Table des matières

Introduction générale

La statistique bayésienne donne un cadre formel séduisant, peut-être l'ultime rationalisation de la statistique classique, elle intervient dans toutes les disciplines scientifiques où se mèlent savoir et données. Elle est donc utilisée par les physiciens, les ingenieurs, les économistes, les géographes, les biologistes, les gestionnaires d'entreprises...etc. Et encore par tous les praticiens soucieux de bâtir sur des fondations solides un pont entre théorie et données experimentales par exemple l'article de **Berger** et **Insua (1998**) donne une vision intéressante et large des possibilités d'application de la statistique bayésienne. C'est aujourd'hui une science mathématique dont l'objectif est de décrire ce qui s'est produit et de faire des projections quant ce qu'il peut advenir dans le futur.

Différents livres ont été écrits sur l'histoire et le développement de cette discipline, notamment **Stigler (1986)**, **Dale (1991)**, **Lad (1996)** et **Hald (1998)**. Elle se base sur le même cadre de pensée de la statistique fréquentiste mais le paramètre θ n'est plus considéré comme étant totalement inconnu ; il est devenu une variable aléatoire à la quelle on associe une loi dite : «*loi a priori*», on utilise pour construire une loi dite : «*loi a posteriori*» un célèbre théorème «*théorème de Bayes»,* il est apparue pour la première fois en **1761** dans le cadre de l'exemple binomial exposé par le révérend **Thomas Bayes** devant la «*Royal Society*», et publié de façon posthume par son ami **R. Price** en **1763**. **Pierre Simon Laplace** redécouvrit ensuite cette formule dans une plus grande généralité en **1773**. Donc l'analyse statistique bayésienne vise à exploiter le plus efficacement possible l'information apportée par X sur le paramètre θ.

Aujourd'hui, la modélisation des relations entre les variables analysées dans un phénomène ou une expérience est plus étudiée. Notre travail traite la regréssion au sens bayésien, il est structuré en trois chapitre :

Le premier chapitre est un rappel sur la statistique bayésienne, on donne quelques notions de bases notament les lois a priori, lois a posteriori, le théorème de bayes.

Le deuxième chapitre consiste à presenter la régression linéaire bayésienne simple, la méthode d'estimationde paramètres par la Moindres Carées (MC) puis définie la loi a priori et a posteriori des paramètres $(\beta, \alpha_{\overline{x}})$.

Le troisieme chapitre est consacré à la régression bayésienne multiple, on traite deux cas (facteur de variance connu et inconu), on précice les lois a posteriori pour le couple (β, σ^2) à partir de vraisemblance et l'a priori.

Enfin, on illustre notre étude par une application réalisé dans le domaine médicale avec l'utilisation de langage R par le package **LearnBayes** et **Leaps**, les résultats apparaissent sous la forme des graphes.

Chapitre 1

Statistique bayésienne

1.1 Intorduction

Le modèle statistique bayésien est le triplet $(X, A, P_\theta, \theta \in \Theta)$. Ayant un a priori sur le paramètre, modélisé par une densité de probabilité que nous noterons $\pi(\theta)$, on "ré-actualise" cet a priori au vu de l'observation, en calculant la densité a posteriori $\pi(\theta/x)$, et c'est à partir de cette loi que l$'$on mène l'inférence.

Le paramètre θ devient donc en quelque sorte une variable aléatoire, à laquelle on associe une loi de probabilité dite *loi a priori*, il y a différentes méthodes existent pour déterminer ces lois a priori.

Dans ce chapitre on va rappeller quelques notions de bases sur la statistique bayésienne.

1.2 La formule de Bayes

Une description générale de l'inversion des probabilités est donnée par le théorème de **Bayes**.

Théorème 1.2.1 *soit deux évènements A et B aux quels sont associés les probabilités $P(A)$ et $P(B)$ avec $P(B) \neq 0$. La probabilité de l'évènment A conditionnellement à l'évènement B est :*

$$P(A/B) = \frac{P(A/B)}{P(B)}$$

Bayes et **Laplace** sont allés plus loin et ont considéré que l'incertitude sur le paramètre θ d'un modèle peut être décrit par une distribution de probabilité π sur θ appelée distribution a priori $\pi(\theta)$, l'inférence est alors fondée sur la distribution de θ sachant x notée $\pi(\theta/x)$ appelée distribution a posteriori est définie par :

$$\pi(\theta/x) = \frac{f(x/\theta)\,\pi(\theta)}{\int_\theta f(x/\theta)\,\pi(\theta)\,d\theta} \tag{1.1}$$

Où $f(x/\theta)$ est la fonction de vraisemblance à l'échantillon de n variables aléatoires iid $x = (x_1, \ldots, x_n)$, elles'écrit sous la forme :

$$f(x/\theta) = \prod_{i=1}^{n} f(x_i/\theta) \tag{1.2}$$

Et on peut construire :

La densité jointe de (x, θ) est :

$$f(x, \theta) = f(x/\theta)\,\pi(\theta) \tag{1.3}$$

La densité de x (prédictive) est :

$$p(x) = \int_\theta f(x/\theta)\,\pi(\theta)\,d\theta \tag{1.4}$$

Parfois, il est possible d'éviter le calcule d'integrale$\int_\theta f(x/\theta)\,\pi(\theta)\,d\theta$, puisque $\frac{1}{p(x)}$ est une constante donc :

$$\pi(\theta/x) \propto f(x/\theta)\,\pi(\theta) \tag{1.5}$$

$\propto$: "proportionelle à "

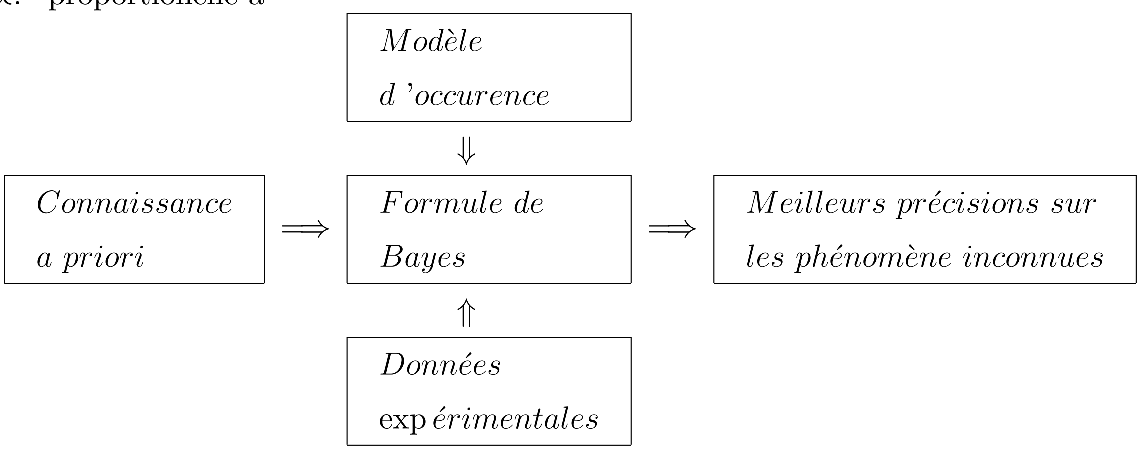


Figure 1.1 : Trajectoire pour trouver les meilleures précisions

Exemple 1.2.2

Soit $x_1, ..., x_n$ *un échantillon de distribution :*

$$f\left(x_i/\theta\right) = \theta(1-\theta)^{x_i-1}$$

Où $\theta \sim U_{[0,1]}$, $i = 1, ..., n$ *et* $\theta > 0$

Donc la loi a priori est :

$$\pi\left(\theta\right) = \mathbf{1}_{[0,1]}(\theta)$$

Calculons la vraisemblance :

$$f\left(x/\theta\right) = \prod_{i=1}^{n} f(x_i/\theta) = \prod_{i=1}^{n} \left(\theta(1-\theta)^{x_i-1}\right) = \theta^n(1-\theta)^{\sum_{i=1}^{n} x_i - n}$$

Calculons la densité marginale :

$$p\left(x\right) = \int_{\theta} f\left(x/\theta\right)\pi\left(\theta\right) d\theta = \int_{\theta} \theta^n(1-\theta)^{\sum_{i=1}^{n} x_i - n}\mathbf{1}_{[0,1]}(\theta)\ d\theta$$
$$= Beta\left(n+1, \textstyle\sum_{i=1}^{n} x_i - n + 1\right)$$

La loi a posteriori est :

$$\pi\left(\theta/x\right) = \frac{f\left(x/\theta\right)\pi\left(\theta\right)}{\int_{\theta} f\left(x/\theta\right)\pi\left(\theta\right) d\theta} = \frac{f\left(x/\theta\right)\pi\left(\theta\right)}{p\left(x\right)} = \frac{\theta^n(1-\theta)^{\sum_{i=1}^{n} x_i - n}}{Beta\left(n+1, \sum_{i=1}^{n} x_i - n + 1\right)}$$

D 'où

$$\theta/x \sim\ Beta(n+1, \sum_{i=1}^{n} x_i - n + 1)$$

1.3 Détermination de lois a priori

Le choix de ces lois a priori est une étape fondamentale, il peut avoir différentes motivations avec des stratégies qui sont diverses, elle peuvent se baser sur des expériences du passé ou sur un intuition, des lois composées comme suites :

1.3.1 Lois subjectives

Précisions tout d'abord que cette démarche n'est pas forcément facile dans la pratique. Quand l'espace des paramètres θ est fini, il est souvent possible d'obtenir une évaluation subjective des probabilités pour différentes valeurs de θ. Par exemple on peut utiliser des expériences précédentes du même type. Par contre s'il n'est pas dénombrable, la détermination subjective de la loi a priori π est beaucoup plus compliquée. Si l'information directe n'est pas disponible sur θ, une alternative est de recourir à la distribution marginale de x.

$$p(x) = \int_{\theta} f(x/\theta)\,\pi(\theta)\,d\theta \tag{1.6}$$

Nous pouvons citer la méthode d'entropie maximale comme une des méthodes utilisables pour obtenir l'information a priori.

1.3.2 Lois a priori conjuguées

Un a priori conjugué d'une vraisemblance est un a priori tel que l'a posteriori ait la même forme analytique que l'a priori. La seule justification pour utiliser de tels a priori est qu'ils facilitent les calculs. Cependant ils sont souvent assez souples pour ne pas rentrer en contradiction.

Définition 1.3.1 *La famille conjuguée*

Une famille $\mathcal{F}$ de lois sur Θ est dite conjuguée si pour tout π appartenant à cette famille, la loi a posteriori $\pi(\theta/x)$ appartient également à celle ci.($\mathcal{F}$)

Définition 1.3.2 *La famille exponentielle*

La famille exponentielle regroupe les lois de probabilités qui admettent une densité de la forme :

$$f(x/\theta) = h(x)\exp\{R(\theta)T(x) - C(\theta)\} \tag{1.7}$$

On pose $R(\theta) \equiv \theta$, donc le modèle exponentielle est dite canonique c'est-à-dire :

$$f(x/\theta) = h(x)\exp\{\theta T(x) - C(\theta)\} \tag{1.8}$$

Telle que

$T(x)$: fonction mesurable à valeurs dans $\mathbb{R}^d$

$C(\theta)$: fonction positive qui ne depend x

$h(x)$: fonction positive qui ne depend θ

Une telle famille est dite :

•*Régulièr si* Θ *est un ouvert telque :*

$$\Theta = \left\{ \theta / \int h(x) \exp(\theta T(x) - \psi(\theta) \prec \infty \right\}$$

•*Naturelle si :*

$$f(x/\theta) = h(x) \exp \{\theta T(x)\} \tag{1.9}$$

Considérons $f(x/\theta)$ *appartenant à une famille exponentielle alors la famille de loi a priori conjuguée est :*

$$\pi_{\mu,\lambda}(\theta) = h(\mu, \lambda) \exp(\theta\mu - \lambda c(\theta) \tag{1.10}$$

Où :

$h(\mu, \lambda)$ *est une constante.*

Et la loi a posteriori est de la forme :

$$\pi(\theta/x) \propto \exp \{(\mu + x)\theta - (\lambda + 1)c(\theta)\} \tag{1.11}$$

Le tableau ci-dessous représent quelques lois a priori conjuguées pour quelques familles exponentielles :

$f(x/\theta)$	$\pi(\theta)$	$\pi(\theta/x)$
$Normale$ $N(\theta, \sigma^2)$	$Normale$ $N(\mu, t^2)$	$Normale N(\rho(\sigma^2\mu + t^2x), \rho\sigma^2t^2))$ $\rho^{-1} = \sigma^2 + t^2$
$Poisson$ $P(\theta)$	$Gamma$ $G(\alpha, \beta)$	$Gamma$ $G(\alpha + x, \beta + 1)$
$Gamma$ $G(v, \theta)$	$Gamma$ $G(\alpha, \beta)$	$Gamma$ $G(\alpha + v, \beta + x)$
$Binomiale$ $B(n, \theta)$	$Beta$ $\beta e\,(\alpha, \beta)$	$Beta$ $\beta\,(\alpha + x, \beta + n - x)$
$Binomiale\,(-)$ $Neg(m, \theta)$	$Beta$ $\beta e\,(\alpha, \beta)$	$Beta$ $\beta e\,(\alpha + m, \beta + x)$
$Multinomiale$ $M_k\,(\theta_1, ..., \theta_k)$	$Dirichet$ $D(\alpha_1, ..., \alpha_k)$	$Dirichet$ $D\,(\alpha_1 + x_1, ..., \alpha_k + x_k)$
$Normale$ $N(\mu, \frac{1}{\theta})$	$Gamma$ $G(\alpha, \beta)$	$Gamma$ $G(\alpha + 0.5, \beta + (\mu - x)^2/2)$

Tableau 1.1 : quelques lois a priori conjuguées pour quelques familles exponentielles.

