'échantillon d'étude. L'évaluation de la qualité de prédiction d'un modèle construit est déterminante lors-

qu'on veut prédire une réponse correspondant aux variables indépendantes en présence dans le modèle. le test de qualité de prédiction consiste à estimer les paramètres sous les $k'_1, k'_1 \approx \frac{n}{2}$, premières observations de l'échantillon d'étude puis d'utiliser le modèle construit pour prédire les valeurs de la variable dépendante des données restantes.

Ce test permet d'évaluer la stabilité des paramètres à estimer sans tenir compte la taille de l'échantillon. Cette démarche permet de comparer la constance des coefficients estimés sous les deux hypothèses :

2. Test de stabilité des paramètres

L'objectif du test de stabilité des paramètres ou du test de changement structurel est d'observer statistiquement des changements de comportement dans l'échantillon étudié. Pour ce faire, il faut caractériser au mieux ce qui permet de définir les sous-échantillons à comparer (en utilisant une information externe ou une variable indépendante disponible dans les données, comme variable de segmentation) et détecter la nature du changement intervenu (modification ou non des coefficients relatifs aux variables indépendantes). Après avoir déterminé les sous-échantillons ou le groupe, nous recherchons un changement dans les coefficients des variables indépendantes. En considérant intra-groupe B de n_i observations ($n_i > p + 1$) le test de stabilité repose sur la comparaison des modèles de régression contrainte et non-contrainte pour déterminer si, sur ces groupes étudiés, les coefficients sont identiques. voire [16] le modèle de régression contraint

$$Q_{y_i}(\tau | x_i) = \beta_0(\tau) + \beta_1(\tau)x_{1i} + \dots + \beta_p(\tau)x_{pi}, i = 1, \dots, n$$

le modèle de régression non-contraint

$$\begin{aligned} Q_{y_i}(\tau | x_i) &= \beta_0^1(\tau) + \beta_1^1(\tau)x_{1i} + \cdots + \beta_p^1(\tau)x_{pi}, i = 1, \dots, n_1 \\ &\vdots \\ Q_{y_i}(\tau | x_i) &= \beta_0^B(\tau) + \beta_1^B(\tau)x_{1i} + \cdots + \beta_p^B(\tau)x_{pi}, i = 1, \dots, n_B \end{aligned}$$

sous les tests d'hypothèses en peut informe sur la nécessité ou non de distinguer les modèles de régression à $(p + 1)$ paramètres dans les B groupes, elles présentent comme suit :

$$H_0 : \begin{pmatrix} \beta_0(\theta) \\ \vdots \\ \beta_p(\theta) \end{pmatrix} = \begin{pmatrix} \beta_0^1(\theta) \\ \vdots \\ \beta_p^1(\theta) \end{pmatrix} = \begin{pmatrix} \beta_0^B(\theta) \\ \vdots \\ \beta_p^B(\theta) \end{pmatrix}$$

contre

H_1 : Au moins un des coefficients diffère des autres.

La statistique de test \hat{T}_1 est donnée par :

$$\hat{T}_1 = \frac{\left[\tilde{V}(\theta) - \sum_{i=1}^K \hat{V}_i(\theta) \right] / \Delta(ddl_{H_0}, ddl_{H_1})}{\sum_{i=1}^K \hat{V}_i(\theta) / ddl_{H_1}} \sim F((k-1)(p+1), n-k(p+1))$$

avec

$$\begin{aligned} \Delta(ddl_{H_0}, ddl_{H_1}) &= n - (p + 1) - \sum_{i=1}^k (n_i - (p + 1)) \\ &= \sum_{i=1}^{k-1} (p + 1) \\ &= (k - 1)(p + 1) \end{aligned}$$

$$\begin{aligned} ddl_{H_1} &= \sum_{i=1}^k [(n_i - (p + 1))] \\ &= n - k(p + 1) \end{aligned}$$

– **Principe d'application du test**

- a** Estimer les paramètres du modèle en utilisant la méthode RQ pour :
- a-1** l'échantillon complet (k' observations) ;
 - a-2** le sous-échantillon de k'_1 premières observations ;
- b** Évaluer la fonction objective pour chacun des deux modèles contraint et non-contraint ;
- c** Calculer la statistique de test T_1 ;
- d** Règle de décision en Rejeter l'hypothèse nulle au seuil α si la valeur calculée de $T_1 > F_{1-\alpha}(k'_2, k'_1 - (p + 1))$.

2.4.4 Cas particuliers

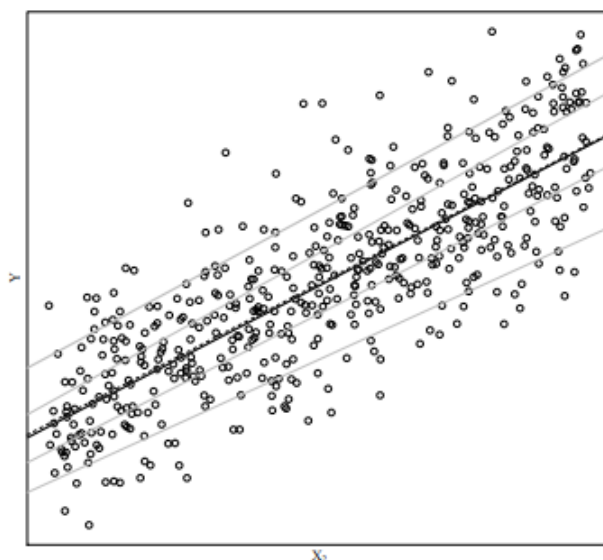
– **Le modèle de translation simple**

Dans translation simple [13] on suppose que les variables explicatives n'ont d'impact que sur la moyenne de la variable d'intérêt (et pas sur sa variance par exemple). Il s'agit du modèle de translation linéaire :

$$Y = \check{X}\zeta + \varepsilon \quad (2, 7)$$

où ε est indépendant de X et de moyenne nulle. Sous cette hypothèse, les résidus sont en particulier homoscédastiques (i.e., $V(\varepsilon | X) = \sigma^2$). Dans ce modèle, les distributions conditionnelles $F_{Y|X=x}$ sont parfaitement parallèles lorsque x varie : les différents quantiles conditionnels dépendent donc linéairement de X , $q_\tau(Y | X) = \check{X}\zeta + q_\tau(\varepsilon)$. On est donc bien dans le cadre de l'hypothèse 1.1, mais ici seul le coefficient correspondant à la constante varie en fonction de τ : $\beta'_{1,\tau} = 1 + q_\tau(\varepsilon)$, tandis que $\beta'_{k,\tau} = \zeta_k$ pour $k > 1$. Dans le cas simple d'une régression univariée où l'on s'intéresse à une seule variable explicative X_2 en plus de la constante. Dans ce cas, les lignes correspondant aux régressions quantiles sont des lignes parallèles : elles ont toutes une pente β'_2 correspondant à la variable X_2 , on parle donc d'une homogénéité des pentes.

les coefficients $(\beta'_{k,\tau})_{k=2,\dots,p}$ dont nous disposons déjà ci-dessus correspondant à la modélisation de la moyenne conditionnelle $E(Y | X) =$



Lecture : chaque point correspond à une observation (X_2, Y) généré par le processus $Y = \beta_1 + \beta_2 X_2 + \epsilon$, avec $\epsilon \mapsto \mathcal{N}(0, \sigma)$. Les droites correspondent aux droites de régressions quantiles pour les déciles d'ordre 1, 3, 7 et 9 (en gris), la médiane (en noir) et à la droite de régression linéaire classique (en noir pointillé).

FIGURE 2.3 – Exemple de données distribuées selon un modèle de translation

$\check{X}\gamma$ pour tout τ . Les résultats obtenus par les régressions (quantile ou linéaire) estiment les mêmes paramètres, par des méthodes différentes. l'utilisation des estimateurs des moindres carrés ordinaires est parfois moins intéressante ; pour cela nous utilisons des estimateurs de régression quantile. Ils sont plus robustes à la présence de valeurs aberrantes, par exemple, et peuvent être plus précis pour certaines distributions de ϵ .

Enfin, comme on ignore généralement que le vrai modèle est un modèle de traduction, on peut les utiliser pour tester cette restriction, ce qui implique que les estimateurs des régressions quantiles effectuées pour différents doivent être très proches. L'utilisation des estimateurs des moindres carrés ordinaires est parfois moins intéressante ; pour cela nous utilisons les estimateurs des régressions quantiles.

– Le modèle de translation-échelle

Le modèle de translation-échelle est plus général que le modèle de translation, les variables d'intérêt déterminantes ont non seulement un impact sur la moyenne mais aussi sur la variance de la variable d'intérêt. Ces modèles, appelés translation-échelle, correspondent à une certaine forme d'hétéroscédasticité :

$$Y = X + (X\theta) \epsilon \quad (2, 8)$$

avec encore une fois ϵ indépendant de X , de moyenne nulle et $\check{X}\theta > 0$. Dans ce modèle, la dispersion de la variable dépendante conditionnelle à X est plus importante pour certaines valeurs de X . Le modèle de translation-échelle correspondant à l'équation (1.4) implique que $Q_\tau(Y | X) = \check{X}(\zeta + Q_\tau(\epsilon)\theta)$. Ainsi, l'hypothèse (a) est bien vérifiée, avec $\beta'_\tau = \zeta + Q_\tau(\epsilon)\theta$ pour chaque différents quantiles L'impact des variables explicatives ne sera pas le même par les quantile.

La figure 2 nous donne le modèle illustré de l'échelle de translation. Ici les pentes $\beta'_{2,\tau}$ dans le cas univarié correspondant aux différentes régressions quantiles sont croissantes avec τ . (c'est-à-dire $\gamma > 0$), ce qui traduit une dispersion d'autant plus grande que X^2 est élevé.

