

Université de Montréal

APPROCHE BAYÉSIENNE DE LA  
RÉGRESSION POISSONNIENNE

par

Marylène Vachon

Département de mathématiques et de statistique  
Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences  
en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en statistique

avril 1998

© MARYLÈNE VACHON, MCMXCVIII



QA  
3  
U54  
1998  
v.019

Université de Montréal

APPROCHE BAYÉSIENNE DE LA  
RÉGRESSION POISSONNIENNE



Université de Montréal

Bibliothèque



Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

APPROCHE BAYÉSIENNE DE LA  
RÉGRESSION POISSONNIENNE

présenté par

Marylène Vachon

a été évalué par un jury composé des personnes suivantes :

*Yves Lepage*

---

(président-rapporteur)

*Jean-François Angers*

---

(directeur de recherche)

*Narayan Giri*

---

(membre du jury)

Mémoire accepté le :

98.08.19

## SOMMAIRE

---

Dans le cadre de ce mémoire, nous étudions la régression poissonnienne à l'aide de deux approches, soit l'approche classique et l'approche bayésienne. Concernant l'approche classique, nous nous concentrons sur deux modèles, l'un avec paramètre de dispersion constant et l'autre avec paramètre de dispersion non constant. Nous utilisons la méthode de Newton-Raphson pour calculer les estimateurs du maximum de vraisemblance.

Concernant l'approche bayésienne, celle-ci est séparée en deux parties. Dans la première, nous considérons une approximation de la distribution *a posteriori* des paramètres tandis que dans la deuxième, nous considérons la distribution exacte. Pour chacune de ces deux approches, nous étudions l'effet d'une valeur aberrante sur les estimateurs des paramètres, ainsi que sur les tests effectués. Nous développons aussi une loi *a priori* conjuguée pour la régression poissonnienne. Or, puisque certains calculs d'intégrales ne peuvent être faits de façon explicite, nous présentons deux méthodes d'approximation.

## REMERCIEMENTS

---

À mes parents, pour le soutien financier et surtout moral qu'ils m'ont apporté. Ils ont toujours été là pour m'encourager et me donner tout le support dont j'avais besoin. Je ne sais vraiment pas ce que je deviendrais sans eux. À mes deux frères, André et Stéphane qui ont toujours été fiers de ce que j'accomplis. Je dis un gros merci à toute ma famille.

À ma meilleure amie Caro, pour son écoute attentive lorsque j'avais des hauts et des bas. Elle a toujours su trouver les mots apaisants pour me donner le goût de continuer. À tous mes amis et compagnons de travail de l'Université qui se reconnaissent, pour l'entraide tout au long de ces années. J'ai été en mesure de passer au travers grâce à notre travail d'équipe et aux nombreuses sorties que nous avons faites tous ensemble.

Finalement, je tiens à remercier sincèrement mon directeur, Jean-François Angers. Sans lui, cet ouvrage n'aurait jamais vu le jour. Je me suis toujours sentie la bienvenue dans son bureau lorsque j'avais besoin d'explications. Je le remercie pour sa grande disponibilité, son intérêt pour mon travail et pour toute l'aide qu'il m'a apportée. Sa patience et la générosité de sa personne font de lui un directeur en or. Un gros merci Jean-François.

# Table des matières

---

Sommaire .....	ii
Remerciements .....	iii
Liste des figures .....	vii
Liste des tableaux .....	ix
Introduction .....	1
<b>Chapitre 1. Approche classique de la régression poissonnienne ...</b>	<b>4</b>
1.1. Modèle avec paramètre de dispersion constant.....	4
1.1.1. Forme de l'estimateur classique du maximum de vraisemblance.	5
1.1.2. Méthode de Newton-Raphson .....	7
1.1.3. Test sur $\beta$ .....	10
1.2. Valeur aberrante pour le modèle ordinaire.....	11
1.3. Modèle avec paramètre de dispersion variable .....	14
1.3.1. Modèle généralisé de régression de Poisson .....	14
1.3.2. Forme de l'estimateur classique du maximum de vraisemblance.	15
1.3.3. Test du modèle généralisé.....	18
1.3.4. Test sur $\beta$ .....	19
1.4. Exemple numérique .....	20

<b>Chapitre 2. Approche bayésienne de la régression poissonienne avec l’approximation normale .....</b>	<b>27</b>
2.1. Introduction .....	28
2.2. Approximation normale .....	30
2.2.1. Estimateurs de Bayes .....	33
2.2.2. Valeur aberrante .....	35
2.2.3. Test sur $\beta$ .....	38
2.3. Exemple numérique .....	40
<b>Chapitre 3. Approche bayésienne de la régression poissonienne avec la distribution exacte .....</b>	<b>44</b>
3.1. Forme des estimateurs de Bayes .....	44
3.1.1. Méthode de Naylor-Smith .....	47
3.1.2. Approximation analytique de Laplace .....	50
3.1.3. Exemple numérique .....	54
3.2. Lois conjuguées .....	55
3.2.1. Détermination de la loi <i>a priori</i> conjuguée .....	56
3.2.2. Forme des estimateurs de Bayes par l’approximation analytique de Laplace .....	59
3.2.3. Test sur $\beta$ .....	61
3.2.4. Valeur aberrante .....	64
3.2.5. Exemple numérique .....	66
<b>Chapitre 4. Exemple numérique .....</b>	<b>68</b>
4.1. Analyse à l’aide des méthodes classiques .....	69
4.2. Analyse à l’aide des méthodes bayésiennes avec l’approximation normale .....	75

4.3. Analyse à l'aide des méthodes bayésiennes avec la distribution exacte	77
4.4. Analyse des résultats .....	78
<b>Conclusion</b> .....	<b>84</b>
<b>Annexe A. Programmes</b> .....	<b>86</b>
<b>Bibliographie</b> .....	<b>111</b>