Exemple 1.3.3 *Loi binomiale*

Soit $X \sim B(n,x)$

Alors $f\,(X/\theta) = p(X = x/\theta) = C_n^x \theta^x (1-\theta)^{n-x}$

$$= \; C_n^x \exp\left\{\log\left(\theta^x (1-\theta)^{n-x}\right)\right\}$$

$$= C_n^x \exp\left\{x \log\theta + (n-x)\log(1-\theta)\right\}$$

$$= C_n^x \exp\left\{x \log\left(\tfrac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\}$$

Donc on a :

$$\begin{cases} R\,(\theta) = \log\left(\tfrac{\theta}{1-\theta}\right) \\ T\,(x) = x \\ C(x) = -n\log(1-\theta) \\ h(x) = C_n^x \end{cases}$$

1.3.3 Lois a priori impropres

La loi a priori peut être impropre ie :

$$\int_\theta \pi\,(\theta)\,d\theta = +\infty \tag{1.12}$$

Ce choix de type de loi n'a donc plus d'intérêt que calculatoire et s'interprète difficilement. Nous verrons pas la suite que la construction de lois non informatives peut conduire à des lois a priori de ce type.

Considérons la loi uniforme sur $\mathbb{R}^+$.

Suppposons

$$f(x/\lambda) = \lambda\exp\{-\lambda x\} \tag{1.13}$$

La loi a posteriori est :

$$\pi(\lambda/x) = \lambda\exp\{-\lambda x\} \tag{1.14}$$

1.3.4 Lois a priori non informatives

Quand on ne dispose pas d'information pertinente pour choisir une loi a priori et on veut cependant utiliser l'approche bayésienne on peut faire appel à des lois a priori non informatives. Pour obtenir une notion plus acceptable de ces lois, il faut citer les lois suivantes :

Lois a priori de Jeffeys

Jeffrey (**1946**, **1961**) propose une approche plus globale fondée sur l'information de Fisher $I(\theta)$ qui évite de prendre en compte une structure invariante particulière de la loi des observations.

Définition 1.3.4 *Cas unidimensionnel*

Les lois a priori de Jeffrey sont définie par :

$$\pi(\theta) \propto [I(\theta)]^{\frac{1}{2}} \tag{1.15}$$

Où :

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log\ f\,(x/\theta)}{\partial\theta}\right)^2\right]$$

Ce qui sous certaines conditions de régularité est égale à :

$$I(\theta) = -E_\theta\left[\left(\frac{\partial^2 \log\ f\,(x/\theta)}{\partial\theta^2}\right)\right] \tag{1.16}$$

Définition 1.3.5 *Cas multidimensionnel ($\theta \in \mathbb{R}^k$)*

La matrice d 'information de Fisher est une généralisation de (1.16) *donc $I(\theta)$ a les éléments suivants :*

$$I_{ij}\,(\theta) = -E_\theta\left[\left(\frac{\partial^2 \log f\,(x/\theta)}{\partial\theta_i\partial\theta_j}\right)\right] \quad (i,j=1,...,k) \tag{1.17}$$

Et la loi non informative de Jeffreys est définie par :

$$\pi(\theta) \propto [\ \det(I(\theta))]^{1/2} \tag{1.18}$$

Exemple 1.3.6 *Soit* $X \sim N(\mu, \sigma^2), \theta = (\mu, \sigma^2)$ *inconnu.*

On a

$$f(x/\mu, \sigma^2) \propto \frac{1}{\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Alors

$$\log f(x/\mu, \sigma^2) \propto -\frac{1}{2}\log\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\begin{cases} \frac{\partial}{\partial\mu} \log\ f(x/\mu, \sigma^2) = \frac{(x-\mu)^2}{\sigma^2} \\ \\ \frac{\partial}{\partial\sigma^2} \log\ f(x/\mu, \sigma^2) = \frac{-1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2} \end{cases}$$

Et

$$\begin{cases} \frac{\partial}{\partial\mu^2} \log\ f(x/\mu, \sigma^2) = -\frac{1}{\sigma^2} \\ \\ \frac{\partial}{\partial(\sigma^2)^2} \log\ f(x/\mu, \sigma^2) = \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3} \\ \\ \frac{\partial}{\partial\mu\partial\sigma^2} \log\ f(x/\mu, \sigma^2) = -\frac{(x-\mu)^2}{2(\sigma^2)^2} \end{cases}$$

On calcule l'esperance mathématique des dérivées secondes.

$$\begin{cases} E(X-\theta) = 0 \\ E((X-\theta)^2) = \sigma^2 \end{cases}$$

Donc :

$$I(\mu, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{bmatrix}$$

Et la loi de Jeffreys donnée par :

$$\pi(\theta) \propto \frac{1}{\sigma^2}$$

1.4 Les bases de la théorie de décision

La densité qu'on trouve par application directe du théorème de Bayes est centrale pour l'inférence bayésienne en ce qu'elle suffit à déterminer les procédures de décision et par extension conduire à toute inférence liée à θ.

1.4.1 Fonctions de pertes

Définition 1.4.1 *Fonction de coût (perte)*

Pour le modèle $X \in (\chi, B, \{P_\theta, \theta \in \Theta\})$*, on définit* D *l'ensemble des décisions possibles. Le but est d'estimer* θ *alors* $D = \Theta$.

On appelle fonction de perte (coût) la fonction mesurable L *de* $(\Theta \times D)$ *à valeurs dans* $\mathbb{R}^+$ *définie par :*

$$\begin{cases} L(\delta, \theta) \geq 0 & \forall (\delta, \theta) \\ L(\delta, \theta) = 0 & si\ \delta = \theta \end{cases}$$

On dit qu'on a pris la bonne décision si cette décision donne une perte nulle.

La fonction $L(\delta(X), \theta)$ *est une variable aléatoire car* X *est un vecteur aléatoire et elle ne nous permet pas de comparer deux règles de décision différentes. Donc on utilise le risque qui sera définie comme l'espérance de la perte pour faire la meilleure décision.*

Définition 1.4.2 *Risque fréquentiste*

On appelle risque fréquentiste le coût moyen (l'espérance mathématique) du coût d'une règle de décision :

$$R(\delta(X), \theta) = E_\theta [L(\delta(X), \theta)] = \int_\chi L[\delta(X), \theta] f(X/\theta) dp_\theta(X) \qquad (1.19)$$

On peut aussi donner la définition suivante :

On dira que δ_1 *est préférable à* δ_2 *si :*

$$R(\delta_1, \theta) \leq R(\delta_2, \theta) \qquad \forall \theta \in \Theta$$

Définition 1.4.3 *Risque a posteriori*

Le risque a posteriori est défini par :

$$\rho(\pi, \delta/X) = E^\pi [L(\delta(X), \theta)/x] = \int_\Theta L(\delta(X), \theta) \pi(\theta/X) d\theta \qquad (1.20)$$

qui moyenne l'erreur (c'est-à-dire le coût) selon la distribution du paramètre conditionnellement à la valeur observée x.

Définition 1.4.4 *Risque intégré (risque de Bayes)*

Pour une fonction de perte donnée, le risque intégré est définie par :

$$r(\pi,\delta) = E^{\pi}(R(\delta,\theta)) = \int_{\Theta} R(\delta,\theta)\,\pi(\theta)d\theta = \int_{\Theta}\int_{\chi} L(\delta(X),\theta)\;\; f(X/\theta)\,\pi(\theta)\,dXd\theta \qquad (1.21)$$

1.4.2 Estimateur bayésien

Un estimateur bayésien associé à une distribution a priori de densité et une fonction de coût L est un estimateur qui minimise δ^{π} ie :

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

Donc

$$\delta^{\pi}(X) = \arg\min_{\delta\in D} r(\pi,\delta/x)$$

δ^{π}est appelée l'estimateur du risque de Bayes minimum.

•Sous le perte quadratique, l'estimateur bayésien est la moyenne de la loi a posteriori.

Preuve. $L : (\Theta\times D) \to \mathbb{R}$, $L(\theta,\delta) = (\theta-\delta)^2$

On pose $f(\delta,X) = \rho(\pi,\delta/X) = E(L(\delta,\theta))$

$$= \int_{\Theta} L(\theta,\delta)\pi(\theta/X)\,d\theta$$

$$= \int_{\Theta} (\theta-\delta)^2\pi(\theta/X)\,d\theta$$

$$= \int_{\Theta} \theta^2\pi(\theta/X)\,d\theta + \delta^2 - 2\delta E(\theta/X)$$

La décision δ *qui minimise* $\rho(\pi,\delta/X)$ *est celle qui vérifie* :

$\frac{\partial f(\delta,X)}{\partial\delta} = 0$

$\Rightarrow 2\delta - 2E(\theta/X) = 0$

$\Rightarrow \delta = E(\theta/X)$

$\frac{\partial^2 f(\delta,X)}{\partial\delta^2} = 2 \succ 0$ ■

•Sous la perte absolue, l'estimateur bayesien est la médiane de la loi a posteriori :

Preuve. $L : (\Theta\times D) \to \mathbb{R}$, $L(\theta,\delta) = \mid\theta-\delta\mid$

$f(\delta,X) = \rho(\pi,\delta/X) = E(L(\delta,\theta)) = \int_{\Theta} \mid\theta-\delta\mid\pi(\theta/X)\,d\theta$

$\int\limits_{-\infty}^{\delta} -(\theta - \delta)\pi(\theta/X)\,d\theta \quad + \int\limits_{\delta}^{+\infty} (\theta - \delta)\pi(\theta/X)\,d\theta$

Nous cherchons à minimiser $f(\delta, X)$, donc il faut résoudre l'équation :

$\frac{\partial\ f(\delta,X)}{\partial\delta} = 0 \Longleftrightarrow \int\limits_{-\infty}^{\delta} (\theta - \delta)\pi(\theta/X)\,d\theta \quad - \int\limits_{\delta}^{+\infty} (\theta - \delta)\pi(\theta/X)\,d\theta = 0$

Ce qui implique :

$\int\limits_{-\infty}^{\delta} \pi(\theta/X)\,d\theta \quad = \int\limits_{\delta}^{+\infty} \pi(\theta/X)\,d\theta$

$\frac{\partial^2\ f(\delta,X)}{\partial\delta^2} = 2\pi(\theta/X) \succ 0$ ■

1.4.3 Estimateur MAP

Un estimateur du maximum a posteriori (MAP) c'est toute estimateur $\delta^{\pi}(X)$, qui maximise l'information sur θ représentée par sa loi a posteriori.

$$\delta^{\pi}(X) \in Arg\max\pi(\theta/x)$$

Chapitre 2

Régression bayésienne simple

2.1 Introduction

Parfois nous espérons modéliser une relation entre deux variables X et Y tell que le couple (x_i, y_i), $i = 1, \ldots, n$ sont les données où la variable prédictive X et la variable de réponse Y sont indépendantes. On veut trouver une équation qui décrit cette relation qui nous permet de prédire la valeur de Y.

Dans ce chapitre, on va analyser la régression simple au sens bayésien, on précise la relation entre X et Y, détermine la loi a priori et la loi a posteriori pour les paramètres.

Ensuite, on va déterminer les estimateurs de paramètres inconnus et la variance par la méthode des Moindres Carrées. Et on va terminer ce chapitre par une application illustrative qui résume les résultats.