– Un cadre général : le modèle à coefficients aléatoires

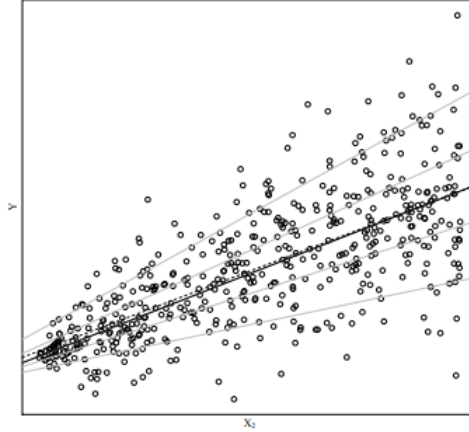
Le modèle à coefficients aléatoires généralise ces modèles. Il s'écrit :

$$Y = X'\beta'_v, \quad v \text{ indépendant de } X \text{ et de loi uniforme sur } [0, 1]$$

où la fonction $v \mapsto x'\beta'_v$ est strictement croissante pour tout x . car Ce modèle vérifie bien l'hypothèse (2.5), comme v et X et croissance de $v \mapsto \check{x}\beta'_v$

$$P(Y \leq x'\beta'_\tau | X = x) = P(X\beta'_v \leq \beta_\tau) = P(v \leq \tau) = \tau$$

Dans ce modèle, v peut s'interpréter comme une composante individuelle inobservée qui positionne l'individu dans la distribution de Y . Ce modèle à coefficients aléatoires dépend des hypothèses très flexibles



Lecture : chaque point correspond à une observation (X_2, Y) généré par le processus $Y = \beta_1 + \beta_2 X_2 + \theta X_2 \epsilon$, avec $\epsilon \mapsto \mathcal{N}(0, \sigma)$. Les droites correspondent aux droites de régressions quantiles pour les déciles d'ordre 1, 3, 7 et 9 (en gris), la médiane (en noir) et à la droite de régression linéaire classique (en noir pointillé).

FIGURE 2.4 – Exemple de données distribuées selon un modèle de translation échelle

sur la dépendance en v , qui peut être non linéaire. Il généralise les deux exemples précédents. Le modèle de translation linéaire correspond à un cas où le coefficient correspondant à la variable explicative $k > 1, \beta'_{k,U}$, est indépendant de v . Dans le modèle de translation-échelle, on a $\beta'_v = \gamma + q_v(\varepsilon)\theta$

– Interprétation des régressions quantiles

La façon dont les distributions conditionnelles d'ordre τ de Y en fonction des variables explicatives soulève plusieurs questions. La première consiste simplement à décrire comment les quantiles conditionnels évoluent en fonction de ces déterminants, sans chercher à savoir s'ils sont (comparables) aux différents quantiles conditionnels.

Remarque 2.4.2 Soit le quantile conditionnel d'ordre τ de Y Y était considéré comme une fonction linéaire de \mathbf{x} ou en peut dire que $Q_Y(\tau | X) = x' \beta(\tau)$, Dès lors

$$P(Y < Q_y(\tau | X) | X = x) = \tau$$

et donc le quantile d'ordre τ de la distribution des résidus doit être

nul. On constate donc que la droite de régression cherchée, $x'\beta(\hat{\tau})$

Les régressions quantiles sont des outils développés pour répondre à cette question. Si l'hypothèse simple mais restrictive d'invariance de rang est faite, elles peuvent être utilisées pour répondre à une deuxième question, un peu plus précise, qui est de déterminer quelle est la variation de la variable d'intérêt correspondant à une variation marginale d'un de ces déterminants,

pour les personnes qui se trouvent à un certain niveau de la distribution conditionnelle de la variable d'intérêt Y . En général, les régressions quantiles ne fournissent pas d'éléments pour répondre à une question encore différente, qui est d'estimer la distribution des effets de ce déterminant X sur la variable d'intérêt Y .

Enfin, l'un des intérêts de la régression quantile étant de ne pas supposer a priori que les variables explicatives ont un effet homogène sur l'ensemble de la distribution de la variable d'intérêt, il est tout à fait possible de tester cette hypothèse à partir des estimations obtenues.

Chapitre 3

Régression quantile bayésienne

3.1 Introduction

Ce chapitre présente l'idée de la régression quantile bayésienne en utilisant une fonction de vraisemblance basée sur la distribution de Laplace asymétrique. Il est démontré que, quelle que soit la distribution originale des données, l'utilisation de la distribution de Laplace asymétrique est un moyen très naturel et efficace de modéliser la régression quantile bayésienne. Les a priori sont uniformes et impropres pour les paramètres inconnus β_τ du modèle produisent un postérieur conjoint correct[57].

3.2 Le modèle de régression quantile bayésienne

La régression quantile bayésienne [18] a été développée en utilisant l'équivalence entre le mode postérieur et l'estimation du maximum de vraisemblance sous des a priori non informatifs. Nous supposons le modèle de régression quantile $Q_\tau(Y | X) = X\beta_\tau$. Comme le montrent Yu et Moyeed (2001)[57], β_τ peut être obtenu comme estimateur du maximum de vraisemblance pour β_τ dans le cadre du modèle entièrement paramétrique

$$y_i = x_i' \beta_\tau + \epsilon_{i\tau}, i = 1, \dots, n$$

avec $\epsilon_{i\tau}$ sont des v.a i.i.d provenant de la distribution de Laplace asymétrique **ALD** (voire l'annexe pour la définition et les propriétés

de cette distribution) tel que $\epsilon_{i\tau} \mid \sigma^2 \sim ALD(0, \sigma^2, \tau)$ de densité donnée par :

$$P(\epsilon_{i\tau} \mid \sigma^2) = \frac{\tau(1-\tau)}{\sigma^2} \exp\left(-\rho_\tau(\epsilon_{i\tau}, 0) \frac{|\epsilon_{i\tau}|}{\sigma^2}\right) \quad (3, 1)$$

Dans ce cas les observation y_i , $i = \overline{1, n}$ sont conditionnellement indépendants et suivant une distribution ALD, tel que $y_i \mid \beta_\tau, \sigma^2 \sim ALD(x'_i \beta_\tau, \sigma^2, \tau)$ de densité donnée par

$$P(y_i \mid \beta_\tau, \sigma^2) = \frac{\tau(1-\tau)}{\sigma^2} \exp\left(-\rho_\tau(y_i, \mathbf{x}'_i \beta_\tau) \frac{|y_i - \mathbf{x}'_i \beta_\tau|}{\sigma^2}\right) \quad (3, 2)$$

En prenant pour la loi a priori uniforme non informative $(\beta_\tau) \propto cste$, alors la loi a posteriori de β_τ et donnée par

$$\begin{aligned} P(\beta_\tau \mid \mathbf{y}, \sigma^2) &\propto \prod_{i=1}^n P(y_i \mid \beta_\tau, \sigma^2) \\ &\propto \exp\left(-\sum_{i=1}^n \rho_\tau(y_i, \mathbf{x}'_i \beta_\tau) \frac{|y_i - \mathbf{x}'_i \beta_\tau|}{\sigma^2}\right) \end{aligned} \quad (3, 3)$$

(3,3) est équivalent à la minimisation du critère d'optimisation (2.6). Si la distribution de Laplace asymétrique permet d'exprimer commodément la régression quantile dans un cadre bayésien. Yue et Rue (2011) [59] suggèrent d'utiliser des conditionnelles complets pour l'échantillonneur de Gibbs. Ils suggèrent d'utiliser les v.a.iid $z_i \mid \sigma^2 \sim \exp(1/\sigma^2)$. Dans ce cas les réponses conditionnelles sont telles que

$$y_i \mid z_i, \beta_\tau, \sigma^2 \sim N(x'_i \beta_\tau + \xi z_i, \sigma^2/w_i)$$

avec

$$\xi = \frac{1-2\tau}{\tau(1-\tau)}, \quad w_i = \frac{1}{\delta^2 z_i}, \quad \delta^2 = \frac{2}{\tau(1-\tau)}$$

Alors la distribution marginale de $y_i \mid \beta_\tau, \sigma^2$ obtenue par intégration sur les z_i suite une distribution de Laplace asymétrique, c'est-à-dire,

$$y_i \mid \beta_\tau, \sigma^2 \sim ALD(x'_i \beta_\tau, \sigma^2, \tau)$$

L'inférence bayésienne peut maintenant être mise en oeuvre efficacement après l'imputation des variables d'échelle z_i comme des inconnues supplémentaires, de façon similaire à l'approche décrite pour les modèles de régression binaire. Le modèle résultant est un modèle de régression conditionnellement gaussien avec des décalages ξz_i et les poids w_i .

Les conditionnelles complètes pour l'échantillonneur de Gibbs sont données par :

$$\mathbf{a} \quad (\beta_\tau | \cdot) \sim N(\mu_{\beta_\tau}, \Sigma_{\beta_\tau}) \quad (3,4)$$

$$\Sigma_{\beta_\tau} = \sigma^2 (X'WX)^{-1}, \quad \mu_{\beta_\tau} = (X'WX)^{-1} X'W(y - \xi z)$$

$$\text{où } W = \text{diag}(w_1, \dots, w_n) \text{ et } z = (z_1, \dots, z_n)'$$

b Conditionnelle complète pour la variance de l'erreur :

$$(\sigma^2 | \cdot) \sim \text{IG} \left(a + \frac{3n}{2}, b + \frac{1}{2} \sum_{i=1}^n w_i (y_i - x'_i \beta_\tau - \xi z_i)^2 + \sum_{i=1}^n z_i \right) \quad (3)$$

$a > 0$ (paramètre de forme) et $b > 0$ (paramètre d'échelle)

c Conditionnelle complète pour les paramètres de l'échelle :

$$(z_i^{-1} | \cdot) \sim \text{InvGauss} \left(\sqrt{\frac{\xi^2 + 2\delta^2}{(y_i - x'_i \beta_\tau)^2}}, \frac{\xi^2 + 2\delta^2}{\sigma^2 \delta^2} \right) \quad (3,6)$$

En effet

a) Le résultat est obtenu en utilisant le fait que

$$\begin{aligned}
P(\beta_\tau | \cdot) &\propto P(y | z, \beta_\tau, \sigma^2) \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} (X\beta_\tau + \xi z - y)' (X\beta_\tau + \xi z - y) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \beta_\tau' \frac{1}{\sigma^2} X'W X \beta_\tau + \frac{1}{\sigma^2} \beta_\tau' X'W (y - \xi z) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \beta_\tau' \Sigma_\beta^{-1} \beta_\tau + \beta_\tau' \Sigma_\beta^{-1} (X'W X)^{-1} X'W (y - \xi z) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \beta_\tau' \Sigma_{\beta_\tau}^{-1} \beta_\tau + \beta_\tau' \Sigma_{\beta_\tau}^{-1} \mu_{\beta_\tau} \right\}
\end{aligned}$$

avec β_τ préservant la condition ci-dessus.

C'est une distribution normale multivariée avec une moyenne μ_{β_τ} et la matrice de covariance Σ_{β_τ} .