## Liste des figures

---

1.4.1	Graphique de $\hat{\alpha}$ et de $\hat{\alpha}_{(K)}$ en fonction de $K$ pour le modèle ordinaire des ensembles de triplets .....	23
1.4.2	Graphique de $\hat{\beta}$ et de $\hat{\beta}_{(K)}$ en fonction de $K$ pour le modèle ordinaire des ensembles de triplets.....	23
1.4.3	Graphique de $\hat{\alpha}$ en fonction de $K$ pour le modèle généralisé des ensembles de triplets.....	24
1.4.4	Graphique de $\hat{\beta}$ en fonction de $K$ pour le modèle généralisé des ensembles de triplets.....	24
1.4.5	Graphique de $\hat{\phi}$ en fonction de $K$ pour le modèle généralisé des ensembles de triplets.....	25
2.3.1	Graphique des erreurs absolues des estimateurs de $\alpha$ et $\beta$ en fonction de $K$ pour le modèle ordinaire des ensembles de triplets.....	42
4.0.1	Graphique des observations par rapport à la covariable.....	69
4.1.1	Graphique de $\hat{\alpha}$ et de $\hat{\alpha}_{(K)}$ en fonction de $K$ pour le modèle ordinaire des espèces de poissons .....	71
4.1.2	Graphique de $\hat{\beta}$ et de $\hat{\beta}_{(K)}$ en fonction de $K$ pour le modèle ordinaire des espèces de poissons .....	72
4.1.3	Graphique de $\hat{\alpha}$ en fonction de $K$ pour le modèle généralisé des espèces de poissons .....	73

4.1.4	Graphique de $\hat{\beta}$ en fonction de $K$ pour le modèle généralisé des espèces de poissons .....	73
4.1.5	Graphique de $\hat{\phi}$ en fonction de $K$ pour le modèle généralisé des espèces de poissons .....	74
4.2.1	Graphique des erreurs absolues des estimateurs de $\alpha$ et $\beta$ en fonction de $K$ pour le modèle ordinaire des espèces de poissons .....	76
4.4.1	Graphique des $\hat{\mu}_{st}$ et des $\hat{\mu}_B$ pour les espèces de poissons.....	82
4.4.2	Graphique des $\hat{\mu}_{gen}$ et des $\hat{\mu}_B$ pour les espèces de poissons .....	82
4.4.3	Graphique des $\hat{\mu}_{norm}$ et des $\hat{\mu}_B$ pour les espèces de poissons .....	83
4.4.4	Graphique des $\hat{\mu}_{ps.B}$ et des $\hat{\mu}_B$ pour les espèces de poissons.....	83

## Liste des tableaux

---

1.4.1	<i>Valeurs des estimateurs de <math>\alpha</math> et <math>\beta</math> pour le modèle standard et le modèle généralisé des ensembles de triplets .....</i>	20
1.4.2	<i>Ensembles de triplets, nés en Norvège, ainsi que les estimations des <math>\mu_i</math></i>	22
1.4.3	<i>Valeurs des estimateurs du modèle ordinaire pour différents <math>K</math> des ensembles de triplets .....</i>	25
1.4.4	<i>Valeurs des estimateurs du modèle général pour différents <math>K</math> des ensembles de triplets .....</i>	26
2.3.1	<i>Ensembles de triplets, nés en Norvège, ainsi que les estimations des <math>\mu_i</math></i>	42
2.3.2	<i>Valeurs des estimateurs de Bayes avec l'approximation normale et des estimateurs de Bayes hiérarchiques pour différents <math>K</math> .....</i>	43
3.1.1	<i>Estimateurs pseudo-bayésiens de <math>\alpha</math> et <math>\beta</math> ainsi que des pertes a posteriori correspondantes .....</i>	54
3.1.2	<i>Ensembles de triplets, nés en Norvège, ainsi que les estimations des <math>\mu_i</math></i>	55
3.2.1	<i>Ensembles de triplets, nés en Norvège, ainsi que les estimations des <math>\mu_i</math></i>	67
4.1.1	<i>Valeurs des estimateurs de <math>\alpha</math> et <math>\beta</math> pour le modèle standard et le modèle généralisé des espèces de poissons .....</i>	70
4.1.2	<i>Valeurs des estimateurs du modèle ordinaire pour différents <math>K</math> des espèces de poissons .....</i>	71

4.1.3	<i>Valeurs des estimateurs du modèle général pour différents <math>K</math> des espèces de poissons</i> .....	72
4.2.1	<i>Valeurs des estimateurs de Bayes avec l'approximation normale et des estimateurs de Bayes hiérarchiques pour différents <math>K</math> des espèces de poissons</i> .....	76
4.3.1	<i>Estimateurs pseudo-bayésiens de <math>\alpha</math> et <math>\beta</math> ainsi que des pertes a posteriori correspondantes des espèces de poissons</i> .....	77
4.4.1	<i>Estimateurs des <math>\mu_i</math> pour toutes les méthodes avec les observations correspondantes ainsi que les valeurs de la covariable pour les espèces de poissons</i> .....	79
4.4.2	<i>Estimateurs des <math>\mu_i</math> pour toutes les méthodes avec les observations correspondantes ainsi que les valeurs de la covariable pour les espèces de poissons (suite)</i> .....	80
4.4.3	<i>Estimateurs des <math>\mu_i</math> pour toutes les méthodes avec les observations correspondantes ainsi que les valeurs de la covariable pour les espèces de poissons (suite et fin)</i> .....	81

## INTRODUCTION

---

La distribution de Poisson est souvent utile pour décrire certains phénomènes de la vie courante. Pour modéliser des événements successifs indépendants, il est assez fréquent de supposer initialement que les données proviennent d'une population ayant une distribution de Poisson avec moyennes apparentées, via une fonction auxiliaire, à une fonction linéaire d'un ensemble de covariables. Une fois la fonction auxiliaire spécifiée, le modèle peut être ajusté en utilisant les techniques de modèles linéaires généralisés (voir McCullagh et Nelder, 1989). Plusieurs applications médicales impliquent des données de Poisson, incluant "the Rand Health Insurance Experiment" par Keeler et Rolph (1988), qui ont estimé les effets de la coassurance sur le coût par traitement et le nombre de traitements. Les taux de maladie et de mortalité ont été analysés par Bernardinelli et Montomoli (1992) en utilisant des méthodes de Gibbs et des méthodes bayésiennes empiriques de Poisson, développées par Clayton et Kaldor (1987). Pour la régression poissonnienne, nous utilisons le modèle log-linéaire comme fonction auxiliaire.

En pratique, les données ne peuvent pas toujours être modélisées par le modèle de base et montrent souvent des degrés de surdispersion ou de sous-dispersion relatifs à un modèle de Poisson. Une façon de prendre en considération ces différents degrés de dispersion est d'utiliser la distribution généralisée de Poisson. Lawless (1987) s'est servi du modèle linéaire généralisé de Poisson et du profil du logarithme de la fonction de vraisemblance, afin d'analyser les effets aléatoires

concernant des données sur les tumeurs. La présentation du modèle ordinaire et du modèle généralisé constitue en grande partie le premier chapitre de ce mémoire.