2.2 Méthode des moindres carrées

Considérons la série statistique (x_i, y_i), $i = 1, \ldots, n$, l'équation de la régression est donnée par deux paramètres essentielles :

L'intercepte à l'origine α et la pente β c'est-à-dire :

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{2.1}$$

Où ε_i sont identiquement independant distribué (iid) et :

$$\varepsilon_i \sim N(0, \sigma^2) \tag{2.2}$$

La somme des carrés des résidus est :

$$SS_{res} = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2 \tag{2.3}$$

Les valeurs α et β sont trouvée à partir à la minimisation de SS_{res} ie :

$$\begin{cases} \frac{dSS_{res}}{d\alpha} = 0 \\ \frac{dSS_{res}}{d\beta} = 0 \end{cases} \Longleftrightarrow \begin{cases} \sum_{i=1}^{n} 2(y_i - (\alpha + \beta x_i))(-1) = 0 \\ \sum_{i=1}^{n} 2(y_i - (\alpha_0 + \beta x_i))(-x_i) = 0 \end{cases}$$

$$\Longleftrightarrow \begin{cases} \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \alpha - \sum_{i=1}^{n} \beta x_i = 0 \\ \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \alpha x_i - \sum_{i=1}^{n} \beta x_i^2 = 0 \end{cases}$$

$$\Longleftrightarrow \begin{cases} \overline{y} - \alpha - \beta \overline{x} = 0 \\ \overline{xy} - \alpha \overline{x} - \beta \overline{x^2} = 0 \end{cases}$$

$$\Longleftrightarrow \begin{cases} \widehat{\alpha} = \overline{y} - \widehat{\beta} \overline{x} \\ \widehat{\beta} = \frac{\overline{xy} - \overline{x}\ \overline{y}}{\overline{x^2} - \overline{x}^2} \end{cases}$$

Donc l'équation du modèle est :

$$y = \widehat{\alpha} + \widehat{\beta} x \tag{2.4}$$

2.3 La forme altertnative pour Moindre carée

La pente β et d'autre point détermine aussi l'équation du l'ajustement, on note $A_{\overline{x}}$ l'intercepte à la ligne verticale en $\overline{x}$ ie :

$$A_{\overline{x}} = \widehat{\alpha} + \widehat{\beta} \overline{x} = \overline{y} \tag{2.5}$$

Ainsi, la droite des moindres carées passe par le point $(\overline{x}, \overline{y})$.

Une equation alternative est :

$$\begin{aligned} y &= \widehat{\alpha} + \widehat{\beta} x \\ &= \widehat{\alpha} + \widehat{\beta} \overline{x} - \widehat{\beta} \overline{x} + \widehat{\beta} x \\ &= \widehat{\alpha} + \widehat{\beta} \overline{x} + \widehat{\beta}(x - \overline{x}) \\ &\overset{B=\widehat{\beta}}{=} A_{\overline{x}} + B(x - \overline{x}) \end{aligned}$$

Donc :

$$y = A_{\overline{x}} + B(x - \overline{x}) \tag{2.6}$$

2.4 Estimation de variance σ^2

L'estimateur de variance σ^2est donnée comme suit :

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - (A_{\overline{x}} + \beta(x_i - \overline{x})))^2}{n-2}$$

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

2.5 Fonction de vraisemblance de $\alpha_{\overline{x}}$ et β

Considérons l'équation du modèle suivante :

$$y_i = \alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right) + \varepsilon_i \qquad i = 1, \ldots, n \tag{2.7}$$

Où :

$\alpha_{\overline{x}}$: la valeure moyenne pour $y(\overline{x} = x)$.

β : la pente

ε_i : l'erreur $\quad \forall i = 1, \ldots, n$

pour toutes les observations, les erreurs sont iid et

$\varepsilon_i \sim N\left(0, \sigma^2\right)$

$y_i \sim N\left(\alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right), \sigma^2\right)$

La fonction de vraisemblance conjointe à la $i\grave{e}me$ observation est la fonction de densité de deux paramètres $\alpha_{\overline{x}}$ et β lorsque $(x_i, y_i), i = 1, \ldots, n$ fixé.

$$f_i\left(\alpha_{\overline{x}}, \beta\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(y_i - \left(\alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right)\right)\right)^2\right\} \tag{2.8}$$

Car on peut ignorer la partie qui contient pas les paramètres $\alpha_{\overline{x}}$ et β, donc :

$$\begin{aligned} f\left(\alpha_{\overline{x}}, \beta\right) &= \prod_{i=1}^{n} \exp\left\{-\tfrac{1}{2\sigma^2}\left(y_i - \left(\alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right)\right)\right)^2\right\} \\ &= \exp\left\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^{n}\left(y_i - \left(\alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right)\right)\right)^2\right\} \end{aligned}$$

On simplifie la formule suivante :

$I = \sum_{i=1}^{n}\left[\left(y_i - \left(\alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right)\right)\right)\right]^2 = \sum_{i=1}^{n}\left[\left(y_i - \overline{y} + \overline{y} - \left(\alpha_{\overline{x}} + \beta\left(x_i - \overline{x}\right)\right)\right)\right]^2$

$= \sum_{i=1}^{n}\left[\left(y_i - \overline{y}\right)^2 \pm 2\left(y_i - \overline{y}\right)\left(\overline{y} - (\alpha_{\overline{x}} + \beta(x_i - \overline{x}))\right) + \left(\overline{y} - (\alpha_{\overline{x}} + \beta(x_i - \overline{x}))\right)^2\right]$

$= \sum_{i=1}^{n}(y_i - \overline{y})^2 + 2\sum_{i=1}^{n}(y_i - \overline{y})(\overline{y} - (\alpha_{\overline{x}} + \beta(x_i - \overline{x}))) + \sum_{i=1}^{n}(\overline{y} - (\alpha_{\overline{x}} + \beta(x_i - \overline{x})))^2$

$= \sum_{i=1}^{n}(y_i - \overline{y})^2 + 2\sum_{i=1}^{n}(y_i - \overline{y})(\overline{y} - (\alpha_{\overline{x}} + \beta(x_i - \overline{x}))) + \sum_{i=1}^{n}(\alpha_{\overline{x}} - \overline{y})^2 - I_1$

Telleque :

$I_1 = 2\beta\sum_{i=1}^{n}(\alpha\overline{x} - \overline{y})(x_i - \overline{x}) + \beta^2\sum_{i=1}^{n}(x_i - \overline{x})^2$

$I = \sum_{i=1}^{n}(y_i - \overline{y})^2 - 2\beta\sum_{i=1}^{n}\left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right) + \beta^2\sum_{i=1}^{n}(x_i - \overline{x})^2 + n\left(\alpha_{\overline{x}} - \overline{y}\right)^2$

Car $(\overline{y} - \alpha_{\overline{x}} = 0)$

Donc :

$I = SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n\left(\alpha_{\overline{x}} - \overline{y}\right)$

Donc :

$f\left(\alpha_{\overline{x}}, \beta\right) = \exp\left[-\frac{1}{2\sigma^2}\left(SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n\left(\alpha_{\overline{x}} - \overline{y}\right)^2\right)\right]$

$\qquad = \exp\left[-\frac{1}{2\sigma^2}\left(SS_y - 2\beta SS_{xy} + \beta^2 SS_x\right)\right]\exp\left[-\frac{n}{2\sigma^2}\left(\alpha_{\overline{x}} - \overline{y}\right)^2\right]$

Donc :

$$f\left(\alpha_{\overline{x}}, \beta\right) = \exp\left[-\frac{1}{2\sigma^2/SS_x}\left(\beta - \frac{SS_{xy}}{SS_x}\right)^2\right]\exp\left[-\frac{n}{2\sigma^2}\left(\alpha_{\overline{x}} - \overline{y}\right)^2\right] \tag{2.9}$$

Notons : $\frac{SS_{xy}}{SS_x} = C$, $A_{\overline{x}} = \overline{y}$

$$f\left(\alpha_{\overline{x}}, \beta\right) \propto f\left(\alpha_{\overline{x}}\right) f\left(\beta\right) \tag{2.10}$$

Où :

$$f\left(\alpha_{\overline{x}}\right) = \exp\left[-\frac{n}{2\sigma^2}\left(\alpha_{\overline{x}} - A_{\overline{x}}\right)^2\right] \tag{2.11}$$

Et

$$f\left(\beta\right) = \exp\left[-\frac{SS_x}{2\sigma^2}\left(\beta - C\right)^2\right] \tag{2.12}$$

Dans le sense bayésienne on a :

$$\alpha_{\overline{x}} \sim N\left(A_{\overline{x}}, \frac{\sigma^2}{n}\right)$$

$$\beta \sim N\left(C, \frac{\sigma^2}{SS_x}\right)$$

2.5.1 La loi a priori conjointe

La loi a priori conjointe entre les paramètres $\alpha_{\overline{x}}$ et β c'est le produit de deux a priori c'est -à- dire :

$$\pi\left(\alpha_{\overline{x}}, \beta\right) = \pi\left(\alpha_{\overline{x}}\right)\pi\left(\beta\right) \tag{2.13}$$

Car $\alpha_{\overline{x}}$ et β sont independant.

2.5.2 La loi a posteriori conjointe

D'aprés le théoréme de Bayes, la loi a posteriori est données comme suite :

$$\pi\left(\alpha_{\overline{x}}, \beta/y\right) = \pi\left(\alpha_{\overline{x}}, \beta\right) f\left(\alpha_{\overline{x}}, \beta\right) \tag{2.14}$$

C'est le produit entre la loi a priori associé et la fonction de vraisemblance, comme on peut aussi la définir par le produit entre la loi a posteriori de chaque paramètre ie :

$$\pi\left(\alpha_{\overline{x}}, \beta/y\right) = \pi\left(\alpha_{\overline{x}}/y\right)\pi\left(\beta/y\right) \tag{2.15}$$

Tell que :

$$\pi\left(\beta/y\right) \propto \pi\left(\beta\right) f\left(\beta\right) \tag{2.17}$$

Si on prend la loi normale comme un a priori pour le paramètre $\beta($ $\beta \sim N\left(m_\beta, \sigma_\beta^2\right))$

La loi a posteriori est aussi la loi normale de moyenne $m_\beta^{'}$ et de variance $(\sigma_\beta^{'})^2$ ie :

$$\left(\beta/y \sim N\left(m_\beta^{'}, (\sigma_\beta^{'})^2\right)\right)$$

En effet :

D'après le tableau dans le chapitre 01 :

$$\beta/y \sim N\left(\rho\left(\frac{\sigma^2}{SS_x}m_\beta + \sigma_\beta^2 B\right), \rho\frac{\sigma^2}{SS_x}\sigma_\beta^2\right)$$
$$\rho = \frac{1}{\frac{\sigma^2}{SS_x}+\sigma_\beta^2}$$

$$\begin{cases} m_\beta^{'} = \rho\left(\frac{\sigma^2}{SS_x}m_\beta + \sigma_\beta^2 B\right) = \frac{1}{\frac{\sigma^2}{SS_x}+\sigma_\beta^2}\left(\frac{\sigma^2}{SS_x}m_\beta + \sigma_\beta^2 B\right) \\ \\ (\sigma_\beta^{'})^2 = \rho\frac{\sigma^2}{SS_x}\sigma_\beta^2 = \frac{1}{\frac{\sigma^2}{SS_x}+\sigma_\beta^2}\frac{\sigma^2}{SS_x}\sigma_\beta^2 \end{cases}$$

$$\Leftrightarrow \begin{cases} m_\beta^{'} = \frac{\frac{\sigma^2}{SS_x}}{\frac{\sigma^2}{SS_x}+\sigma_\beta^2}m_\beta + \frac{\sigma_\beta^2}{\frac{\sigma^2}{SS_x}+\sigma_\beta^2}B = \frac{1}{1+\sigma_\beta^2\frac{SS_x}{\sigma^2}}m_\beta + \frac{1}{\frac{\sigma^2}{\sigma_\beta^2 SS_x}+1}B \\ \\ (\sigma_\beta^{'})^2 = \frac{\frac{\sigma^2}{SS_x}\sigma_\beta^2}{\frac{\sigma^2}{SS_x}+\sigma_\beta^2} = \frac{\sigma_\beta^2}{1+\sigma_\beta^2\frac{SS_x}{\sigma^2}} \end{cases}$$

$$\Leftrightarrow \begin{cases} m_\beta^{'} = \frac{1}{\sigma_\beta^2(\frac{1}{\sigma_\beta^2}+\frac{SS_x}{\sigma^2})}m_\beta + \frac{1}{\frac{\sigma^2}{SS_x}\left(\frac{1}{\sigma_\beta^2}+\frac{SS_x}{\sigma^2}\right)}B \\ \\ (\sigma_\beta^{'})^2 = \frac{1}{(\frac{1}{\sigma_\beta^2}+\frac{SS_x}{\sigma^2})} \end{cases}$$

$$\Leftrightarrow \begin{cases} m_\beta^{'} = \frac{\frac{1}{\sigma_\beta^2}}{\frac{1}{(\sigma_\beta^{'})^2}}m_\beta + \frac{\frac{SS_x}{\sigma^2}}{\frac{1}{(\sigma_\beta^{'})^2}}B \\ \\ \frac{1}{(\sigma_\beta^{'})^2} = \frac{1}{\sigma_\beta^2} + \frac{SS_x}{\sigma^2} \end{cases}$$