- b) Le résultat est obtenu en utilisant la loi a priori de σ^2 La conditionnelle complète pour $\sigma^2, \sigma^2 \sim GI(a, b)$ de densité donnée par :

$$P(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{a-1} \exp\left(-\frac{b}{\sigma^2}\right)^{a-1}$$

Dans ce cas

$$\begin{aligned}
P(\sigma^2 | \cdot) &\propto P(y | z, \beta_t, \sigma^2) p(z | \sigma^2) p(\sigma^2) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{\sigma^2} \frac{1}{2} (X\beta_r + \xi z - y)' W (X\beta_r + \xi z - y)\right\} \\
&\quad \prod_{i=1}^n \frac{1}{\sigma^2} \exp\left(-\frac{1}{\sigma^2} z_i\right) \\
&\quad \left(\frac{1}{\sigma^2}\right)^{a-1} \exp\left(-\frac{1}{\sigma^2} b\right) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n+\frac{n}{2}+a-1} \exp\left\{-\frac{1}{\sigma^2} \left(b + \frac{1}{2} (X\beta_r + \xi z - y)'\right.\right. \\
&\quad \times \exp W(X\beta_r + \xi z - y) \\
&\quad \left.\left. \times \exp\left\{\sum_{i=1}^n z_i\right\}\right.\right.
\end{aligned}$$

$$a' = a + \frac{3n}{2} \text{ and } b' = b + \frac{1}{2} (X\beta_r + \xi z - y)' W (X\beta_r + \xi z - y)$$

Alors $(\sigma^2 | \cdot) \sim IG(a', b')$

c) Le résultat est obtenu en utilisant le fait que :

$$\begin{aligned}
P(z_i | \cdot) &\propto P(y_i | z_i, \boldsymbol{\beta}_\tau, \sigma^2) p(z_i | \sigma^2) \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau - \xi z_i)^2}{\delta^2 z_i \sigma^2} \right\} \exp \left(-\frac{1}{\sigma^2} z_i \right) \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau)^2 - 2(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau) \xi z_i + z_i^2 (\xi^2 + 2\delta^2)}{\delta^2 z_i \sigma^2} \right\} \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau)^2}{2\sigma^2 \delta^2} \cdot \frac{1 - \frac{2\xi z_i}{y - \mathbf{x}'_i \boldsymbol{\beta}_\tau} + z_i^2 \frac{\xi^2 + 2\delta^2}{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau)^2}}{z_i} \right\} \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau)^2}{2\sigma^2 \delta^2} \cdot \frac{(z_i^{-1})^2 - \frac{2\xi}{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau) z_i} z_i^{-1} + \mu_i^2}{z_i^{-1}} \right\} \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau)^2 (\xi^2 + 2\delta^2)}{2\sigma^2 \delta^2 (\xi^2 + 2\delta^2)} \cdot \frac{(z_i^{-1})^2 - \frac{2\xi}{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau) z_i} z_i^{-1} + \mu_i^2}{z_i^{-1}} \right\} \\
&= \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{\lambda}{2\mu_i^2} z_i \cdot \left(\frac{1}{z_i^2} - \frac{2\xi}{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau) z_i} + \mu_i^2 \right) \right\} \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{\lambda}{2\mu_i^2} z_i \cdot \left(\frac{1}{z_i^2} - 2\mu_i \frac{1}{z_i} + \mu_i^2 \right) \right\} \\
&\propto \frac{1}{\sqrt{z_i}} \exp \left\{ -\frac{\lambda}{2\mu_i^2} z_i \cdot \left(\frac{1}{z_i} - \mu_i \right)^2 \right\}
\end{aligned}$$

avec

$$\mu_i = \left(\frac{\xi^2 + 2\delta^2}{(y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau)^2} \right)^{1/2} \quad \text{and} \quad \lambda = \frac{\xi^2 + 2\delta^2}{\sigma^2 \delta^2}$$

Alors $(z_i | \cdot) \sim \text{InvGauss}(\mu_i, \lambda)$

La distribution de $1/z_i$ est gaussienne inverse avec un paramètre de localisation μ_i et le paramètre d'échelle λ . La mise à jour de z_i est alors obtenue par inversé les résultats. Pour dériver la distribution $g(z_i) = 1/z_i$, $g^{-1}(z_i) = z_i$ et $g'(z_i) = -1/z_i^2$.

3.3 APPLICATION :

3.3.1 Présentation des données

L'échantillon (stackloss) est composé de 21 données qui sont faites sur 21 jours consécutifs durant lesquels fut observé la transformation par une plante de l'ammoniac en acide nitrique, l'acide ainsi produit étant ensuite absorbé dans un tube [57]. Les variables étudiées sont les suivantes :

1. le flux d'air vers la plante, x_1 ,
 2. la température d'entrée de l'eau de refroidissement, x_2
 3. la concentration d'acide circulant dans le tube, x_3
- La variable de réponse, Y , est le pourcentage d'ammoniac perdu (fois 10).

Le but est d'estimer les paramètres de modèle RQ

$$Q_\tau(y | x) = \beta_0(\tau) + \beta_1(\tau)x_1 + \beta_2(\tau)x_2 + \beta_3(\tau)x_3,$$

3.3.2 Étude du quantile conditionnel d'ordre $\tau = 0.75$

Les estimations des paramètres de régression données par la méthode de Koenker sont résumées dans le tableau (3.1). On constate à part pour la première covariable (flux d'air x_1), est une valeur possible pour les trois paramètres de régression (notons en outre que le paramètre de régression correspondant à la covariable représentant la concentration en acide x_3 est d'ailleurs estimé à 0).

Considérons comme valeurs initiales des paramètres les estimations de données par la méthode Koenker et choisissons également une matrice de variance-covariance fournie par cette méthode et obtenue par bootstrap [20] 1000 itérations sont effectuées comme nous pouvons le constater sur le graphique (3, 1), en considérant ces valeurs initiales et cette

Quantiles de régression	Estimation	Borne inf.	Borne sup.
$\hat{\beta}_0(0.75)$	-54.190	-61.163	8.484
$\hat{\beta}_1(0.75)$	0.871	0.533	1.206
$\hat{\beta}_2(0.75)$	0.983	-0.538	1.782
$\hat{\beta}_3(0.75)$	0.000	-0.517	0.053

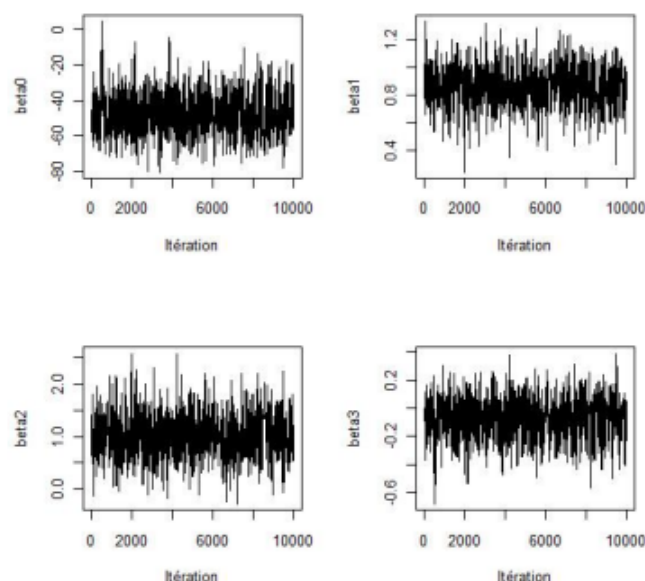
TABLE 3.1 – Estimation par la méthode Yu et al des quantiles de régression pour $\tau = 0.75$ 

FIGURE 3.1 – Chaînes des itérations

matrice de variance-covariance , nous ne visualisons pas de phase de (burn-in) , et on peut constater que l'auto-corrélation diminue avec le nombre d'itérations figure (3,2). Enfin, le graphique (3,3) représente quant à lui les distributions jointes des paramètres de la régression quantile .

Pour éviter la phase (burn-in) , nous devons ignorer les premières valeurs itérées (les 1000 premières valeurs générées). Le tableau (3,2) donne des informations concernant la liste de chaque paramètre de RQ. On constate à présent que seule la covariable relative à la concentration en acide x_3 ne semble pas être indispensable au modèle vu que 0 est à nouveau une valeur plausible pour le paramètre de régression relatif à cette variable explicative. Le graphique (3,4) nous permet

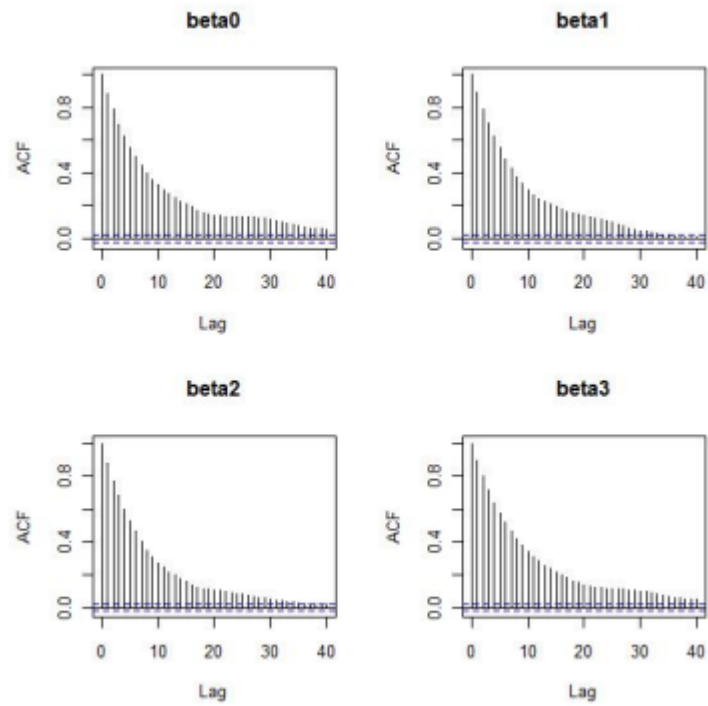


FIGURE 3.2 – Auto-corrélation

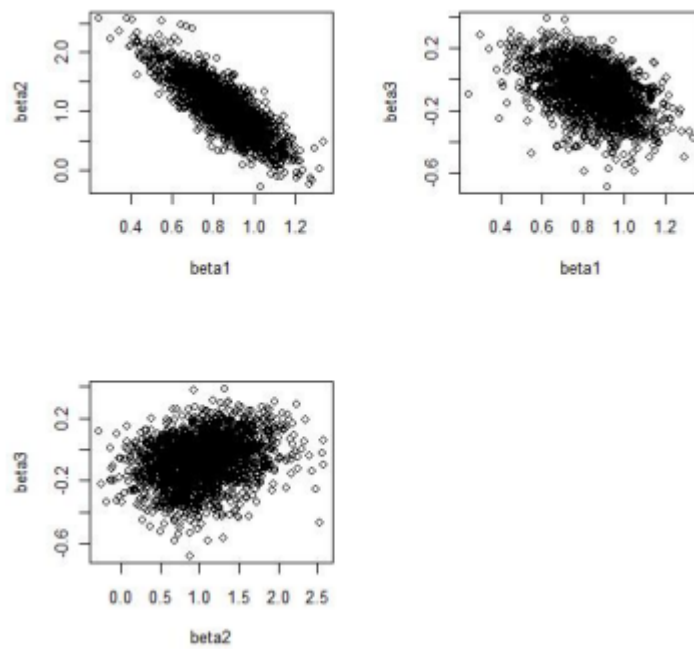


FIGURE 3.3 – Distributions jointes.