Afin d'estimer les paramètres du modèle de régression, nous employons des techniques classiques telles que le maximum de vraisemblance. Nous utilisons la méthode de Newton-Raphson pour calculer les estimés. Nous discutons aussi d'un test nous indiquant lequel des deux modèles est le plus approprié, ainsi que les tests déterminant s'il y a une tendance particulière pour chacun de ces modèles. Finalement, nous intégrons dans ce premier chapitre, une étude du comportement de l'estimateur du maximum de vraisemblance en présence d'une valeur aberrante.

Dans le deuxième chapitre, nous utilisons les éléments de la théorie bayésienne afin d'étudier le modèle de régression poissonienne standard. Nous introduisons, en premier lieu, les concepts bayésiens qui seront nécessaires à notre étude, tels que estimateur de Bayes, famille conjuguée et cote de Bayes. En se basant sur les résultats de Lindley (1965), nous approximons la distribution *a posteriori* des paramètres par la loi normale. Aussi, à partir de ce modèle, nous développons un estimateur hiérarchique qui est robuste par rapport à la présence de valeurs aberrantes. Récemment, Christiansen et Morris (1997) ont proposé une nouvelle approche avec le modèle de régression poissonienne interactive à plusieurs niveaux. Pour développer cette approche, ils ont utilisé l'analyse bayésienne hiérarchique. Enfin, nous utilisons la cote de Bayes pour vérifier s'il y a une tendance particulière au niveau des paramètres.

Le troisième chapitre reprend quelques éléments du second chapitre, mais en considérant la distribution *a posteriori* exacte des paramètres. Les estimateurs de Bayes et les cotes de Bayes nécessitent un calcul d'intégrales qui ne peut pas se faire de façon analytique. Ainsi, nous présentons deux méthodes qui nous permettront d'approximer ces intégrales, soient la méthode de Naylor-Smith et

l'approximation de Laplace. Pour exprimer la cote de Bayes, nous devons avoir une loi *a priori* qui soit propre et c'est pour cette raison que nous présentons une loi conjuguée pour un modèle de régression de Poisson. De plus, nous étudions le comportement de la cote de Bayes en présence d'une valeur aberrante.

Finalement, dans le dernier chapitre, nous appliquons toutes les techniques étudiées à un exemple numérique, afin de pouvoir comparer les résultats de chacune d'entre elles et de voir leurs comportements en présence d'une valeur aberrante.

# Chapitre 1

---

## APPROCHE CLASSIQUE DE LA RÉGRESSION POISSONNIENNE

Dans ce premier chapitre, nous nous intéressons à deux modèles de régression poissonnienne, soient le modèle standard et le modèle généralisé. Dans le premier modèle, le paramètre de dispersion est supposé constant tandis que dans le deuxième, il est supposé non constant. Pour chacun de ces modèles, nous présentons l'estimateur classique du maximum de vraisemblance. Nous discutons aussi de la méthode de Newton-Raphson qui nous permet d'évaluer ces derniers.

### 1.1. MODÈLE AVEC PARAMÈTRE DE DISPERSION CONSTANT

Dans cette première section, nous introduisons le modèle log-linéaire pour lequel nous développons l'estimateur du maximum de vraisemblance lorsque le paramètre de dispersion est supposé constant. Notons que sous des conditions expérimentales idéales, lorsque des événements successifs se produisent indépendamment les uns des autres et au même taux, le modèle de Poisson est approprié pour décrire le nombre d'événements observés. Soient  $Y_1, \dots, Y_n$  des observations provenant d'une loi de Poisson ayant comme moyennes respectives  $\mu_1, \dots, \mu_n$  et soient  $\mathbf{x}_1, \dots, \mathbf{x}_n$  les vecteurs des covariables correspondantes. Posons

$$\text{Var}(Y_i) = \sigma^2 \mathbb{E}[Y_i] = \sigma^2 \mu_i(\mathbf{x}_i),$$



où  $\sigma^2$ , le paramètre de dispersion, est supposé constant à travers les données. Dans les modèles log-linéaires, la dépendance de  $\mu_i(\mathbf{x}_i) = \mathbb{E}[Y_i]$  sur le vecteur de covariables  $\mathbf{x}_i$ , est supposée multiplicative et s'écrit habituellement sous la forme logarithmique, c'est-à-dire

$$\log \mu_i(\mathbf{x}_i) = \eta_i = \mathbf{x}_i' \boldsymbol{\beta},$$

pour  $i = 1, \dots, n$ , où  $\mathbf{x}_i' = (x_{i0}, x_{i1}, \dots, x_{ip})$  est le vecteur de  $(p + 1)$  variables explicatives linéairement indépendantes ( $p + 1 \leq n$ ) avec  $x_{i0} = 1$  pour  $i = 1, \dots, n$  et où  $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_p)$  est le vecteur de paramètres inconnus.

### 1.1.1. Forme de l'estimateur classique du maximum de vraisemblance

Si nous avons  $Y_1, \dots, Y_n$ ,  $n$  variables aléatoires de Poisson, ayant pour moyennes  $\mu_1, \dots, \mu_n$  respectivement, avec  $\mu_i = \mu_i(\mathbf{x}_i)$ , la fonction de vraisemblance est donnée par

$$L(\boldsymbol{\mu}, \mathbf{y}) = \prod_{i=1}^n e^{-\mu_i} \frac{\mu_i^{y_i}}{y_i!}.$$

Afin de simplifier les calculs, nous prenons le logarithme de la fonction de vraisemblance qui est donné par

$$\begin{aligned} l(\boldsymbol{\mu}) &= \log L(\boldsymbol{\mu}, \mathbf{y}) \\ &= \sum_{i=1}^n (y_i \log \mu_i - \mu_i - \log(y_i!)). \end{aligned} \quad (1.1.1)$$

Puisque nous travaillons avec le modèle log-linéaire, nous substituons  $\mu_i$  par  $e^{\mathbf{x}_i' \boldsymbol{\beta}}$  dans l'équation (1.1.1), afin d'obtenir

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \left( y_i \mathbf{x}_i' \boldsymbol{\beta} - e^{\mathbf{x}_i' \boldsymbol{\beta}} - \log(y_i!) \right).$$

Comme nous sommes intéressés à trouver l'estimateur du maximum de vraisemblance de  $\beta$ , nous dérivons  $l(\beta)$  par rapport à  $\beta$  et égalisons cette dérivée à zéro. Nous obtenons donc

$$\frac{\partial}{\partial \beta} l(\beta) = \sum_{i=1}^n (y_i - e^{\mathbf{x}'_i \beta}) \mathbf{x}'_i.$$