Si on prend la loi normale comme un a priori pour le paramètre $\alpha_{\overline{x}}$, la loi a posteriori est la loi normale de moyenne $m_{\alpha_{\overline{x}}}^{'}$ et de variance $(\sigma_{\alpha_{\overline{x}}}^{'})^2$ ie :

$$\left(\alpha_{\overline{x}} \sim N\left(m_{\alpha_{\overline{x}}}, \sigma_{\alpha_{\overline{x}}}^2\right),\ \alpha_{\overline{x}}/y \sim N\left(m_{\alpha_{\overline{x}}}^{'}, (\sigma_{\alpha_{\overline{x}}}^{'})^2\right)\right)$$
$$\rho = \frac{1}{\frac{\sigma^2}{n}+\sigma_{\alpha_{\overline{x}}}^2}$$

$$\begin{cases} m_{\alpha_{\overline{x}}}^{'} = \rho\left(\frac{\sigma^2}{n}m_{\alpha_{\overline{x}}} + \sigma_{\alpha_{\overline{x}}}^2 A_{\overline{x}}\right) = \frac{1}{\frac{\sigma^2}{n}+\sigma_\beta^2}\left(\frac{\sigma^2}{n}m_{\alpha_{\overline{x}}} + \sigma_{\alpha_{\overline{x}}}^2 A_{\overline{x}}\right) \\ \\ (\sigma_{\alpha_{\overline{x}}}^{'})^2 = \rho\frac{\sigma^2}{n}\sigma_{\alpha_{\overline{x}}}^2 = \frac{1}{\frac{\sigma^2}{n}+\sigma_\beta^2}\frac{\sigma^2}{n}\sigma_{\alpha_{\overline{x}}}^2 \end{cases}$$

$$\Leftrightarrow \begin{cases} m'_{\alpha_{\overline{x}}} = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}+\sigma^2_{\alpha_{\overline{x}}}} m_{\alpha_{\overline{x}}} + \frac{\sigma^2_{\alpha_{\overline{x}}}}{\frac{\sigma^2}{n}+\sigma^2_{\alpha_{\overline{x}}}} A_{\overline{x}} = \frac{1}{1+\sigma^2_{\alpha_{\overline{x}}}\frac{n}{\sigma^2}} m_{\alpha_{\overline{x}}} + \frac{1}{\frac{\sigma^2}{n\sigma^2_{\alpha_{\overline{x}}}}+1} A_{\overline{x}} \\ \\ (\sigma'_{\alpha_{\overline{x}}})^2 = \frac{\frac{\sigma^2}{n}\sigma^2_{\alpha_{\overline{x}}}}{\frac{\sigma^2}{n}+\sigma^2_{\alpha_{\overline{x}}}} = \frac{\sigma^2_{\alpha_{\overline{x}}}}{1+\sigma^2_{\alpha_{\overline{x}}}\frac{n}{\sigma^2}} \end{cases}$$

$$\Leftrightarrow \begin{cases} m'_{\alpha_{\overline{x}}} = \frac{1}{\sigma^2_{\alpha_{\overline{x}}}(\frac{1}{\sigma^2_{\alpha_{\overline{x}}}}+\frac{n}{\sigma^2})} m_{\alpha_{\overline{x}}} + \frac{1}{\frac{\sigma^2}{n}\left(\frac{1}{\sigma^2_{\alpha_{\overline{x}}}}+\frac{n}{\sigma^2}\right)} A_{\overline{x}} \\ \\ (\sigma'_{\alpha_{\overline{x}}})^2 = \frac{1}{(\frac{1}{\sigma^2_{\alpha_{\overline{x}}}}+\frac{n}{\sigma^2})} \end{cases}$$

$$\Leftrightarrow \begin{cases} m'_{\alpha_{\overline{x}}} = \frac{\frac{1}{\sigma^2_{\alpha_{\overline{x}}}}}{\frac{1}{(\sigma'_{\alpha_{\overline{x}}})^2}} m_{\alpha_{\overline{x}}} + \frac{\frac{n}{\sigma^2}}{\frac{1}{(\sigma'_{\alpha_{\overline{x}}})^2}} A_{\overline{x}} \\ \\ \frac{1}{(\sigma'_{\alpha_{\overline{x}}})^2} = \frac{1}{\sigma^2_{\alpha_{\overline{x}}}} + \frac{n}{\sigma^2} \end{cases}$$

2.6 Application

Ici, nous voulons appliquer tout ce que nous avons parlé dans la partie théorique du côté médicale, exactement le sujet des crises cardiaques qui est directement responsable d'environ 10% de toutes les morts chaque année.

Les données proposées présentent le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans dans différents pays.

Les variables sont les suivantes :

Y : 100 [log(nombre de décès par crise cardiaque pour 100000 hommes de 55 à 59 ans) -2].

X : calories provenant de protéines animales en pourcentage du total des calories.(voir bibiliographie : référence nm 9)

Obs	*Pays*	*X*	*Y*	$\widehat{Y}$	*Obs*	*Pays*	*X*	*Y*	$\widehat{Y}$
1	*Australie*	8	81	71.88060	12	*Italie*	3	45	39.92164
2	*Autriche*	6	55	59.09701	13	*Japon*	3	24	39.92164
3	*Canada*	8	80	71.88060	14	*Mexique*	3	43	39.92164
4	*Ceylan*	2	24	33.52985	15	$Pays - Bas$	6	38	59.09701
5	*Chili*	4	78	46.31343	16	$Nouvelle - Zélande$	8	72	71.88060
6	*Danemark*	6	52	59.09701	17	*Norvège*	6	41	59.09701
7	*Finlande*	7	88	65.48881	18	*Portugal*	4	38	46.31343
8	*France*	7	45	65.48881	19	*Suède*	7	52	65.48881
9	*Allemagne*	6	50	59.09701	20	*Suisse*	7	52	65.48881
10	*Irlande*	5	69	52.70522	21	$Grande - Bretagne$	6	66	59.09701
11	*Brazile*	4	66	46.31343	22	$E'tats - Unis$	8	89	71.88060

Tableau 2.1 : Le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans et les calories provenant de protéines animales.

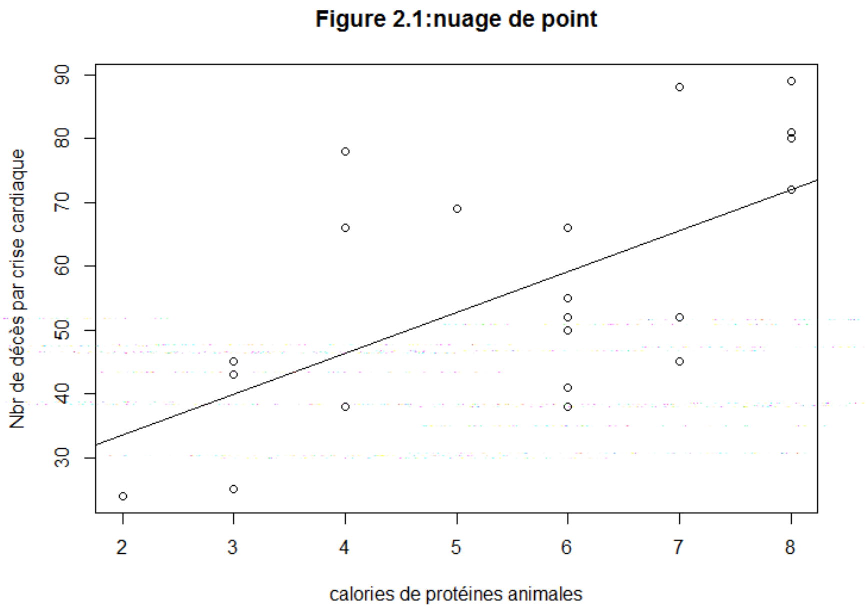
D'aprés les calcules qui on fait a l'aide de logiciel R (voir programme 1), on trouve :

$$\begin{cases} \widehat{\beta} = 6.391791 \\ \widehat{\alpha} = 20.74627 \end{cases}$$

L'équation alternative de la droite est :

$y \ = \overline{y} + B(x - \overline{x})$

$y = 56.77273 + 6.391791(x - 5.636364)$



Suppossons : $\beta \sim N\left(1.02, (0.25)^2\right), \alpha_{\overline{x}} \sim N\left(15, (1)^2\right)$

La loi a posteriori de β et $\alpha_{\overline{x}}$:

$$\begin{cases} m'_{\beta} = 1.101171 \\ \sigma'_{\beta} = 0.2476434 \end{cases}$$

D'où :

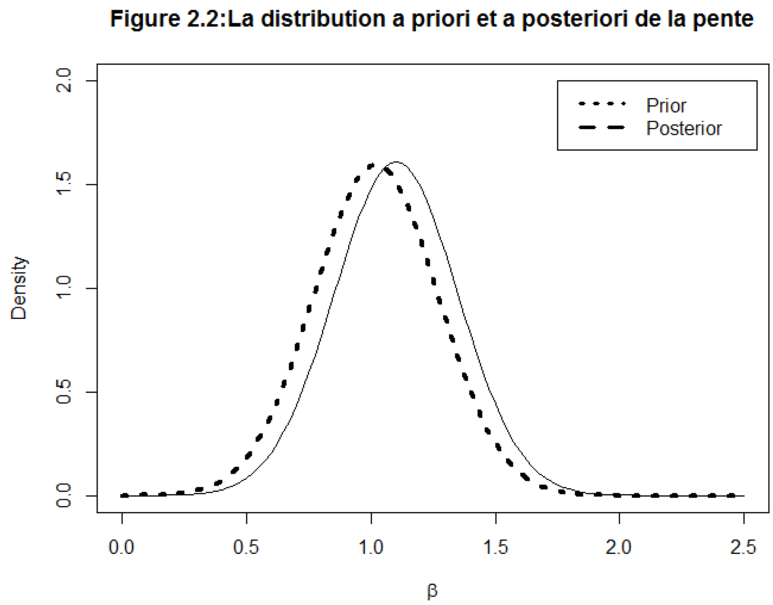
$$\beta/Y \sim N(1.101171, (0.247643)^2)$$

$$\begin{cases} m'_{\alpha_{\overline{x}}} = 18.5226 \\ \sigma'_{\alpha_{\overline{x}}} = 0.9569076 \end{cases}$$

D'où :

$$\alpha_{\overline{x}}/Y \sim N(18.5226, (0.9569076)^2)$$

D'aprés La figure (2.2), on remarque que la distribution a posteriori de la pente marche avec la distribution a priori (loi normale).



Chapitre 3

Régression bayésienne multiple

3.1 Introduction

Le modéle d'analyse est déterminé à partir des données préliminaires, elles sont réccueillies pour détecter et décrire la relation entre les paramètres inconnus et les observations pour précicer les proprietés statistiques. Ce qui conduit à estimer ces paramètres par les méthodes d'estimations connus.

Considérons X , τ deux variables ont une distribution *gamma normale* c'est -à-dire :

$$X, \tau \sim NG\left(m, W, a, b\right)$$

Où

W : est une matrice.

m, a, b : sont des paramètres données.

•La ditribution marginale a priori de x est la distribution $t - multivarié$ ie :

$$X \sim t\left(m, \frac{aW}{b}, 2b\right)$$

Où :

$$\begin{cases} E\left(X\right) = m \\ V\left(X\right) = 2b\left(2b-2\right)^{-1}\frac{aW}{b} = \left(b-1\right)^{-1}aW \end{cases}$$

•La distribution marginale a priori de τ est la distribution *gamma* ie :

$$\tau \sim G\left(a, b\right)$$

$\tau = \frac{1}{\sigma^2}$, donc la distribution marginale de σ^2 est l'inverse de gamma ie :

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

Où :

$$\begin{cases} E(\sigma^2) = \frac{a}{b-1} \\ V(\sigma^2) = \frac{a^2}{(b-1)^2(b-2)} \end{cases}$$

3.2 Fonction de vraisemblance

Soit $X_{(n.p)}$ une matrice avec un rang complet c'est-à-dire rang $(X) = p$.

$\beta_{(1.p)}$: un vecteur aléatoire de paramétres inconnus.