Quantiles de régression	Moyenne	Quantile 0.025	Médiane .	Quantile 0.975.
$\hat{\beta}_0(0.75)$	-48.325	-66.920	-49.666	-20.938
$\hat{\beta}_1(0.75)$	0.856	0.587	0.851	1.150
$\hat{\beta}_2(0.75)$	1.042	0.250	1.045	1.855
$\hat{\beta}_3(0.75)$	-0.065	-0.410	-0.043	0.178

TABLE 3.2 – Estimation par la méthode Yu et al des quantiles de régression pour $\tau = 0.75$

de visualiser l'auto-correlation après avoir retiré ces 1000 premières valeurs. On constate dès lors que l'auto-correlation diminue avec le nombre d'itérations.

Enfin, les graphiques (3,5) représentent la distribution de chaque paramètre de la régression quantile, toujours en ayant mis de côté les 1000 premières valeurs générées. On constate que les distributions pour $\hat{\beta}_1$ et $\hat{\beta}_2$ sont plutôt symétriques, ce qui n'est guère étonnant vu les valeurs très proches de la médiane et de la moyenne. On remarque par contre une légère dissymétrie pour $\hat{\beta}_0$ et $\hat{\beta}_3$, dissymétrie à gauche pour $\hat{\beta}_0$ avec donc une valeur moyenne supérieure à celle correspondant la médiane, et une dissymétrie à droite où la valeur moyenne est dans ce cas inférieure à celle de la moyenne, comme nous pouvions déjà le constater dans le tableau (3,2).

Interprétation

Enfin, pour terminer cet exemple, la distribution du quantile conditionnel d'ordre 0,75 pour des valeurs fixes des covariables. Pour trouver cette distribution, à savoir estimer le quantile conditionnel d'ordre 0,75 pour chaque paramètre généré, en omettant les 1000 premières valeurs, pour des valeurs fixées à 50, 20 et 80 pour les covariables et en tenant compte d'un intercept. Nous obtenons alors la distribution représentée sur la figure (3.5). Nous pouvons voir que cette distribution est asymétrique à gauche avec une valeur moyenne de 10.097 et une médiane de 9.955. Nous constatons également que 95% des valeurs plausibles pour ce quantile conditionnel sont comprises entre 8,163 et 12,919. Nous concluons que, pour ces valeurs fixes des covariables, il y a 75% de chances que le pourcentage d'ammoniac soit en moyenne

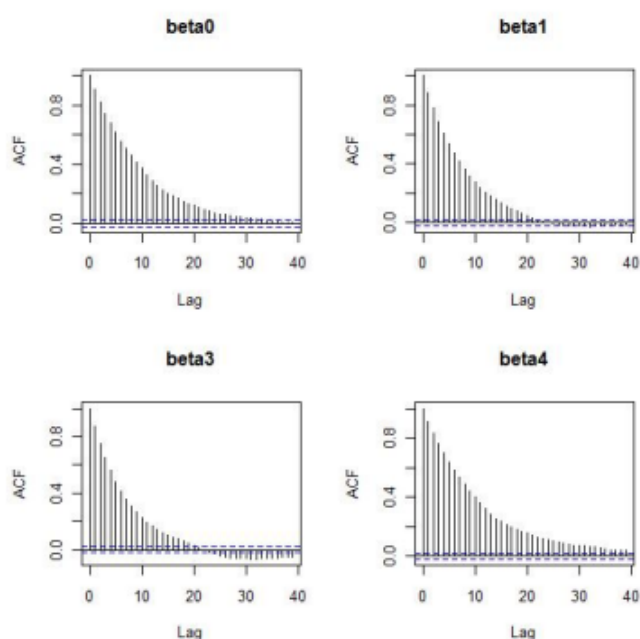


FIGURE 3.4 – Auto-corrélation avec 1 comme valeur initiale pour chaque paramètre..

Paramère	2.5% quantile	97.5% quantile	Moyenne .	Médiane.
$\hat{\beta}_0(0.95)$	-92 : 546	34.332	-44 : 259	-50.269
$\hat{\beta}_1(0.95)$	0.180	1.453	0.751	0.737
$\hat{\beta}_2(0.95)$	-0.165	2.731	1.478	1.549
$\hat{\beta}_3(0.95)$	-1.016	0.601	-0.098	-0.045

TABLE 3.3 – Estimation et IC pour 'échantillons empiriques de RQ , $v = 0.95$

inférieur à 0,101%.

3.3.3 Étude du quantile conditionnel d'ordre $\tau = 0,95$

De manière simulatoire que dans le cas du quantile 0,75 . les résultats de l'étude pour quantile $\tau = 0,95$ sont résumés dans le tableau (3,3) [57] . Pour les échantillons empiriques de la distribution a posteriori conjointe coefficients de RQ , pour $v=0.95$ que nous donnerons dans la figure (3,6)

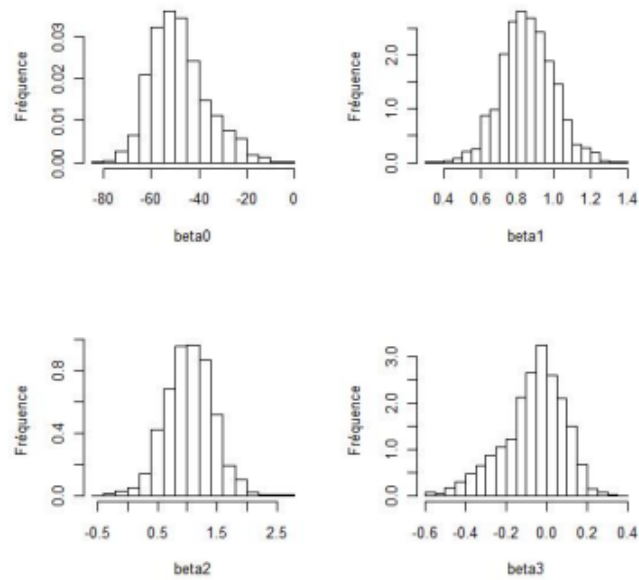


FIGURE 3.5 – Distribution de chaque quantile de régression avec 1 comme valeur initiale.

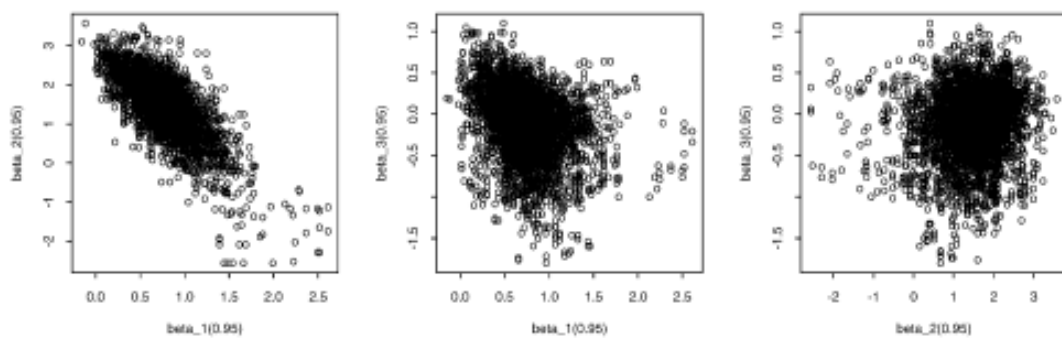
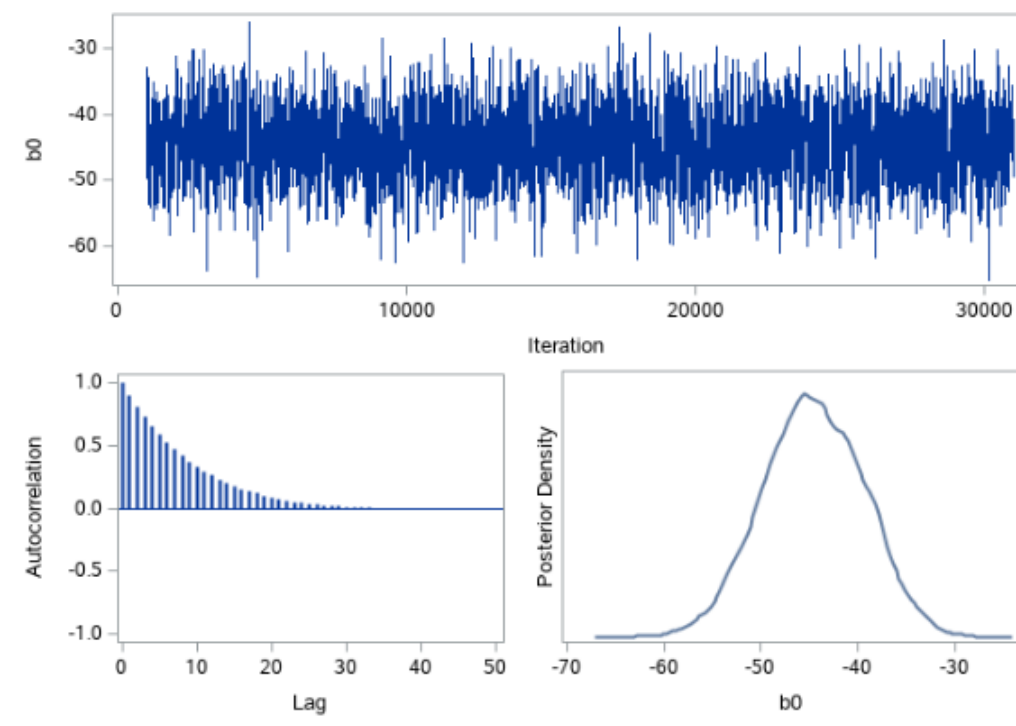


FIGURE 3.6 – Les échantillons empiriques des distributions conjointes des paramètres de régression quantile pour $v = 0.95$

Quantiles de régression	Estimation	Borne inf.	Borne sup.
$\hat{\beta}_0(0.5)$	-44.6143	-54.3075	-34.1243
$\hat{\beta}_1(0.5)$	0.764	0.6529	0.8819
$\hat{\beta}_2(0.5)$	0.5908	0.2101	1.0121
$\hat{\beta}_3(0.5)$	0.0383	-0.0587	0.1266

TABLE 3.4 – Estimation des quantiles de régression pour $\tau = 0.5$ FIGURE 3.7 – diagnostic pour β_0 .