Donc

$$\begin{aligned} \frac{\partial}{\partial \beta} l(\beta) \Big|_{\beta=\hat{\beta}} &= 0 \\ \iff \sum_{i=1}^n y_i \mathbf{x}'_i &= \sum_{i=1}^n e^{\mathbf{x}'_i \hat{\beta}} \mathbf{x}'_i \end{aligned} \tag{1.1.2}$$

$$\iff X' \mathbf{y} = X'(\exp(X \hat{\beta})) = X' \hat{\boldsymbol{\mu}}, \tag{1.1.3}$$

où

$$\begin{aligned} X &= (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)' \\ &= \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \end{aligned}$$

qui est une matrice  $(n \times (p+1))$  et où nous définissons

$$\hat{\boldsymbol{\mu}} = (\exp(X \hat{\beta})) = \begin{pmatrix} e^{\mathbf{x}'_1 \hat{\beta}} \\ e^{\mathbf{x}'_2 \hat{\beta}} \\ \vdots \\ e^{\mathbf{x}'_n \hat{\beta}} \end{pmatrix}.$$

La matrice de covariance asymptotique de l'estimateur du maximum de vraisemblance de  $\hat{\beta}$  est la matrice pour laquelle l'élément sur la ligne  $j$  et la colonne  $m$  de son inverse est  $-\mathbb{E} \left[ \frac{\partial^2 l(\beta)}{\partial \beta_m \partial \beta_j} \right]$ . Pour le modèle log-linéaire, nous avons donc

$$\begin{aligned} \frac{\partial^2 l(\beta)}{\partial \beta_m \partial \beta_j} &= - \sum_{i=1}^n x_{ij} \left( \frac{\partial \mu_i}{\partial \beta_m} \right) \\ &= - \sum_{i=1}^n x_{ij} x_{im} e^{x_i' \beta}. \end{aligned}$$

Une fois  $\hat{\beta}$  calculé, nous pouvons obtenir  $\hat{\mu}$  par  $\hat{\mu} = \exp(X\hat{\beta})$  et estimer la matrice de covariance de  $\hat{\beta}$  par  $(X'\hat{U}X)^{-1}$ , où  $\hat{U}$  est la matrice diagonale avec les éléments de  $\hat{\mu}$  sur la diagonale.

L'équation (1.1.3) n'a pas toujours de solution explicite et c'est pourquoi, il est souvent nécessaire d'utiliser des méthodes numériques pour calculer  $\hat{\beta}$ .

### 1.1.2. Méthode de Newton-Raphson

La méthode de Newton-Raphson (voir Stoer et Burlirsch, 1980, section 5.1) est une procédure pour trouver le maximum d'une fonction. Pour l'utiliser, nous devons faire une première estimation de l'emplacement du maximum. Ensuite, la fonction est approximée, dans un voisinage de l'estimé initial, par un polynôme du second degré. Nous pouvons donc calculer la valeur maximale de ce polynôme et celle-ci devient la deuxième approximation. En répétant ce processus, nous convergions vers le maximum de la fonction si le point de départ est choisi suffisamment près de l'emplacement du maximum. Dans Stoer et Burlirsch (1980), section 5.2, nous retrouvons les conditions de convergence pour les procédures itératives, ainsi qu'un théorème très utile.

Soient  $\phi$  une fonction itérative sur  $\mathbb{R}^n$  et  $\xi$ , un point fixe de la fonction  $\phi$ , c'est-à-dire que  $\xi = \phi(\xi)$ . Si, pour tout vecteur de départ  $w_0$ , pris dans un voisinage de  $\xi$  et pour la suite  $w_{i+1} = \phi(w_i)$ ,  $i = 0, 1, \dots$ , l'inégalité suivante

$$\|w_{i+1} - \xi\| \leq C \|w_i - \xi\|^q,$$

est vérifiée pour tout  $i \geq 0$  et où  $C < 1$  si  $q = 1$ , alors la méthode itérative définie par  $\phi$  sera au moins d'ordre  $q$  pour déterminer  $\xi$ . Pour  $q = 2, 3, \dots$ , la méthode est précisément d'ordre  $q$ . Une méthode est de premier ordre si, en plus que  $q = 1$ , il est vrai que  $|\phi'(\xi)| < 1$ . Nous avons le théorème suivant.

**Théorème 1.1.** *Toute méthode d'au moins d'ordre  $q$  pour déterminer un point fixe  $\xi$  est localement convergente dans le sens qu'il existe un voisinage  $N(\xi)$  de  $\xi$  avec la propriété que  $\forall w_0 \in N(\xi)$ , la suite  $w_i$  générée par  $\phi$ , converge vers  $\xi$ .*

Pus précisément, si nous voulons trouver la valeur  $\hat{\beta}$  de  $\beta$  qui maximise une fonction  $l$ , nous procédons de la façon suivante: soient  $\mathbf{q}$  et  $H$ , le gradient et la matrice hessienne de la fonction  $l(\beta)$ , donnés respectivement par

$$\mathbf{q}' = \left( \frac{\partial l}{\partial \beta_0}, \frac{\partial l}{\partial \beta_1}, \dots, \frac{\partial l}{\partial \beta_p} \right),$$

$$H = \left( h_{jm} \right)_{1 \leq j, m \leq p},$$

où

$$h_{jm} = \frac{\partial^2 l}{\partial \beta_m \partial \beta_j}.$$

Notons que si les dérivées premières  $l_j = \frac{\partial l}{\partial \beta_j}$  et  $l_m = \frac{\partial l}{\partial \beta_m}$  existent et si  $l_j$  et  $l_m$  sont différentiables, alors  $l_{jm} = \frac{\partial^2 l}{\partial \beta_m \partial \beta_j} = \frac{\partial^2 l}{\partial \beta_j \partial \beta_m} = l_{mj}$  et nous pouvons écrire  $H$

sous la forme suivante.

$$H = \begin{pmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_0} & \cdots & \frac{\partial^2 l}{\partial \beta_p \partial \beta_0} \\ \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 l}{\partial \beta_1^2} & \cdots & \frac{\partial^2 l}{\partial \beta_p \partial \beta_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 l}{\partial \beta_0 \partial \beta_p} & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_p} & \cdots & \frac{\partial^2 l}{\partial \beta_p^2} \end{pmatrix}.$$