$Y_{(n.1)}$: un vecteur d'observation.

$V(Y/\sigma^2) = \sigma^2 P^{-1}$: la matrice de covariance $(n.n)$.

Où :

σ^2 : la variable aléatoire inconnu qui appellée facteur de variance où variance du poids.

P : la matrice du poids définie positive d'observation.

$\varepsilon_{(n.1)}$: un vecteur désigne l'erreur.

On considère le modèle suivant :

$$Y = X\beta + \varepsilon \tag{3.1}$$

Où :

$$\begin{cases} E(\varepsilon) = 0 \\ V(Y) = V(\varepsilon) = X\beta \end{cases}$$

Car les observations sont supposées comme normalement distribuées, donc la densité est donnée par :

$$f\left(Y, \beta, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} (\det P)^{-\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)' P (Y - X\beta)\right\} \tag{3.2}$$

3.3 La régression linéaire avec un facteur de variance connu

Dans ce cas le vecteur β est uniquement inconnu et on a deux cas :

3.3.1 A priori informative

Accord de théoréme de Bayes et la fonction de densité (3.2), on trouve la fonction de C'est -à-dire on a une information sur le vecteur β.

Soit :

$\beta \sim N(\mu, \sigma^2\Sigma\ )$

Où :

$$\begin{cases} E\left(\beta\right) = \ \mu \\ V\left(\beta\right) = \sigma^2\Sigma \end{cases}$$

Soit le vecteur Y suit une distribution normale et le facteur de variance σ^2 donné.

•La fonction de densité a priori de β donné par :

$$\pi\left(\beta\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{n}{2}}\left(\det\Sigma\right)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\left(\beta-\mu\right)\Sigma^{-1}\left(\beta-\mu\right)\right\} \tag{3.3}$$

On utilise le théorème de Bayes pour trouver la fonction de densité a posteriori.

$$\pi\left(\beta/Y\right) \propto \pi\left(\beta\right) f\left(Y, \beta, \sigma^2\right)$$

$$\pi\left(\beta/Y\right)\ \propto \frac{1}{\left(2\pi\sigma^2\right)^{\frac{n}{2}}\left(\det\Sigma\right)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\left(\beta-\mu\right)'\Sigma^{-1}\left(\beta-\mu\right)\right\} I$$

Tell que :

$I = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{n}{2}}\left(\det P\right)^{-\frac{1}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\left(Y-X\beta\right)'P\left(Y-X\beta\right)\right\}$

Donc :

$$\pi\left(\beta/Y\right)\ \ \propto \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\beta-\mu\right)'\Sigma^{-1}\left(\beta-\mu\right)+\left(Y-X\beta\right)'P\left(Y-X\beta\right)\right]\right\} \tag{3.4}$$

On simplifie l'exposant :

$\left(\beta-\mu\right)'\Sigma^{-1}\left(\beta-\mu\right)+\left(Y-X\beta\right)'P\left(Y-X\beta\right)$

$= \beta'\Sigma^{-1}\beta - \beta'\Sigma^{-1}\mu - \mu'\Sigma^{-1}\beta + \mu'\Sigma^{-1}\mu + Y'PY - 2\beta'X'PY + \beta'X'PX\beta$

$= Y'PY + \mu'\Sigma^{-1}\mu - 2\beta'(X'PY + \Sigma^{-1}\mu) + \beta'(X'PX + \Sigma^{-1})\beta$

$= Y'PY + \mu'\Sigma^{-1}\mu - \mu_0'\left(X'PX + \Sigma^{-1}\right)\mu_0 + (\beta - \mu_0)'\left(X'PX + \Sigma^{-1}\right)(\beta - \mu_0)$

Où :

$\mu_0 = \left(X'PX + \Sigma^{-1}\right)^{-1}\left(X'PY + \Sigma^{-1}\mu\right)$

Donc :

$$\pi\left(\beta/Y\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\left[Y'PY + \mu'\Sigma^{-1}\mu - \mu_0'H\mu_0 + (\beta - \mu_0)'H(\beta - \mu_0)\right]\right\}$$

Tell que : $H = (X'PX + \Sigma^{-1})$

La constante est négligeable :

$$\pi\left(\beta/Y\right) \propto \exp\left\{-\frac{1}{2\sigma^2}(\beta - \mu_0)'\left(X'PX + \Sigma^{-1}\right)(\beta - \mu_0)\right\} \tag{3.5}$$

Ie :

$$\beta/Y \sim N\left(\mu_0, \sigma^2\left(X'PX + \Sigma^{-1}\right)^{-1}\right) \tag{3.6}$$

Estimation par la méthode des moindres carrées

Les observations Y contiennent des informations sur les paramètres β.

On utilise la fonction de coût pour trouvé les paramètres estimé du vecteur β du sorte que la forme quadratique est minimisé :

$S(\beta) = (\beta - \mu_0)'\left(X'PX + \Sigma^{-1}\right)(\beta - \mu_0)$

$= (\beta^{'}(X^{'}PX + \Sigma^{-1})\beta - \mu_0^{'}(X^{'}PX + \Sigma^{-1}))(\beta - \mu_0)$

$= \beta^{'}(X^{'}PX+\Sigma^{-1})\beta - \mu_0^{'}(X^{'}PX+\Sigma^{-1})\beta - \beta^{'}(X^{'}PX+\Sigma^{-1})\mu_0 + \mu_0^{'}(X^{'}PX+\Sigma^{-1})\mu_0$

$\frac{\partial S(\beta)}{\partial \beta} = 2(X^{'}PX + \Sigma^{-1})\widehat{\beta} - 2(X^{'}PX + \Sigma^{-1})\mu_0$

$\frac{\partial S(\beta)}{\partial \beta} = 0 \iff \widehat{\beta} = (X^{'}PX + \Sigma^{-1})^{-1}(X^{'}PX + \Sigma^{-1})\mu_0$

$\widehat{\beta} = (X^{'}PX + \Sigma^{-1})^{-1}(X^{'}PX + \Sigma^{-1})(X^{'}PX + \Sigma^{-1})^{-1}(X^{'}PY + \Sigma^{-1}\mu)$

Donc :

$$\widehat{\beta} = (X^{'}PX + \Sigma^{-1})^{-1}(X^{'}PY + \Sigma^{-1}\mu) \tag{3.7}$$

Estimateur de Bayes

On a déja vu dans le premier chapitre que l'estimateur de Bayes notée $\widehat{\beta}_B$ est l'esperance mathématique :

$$\widehat{\beta}_B = \mu_0 = \left(X'PX + \Sigma^{-1}\right)^{-1} \left(X'PY + \Sigma^{-1}\mu\right) \tag{3.8}$$

Et la matrice de variance associée est :

$$V\left(\beta/Y\right) = \sigma^2 \left(X'PX + \Sigma^{-1}\right)^{-1}$$

Estimateur de MAP

L'estimateur de maximun a posteriori notée $\widehat{\beta}_M$ est :

$$\widehat{\beta}_M = \left(X'PX + \Sigma^{-1}\right)^{-1} \left(X'PY + \Sigma^{-1}\mu\right) \tag{3.9}$$

Exemple 3.3.1

Soit la quantité inconnu β qu'elle a mesuré n fois. Les observations obtenus $Y = (y_1, ..., y_n)$ sont indépendentes de différents poids.
$X = (1, ..., 1), P^{-1} = diag\left(\frac{1}{p_1}, ..., \frac{1}{p_n}\right)$.

Supposons que l'information a priori est existe telle que :

$$\begin{cases} E\left(\beta\right) = \mu \\ V\left(\beta\right) = \sigma^2\sigma_s^2 \end{cases}$$

Où :

$$\begin{cases} \Sigma = \sigma_s^2 \\ \Sigma^{-1} = \frac{1}{\sigma_s^2} \\ \widehat{\beta} = \widehat{\beta}_M = \widehat{\beta}_B = \left(X'PX + \Sigma^{-1}\right)\left(X'PY + \Sigma^{-1}\mu\right) \\ X'P = \left((1, ..., 1)'\right)' (p_1, ..., p_n) = \sum_{i=1}^{n} p_i \end{cases}$$

Et

$$\begin{cases} X'PX = \sum_{i=1}^{n} p_i \\ XPY = \sum_{i=1}^{n} p_i y_i \\ \Sigma^{-1}\mu = \frac{\mu}{\sigma_s^2} \end{cases}$$

D'ou :

$$\left\{\begin{array}{c} \widehat{\beta} = \frac{1}{\sum_{i=1}^{n} P_i + \frac{1}{\sigma_s^2}} \left(\sum_{i=1}^{n} p_i y_i + \frac{\mu}{\sigma_s^2} \right) \\ V(\beta \diagup y) = \sigma^2 \left(X' P X + \Sigma^{-1} \right)^{-1} = \sigma^2 \left(\sum_{i=1}^{n} p_i + \frac{1}{\sigma_s^2} \right)^{-1} \end{array}\right.$$

3.3.2 A priori non informative

Accord de théoréme de Bayes et la fonction de densité (3.2), on trouve la fonction de densité a posteriori de vecteur β sachant que :

$\pi(\beta) \propto const$

$\pi(\beta/Y) \propto \pi(\beta) * f(Y, \beta, \sigma^2)$

Donc :

$$\pi(\beta/Y) \ \propto \ \exp\left\{ -\frac{1}{2\sigma^2} (Y - X\beta)' P (Y - X\beta) \right\} \qquad (3.10)$$

On simplifie la formule suivante :

$I = (Y - X\beta)' P (Y - X\beta)$

$= Y'PY - 2\beta' X' PY + \beta' X' PX\beta$

$= Y'PY - \mu_0' (X'PX) \mu_0 + (\beta - \mu_0)' (X'PX) (\beta - \mu_0)$

Où :

$\mu_0 = (X'PX)^{-1} (X'PY)$

Donc :

$$\pi(\beta/Y) \ \propto \exp\left\{ -\frac{1}{2\sigma^2} \left(Y'PY - \mu_0' (X'PX) \mu_0 + (\beta - \mu_0)' (X'PX) (\beta - \mu_0) \right) \right\}$$

On prend le terme où il y'a β :

$$\pi(\beta/Y) \ \propto \exp\left\{ -\frac{1}{2\sigma^2} (\beta - \mu_0)' (X'PX) (\beta - \mu_0) \right\} \qquad (3.11)$$

On remarque qu'est une distribution du loi normale ie :

$$\beta/Y \sim N\left(\mu_0, \sigma^2 (X'PX)^{-1} \right) \qquad (3.12)$$

$$\begin{cases} E\left(\beta/Y\right) = \mu_0 \\ V\left(\beta/Y\right) = \sigma^2\left(X'PX\right)^{-1} \end{cases}$$

En effet :

$rang\left(X\right) = p \Longrightarrow rang\left(X'PX\right) = p \Longrightarrow \left(X'PX\right)^{-1}$ existe.

Donc le terme $\sigma^2\left(X'PX\right)^{-1}$ existe.

Estimation par la méthode des moindres carrées

Les observations Y contiennent des informations sur les paramètres β.

On utilise la fonction de coût par le choix du la différence $(Y - X\beta)$ et l'inverse de la matrice de covariance $V(Y/\beta) = (\sigma^2 P^{-1})^{-1}$.