3.3.4 Étude du quantile conditionnel d'ordre $\tau = 0,5$

Les résultats associés à l'étude du quantile conditionnel d'ordre $\tau = 0,5$ pour 30000 itération sont résumés dans le tableau (3.4) et les figures (3,7),(3,8),(3,9) et (3,10) :

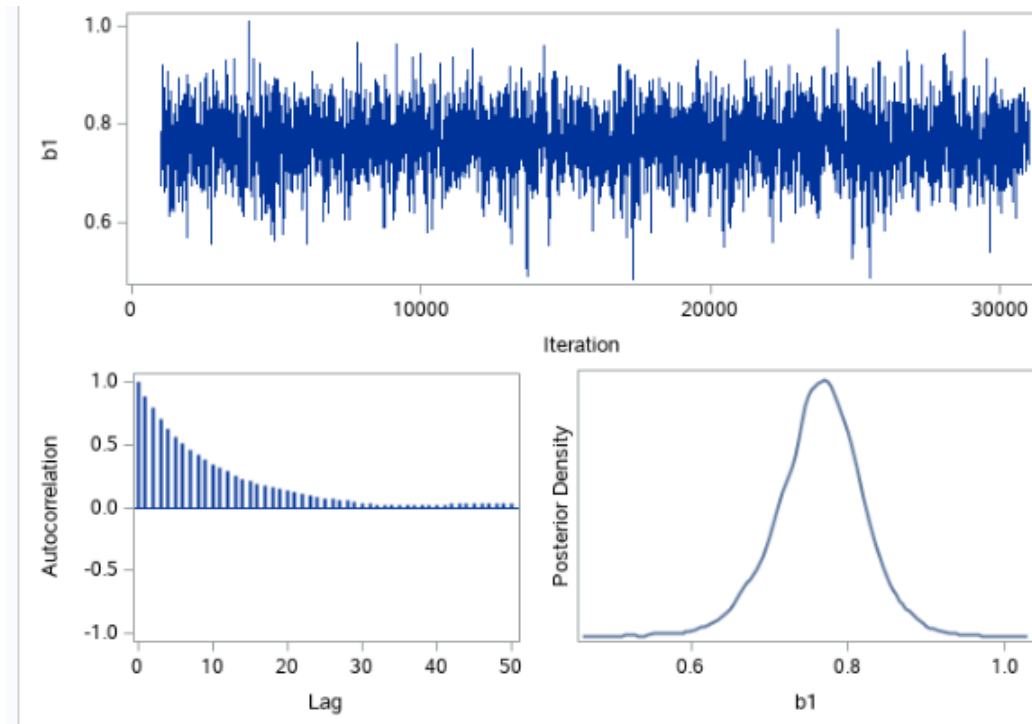


FIGURE 3.8 – diagnostic pour β_1 .

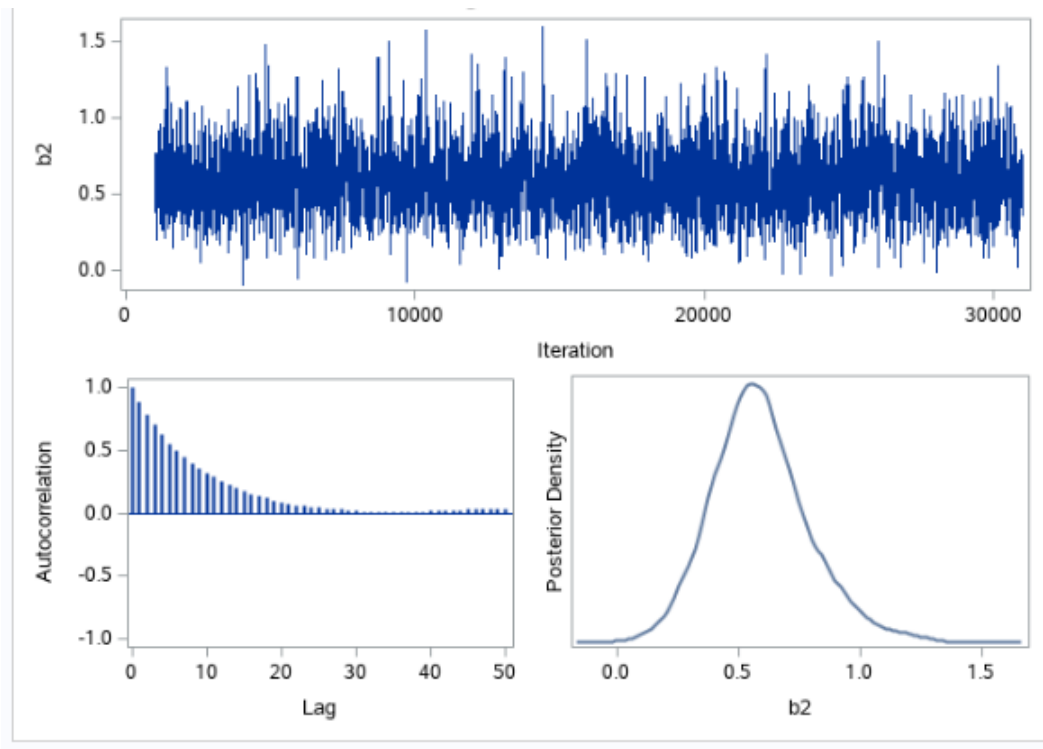
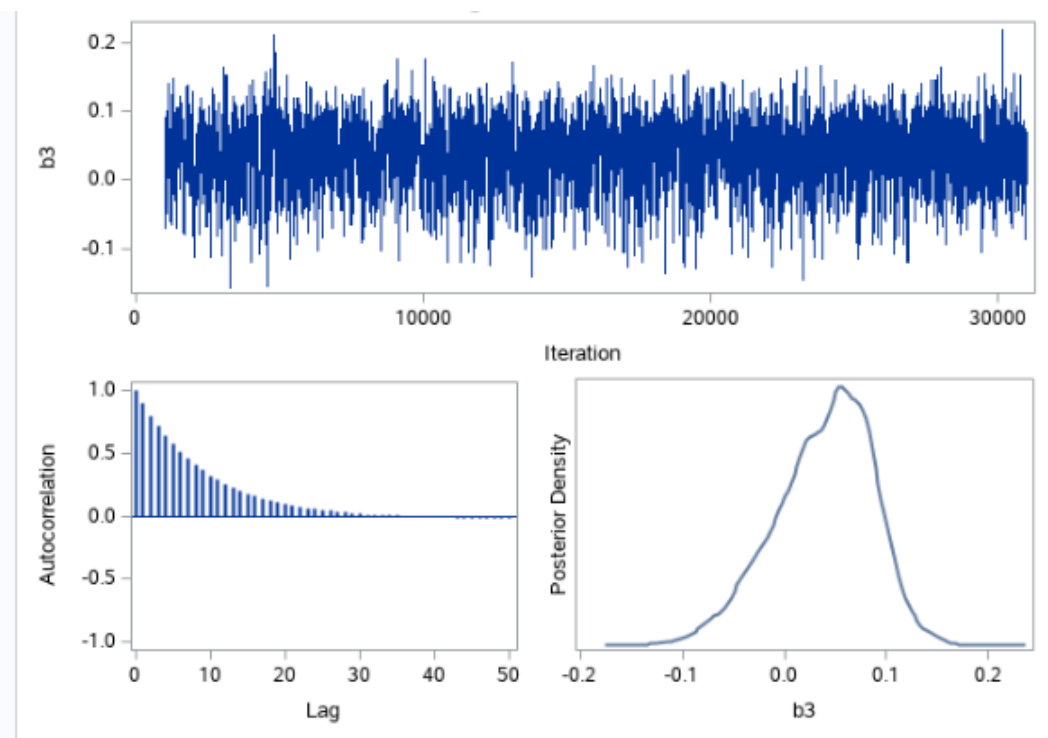


FIGURE 3.9 – diagnostic pour β_2 .

FIGURE 3.10 – diagnostic pour β_3 .

Conclusion

Nous avons présenté dans cet mémoire différentes approches de la régression quantile, où nous avons le cas bayésien, pour un certain nombre d'entrées elles ont été appliquées sur différents échantillons de données représentant différents cas (relation linéaire entre la variable dépendante et la covariable), et ce afin de conclure aux avantages et inconvénients de ces méthodes selon le cas considéré.

La régression quantile est un outil précieux pour faire face à l'hétéroscédasticité, et fournit une méthode pour modéliser les taux de changement de la variable de réponse en plusieurs points de la distribution lorsque ces taux de changement sont différents. de la distribution lorsque ces taux de changement sont différents. de la distribution lorsque ces taux de changement sont différents. Cependant, elle est également utile dans le cas de modèles de régression homogènes en dehors du modèle de régression normal classique, et dans le cas où l'hypothèse d'indépendance de l'erreur est violée, car aucune hypothèse de distribution paramétrique n'est requise pour la distribution des erreurs.

En cas de normalité, l'estimateur des MCO est plus efficace, mais lorsque les distributions sont non-normales, la précision de l'estimateur QR s'améliore par rapport aux MCO. L'estimateur QR est asymptotiquement normal comme le montre l'analyse des distributions empiriques du coefficient de pente dans un modèle de régression linéaire simple.