Soient  $\mathbf{q}^{(s)}$  et  $\mathbf{H}^{(s)}$ , les termes définis ci-haut évalués en  $\boldsymbol{\beta}^{(s)}$ , la  $s$ -ième itération dans la série d'approximations pour  $\hat{\boldsymbol{\beta}}$ . À la  $s$ -ième itération ( $s = 0, 1, 2, \dots$ ),  $l(\boldsymbol{\beta})$  est approximé, près de  $\boldsymbol{\beta}^{(s)}$ , par les termes de sa série de Taylor d'ordre 2, développée autour de  $\boldsymbol{\beta}^{(s)}$ , c'est-à-dire

$$Q^{(s)}(\boldsymbol{\beta}) = l(\boldsymbol{\beta}^{(s)}) + (\mathbf{q}^{(s)})'(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)})' \mathbf{H}^{(s)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)}).$$

Résoudre  $\frac{\partial}{\partial \boldsymbol{\beta}} Q^{(s)} = \mathbf{q}^{(s)} + \mathbf{H}^{(s)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)}) = 0$  pour  $\boldsymbol{\beta}$  mène à la prochaine approximation,

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - (\mathbf{H}^{(s)})^{-1} \mathbf{q}^{(s)},$$

en supposant que  $\mathbf{H}^{(s)}$  soit non singulière, ce qui devrait être le cas si  $\boldsymbol{\beta}^{(s)}$  est près de la solution (voir Stoer et Burlirsch, 1980, section 5.1). Pour le modèle log-linéaire, nous retrouvons les expressions suivantes.

$$\mathbf{q}^{(s)} = \left[ \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}) \right]_{\boldsymbol{\beta}^{(s)}} = \sum_{i=1}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}^{(s)}}) \mathbf{x}_i, \quad (1.1.4)$$

$$\mathbf{H}^{(s)} = \left[ \frac{\partial^2}{\partial \boldsymbol{\beta}^2} l(\boldsymbol{\beta}) \right]_{\boldsymbol{\beta}^{(s)}} = - \sum_{i=1}^n e^{\mathbf{x}'_i \boldsymbol{\beta}^{(s)}} \mathbf{x}_i \mathbf{x}'_i. \quad (1.1.5)$$

Une façon de commencer le processus itératif est de prendre l'estimateur classique des moindres carrés pour  $\boldsymbol{\beta}^{(0)}$ . Nous avons le modèle suivant.

$$y_i = e^{\mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i},$$

$$\log y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i.$$

En posant  $\mathbf{Z} = \log \mathbf{y}$ , nous retrouvons le modèle linéaire habituel, qui est donné par  $\mathbf{Z} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ . En supposant que la matrice  $X$  soit de plein rang, nous retrouvons l'estimateur des moindres carrés pour  $\boldsymbol{\beta}^{(0)}$ , c'est-à-dire

$$\boldsymbol{\beta}^{(0)} = (X'X)^{-1}X'\mathbf{Z} = (X'X)^{-1}X'(\log \mathbf{Y}).$$

À partir de cette première estimation, nous pouvons maintenant appliquer la méthode de Newton-Raphson.

### 1.1.3. Test sur $\boldsymbol{\beta}$

Nous voulons tester, de façon classique, s'il y a présence d'une tendance apparente dans le modèle. Nous utilisons le test du rapport de vraisemblance afin de confronter les hypothèses

$$H_0 : \boldsymbol{\beta} = 0 \text{ vs } H_1 : \boldsymbol{\beta} \neq 0.$$

Le test du rapport de vraisemblance est donné par (voir Consul et Famoye, 1992)

$$\begin{aligned} l_1 &= -2 \log \left( \frac{L(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\beta} = 0)}{L(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})} \right) \\ &= 2 \left( l(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) - l(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\beta} = 0) \right) \\ &= 2 \sum_{i=1}^n \left\{ y_i \log \left( \widehat{\mu}_i / \widehat{\mu}_i \right) + (\widehat{\mu}_i - \widehat{\mu}_i) \right\}, \end{aligned}$$

où  $\widehat{\mu}_i = \exp\{\widehat{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\beta}}x_i\}$  sont les estimateurs du maximum de vraisemblance des  $\mu_i$ , sans aucune spécification des paramètres et  $\widehat{\mu}_i = \exp\{\widehat{\boldsymbol{\alpha}}\}$  sont les estimateurs du maximum de vraisemblance lorsque l'hypothèse  $H_0$  est vraie.

Nous rejetons l'hypothèse  $H_0$  au niveau  $1 - \alpha$  lorsque  $n$  est grand si la valeur de  $l_1$  est plus petite que le percentile d'ordre  $(1 - \alpha) \times 100\%$  d'une statistique du khi-deux à un degré de liberté.

## 1.2. VALEUR ABERRANTE POUR LE MODÈLE ORDINAIRE

Nous voulons maintenant étudier le comportement de l'estimateur pour le modèle ordinaire, lorsqu'il y a présence d'une valeur aberrante dans le jeu de données. Une valeur aberrante est une observation qui se tient à l'écart du nuage de points de toutes les autres observations. Cette dernière peut avoir une grande influence sur l'ajustement du modèle et sur le calcul des estimés.

Pour ce modèle, nous avons vu que le logarithme de la fonction de vraisemblance est donné par l'équation (1.1.1). Afin de voir comment se comporte l'estimateur du maximum de vraisemblance de  $\beta$  en présence d'une valeur aberrante, nous introduisons la notation suivante pour le vecteur de données. Soit  $\mathbf{y}_{(K)} = (y_1, y_2, \dots, Ky_I, y_{I+1}, \dots, y_n)$  où  $K$  est une constante, soit  $\beta_{(K)}$  le nouveau  $\beta$  pour ce nouveau  $\mathbf{y}$  et  $I$ , la position de l'observation aberrante. Avec cette nouvelle notation, nous reprenons les calculs de la section 1.1. La fonction de vraisemblance devient alors,

$$\begin{aligned} l(\beta_{(K)}) &= Ky_I \mathbf{x}'_I \beta_{(K)} + \sum_{i \neq I}^n y_i \mathbf{x}'_i \beta_{(K)} - e^{\mathbf{x}'_I \beta_{(K)}} \\ &\quad - \sum_{i \neq I}^n e^{\mathbf{x}'_i \beta_{(K)}} - \log((Ky_I)!) - \sum_{i \neq I}^n \log(y_i!). \end{aligned}$$