Donc les paramètres de vecteur β sont estimés de telle sorte que la forme quadratique est minimisée :

$S\left(\beta\right) = \left(Y - X\beta\right)' P\left(Y - X\beta\right)$

$\frac{\partial S(\beta)}{\partial \beta} = -2X'PY + 2X'PX\ \widehat{\beta}$

$\frac{\partial S(\beta)}{\partial \beta} = 0 \Longleftrightarrow -2X'PY + 2X'PX\widehat{\beta} = 0$

Donc :

$$\widehat{\beta} = \left(X'PX\right)^{-1} X'PX \qquad (3.13)$$

Et on a :

$$\begin{aligned} E\left(\widehat{\beta}\right) &= \left(X'PX\right)^{-1} X'P\ E(Y) \\ &= \left(X'PX\right)^{-1} X'PX\beta \\ &= \beta \end{aligned}$$

$$\begin{aligned} V\left(\widehat{\beta}\right) &= \left(X'PX\right)^{-1} X'PD(Y)\left(\left(X'PX\right)^{-1} X'P\right)' \\ &= \left(X'PX\right)^{-1} X'P\ \sigma^2 P^{-1} PX\left(X'PX\right)^{-1} \\ &= \sigma^2\left(X'PX\right)^{-1} \end{aligned}$$

Estimateur de Bayes

L'estimateur de Bayes associé est :

$$\widehat{\beta}_B = \mu_0 = \left(X'PX\right)^{-1} X'PY \qquad (3.14)$$

Estimateur de MAP

L'estimateur de maximun a posteriori (MAP) est la fonction qui minimise $S(\beta)$ où :
$S(\beta) = (Y - X\beta)^{'} P (Y - X\beta) = Y'PY - 2\beta' X'PY + \beta' X'PX\beta$
$\frac{\partial S(\beta)}{\partial \beta} = -2X'PY + 2X'PX\ \widehat{\beta}$
$\frac{\partial S(\beta)}{\partial \beta} = 0 \Longleftrightarrow -2X'PY + 2X'PX\widehat{\beta} = 0$

$$\widehat{\beta} = (X'PX)^{-1} X'PY \tag{3.15}$$

Exemple 3.3.2

On prend l'exemple précedent (le cas informative) et supposons qu'il n'existe pas l'information a priori.

$$\begin{cases} X = (1, ..., 1) \\ P^{-1} = diag\left(\frac{1}{p_1}, ..., \frac{1}{p_n}\right) \\ X'P = ((1, ..., 1)')' (p_1, ..., p_n) = \sum_{i=1}^{n} p_i \\ X'PX = \sum_{i=1}^{n} p_i \\ X'PY = \sum_{i=1}^{n} p_i y_i \end{cases}$$

$\widehat{\beta} = \left(X'PX\right)^{-1} X'PY = \frac{1}{\sum_{i=1}^{n} P_i} \sum_{i=1}^{n} p_i y_i$

$V(\beta/y) = \sigma^2 (X'PX)^{-1} = \sigma^2 \left(\sum_{i=1}^{n} p_i\right)^{-1} = \frac{\sigma^2}{\sum_{i=1}^{n} p_i}$

3.4 La régression linéaire avec un facteur de variance inconnu

Contrairement à la partie précedente, maintenant le facteur de variance σ^2 est supposé comme un variable aléatoire inconnu et on remplace σ^2 par le paramètre de poids τ où $\tau = \frac{1}{\sigma^2}$, tell que :

$$\tau \sim G(b, p) \qquad b, p \succ 0$$

On sait que :

$$(Y, \beta, \tau) \sim N\left(X\beta, \tau^{-1} P^{-1}\right)$$

Car

$$\sigma^2 = \tau^{-1}$$

Supposons le vecteur β suit la loi normale ie :

$$\beta \sim N\left(\mu, \tau^{-1} V\right)$$

•La fonction de vraisemblance est donnée par :

$$f\left(Y, \beta, \tau\right) \quad = (2\pi)^{-\frac{n}{2}} \tau^{\frac{n}{2}} \left(\det P\right)^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left(Y - X\beta\right)' P\left(Y - X\beta\right)\right\} \qquad (3.16)$$

La fonction de densité jointe est :

$$\pi\left(\beta, \tau\right) = \pi\left(\beta\right) \pi\left(\tau\right) \qquad (3.17)$$

C'est une distribution de *gamma normale* notée $NG(u, V, b, p)$ où la fonction de densité donnée par :

$$\pi\left(\beta, \tau\right) = (2\pi)^{-\frac{n}{2}} \left(\det v\right)^{-1} b^p \left(\Gamma\left(p\right)\right)^{-1} \tau^{\frac{n}{2}+p-1} \exp\left\{-\frac{\tau}{2}(2b + \left(\beta - \mu\right)' V^{-1}\left(\beta - \mu\right))\right\} \quad (3.18)$$

$b, \tau \succ 0, p \succ 0, \tau \succ 0.$

Ie :

$$\beta, \tau \sim NG\left(\mu, V, b, p\right) \qquad (3.19)$$

3.4.1 A priori informative

Pour estimer les paramètres du modèle où les informations sont disponibles sur le vecteur β.

Notons :

$E\left(\beta\right) = \mu$: les valeurs attendus.

$V\left(\beta\right) = \sigma_p^2 \Sigma$: la matrice de covariance.

Et n'oublions pas que le facteur de variance σ^2 est également inconnu.

Supposons le vecteur attendus :

$$\begin{cases} E\left(\sigma^2\right) = \sigma_p^2 \\ V\left(\sigma^2\right) = V_{\sigma^2} \end{cases}$$

D'aprés la formule (3.19) on a :

•La distribution marginale pour le vecteur β est une distribution $t-multivari\acute{e}$ ie :

$$\beta \sim t\left(\mu, \frac{bv}{p}, 2p\right) \qquad (3.20)$$

Où :

$$\begin{cases} E(\beta) = \mu \\ V(\beta) = 2p(2p-1)^{-1}\frac{bV}{p} = \frac{b}{p-1}V \end{cases}$$

Par identification à l'information a priori on trouve :

$$\frac{b}{p-1}V = \sigma_p^2\Sigma \qquad (3.21)$$

•La distribution marginale pour le paramètre σ^2 est une distribution d'*inverse gamma* ie :

$$\sigma^2 \sim IG(b, p) \qquad (3.22)$$

Où :

$$\begin{cases} E(\sigma^2) = \frac{b}{p-1} = \sigma_p^2 & (3.23) \\ V(\sigma^2) = \frac{b^2}{(p-1)^2(p-2)} = V_{\sigma^2} & (3.24) \end{cases}$$

D'aprés (3.23) et (3.21) on remarque :

$$V = \Sigma$$

D'aprés (3.24) :

$$V(\sigma^2) = \frac{(\sigma_p^2)^2}{(p-2)} \Longrightarrow p = \frac{(\sigma_p^2)^2}{V_{\sigma^2}} + 2$$

D'aprés (3.23) :

$$b = \sigma_p^2\ (p-1) = \sigma_p^2\left(\frac{(\sigma_p^2)^2}{V_{\sigma^2}} + 2 - 1\right) = \sigma_p^2\left(\frac{(\sigma_p^2)^2}{V_{\sigma^2}} + 1\right)$$

D'où :

$$\beta, \tau \sim NG(\mu, \Sigma, b, p) \qquad (3.25)$$

•La loi a posteriori de vecteur β et le paramètre τ notée $\pi(\beta, \tau/Y)$ est :

$\pi\left(\beta,\tau/Y\right) \propto \pi\left(\beta,\tau\right) f\left(Y,\beta,\tau\right)$

$$\propto \tau^{\frac{p}{2}+p-1}\exp\left\{-\tfrac{\tau}{2}(2b+H_1)\right\}\tau^{\frac{n}{2}}\exp\left\{-\tfrac{\tau}{2}H_2\right\}$$

$$\propto \tau^{\frac{p}{2}+p+\frac{n}{2}-1}\exp\left\{-\tfrac{\tau}{2}\left[2b+H_1+H_2\right]\right\}$$

Tell que :

$H_1 = (\beta-\mu)'\,\Sigma^{-1}\,(\beta-\mu)$

$H_2 = (Y-X\beta)'\,P\,(Y-X\beta)$

On simplifie l'expression suivante :

$I = 2b + H_1 + H_2$

$= 2b + (\beta-\mu)'\,\Sigma^{-1}\,(\beta-\mu) + (Y-X\beta)'\,P\,(Y-X\beta)$

$= 2b + Y'PY + \mu'\Sigma^{-1}\mu - 2\beta'(X'PY+\Sigma^{-1}\mu) + \beta'(X'PX+\Sigma^{-1})\beta$

$= 2b + Y'PY + \mu'\Sigma^{-1}\mu - \mu_0'\left(X'PX+\Sigma^{-1}\right)\mu_0 + I_1$

Tell que :

$\mu_0 = (X'PX+\Sigma^{-1})^{-1}(X'PY+\Sigma^{-1}\mu)$

$I_1 = (\beta-\mu_0)'\left(X'PX+\Sigma^{-1}\right)(\beta-\mu_0)$

Donc

$I = 2b + Y'PY + \mu'\Sigma^{-1}\mu - 2\mu_0'\left(X'PX+\Sigma^{-1}\mu\right) + \mu_0'\left(X'PX+\Sigma^{-1}\right)\mu_0 + I_1$

$= 2b + (\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0) + (Y-X\mu_0)'\,P\,(Y-X\mu_0) + I_1$

$= 2b + I_2 + I_3 + I_1$

Tell que :

$I_2 = (\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)$

$I_3 = (Y-X\mu_0)'\,P\,(Y-X\mu_0)$

$$\pi(\beta,\tau/Y) \propto \tau^{\frac{p}{2}+p+\frac{n}{2}-1}\exp\left\{-\frac{\tau}{2}\left[2b+I_2+I_3+I_1\right]\right\} \tag{3.26}$$

D'où :

$$(\beta,\tau/Y) \sim NG(\mu_0, V_0, b_0, p_0) \tag{3.27}$$

$$\begin{cases} \mu_0 = (X'PX+\Sigma^{-1})^{-1}(X'PY+\Sigma^{-1}\mu) \\ V_0 = \left(X'PX+\Sigma^{-1}\right)^{-1} \\ b_0 = \frac{2b+(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)+(Y-X\mu_0)'P(Y-X\mu_0)}{2} \\ p_0 = \frac{n+2p}{2} \end{cases}$$

$$\Longleftrightarrow \begin{cases} \mu_0 = (X'PX + \Sigma^{-1})^{-1}(X'PY + \Sigma^{-1}\mu) \\ V_0 = (X'PX + \Sigma^{-1})^{-1} \\ b_0 = \frac{\left[2\sigma_p^2\left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+1\right)+(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)+(Y-X\mu_0)'P(Y-X\mu_0)\right]}{2} \\ p_0 = \frac{n+2\left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+2\right)}{2} = \frac{n+2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+4}{2} \end{cases}$$

•La distribution marginale a posteriori pour le vecteur β est donnée par :

$$\beta/Y \sim t\left(\mu_0, \frac{b_0V_0}{p_0}, 2p_0\right) \tag{3.28}$$

Avec

$$\begin{cases} E\left(\beta/Y\right) = \mu_0 \\ V\left(\beta/Y\right) = \left(p_0 - 1\right)^{-1} b_0V_0 \end{cases}$$

En effet :

$$V(\beta/Y) = \left(\frac{n+2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+4}{2} - 1\right)^{-1} \frac{\left[2\sigma_p^2\left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+1\right)+(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)+(Y-X\mu_0)'P(Y-X\mu_0)\right]}{2} I$$

Tell que :

$I = (X'PX + \Sigma^{-1})^{-1}$

Donc :

$$V(\beta/Y) = \left(n + 2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}} + 2\right)^{-1} I_1 I$$

Tell que :

$$I_1 = \left[2\sigma_p^2\left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}} + 1\right) +) \left(\mu - \widehat{\beta}\right)' \Sigma^{-1} \left(\mu - \widehat{\beta}\right) + \left(Y - X\widehat{\beta}\right)' P \left(Y - X\widehat{\beta}\right)\right]$$

Car $\widehat{\beta} = \mu_0$.

Estimation par la méthode des moindres carrées

On minimise $S\left(\beta\right)$:

$S\left(\beta\right) = \left(\beta - \mu_0\right)' \left(X'PX + \Sigma^{-1}\right) \left(\beta - \mu_0\right)$

$\frac{\partial S(\beta)}{\partial \beta} = \Longleftrightarrow \widehat{\beta} = (X'PX + \Sigma^{-1})^{-1}(X'PX + \Sigma^{-1})\mu_0$

Donc

$$\widehat{\beta} = (X'PX + \Sigma^{-1})^{-1}(X'PY + \Sigma^{-1}\mu) \tag{3.29}$$

Estimateur de Bayes

L'estimateur de Bayes $\widehat{\beta}_B$ est :

$$\widehat{\beta}_B = E(\beta/Y) = \mu_0 = (X'PX + \Sigma^{-1})^{-1}(X'PY + \Sigma^{-1}\mu) \tag{3.30}$$

Estimateur de MAP

L'estimateur MAP est :

$$\widehat{\beta}_M = (X'PX + \Sigma^{-1})^{-1}(X'PY + \Sigma^{-1}\mu) \tag{3.31}$$

•La ditribution marginale a posteriori pour le paramètre τ $\left(\tau = \frac{1}{\sigma^2}\right)$ est la distribution *gamma* ce qui implique que la distribution pour le facteur σ^2 est $l'inverse\ gamma$ ie :

$$\sigma^2/Y \sim IG(b_0, p_0) \tag{3.32}$$

Où :

$E(\sigma^2/Y) = \frac{b_0}{p_0-1}$

$$V(\sigma^2/Y) = \frac{b_0^2}{(p_0-1)^2(p_0-2)}$$

$$= \frac{\left(\frac{\left[2\sigma_p^2\left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+1\right)+(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)+(Y-X\mu_0)'P(Y-X\mu_0)\right]}{2}\right)^2}{\left(\frac{n+2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+4}{2}-1\right)^2\left(\frac{n+2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+4}{2}-2\right)}$$