La comparaison entre les différentes estimations de QR est mise en

oeuvre en examinant les différences inter-quantiles. Il s'agit d'une question pertinente car les QR peuvent différer uniquement en ordonnée ou à la fois en ordonnée et en pente, permettent de déterminer lequel des deux est le cas. Dans le premier cas, la distribution conditionnelle de la variable dépendante est i.i.d., inversement, dans le second cas, l'hypothèse i.i.d. est irréaliste. Nous avons étudié dans ce mémoire la classe des modèles GLM qui contient les modèles linéaire classiques simple et donnerons une définition générale sur sa propriété. Aussi nous définirons le modèle RQ et leur propriétés et théorèmes. Il est important de mentionner que le fil conducteur des méthodes bayésiennes présentées ici était de supposer une distribution asymétrique de Laplace de paramètre de localisation μ nul, de paramètre d'échelle σ et de paramètre d'asymétrie τ fixé selon l'ordre du quantile conditionnel souhaité, cette distribution vérifiant l'hypothèse faite sur la distribution des résidus du modèle. Concernant les méthodes bayésiennes dites paramétriques, nous avons mentionné les définitions et les thèmes généraux de RQB tels qu'ils ont été présentés dans ce mémoire. Néanmoins, toutes les techniques de la régression quantile que nous venons de citer présentent le même désavantage qui est de devoir résoudre autant de problèmes de minimisation. Ou de réappliquer l'algorithme de Metropolis ou de Gibbs (en bayésienne) autant de fois que de quantiles conditionnelles d'ordre τ souhaités, pour une valeur à chaque fois différente et fixée de τ , ce qui peut s'avérer fastidieux. De plus, bien que théoriquement les droites de régressions ne devraient pas se croiser, il se pourrait que leurs estimations se croisent ce qui semble contraire à toute logique.

La distribution de Laplace comme distribution auxiliaire des erreurs [18] permet d'obtenir une équivalence formelle entre les modes postérieurs et les estimations habituelles de la régression quantile. Cependant, l'approche de régression quantile bayésienne présente encore des avantages considérables en ce qui concerne les spécifications de modèle généralisées. D'autres approches de régression quantile bayésienne tentent de contourner le problème de la distribution d'erreurs

auxiliaire en incluant la distribution d'erreurs dans le problème d'estimation et en déterminer les quantiles à partir de la distribution des erreurs estimée. Des exemples et des applications sont présentés afin d'illustrer la modélisation de régression quantile en utilisant R et SAS qui permettent d'implémenter l'approximation de ALD et les lois a priori.

La bibliographie jointe démontre par son volume que les modèles de régression ont ces dernières années su séduire un nombre grandissant de statisticiens et de chercheurs pour développer et à chercher la solution plus simple et facile de la régression.

Enfin, on a pu constater, au vu des nombreuses références citées dans ce mémoire, que la théorie des quantiles de régression est vaste, et est toujours en développement en particulier en statistique bayésienne.

Comme perspective nous proposons :

- Utiliser d'autres lois a priori et de vraisemblance pour l'étude bayésienne.
- Étudier l'estimation non paramétrique de la régression quantile par la méthode de spline et la méthode du noyau.

Annexe

A] Lois utilisés

La loi inverse gamma

Considérons la distribution Gamma Inverse $GI(\alpha_1, \beta_1)$ [41], de densité

$$f(x) = \frac{\beta_1^{\alpha_1} x^{-\alpha_1-1}}{\Gamma(\alpha_1)} \exp \frac{-\beta_1}{x}$$

et

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} \exp^{-x} dx \quad , a > 0$$

et $\alpha_1, \beta_1 \in \mathbb{R}^+$, Le Gamma inverse est défini sur le support $x \in]0, +\infty[$

.

On note X une v.a. de la loi $GI(\alpha_1, \beta_1)$ alors :

$$E(X) = \frac{\beta_1}{\alpha_1 - 1} \quad \text{et} \quad V(X) = \frac{\beta_1^2}{(\alpha_1 - 1)^2(\alpha_1 - 2)}$$

l'antériorité conjuguée pour β_1 le paramètre d'échelle de la distribution Gamma Inverse, est une distribution Gamma paramétrée à l'aide des paramètres d_1 et d'échelle e_1

$$p(\beta_1) = \mathcal{G}(\beta_1 \mid d_1, e_1) = \frac{e_1^{d_1} \beta_1^{d_1-1}}{\Gamma(d_1)} \exp(-e_1 \beta_1)$$

Étant donné les observations x , nous multiplions la vraisemblance des données de la distribution Gamma Inverse et la priorisation de l'échelle pour obtenir sa postérieure

$$q(\beta_1) = p(\beta_1 | x) \propto \left(\prod_{i=1}^n \frac{\beta_1^{\alpha_1} x_i^{-\alpha_1-1}}{\Gamma(\alpha_1)} \exp\left(\frac{-\beta_1}{x_i}\right) \right) \frac{e^d \beta_1^{d-1}}{\Gamma(d_1)} \exp(-e_1 \beta_1)$$

En ne gardant que les termes dépendant de β_1

$$q(\beta_1) \propto e^{d_1} \beta_1^{d_1-1+n\alpha_1} \exp\left(-\beta_1 \left(e_1 + \sum_{i=1}^n x_i^{-1}\right)\right)$$

donc la postérieure est une Gamma $q(\beta_1) = G(\beta_1 | \hat{d}_1, \hat{e}_1)$ avec des valeurs de paramètres

$$\hat{d}_1 = d_1 + n\alpha_1, \quad \hat{e}_1 = e_1 + \sum_{i=1}^n x_i^{-1}$$

la loi normale

Soit X une variable aléatoire est distribuée selon une loi normal [45] de paramètres $\mu_1 \in \mathbf{R}$ et $\sigma_1 \in]0, +\infty[$, $X \sim N(\mu_1, \sigma_1)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2} \left(\frac{x - \mu_1}{\sigma_1}\right)^2\right], \quad -\infty < x < +\infty$$

La fonction de densité de probabilité de la variable $Z = (X - \mu_1)/\sigma_1$ est :

$$g(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_1^2}{2}\right), \quad -\infty < z_1 < +\infty$$

et ne dépend pas des paramètres μ_1 et σ_1 . Cette distribution est appelée la loi normale standardisée $N(0, 1)$. Puisque

$$P\{X \leq x\} = P\left\{z_1 \leq \frac{x - \mu_1}{\sigma_1}\right\}$$

de telles probabilités peuvent être calculées à partir des tables de la fonction de distribution de Z , qui est :

$$\Phi(z_1) = \text{Prob}\{Z \leq z_1\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{t^2}{2}} dt$$

Si la loi normale standardisée est symétrique par rapport à l'origine, c'est-à-dire :

$$\Phi(z_1) = 1 - \Phi(-z_1)$$

L'estimation de paramètre $\hat{\mu}_1$ et $\hat{\sigma}_1$

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

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

La loi exponentielle

Soit X une v.a.r. de loi exponentielle [14] de paramètre $\lambda_1 > 0$, $X \sim \exp(\lambda_1)$

1. densité : $f_X(x) = \lambda_1 e^{-\lambda_1 x} 1_{[0, +\infty[}(x)$
2. fonction de répartition : $F_X(x) = \mathbb{P}(X \leq x) = (1 - e^{-\lambda_1 x}) 1_{[0, +\infty[}(x)$
3. espérance : $\mathbb{E}[X] = 1/\lambda_1$
4. variance : $\text{Var}(X) = 1/\lambda_1^2$

– Gaussienne inverse

La fonction de densité de probabilité d'une variable aléatoire X à distribution gaussienne inverse est

$$f(x; \mu_1, \lambda_1) = \sqrt{\frac{\lambda_1}{2\pi}} x^{-3/2} \exp\left(-\frac{\lambda_1(x - \mu_1)^2}{2\mu_1^2 x}\right), \quad x > 0$$

où $\mu_1 > 0$ et $\lambda_1 > 0$. Le paramètre μ_1 est la moyenne de la distribution et λ_1 est le paramètre d'échelle.

Les caractéristiques de modèle :

La fonction caractéristique de $X \sim \text{IG}(\mu_1, \lambda_1)$,

$$E[e^{itx}] = \int_0^\infty e^{itx} f(x; \mu_1, \lambda_1) dx$$

peut être facilement obtenu en combinant les deux termes de l'intégrante et en simplifiant le côté droit sous la forme

$$E[e^{itx}] = \exp \left\{ \frac{\lambda_1}{\mu_1} \left[1 - \left(1 - \frac{2it\mu_1^2}{\lambda_1} \right)^{1/2} \right] \right\} \\ \times \int_0^\infty f \left(x; \mu_1 \left(1 - \frac{2it\mu_1^2}{\lambda_1} \right)^{-1/2}, \lambda_1 \right) dx$$

Ainsi, la fonction caractéristique, désignée par $C_X(t)$, est donné par

$$C_X(t) = \exp \left\{ \frac{\lambda_1}{\mu_1} \left[1 - \left(1 - \frac{2i\mu_1^2 t}{\lambda_1} \right)^{1/2} \right] \right\}$$

Tous les moments positifs et négatifs existent. Les moments positifs d'ordre τ peuvent être obtenus en différenciant la fonction caractéristique sur $C_X(t)$ et les moments négatifs en intégrant la r ème dérivée de $C_X(t)$ et de l'évaluer à $t = 0$, on obtient

$$E[X^r] = \mu_1^r \sum_{s=0}^{r-1} \frac{(r-1+s)!}{s!(r-1-s)!} \left(2 \frac{\lambda_1}{\mu_1} \right)^{-s}$$

Loi asymétrique de Laplace

1. Densité

En dit que $U \sim ALD(\mu, \sigma, \tau)$ [58]

la fonction densité de probabilité est donné par :

$$f(u; \mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\rho_\tau \left(\frac{u-\mu}{\sigma} \right) \right\}$$

où $\rho_\tau(x)$ n'est rien d'autre que la fonction de perte déjà définie dans le chapitre 2 , définition (2.1), par

$$\rho_p(x) = x(\tau - I(x < 0))$$

avec

- τ : paramètre d'asymétrie tel que $\tau \in]1, 0[$,
- σ : paramètre d'échelle tel que $\sigma \in]0, +\infty[> 0$,

– μ : paramètre de localisation tel que $\mu \in \mathbb{R}$

en remarque que la fonction de densité définie ci-dessus vérifie les 3 conditions .

- (a) $f(u; \mu, \sigma, \tau) \geq 0$ car cette fonction est, \tilde{A} une constante positive près, une fonction exponentielle,
- (b) $f(u, \mu, \sigma, \tau)$ est continue et ce pour la même raison évoquée au point (a) ,
- (c) $\int_{-\infty}^{+\infty} f(u; \mu, \sigma, \tau) du = 1$, ce qui est vérifié ci-dessous.