Nous avons vu que l'équation de l'estimateur du maximum de vraisemblance n'avait pas de solution explicite et c'est pourquoi, nous avons utilisé la méthode de Newton-Raphson pour l'approximer. Nous substituons maintenant notre nouveau vecteur de données  $\mathbf{y}_{(K)}$  dans les formules (1.1.4) et (1.1.5) et nous obtenons,

$$\begin{aligned} \mathbf{q}^{(s)} &= Ky_I \mathbf{x}_I + \sum_{i \neq I}^n y_i \mathbf{x}_i - e^{\mathbf{x}'_I \beta_{(K)}^{(s)}} \mathbf{x}_I - \sum_{i \neq I}^n e^{\mathbf{x}'_i \beta_{(K)}^{(s)}} \mathbf{x}_i \\ &= (Ky_I - e^{\mathbf{x}'_I \beta_{(K)}^{(s)}}) \mathbf{x}_I + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \beta_{(K)}^{(s)}}) \mathbf{x}_i, \end{aligned}$$

$$H^{(s)} = - \sum_{i=1}^n e^{\mathbf{x}'_i \boldsymbol{\beta}_{(K)}^{(s)}} \mathbf{x}_i \mathbf{x}'_i.$$

La formule itérative devient

$$\boldsymbol{\beta}_{(K)}^{(s+1)} = \boldsymbol{\beta}_{(K)}^{(s)} - (H^{(s)})^{-1} \left[ (K y_I - e^{\mathbf{x}'_I \boldsymbol{\beta}_{(K)}^{(s)}}) \mathbf{x}_I + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}_{(K)}^{(s)}}) \mathbf{x}'_i \right].$$

Pour calculer l'estimateur des moindres carrés pour  $\boldsymbol{\beta}_{(K)}^{(0)}$ , qui sert de point de départ à la méthode de Newton-Raphson, nous posons

$$\begin{aligned} \mathbf{Z}_{(K)} &= \log \mathbf{Y}_{(K)} \\ &= \log \mathbf{Y} + (\log K) \mathbf{e}_I \\ &= \mathbf{Z} + (\log K) \mathbf{e}_I, \end{aligned}$$

où  $\mathbf{Y}$  est le vecteur du jeu de données sans valeur aberrante et  $\mathbf{e}_I$ , représente un vecteur avec des 0 partout, sauf à la  $I$ -ème position. Nous retrouvons le théorème suivant.

**Théorème 1.2.**

$$\boldsymbol{\beta}_{(K)}^{(0)} = \boldsymbol{\beta}_0 + (\log K) \boldsymbol{\beta}_* + \begin{cases} o(K) & \text{si } \mathbf{x}'_* \boldsymbol{\beta}_* < 1, \\ \mathcal{O}(K^{-(\mathbf{x}'_* \boldsymbol{\beta}_* - 1)}) & \text{sinon,} \end{cases}$$

où

$$\begin{aligned} \boldsymbol{\beta}_0 &= (X'X)^{-1} X'Z, \\ \boldsymbol{\beta}_* &= (X'X)^{-1} X' \mathbf{e}_I = (X'X)^{-1} \mathbf{x}_I, \\ \mathbf{x}_I &= (1, x_{I1}, \dots, x_{Ip})' \end{aligned}$$

et

$$\mathbf{x}'_* \boldsymbol{\beta}_* = \max_i \mathbf{x}'_i \boldsymbol{\beta}_*.$$



**Démonstration 1.1.** Nous avons

$$\begin{aligned} e^{\mathbf{x}'_I \boldsymbol{\beta}_{(K)}^{(0)}} &= e^{\mathbf{x}'_I \boldsymbol{\beta}_0} e^{\mathbf{x}'_I \boldsymbol{\beta}_*} \\ &= K^{\mathbf{x}'_I \boldsymbol{\beta}_*} e^{\mathbf{x}'_I \boldsymbol{\beta}_0}, \\ e^{\mathbf{x}'_i \boldsymbol{\beta}_{(K)}^{(0)}} &= K^{\mathbf{x}'_i \boldsymbol{\beta}_*} e^{\mathbf{x}'_i \boldsymbol{\beta}_0} \text{ pour } i \neq I. \end{aligned}$$

Pour la première itération, nous obtenons

$$\begin{aligned} \boldsymbol{\beta}_{(K)}^{(1)} &= \boldsymbol{\beta}_{(K)}^{(0)} - (H^{(0)})^{-1} \left[ (K y_I - e^{\mathbf{x}'_I \boldsymbol{\beta}_{(K)}^{(0)}}) \mathbf{x}_I + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}_{(K)}^{(0)}}) \mathbf{x}_i \right] \\ &= \boldsymbol{\beta}_0 + (\log K) \boldsymbol{\beta}_* - (H^{(0)})^{-1} \left[ (K y_I - e^{\mathbf{x}'_I \boldsymbol{\beta}_0} K^{\mathbf{x}'_I \boldsymbol{\beta}_*}) \mathbf{x}_I \right. \\ &\quad \left. + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}_0} K^{\mathbf{x}'_i \boldsymbol{\beta}_*}) \mathbf{x}_i \right] \\ &\simeq \boldsymbol{\beta}_0 + (\log K) \boldsymbol{\beta}_* - (H^{(0)})^{-1} \\ &\quad \times \begin{cases} K y_I \mathbf{x}_I & \text{si } \mathbf{x}'_* \boldsymbol{\beta}_* = \max_i \mathbf{x}'_i \boldsymbol{\beta}_* < 1, \\ -e^{\mathbf{x}'_I \boldsymbol{\beta}_0} K^{\mathbf{x}'_* \boldsymbol{\beta}_*} \mathbf{x}_* & \text{sinon.} \end{cases} \\ &\quad \left. + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}_0} K^{\mathbf{x}'_i \boldsymbol{\beta}_*}) \mathbf{x}_i \right\}. \end{aligned}$$

Puisque  $H^{(1)} = - \sum_{i=1}^n e^{\mathbf{x}'_i \boldsymbol{\beta}_{(K)}^{(1)}} \mathbf{x}_i \mathbf{x}'_i$ , nous avons

$$H^{(1)} \simeq - \sum_{i=1}^n e^{\mathbf{x}'_i [\boldsymbol{\beta}_0 + (\log K) \boldsymbol{\beta}_* + (H^{(0)})^{-1} (e^{\mathbf{x}'_I \boldsymbol{\beta}_0} K^{\mathbf{x}'_* \boldsymbol{\beta}_*}) \mathbf{x}_* + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}_0} K^{\mathbf{x}'_i \boldsymbol{\beta}_*}) \mathbf{x}_i]} \mathbf{x}_i \mathbf{x}'_i$$

si  $\mathbf{x}'_* \boldsymbol{\beta}_* < 1$  et

$$H^{(1)} \simeq - \sum_{i=1}^n e^{\mathbf{x}'_i [\boldsymbol{\beta}_0 + (\log K) \boldsymbol{\beta}_* - (H^{(0)})^{-1} K y_I \mathbf{x}_I + \sum_{i \neq I}^n (y_i - e^{\mathbf{x}'_i \boldsymbol{\beta}_0} K^{\mathbf{x}'_i \boldsymbol{\beta}_*}) \mathbf{x}_i]} \mathbf{x}_i \mathbf{x}'_i$$

si  $\mathbf{x}'_* \boldsymbol{\beta}_* \geq 1$ .