Donc :

$$V\left(\sigma^2/Y\right) = \frac{2\left(\widehat{\sigma}_\beta^2\right)^2}{n + \frac{2\left(\sigma_p^2\right)^2}{V_{\sigma^2}}}$$

Où $\widehat{\sigma}_\beta^2$ est l'estimateur de Bayes.

$$\widehat{\sigma}_\beta^2 = E\left(\sigma^2/Y\right) = \frac{b_0}{p_0-1} = \frac{\frac{\left[2\sigma_p^2\left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+1\right)+(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)+(Y-X\mu_0)'P(Y-X\mu_0)\right]}{2}}{\left(\frac{n+2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}}+4}{2}-1\right)}$$

On a :

$\mu_0 = \widehat{\beta}$

$\widehat{\sigma}^2_\beta = \left(n + 2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}} + 2\right)^{-1} \left[2\sigma_p^2 \left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}} + 1\right) + \left(\mu - \widehat{\beta}\right)' \Sigma^{-1} \left(\mu - \widehat{\beta}\right) + I\right]$

Tell que

$I = (Y - X\ \widehat{\beta})'P(Y - X\ \widehat{\beta})$

On remarque :

$$V\left(\beta/Y\right) = \widehat{\sigma}^2_\beta \left(X'PX + \Sigma^{-1}\right)^{-1}$$

Exemple 3.4.1

Nous supposons maintenant pour le même exemple précédent, $y = (y_1, ..., y_n)'$ avec les variance $V\left(y_i\right) = \frac{\sigma^2}{p_i}$

Pour determiner la quantité β on donne l'a priori par les valeurs attendus :

$$\begin{cases} E\left(\beta\right) = \mu \\ V\left(\beta\right) = \sigma_p^2 \sigma_s^2 \end{cases}$$

Où :

$$\begin{cases} E\left(\sigma^2\right) = \sigma_p^2 \\ V\left(\sigma^2\right) = V_{\sigma^2} \end{cases}$$

$\Sigma = \sigma_s^2$

Avec

$$\widehat{\beta} = \widehat{\beta}_M = \widehat{\beta}_B$$

L'estimateur de Bayes pour la variance $\widehat{\sigma}^2$ est :

$$\widehat{\sigma}^2_B = \left(n + 2\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}} + 2\right)^{-1} \left[2\sigma_p^2 \left(\frac{\left(\sigma_p^2\right)^2}{V_{\sigma^2}} + 1\right) + \frac{\left(\widehat{\beta} - \mu\right)^2}{\sigma_s^2} + \sum_{i=1}^{n} p_i \left(\widehat{\beta} - y_i\right)^2\right]$$

Et

$$V\left(\sigma^2/Y\right) = \frac{2\left(\widehat{\sigma}^2_B\right)^2}{n + \frac{2\left(\sigma_p^2\right)^2}{V_{\sigma^2}}}$$

3.4.2 A priori non informative

On choisi la fonction de densité a priori non informative qui est déterminé par une constante pour le vecteur β c'est-à-dire $\pi(\beta) \propto const$ et la densité a priori non informative $\pi(\tau) \propto \frac{1}{\tau}$ qui est proportionnelle pour le paramètre de poids τ.

Le théorème de Bayes donne avec la fonction de vraissemblance (3.2) la fonction de densité a posteriori pour β et τ notée $\pi(\beta, \tau/Y)$ telleque les termes dependant de β et τ doivent être considérée seulement et :

$\pi(\beta, \tau/Y) \propto \pi(\tau) f(Y, \beta, \tau)$

$\pi(\beta, \tau/Y) \propto \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}(Y - X\beta)' P (Y - X\beta)\right\}$

L'exposant peut être transformé comme suit :

$I = (Y - X\beta)' P (Y - X\beta)$

$= Y'PY - 2\beta'X'PY + \beta'X'PX\beta$

$= Y'PY - 2\mu_0'(X'PY) + \mu_0'(X'PX)\mu_0 + (\beta - \mu_0)'(X'PX)(\beta - \mu_0)$

$= (Y - X\mu_0)' P (Y - X\mu_0) + (\beta - \mu_0)'(X'PX)(\beta - \mu_0)$

$= I_1 + (\beta - \mu_0)'(X'PX)(\beta - \mu_0)$

Avec

$\mu_0 = (X'PX)^{-1}(X'PY)$

$I_1 = (Y - X\mu_0)' P (Y - X\mu_0)$

Donc :

$$\pi(\beta, \tau/Y) \propto \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\left[I_1 + (\beta - \mu_0)'(X'PX)(\beta - \mu_0)\right]\right\} \tag{3.33}$$

La fonction de densité a posteriori est une distribution *gamma normale* c'est-à-dire :

$$\beta, \tau/Y \sim NG\left(\mu_0, (X'PX)^{-1}, \frac{n-p}{2}\widehat{\sigma}^2, \frac{n-p}{2}\right) \tag{3.34}$$

•La distribution marginale a posteriori pour le vecteur β est une distribution $t-$ *multivarié* ie :

$$\beta/Y \sim t\left(\mu_0, \widehat{\sigma}^2(X'PX)^{-1}, n-p\right) \tag{3.35}$$

En effet :

$$\begin{cases} \frac{aW}{b} = \frac{\frac{n-p}{2}\widehat{\sigma}^2 * (X'PX)^{-1}}{\frac{n-p}{2}} = \frac{n-p}{2}\widehat{\sigma}^2(X'PX)^{-1}\frac{2}{n-p} = \widehat{\sigma}^2(X'PX)^{-1} \\ 2b = 2\left(\frac{n-p}{2}\right) = n-p \end{cases}$$

$$\begin{cases} E\left(\beta/Y\right) = \mu_0 = (X'PX)^{-1}(X'PY) \\ V\left(\beta/Y\right) = \left(\frac{n-p}{2} - 1\right)^{-1}\left(\frac{n-p}{2}\widehat{\sigma}^2\right)(X'PX)^{-1} = \frac{n-p}{n-p-2}\widehat{\sigma}^2(X'PX)^{-1} \end{cases}$$

Estimation par la méthode des moindres carrées

On minimise $S\left(\beta\right)$:

$S\left(\beta\right) = \left(\beta - \mu_0\right)'\left(X'PX\right)\left(\beta - \mu_0\right)$

$\frac{\partial S(\beta)}{\partial \beta} = \Longleftrightarrow \ \widehat{\beta} = (X'PX)^{-1}(X'PX)\mu_0$

Donc

$$\widehat{\beta} \ = (X'PX)^{-1}(X'PY) \tag{3.36}$$

Estimateur de Bayes

L'estimateurde Bayes est :

$$\widehat{\beta}_B = \mu_0 = (X'PX)^{-1}(X'PY) \tag{3.37}$$

Estimateur de MAP

L'estimateur de maximun a posteriori (MAP) notée est :

$$\widehat{\beta}_M = (X'PX + \Sigma^{-1})^{-1}(X'PY + \Sigma^{-1}\mu) \tag{3.38}$$

•La distribution marginale a posteriori de paramètre τ est :

$$\tau/Y \sim G\left(\frac{n-p}{2}\widehat{\sigma}^2, \frac{n-p}{2}\right) \tag{3.39}$$

On sait que $\tau = \frac{1}{\sigma^2}$, d'où la distribution marginale a posteriori pour le facteur de variance σ^2 est :

$$\sigma^2/Y \sim IG\left(\frac{n-p}{2}\widehat{\sigma}^2, \frac{n-p}{2}\right) \tag{3.40}$$

Avec

$$\left\{\begin{array}{r} E\left(\sigma^2/Y\right)=\frac{\frac{n-p}{2}\widehat{\sigma}^2}{\frac{n-p}{2}-1}=\frac{n-p}{n-p-2}\widehat{\sigma}^2 \\ V\left(\sigma^2/Y\right)=\frac{\left(\frac{n-p}{2}\widehat{\sigma}^2\right)^2}{\left(\frac{n-p}{2}-1\right)^2\left(\frac{n-p}{2}-2\right)}=\frac{2(n-p)^2\left(\widehat{\sigma}^2\right)^2}{(n-p-2)^2(n-p-4)} \end{array}\right.$$

Estimateur de Bayes

L'estimateur de Bayes est :

$$\widehat{\sigma}_B^2=E\left(\sigma^2/Y\right)=\frac{n-p}{n-p-2}\widehat{\sigma}^2 \tag{3.41}$$

Estimateur de MAP

L'estimateur de maximun a posteriori (MAP) est :

$$\widehat{\sigma}_M^2=\frac{n-p}{n-p-2}\widehat{\sigma}^2 \tag{3.42}$$

3.5 Application

D'aprés l'application de chapitre 2, on considère d'autres variables :

X_1 : téléphones par millier d'habitants, noté TMH.

X_2 : calories grasses en pourcentage du total des calories, noté CGP.

X_3 : calories provenant de protéines animales en pourcentage du total des calories, noté CPA.

Obs	*pays*	X_1	X_2	X_3	Y	*Obs*	*pays*	X_1	X_2	X_3	Y
1	*Australie*	124	33	8	81	12	*Italie*	22	21	3	45
2	*Autriche*	49	31	6	55	13	*Japon*	16	8	3	24
3	*Canada*	181	38	8	50	14	*Mexique*	10	23	3	43
4	*Ceylan*	4	17	2	24	15	*Pays − Bas*	63	37	6	38
5	*Chili*	22	20	4	78	16	*Nouvelle− Zélande*	170	40	8	72
6	*Danemark*	152	39	6	52	17	*Norvège*	125	38	6	41
7	*Finlande*	75	30	7	88	18	*Portugale*	12	25	4	38
8	*France*	54	29	7	45	19	*Suède*	221	39	7	52
9	*Allemagne*	43	35	6	50	20	*Suisse*	171	33	7	52
10	*Irlande*	41	31	5	69	21	*Grande− Bretagne*	97	38	6	66
11	*Brazil*	17	23	4	66	22	*E′tats − Unis*	254	39	8	89

Tableau 3.1 : Le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans et les variables X_1, X_2, X_3 .

Pour simplifier les calcules supposons que :

$Y \sim N\left(X\beta, \sigma^2 I_n\right)$

$\beta/Y \sim N\left(\widehat{\beta}, \sigma^2 V_\beta\right)$

Tell que :

$\widehat{\beta} = (X'X)^{-1}X'Y$

$V_\beta = (X'X)^{-1}$

On prend le cas non informative sachant que :

$\sigma^2/Y \sim IN(\frac{n-p}{2}\widehat{\sigma}^2, \frac{n-p}{2}), où : n = 22, p = 4$

Pour les calcules voir le programme 2.

La distribution est représenté par un histogramme pour chaque coéfficient avec l'erreur sd à partir N simulation (N=22, N=500, N=10000)

Notre but est de préciser l'influence de chaque variable pour déterminer le meilleur modèle, celà on a besoin d'un critère nommé BIC.

Les graphes sont représentées comme suits :

Figure 3.1 : La distribution a posteriori jointe de β et $\sigma^2 (N = 22)$

Figure 3.2 : La distribution a posteriori jointe de β et $\sigma^2 (N = 500)$

Figure 3.3 : La distribution a posteriori jointe de β et $\sigma^2 (N = 10000)$

Interprétation des résultats :

On remarque que :

Quand on augmente le nombre de simulation la distribution des coéfficients suit une loi normale.

Figure 3.4 : Représentation graphique de BIC

Interprétation des résultats :

D'apré la figure 3.4 :

Le meilleur modèle est celui qui contient l'intercepte avec le variable X_3 $(BIC = -4.5)$, ensuite le modèle avec l'intercept et les variables X_3 et $X_2 (BIC = -2)$.

C-à-d :

L'effet de variable X_3 est statistiquement significatif ie : l'alimentation qui est trés riche au calories de protéines animales en pourcentage du total des calories provoque l'incidence d'attaque cardiaque.

Conclusion générale

Dans ce travail, nous avons étudié la théorie mathématique de la régression lineáre mais au sens d'une nouvelle approche ; c'est l'approche bayésienne qui possède plusieurs avantages par rapport à une analyse classique. Nous avons montré l'intérêt de la régression bayésienne et les conclusions significatives, elle permet aussi de travailer à des grande échantillon comme on a déja fait sur le critère BIC à l'aide de logiciel R et les résultats apparaissent sous la forme des graphes. De plus, nous avons distingué deux régression ; la régression bayésienne simple et la régression bayésienne multiple.

Dans la régression bayésienne simple, nous avons exposé la méthode des moindres carrées et essayé de donner sa forme alternative avec la représentation graphique de la distribution a priori et a posteriori des données.