On remarque aussi que pour une valeur de $\tau = 1/2$, on retrouve la fonction de densité de la loi standard de Laplace

$$f(u) = \frac{1}{4\sigma} \exp\left(\frac{-|u-\mu|}{2\sigma}\right)$$

qui est symétrique par rapport au paramètre de localisation μ . En fonction des valeurs de τ , la loi est asymétrique, avec une asymétrie vers la gauche, c'est-à-dire un étalement des observations vers la droite de la distribution pour des valeurs de τ inférieures à $\frac{1}{2}$ et une asymétrie vers la droite, c'est-à-dire un étalement des observations vers la gauche, lorsque τ est supérieur à $\frac{1}{2}$ ce qui est représenté respectivement sur les graphiques (3, 7) et (3, 8) en considérant un paramètre de localisation nul et un paramètre d'échelle égal à 1.

Le graphique(3, 9)représente quant à lui la fonction de densité pour des valeurs du paramètre d'échelle variant de 0.1 à 1, et ce pour un paramètre d'asymétrie et de localisation respectivement de 0.75 et 0.

2. Fonction de répartition

Soit une variable aléatoire $X \sim ALD(\mu, \sigma, \tau)$. Sa fonction de

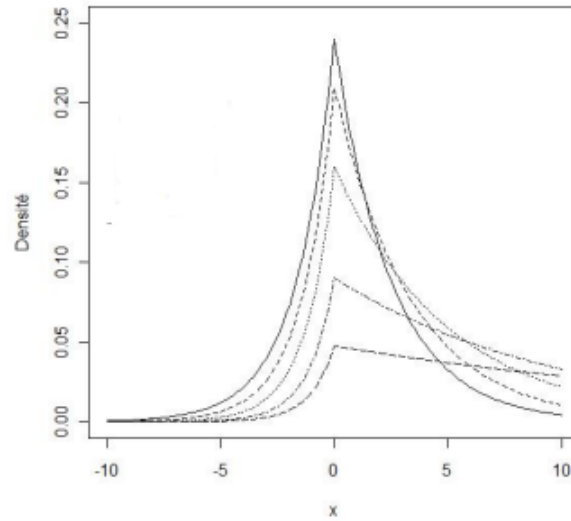


FIGURE 3.11 – Représentation de la fonction de densité pour des valeurs de τ inférieures à 0.5

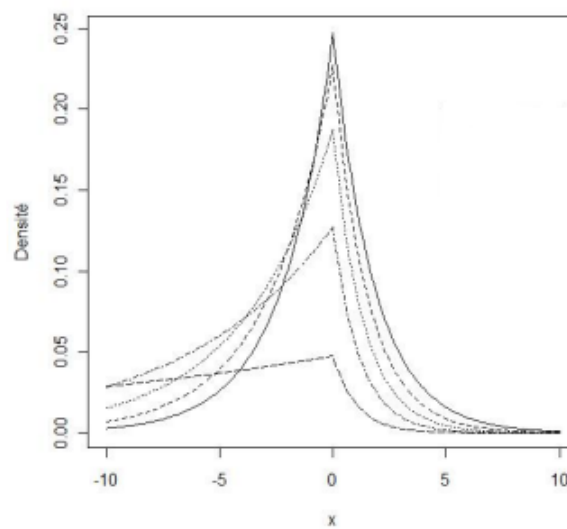


FIGURE 3.12 – Représentation de la fonction de densité pour des valeurs de τ supérieures à 0.5

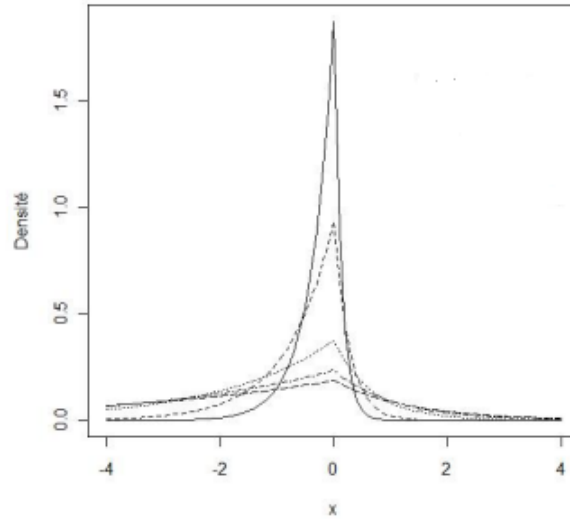


FIGURE 3.13 – Représentation de la fonction de densité pour différentes valeurs du paramètre d'échelle.

répartition est donnée par :

$$F(x; \mu, \sigma, \tau) = \begin{cases} \tau \exp\left(\frac{1-\tau}{\sigma}(x - \mu)\right) & \text{si } x \leq \mu, \\ 1 - (1 - \tau) \exp\left(\frac{-\tau}{\sigma}(x - \mu)\right) & \text{si } x > \mu. \end{cases} \quad (3.1)$$

En effet, par définition

$$F(x; \mu, \sigma, \tau) = \int_{-\infty}^x f(t; \mu, \sigma, \tau) dt$$

3. Propriétés des ALD

Proposition 3.3.1 *si l'on veut obtenir les valeurs de la moyenne, de la variance et par extension d'autres moments moments d'un certain ordre, une possibilité est de dériver successivement la fonction fonction génératrice des moments $E(e^{tx})$ et calculer cette dérivée en $t = 0$.*

La fonction génératrice est donnée par :

$$E[e^{tX}] = \tau(1 - \tau) \frac{e^{\mu t}}{(\tau - \sigma t)(\sigma t + 1 - \tau)}.$$

En dérivant cette fonction successivement et en l'évaluant ensuite à $t = 0$, les valeurs de la moyenne et de la variance sont :

$$E[X] = \mu + \frac{\sigma(1 - 2\tau)}{p(1 - \tau)}$$

$$\text{et } V[X] = \frac{\sigma^2 (1 - 2\tau + 2\tau^2)}{(1 - \tau)^2 \tau^2}.$$

Pour toutes les valeurs de τ supérieures à 0,5, on a un coefficient négatif alors que ce coefficient est positif pour les valeurs de τ inférieures à 0,5. Notons qu'il est égal à zéro lorsque τ est égal à 0,5. Le graphique (3,11) montre cette tendance .

Les coefficient d'asymétrie

Considérons l'échantillon aléatoire X de taille n tel que $X_i \stackrel{iid}{\sim} ALD(\mu, \sigma, \tau)$. L'estimateur par maximum de vraisemblance de μ est

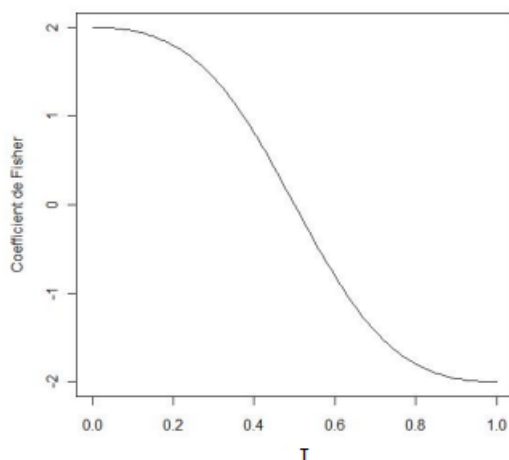
$$\hat{\mu} = \arg \max_{\mu \in \mathbb{R}} \frac{\tau^n (1 - \tau)^n}{\sigma^n} e^{-\sum_{i=1}^n \rho_\tau(x_i - \mu)}$$

$$= \arg \min_{\mu \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(x_i - \mu).$$

On retrouve la définition du quantile d'ordre τ telle que défini dans le chapitre 2. Ce résultat n'a rien d'étonnant étant donné que μ est, comme nous l'avons remarqué, le quantile d'ordre τ de la distribution.

Proposition 3.3.2 *Si $X \sim ALD(0, 1, p)$, alors $Y = \mu + \sigma X \sim ALD(\mu, \sigma, p)$. De plus, si $X \sim ALD(\mu, \sigma, p)$, alors $Y = \alpha + \beta X \sim ALD(\alpha + \beta\mu, \beta\sigma, p)$ Démonstration. Ces résultats s'obtiennent par un simple changement de variables. Il suffit donc de simuler deux variables aléatoires indépendantes selon une loi exponentielle standard pour obtenir, par combinaison linéaire, la loi asymétrique de Laplace voulue.*

Proposition 3.3.3 *Considérons deux variables aléatoires α et β indépendantes et distribuées tel que $\alpha \sim \text{Exp}(1)$ et $\beta \sim \text{Exp}(1)$. Dans ce cas, il vient que $\frac{\alpha}{\tau} - \frac{\beta}{1-\tau} \sim ALD(0, 1, \tau)$.*

FIGURE 3.14 – Evolution du coefficient de Fisher en fonction de τ

B] Algorithme de simulation

Algorithme de SAS

```
data by_stackloss ;  
set stackloss ;  
do    $\tau = 0.5$  ;  
output ;  
end ;  
run ;  
proc sort data=by_stackloss ;  
by  $\tau$  ;  
run ;  
ods output postsummaries=by_ps postintervals=by_pi ;  
proc mcmc data=by_stackloss  
seed=73625  
propcov=congra  
ntu=1000  
nmc=30000
```

```

mintune=17;
by  $\tau$  ; parms (b0-b3) 0;
prior b : ~ general(0);
mu= b0 + b1*airflow + b2*watertemp + b3*acidconc;
u = stackloss - mu;
ll = log(p)+log(1-p) - 0.5*(abs(u)+(2*p-1)*u);
model stackloss ~ general(ll);

run;
data process;
merge by_ps by_pi;
run;
proc sort data=process out=process;
by parameter  $\tau$  ;
run;
proc sgplot data=process(where=(parameter="b0"));
title "Estimated Parameter by Quantile";
title2 "With 95% HPD Interval";
series x= $\tau$  y=mean / markers legendlabel="Intercept (b0)";
band x= $\tau$  lower=hpdlower upper=hpdupper /
transparency=.5 legendlabel="HPD Interval";
yaxis label="Intercept (b0)";
xaxis label="Quantile";
refline 0 / axis=y;
run;
proc print data=process(where=(parameter="b0")) noobs;
title "Estimated Parameter b0 by Quantile";
title2 "with 95% HPD Interval";
var  $\tau$  mean hpdlower hpdupper;
run;