Nous voyons que  $H^{(1)}$  croît de façon exponentielle en  $K$  et nous pouvons donc négliger le deuxième terme de la formule itérative. Nous obtenons alors, lorsque  $K \rightarrow \infty$ ,

$$\beta_{(K)} \rightarrow \begin{cases} \beta_0 + (\log K) \beta_* + o(K) & \text{si } \mathbf{x}'_* \beta_* < 1, \\ \beta_0 + (\log K) \beta_* + \mathcal{O}(K^{-(\mathbf{x}'_* \beta_* - 1)}) & \text{sinon.} \end{cases}$$

### 1.3. MODÈLE AVEC PARAMÈTRE DE DISPERSION VARIABLE

Jusqu'à maintenant, nous avons supposé que le paramètre de dispersion était constant à travers les données. Par contre, il arrive souvent qu'il ne le soit pas. Nous disons qu'il y a surdispersion, lorsque la variance échantillonnale est supérieure à la moyenne échantillonnale et sous-dispersion, lorsqu'elle est inférieure. Ce dernier cas se produit souvent lorsque l'échantillonnage se fait en grappes (voir McCullagh et Nelder, 1989, section 6.1).

#### 1.3.1. Modèle généralisé de régression de Poisson

La distribution généralisée de Poisson est utile pour ajuster des données de dénombrement avec surdispersion ou sous-dispersion. Nous utilisons donc le modèle généralisé de régression de Poisson (GPR) qui peut s'appliquer à tous les cas, c'est-à-dire à la sur et la sous-dispersion ainsi qu'à la dispersion constante. Nous introduisons donc le modèle GPR (voir Consul et Famoye, 1992).

**Définition 1.1.** Soit  $\mathbf{x}'_i = (x_{i0}, x_{i1}, \dots, x_{ip})$ , un vecteur de  $(p + 1)$  variables explicatives, avec  $x_{i0} = 1, \forall i = 1, \dots, n$  et soit  $Y$ , la variable réponse qui est un dénombrement ayant comme valeurs  $y_1, \dots, y_n$ . Alors,  $Y_i | \mathbf{x}_i$  aura une distribution généralisée de Poisson si

$$P(Y_i|\mathbf{x}_i) = \begin{cases} \mu_i[\mu_i + (\phi - 1)y_i]^{y_i-1} \phi^{-y_i} \\ \times e^{-[(\mu_i+(\phi-1)y_i)/\phi]}/y_i!; & \text{pour } y_i = 0, 1, 2, \dots, m, \\ 0; & \text{pour } y_i > m, \text{ quand } \phi < 1, \end{cases}$$

où  $\mu_i \equiv \mu_i(\mathbf{x}_i) > 0$ ,  $\phi \geq \max\left(\frac{1}{2}, 1 - \frac{\mu_i}{m}\right)$  représente la racine carrée de l'index de dispersion et  $m(\geq 4)$  est le plus grand entier positif pour lequel  $\mu_i + m(\phi - 1) > 0$  quand  $\phi < 1$  et il est égale à  $\infty$  si  $\phi \geq 1$ .

Les deux premiers moments centrés de  $Y$ , pour ce modèle, sont donnés par

$$\begin{aligned} \mathbb{E}[Y_i|\mathbf{x}_i] &= \mu_i, \\ \text{Var}(Y_i|\mathbf{x}_i) &= \phi^2 \mu_i > 0. \end{aligned}$$

Notons que lorsque  $\phi = 1$ , le modèle GPR se réduit au modèle de régression de Poisson ordinaire, lorsque  $\phi > 1$ , il y a surdispersion et lorsque  $\frac{1}{2} \leq \phi < 1$ , il y a sous-dispersion pour  $\mu_i > 2$  (voir Consul et Famoye, 1992).

### 1.3.2. Forme de l'estimateur classique du maximum de vraisemblance

Pour ce modèle, nous utilisons encore l'estimateur du maximum de vraisemblance pour estimer  $\beta$ . Par contre, il faut estimer un paramètre de plus que dans le modèle de Poisson ordinaire qui est  $\phi$ . Le logarithme de la fonction de vraisemblance pour le modèle GPR est donné par

$$l(\beta, \phi) \equiv l = \sum_{i=1}^n \left\{ \log \mu_i + (y_i - 1) \log[\mu_i + (\phi - 1)y_i] - y_i \log \phi - \frac{\mu_i + (\phi - 1)y_i}{\phi} - \log(y_i!) \right\}.$$

Les équations du maximum de vraisemblance pour les estimations de  $\beta$  et  $\phi$  sont données par

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^n \left( \frac{y_i(y_i - 1)}{\mu_i + (\phi - 1)y_i} - \frac{y_i}{\phi} + \frac{\mu_i - y_i}{\phi^2} \right) = 0, \quad (1.3.1)$$

$$\frac{\partial l}{\partial \beta_r} = \sum_{i=1}^n \left( 1 - \frac{\mu_i}{\phi} + \frac{(y_i - 1)\mu_i}{\mu_i + (\phi - 1)y_i} \right) x_{ir} = 0, \quad (1.3.2)$$

pour  $r = 0, 1, \dots, p$ . En particulier, nous remarquons que si  $p = 0$ , l'équation (1.3.2) se réduit à

$$\frac{\partial l}{\partial \beta_0} = n - \frac{1}{\phi} \sum_{i=1}^n \mu_i + \sum_{i=1}^n \left( \frac{(y_i - 1)\mu_i}{\mu_i + (\phi - 1)y_i} \right) = 0. \quad (1.3.3)$$

Si  $\phi \neq 1$  et si nous multiplions l'équation (1.3.1) par  $(\phi - 1)$  et que nous additionnons cette dernière à l'équation (1.3.3), nous obtenons  $\sum_{i=1}^n (\mu_i - y_i) = 0$  ou encore,  $n\bar{y} = \sum_{i=1}^n \mu_i$ .