A travers de ce travail, on a fait la régression bayésienne multiple avec un facteur de variance connu, dans ce cas le vecteur β est uniquement inconnu et on a ici deux cas :

A priori informative et a priori non informative, dans les deux a priori on a utilisé la méthode des moindres carrées pour calculer l'estimateur de Bayes et l'estimateur de MAP. On a traité aussi le cas d'un facteur de variance inconnu avec les mêmes étapes.

Résumé

L'objectif de ce mémoire est d'étudier la régression linéaire de l'inférence bayésienne parcequ'elle offre plus de souplesse et d'objectivité dans les données statistiques.

Nous avons essayé d'exposer des notions qui sont utilisées pour étudier la régression bayésienne (l'a priori, l'a posteriori, la vraisemblance...etc).

Des applications numériques et les simulations ont illustrées la méthodologie dans différents modèles pour calculer l'a posteriori dans la régression bayésienne simple et la régression bayésienne multiple.

Mots clés : inférence bayésienne, régression bayésienne simple, régression bayésienne multiple.

Abstract

The objective of this thesis is to study the linear regression of bayesian inference because it offers more flexibility and objectivity in the statistical data.

We have tried to expose concepts that are used to study bayesian regression (a priori, a posteriori, likelihood...etc).

Numerical applications and simulations illustrated the methodology in different models to compute a posteriori in simple bayesian regression and multiple bayesian regression.

Key words : bayesian inference, simple bayesian regression, multiple bayesian regression.

[redacted]

[redacted].

Annexe

$SS_x = \sum_{i=1}^{n}(x_i - \overline{x})^2$
$SS_y = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$
$SS_{xy} = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right)$

Les distributions :

1. **Distribution normale multivarié :**

   Soit x un vecteur , si $x \sim N(\mu, \Sigma)$

   La densité associé est :

   $f\left(x/\mu, \Sigma\right) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\left(x-\mu\right)'\Sigma^{-1}\left(x-\mu\right)\right\}$
   $E\left(x\right) = \mu \quad , \quad V\left(x\right) = \Sigma$

2. **Distribution t-multivarié :**

   Si $Z \sim N(0, W^{-1})$

   H est un variable de valeur h et $H \sim \chi_{(v)}$, Z et H sont iid .

   $x = z_i\left(\frac{h}{v}\right)^{-\frac{1}{2}} + \mu_i \qquad i = 1, ..., k$, $x \sim t\left(\mu, W^{-1}, v\right)$

   La densité associé est :

   $f\left(x/\mu, W^{-1}, v\right) = \frac{v^{\frac{v}{2}}\Gamma\left(\frac{k+v}{2}\right)(\det W)^{\frac{1}{2}}}{\pi^{\frac{k}{2}}\Gamma\left(\frac{v}{2}\right)}\left(v + \left(x-\mu\right)'W\left(x-\mu\right)\right)^{-\frac{k+v}{2}}$

   $E\left(x\right) = \mu, v \succ 1$

   $V\left(x\right) = v\left(v-2\right)^{-1}W^{-1}, v \succ 2$

3. **Distribution gamma normale :**

   Soit x un vecteur : $x \sim N(\mu, \tau^{-1}V)$

   τ un variable : $\tau \sim G(b, p)$

   $x, \tau \sim NG(\mu, V, b,\ p)$

   La densité associé de x et τ est :

   $f\left(x, \tau/\mu, V, b, p\right) = f\left(x/\mu, \tau^{-1}V\right)f\left(\tau/b, p\right)$

$$= (2\pi)^{-\frac{n}{2}} (\det V)^{-\frac{1}{2}} b^p (\Gamma(p))^{-1} \tau^{\frac{n}{2}+p-1} \exp\left\{-\frac{\tau}{2}\left[(x-\mu)' V^{-1}(x-\mu)\right]\right\}$$

$b \succ 0,\ p \succ 0,\ 0 \prec \tau \prec +\infty, -\infty \prec x_i \prec +\infty$

La distribution marginale de x est : $x \sim t(\mu, \frac{bV}{p}, 2p)$

La distribution marginale de x est : $\tau \sim G(b, p)$

4. **Distribution inverse-gamma :**

   Si $x \sim G(b,p)\, z = \frac{1}{x},\ z \sim IG(b,p)$

   La densité associé de x est :

   $f(z/b,p) = \frac{b^p}{\Gamma(p)} \left(\frac{1}{z}\right)^{p+1} \exp\left(-\frac{b}{z}\right)$

   $b \succ 0, p \succ 0, 0 \prec z \prec +\infty$

   $E(z) = \frac{b}{p-1}$

   $V(z) = \frac{b^2}{(p-1)^2(p-2)} \qquad p \succ 2$

•Le critère d'information bayésien (BIC) :

BIC est un critère d'information dérivé du critère d'information d'**Akaike** proposé par **Gideon Schwarz** en **1978,** à la différence, la pénalité dépend de la taille de l'échantillon et pas seulement du nombre du paramètre. Il s'écrit :

$BIC = -2ln(L) + k * ln(N)$

L : la vraisemblance du modèle éstimée.

N : le nombre d'observation dans l'échantillon.

k : le nombre de paramètre.

Code sous R

Programme 1

```
X=c(8,6,8,2,4,6,7,7,6,5,4,3,3,3,6,8,6,4,7,7,6,8)
Y=c(81,55,80,24,78,52,88,45,50,69,66,45,25,43,38,72,41,38,52,52,66,89)
reg=lm(Y~X)
#nuage de point :
plot(X,Y,xlab="calories de protéines animales", ylab="Nbr de décès par crise cardiaque",main="Fi
2.1 : nuage de point", lty=3)
#droit de regression :
abline(reg)
#moyenne de X :
mean(X)
#moyenne de Y :
mean(Y)
#moyenne de (X*Y) :
mean(X*Y)
#moyenne de (X^2) :
 Z=X^2
 Z
 mean(Z)
#(moyenne de (X))^2 :
mean(X)^2
#(moyenne de (Y))^2 :
 mean(Y)^2
#Les valeurs Ax, B et A0 :
Ax=mean(Y)
```

```
Ssx=(mean(Z)-mean(X)^2)
Ssx
SSxy=mean(X*Y)-(mean(X)*mean(Y))
SSxy
B=SSxy/Ssx
B
A0=mean(Y)-B*mean(X)
A0
#Ychapeau :
Ychapeau=A0+B*X
Ychapeau
#Résiduel :
residuel=Y-Ychapeau
residuel
residuel^2
#le tableau des données :
tableau=data.frame(X,Y,Ychapeau,residuel,residuel^2)
tableau
n=22
#la somme des résidus :
s=sum(residuel^2)/(n-2)
s
#la valeur de sd :
S=sqrt(sum(residuel^2)/(n-2))
S
```

#la lois a priori pour β et $\alpha_{\overline{x}}$:

```
sbeta=0.25
mbeta=1.02
salpha=1
malpha=15
```

#la lois a posteriori pour β et $\alpha_{\overline{x}}$:

```
SSx=n*(mean(Z)-mean(X)^2)
```

```
SSx
F=(1/sbeta^2)+(SSx/s)
F
sprimbeta=(F)^(-1/2)
sprimbeta
c=((1/sbeta^2))/F
c
d=((SSx/s)/F)
d=((SSx/s)/F)*B
d
mprimbeta=c*1+d
mprimbeta
k=(1/salpha^2)+(n/s)
k
sprimalpha=(k)^(-1/2)
sprimalpha
L=((n/s)/k)
L
w=((1/salpha^2))/k
w
mprimalpha=w*15+L*(mean(Y))
mprimalpha
#la courbe de densité a priori et a posteriori de la pente β :
curve(dnorm(x,mbetaprim,sprimbeta),xlim=c(0,2.5),
+ylim=c(0,2), xlab=expression(beta),ylab="Density",
+main="Figure 1.2 : La distribution a priori et a posteriori de la pente")
curve(dnorm(x,mbeta,sbeta),add=TRUE,lty=3,lwd=4)
legend(1.75,2,c("Prior","Posterior"), lty=c(3,2,1),lwd=c(3,3,3))
```

Programme 2

```
X1=c(124,49,181,4,22,152,75,54,43,41,17,22,16,10,63,170,125,12,221,171,97,254)
X2=c(33,31,38,17,20,39,30,29,35,31,23,21,8,23,37,40,38,25,39,33,38,39)
X3=c(8,6,8,2,4,6,7,7,6,5,4,3,3,3,6,8,6,4,7,7,6,8)
```

```
Y=c(81,55,80,24,78,52,88,45,50,69,66,45,25,43,38,72,41,38,52,52,66,89)
 #tableau des données :
tab=data.frame(Y,X1,X2,X3)
tab
reg=lm(Y~X1+X2+X3,data=tab,x=TRUE,y=TRUE)
reg
#simulation N fois :
 theta.sample=blinreg(reg$y,reg$x,500)
theta.sample
#somme des résidus :
S=sum(reg$residual^2)
S
shape=reg$df.residual/2
 shape
rate=S/2
rate
#simulation à partir d'une distribution inverse gamma :
sigma2=rigamma(1,shape,rate)
sigma2
MSE = sum(reg$residuals^2)/reg$df.residual
MSE
vbeta=vcov(reg)/MSE
vbeta
#simulation à partir d'une distribution normale multivarié :
beta=rmnorm(,mean=fit$coef,varcov=vbeta*sigma2)
beta
#histogramme de chaque coéfficient plus ereur sd :
par(mfrow=c(2,2))
 hist(theta.sample$beta[,2],main="TMH",xlab=expression(beta[1]))
 hist(theta.sample$beta[,3],main="CGP",xlab=expression(beta[2]))
 hist(theta.sample$beta[,4],main="CPA",xlab=expression(beta[3]))
hist(theta.sample$sigma,main="ERREUR SD",xlab=expression(sigma))
```

```
#le critère BIC :
tableau=data.frame(X1,X2,X3)
tableau
subsets=regsubsets(Y~X1+X2+X3,tableau)
plot(subsets)
```

Bibliographie

[1] **Ali Mohammad-Djafari** (2009 and 2010) **:** Information a priori et lois a priori (Francais-English).

[2] **Andrew D. Taylor**, Department of Biology, University of Hawai at Manoa.

[3] **Boreux, J, Parent. E, Bernier. J.** (2010). Pratique du calcul Bayesien. Springer-Verlag. France, Paris.

[4] **Christian P.Robert** (2006). Le choix bayésien : principe et pratique. Springer-Verlag. France, Paris .

[5] **Djeridi, F.** (2011). Estimation bayésienne. Memoire de fin d'étude pour l'obtention de licence en mathématiques. university de Jijel.

[6] **Djeridi, F.**(2012). Choix de la loi a priori dans la statistique bayésienne. Memoire de Master. university de Jijel.

[7] **Djeridi, F.**(2014)**.** Les méthodes adaptatives bayésiennes dans les éssais cliniques. Memoire de Master. university de Jijel.

[8] **Dominique Soudant** (1997). Application des modèles dynamiques bayésien au séries tomporèlles de dinophysis à antiffer, thèse de doctorat (Normandie, France).

[9] **Frédéric. Bertrand** (2009/2010). Master 1 MCB.TD n°2. Régression Linéaire Multiple.

[10] **Jean-Michel Marin, Christian P. Robert.** Les bases de la statistique bayésienne.

[11] **Jim, A**. (2009). Bayesian Computation with R. Springer+Busness Media, LLC.

[12] **Judith. Rousseau** (2010). Statistique bayésienne : note de cours.

[13] **Karl. Rudolf.Koch** (2007). Introduction to bayesian statitistics(2nd ed). Springer-Verlag. Berlin Heidelberg.

[14] **Michael S. Hamada, Alyson G. Wilson,C. Shane Reese et Harry F. Martz** (2008). Bayesian Reliability. Springer Science+Busness Media, LLC.

[15] **Parent, E. et Bernier, J.** (2007). Le raisonnement Bayésien : Modélisation Inférence. Springer-Verlag France, Paris.

[16] **Pierre-Charles Dangauthier** (2007). Fondations, méthode et applications de l'apprentissage bayésien, thèse de doctorat.

[17] **Qun Ying** (1991). Modèle bayésiens et application à l'éstimation des caractéristiques de produits finis et au controle de la qualité, thèse de doctorat.

[18] **William M.Bolstad** (2007). Intoduction to bayesian statistics(2nd ed).

[19] **Site : https ://fr.wikipedia.org/wiki/Critère_d% 27 information.bayésien.**

[20] **Site : web.univ_ubs.fr/Imam/gouno/BAYES/COURS/cours 1.pdf.**

[21] **Site : web.univ_ubs.fr/Imam/gouno/BAYES/COURS/cours 3.pdf.**