```

Bibliographie

- [1] *Christèle Bioche*. Approximation de lois impropres et applications. *PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2015*.
- [2] *Leo Breiman and Jerome H Friedman*. *Estimating optimal transformations for multiple regression and correlation*. *Journal of the American statistical Association*, pages 580–598, 1985.
- [3] *Brian S Cade and Barry R Noon*. *A gentle introduction to quantile regression for ecologists*. *Frontiers in Ecology and the Environment*, pages 412–420, 2003.
- [4] *Fabien Campillo, Philippe Cantet, Rivo Rakotozafy, and Vivien Rossi*. *Méthodes mcmc en interaction pour l'évaluation de ressources naturelles*. 2008.
- [5] *Probal Chaudhuri*. *Nonparametric estimates of regression quantiles and their local bahadur representation*. *The Annals of statistics*, pages 760–777, 1991.
- [6] *Siddhartha Chib and Edward Greenberg*. *Understanding the metropolis-hastings algorithm*. *The american statistician*, pages 327–335, 1995.
- [7] *Gregory C Chow*. *Tests of equality between sets of coeffi-*

- cients in two linear regressions. Econometrica : Journal of the Econometric Society, pages 591–605, 1960.*
- [8] *Pierre-André Cornillon and Eric Matzner-Løber. Régression : théorie et applications. Springer, 2007.*
- [9] *Pierre-André Cornillon and Eric Matzner-Løber. La régression linéaire simple, pages 1–28. Springer Paris, Paris, 2011.*
- [10] *Nadia Dardenne. Régression quantile bayésienne. PhD thesis, Université de Liège, Belgique, 2009.*
- [11] *Cristina Davino, Marilena Furno, and Domenico Vistocco. Quantile regression : theory and applications. John Wiley & Sons, 2013.*
- [12] *Pierre Lafaye de Micheaux, Rémy Drouilhet, and Benoît Liquet. Régression linéaire simple et multiple. In Le logiciel R, pages 375–422. Springer, 2011.*
- [13] *Xavier D’hault fœuille and Pauline Givord. La régression quantile en pratique. Economie et statistique, pages 85–111, 2014.*
- [14] *Alexandre Dumas. Nouvelles extensions de la loi exponentielle bivariée de Raftery. PhD thesis, Université du Québec à Trois-Rivières, 2019.*
- [15] *Jerôme Dupuis. Statistique bayésienne et algorithmes mcmc. Cours de Master, 1, 2007.*
- [16] *Eric Eboulet and Alina Matei. Application de la régression quantile et ses tests statistiques sur les données pisa en suisse romande. Technical report, Institut de recherche et de documentation pédagogique (IRDPA), 2013.*
- [17] *PHilippe Choné-Franck Evain and Lionel Wilner-Engin Yilmaz. Direction des études et synthèses économiques. 2013.*

- [18] Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian Marx. *Regression*. Springer, 2007.
- [19] James Fenske. *Land tenure and investment incentives : Evidence from west africa*. Journal of Development Economics, pages 137–156, 2011.
- [20] David A Freedman. *Bootstrapping regression models*. The Annals of Statistics, pages 1218–1228, 1981.
- [21] Marilena Furno. *Quantile regression analysis of the italian school system*. Statistical Modelling, pages 333–351, 2010.
- [22] Yémalin Gabin Gbedo. *Les techniques Monte Carlo par chaînes de Markov appliquées à la détermination des distributions de partons*. PhD thesis, Université Grenoble Alpes, 2017.
- [23] Djoweyda Ghouil. *Aspects de la robustesse Bayésienne dans les modèles*. PhD thesis, UMMTO, 2011.
- [24] Abdelhak Ghouil. *Contrôle optimal stochastique et mathématique financière*. PhD thesis, Université de Mohamed Ohammed Khider Biskra, 2011.
- [25] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.
- [26] Arnaud Guyader. *Régression linéaire*. Université Rennes, pages 60–61, 2011.
- [27] W Keith Hastings. *Monte carlo sampling methods using markov chains and their applications*. 1970.
- [28] Peter J Huber. *Robust regression : asymptotics, conjectures and monte carlo*. The annals of statistics, pages 799–821, 1973.

- [29] *Christophe Hurlin. Advanced econometrics hec lausanne.* University of Orléans, France, page 243, 2013.
- [30] *John Johnston and John DiNardo. Econometric methods.* 1963.
- [31] *Rima Khalfi, Souad Mesbout, and Zeyneb Encadreur Abdi. Propriétés asymptotiques des facteurs de Bayes. PhD thesis, Université jijel, 2018.*
- [32] *Muhammad Anas Knefati. Estimation non-paramétrique du quantile conditionnel et apprentissage semi-paramétrique : applications en assurance et actuariat. PhD thesis, Université de Poitiers, 2015.*
- [33] *Roger Koenker. Quantile regression. edited by. Econometric Society Monographs, pages 13–25, 2005.*
- [34] *Roger Koenker and Gilbert Bassett Jr. Regression quantiles. Econometrica : journal of the Econometric Society, pages 33–50, 1978.*
- [35] *Roger Koenker and Kevin F Hallock. Quantile regression. Journal of economic perspectives, pages 143–156, 2001.*
- [36] *Roger Koenker and José AF Machado. Gmm inference when the number of moment conditions is large. Journal of Econometrics, pages 327–344, 1999.*
- [37] *Roger Koenker and Jose AF Machado. Goodness of fit and related inference processes for quantile regression. Journal of the american statistical association, pages 1296–1310, 1999.*
- [38] *Roger Koenker and Quanshui Zhao. L-estimattion for linear heteroscedastic models. Journaltitle of Nonparametric Statistics, pages 223–235, 1994.*

- [39] Alexis Lechervy, Philippe-Henri Gosselin, and Frédéric Precioso. *Boosting actif pour la recherche interactive d'images*. In *Reconnaissance des Formes et Intelligence Artificielle*, page 1, 2010.
- [40] Oliver Linton and Yoon-Jae Whang. *The quantilogram : With an application to evaluating directional predictability*. *Journal of Econometrics*, pages 250–282, 2007.
- [41] A Llera and CF Beckmann. *Estimating an inverse gamma distribution*. arXiv preprint arXiv :1605.01019, 2016.
- [42] Himel Mallick and Nengjun Yi. *A new bayesian lasso*. *Statistics and its interface*, pages 571–582, 2014.
- [43] Ali Mohammad-Djafari. *Maximum entropy and linear inverse problems. a short review*. *Maximum Entropy and Bayesian Methods*, pages 253–264, 1993.
- [44] Trevor Park and George Casella. *The bayesian lasso*. *Journal of the American Statistical Association*, pages 681–686, 2008.
- [45] Luc Perreault and Bernard Bobée. *Loi normale : Propriétés mathématiques et statistiques : Estimation des paramètres et des quantiles x_t de période de retour t* . 1992.
- [46] Walter W Piegorsch and A John Bailer. *Minimum mean-square error quadrature*. *Journal of statistical computation and simulation*, pages 217–234, 1993.
- [47] Sandrine Poiraud-Casanova and Christine Thomas-Agnan. *Quantiles conditionnels*. *Journal de la société française de statistique*, 1998.
- [48] Dimitris N Politis and Joseph P Romano. *The stationary bootstrap*. *Journal of the American Statistical association*, pages 1303–1313, 1994.

- [49] *Judith Rousseau. Statistique bayésienne notes de cours.* ENSAE, Paris Tech, 2009.
- [50] *Ibrahim Sidi Zakari. Sélection de variables et régression sur les quantiles.* PhD thesis, Lille 1, 2013.
- [51] *Matthew A Taddy and Athanasios Kottas. A bayesian non-parametric approach to inference for quantile regression.* Journal of Business & Economic Statistics, pages 357–369, 2010.
- [52] *Wei Xiong and Maozai Tian. Weighted quantile regression theory and its application.* Journal of Data Science, pages 145–160, 2019.
- [53] *Ilker Yildirim. Bayesian inference : Metropolis-hastings sampling.* Dept. of Brain and Cognitive Sciences, Univ. of Rochester, Rochester, NY, 2012.
- [54] *Keming Yu and MC1614628 Jones. Local linear quantile regression.* Journal of the American statistical Association, pages 228–237, 1998.
- [55] *Keming Yu and Zudi Lu. Local linear additive quantile regression.* Scandinavian Journal of Statistics, pages 333–346, 2004.
- [56] *Keming Yu, Zudi Lu, and Julian Stander. Quantile regression : applications and current research areas.* Journal of the Royal Statistical Society : Series D (The Statistician), pages 331–350, 2003.
- [57] *Keming Yu and Rana A Moyeed. Bayesian quantile regression.* Statistics & Probability Letters, pages 437–447, 2001.
- [58] *Keming Yu and Jin Zhang. A three-parameter asymmetric laplace distribution and its extension.* Communications in Statistics-Theory and Methods, pages 1867–1879, 2005.

- [59] Yu Ryan Yue and Håvard Rue. *Bayesian inference for additive mixed quantile regression models*. Computational Statistics & Data Analysis, pages 84–96, 2011.
- [60] Qun Ying Zhu. *Modèles bayésiens et application à l'estimation des caractéristiques de produits finis et au contrôle de la qualité*. PhD thesis, Ecole Nationale des Ponts et Chaussées, 1991.

Abstract

The advantage of the quantile regression method is that it allows us to understand the relationships between the variables outside of the conditional mean of the response. Thus, in this document we focus on the standard regression for which the quantiles of the response variable Y are linear with the covariates. The estimation of the parameters of this model is given by an optimization problem of an objective function defined by the asymptotic weighted absolute error. The Bayesian estimator is obtained on the basis of the asymmetric Laplace distribution for well-defined a priori.

Résumé

L'avantage de la méthode de régression quantile est qu'elle nous permet de comprendre les relations entre les variables en dehors de la moyenne conditionnelle de la réponse. Ainsi, dans ce mémoire, nous nous concentrons sur la régression standard pour laquelle les quantiles de la variable réponse Y sont linéaires avec les covariables. L'estimation des paramètres de ce modèle est donnée par un problème d'optimisation d'une fonction objective définie par l'erreur absolue pondérée asymptotique. L'estimateur bayésienne est obtenu sur la base de la distribution asymétrique de Laplace pour des a priori bien définis.