Nous pouvons voir que les équations du maximum de vraisemblance (1.3.1) et (1.3.2) sont non linéaires et nous aurons besoin de méthodes par approximations successives pour obtenir les estimateurs du maximum de vraisemblance des paramètres  $\beta$  et  $\phi$ . Comme pour le modèle avec paramètre de dispersion constant, l'estimation initiale pour  $\beta^{(0)}$  est obtenue par la méthode des moindres carrés. Dans le cas de l'estimation initiale du paramètre de dispersion  $\phi^{(0)}$ , Breslow (1984) propose d'utiliser la méthode des moments. Cela nous mène à égaliser la statistique du khi-deux à son degré de liberté, c'est-à-dire,  $\sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i \phi^2} = n - k$ , où  $k = p + 2$ . Cette équation nous donne la valeur de  $\phi$  en terme des  $\mu_i$ ,

$$\tilde{\phi} = \left( \frac{1}{n - k} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i} \right)^{1/2}. \quad (1.3.4)$$

Cependant, nous avons besoin de l'estimation des différents  $\mu_i$  par approximations successives, où les valeurs initiales des  $\mu_i$  proviennent du modèle de régression de

Poisson ordinaire (voir équation (1.1.3)). Premièrement, nous ajustons le modèle de Poisson ( $\phi = 1$ ) pour obtenir les  $\mu_i$  initiaux et ensuite, nous résolvons l'équation (1.3.4) pour obtenir  $\tilde{\phi}$ . Si  $\tilde{\phi} \simeq 1$ , cela implique que le modèle de régression de Poisson ordinaire est approprié et aucune estimation supplémentaire est nécessaire. Dans ce cas, nous procédons de la même façon qu'à la section 1.1.1 pour estimer  $\beta$ . Cependant, si  $\tilde{\phi} \neq 1$ , nous utilisons la valeur de l'estimation de  $\phi$  pour obtenir un nouveau  $\hat{\beta}$  à l'aide de l'équation (1.3.2). Après avoir obtenu un nouveau  $\hat{\beta}$ , nous résolvons l'équation (1.3.1) pour trouver un nouveau  $\hat{\phi}$  et (1.3.2) pour trouver un nouveau  $\hat{\beta}$ , et ainsi de suite jusqu'à ce que nous obtenions une solution stable. Nous disons que la solution est stable, lorsque le changement apporté par la nouvelle itération est plus petit qu'une certaine valeur. Dans notre cas, nous prenons comme mesure la norme quadratique,  $\frac{\|\theta^{(s+1)} - \theta^{(s)}\|}{\|\theta^{(s)}\|} \leq 0.1 \times 10^{-10}$ , où  $\theta = (\beta, \phi)'$ . D'après le théorème 1.1, si nous sommes dans le bon voisinage, les équations itératives devraient converger. Si nos calculs ne semblent pas converger, il faut utiliser la méthode de Newton-Raphson dite "amortie". Cette méthode peut être décrite de la façon suivante (voir Conte et de Boor, 1980, section 5.2). Notons  $\beta^{(s+1)} = \beta^{(s)} + \mathbf{h}$  où  $\mathbf{h} = -H^{-1}(\beta^{(s)}) \mathbf{q}(\beta^{(s)})$ . Cette méthode n'accepte pas l'estimé  $\beta^{(s+1)}$  si  $\|\mathbf{q}(\beta^{(s+1)})\| \geq \|\mathbf{q}(\beta^{(s)})\|$ . À chaque itération nous considérons les vecteurs  $\beta^{(s+1)} = \beta^{(s)} + (\mathbf{h}/2^i)$ , pour  $i = 1, 2, \dots$  et nous prenons pour  $\beta^{(s+1)}$  le premier de ces vecteurs qui vérifie  $\|\mathbf{q}(\beta^{(s+1)})\| < \|\mathbf{q}(\beta^{(s)})\|$ . Nous utilisons cette méthode dans l'exemple présenté à la section 1.4.

Pour résoudre les équations (1.3.1) et (1.3.2), nous nous référons à la section 1.1.2, afin d'utiliser la méthode de Newton-Raphson. Pour ce cas, nous posons  $\theta = (\beta, \phi)'$ . Nous avons alors,

$$\theta^{(s+1)} = \theta^{(s)} - (H^{(s)})^{-1} \mathbf{q}^{(s)}.$$

Le gradient  $\mathbf{q}^{(s)}$ , évalué en  $\boldsymbol{\theta}^{(s)}$ , est donné par

$$\mathbf{q}^{(s)} = \left[ \frac{\partial l}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}^{(s)}} = \left[ \left( \frac{\partial l}{\partial \theta_t} \right) \right]_{\boldsymbol{\theta}^{(s)}},$$

avec  $\frac{\partial l}{\partial \theta_t} = \frac{\partial l}{\partial \beta_r}$  donné par l'équation (1.3.2) pour  $t = 0, 1, \dots, p$  et  $\frac{\partial l}{\partial \theta_{p+1}} = \frac{\partial l}{\partial \phi}$  donné par l'équation (1.3.1). La matrice hessienne  $H^{(s)}$ , évaluée en  $\boldsymbol{\theta}^{(s)}$ , est donnée par

$$H^{(s)} = \left[ (h_{tm}) \right]_{\boldsymbol{\theta}^{(s)}},$$

avec  $h_{rm}$  défini de la façon suivante:

$$\begin{aligned} h_{tm} &= \frac{\partial^2 l}{\partial \beta_m \partial \beta_t} = \frac{\partial^2 l}{\partial \beta_t \partial \beta_m} \\ &= \sum_{i=1}^n \left( \frac{(y_i - 1)(\phi - 1)y_i}{[\mu_i + (\phi - 1)y_i]^2} - \frac{1}{\phi} \right) \mu_i x_{it} x_{im} \text{ pour } t, m = 0, 1, \dots, p, \\ h_{i(p+1)} &= \frac{\partial^2 l}{\partial \phi \partial \beta_t} = \frac{\partial^2 l}{\partial \beta_t \partial \phi} \\ &= \sum_{i=1}^n \left( \frac{1}{\phi^2} - \frac{y_i(y_i - 1)}{[\mu_i + (\phi - 1)y_i]^2} \right) \mu_i x_{it} \text{ pour } t = 0, 1, \dots, p \\ \text{et } h_{(p+1)(p+1)} &= \frac{\partial^2 l}{\partial \phi^2} \\ &= \sum_{i=1}^n \left( \frac{y_i}{\phi^2} - \frac{y_i^2(y_i - 1)}{[\mu_i + (\phi - 1)y_i]^2} - \frac{2(\mu_i - y_i)}{\phi^3} \right). \end{aligned}$$

Nous avons maintenant toutes les équations nécessaires pour estimer  $\boldsymbol{\beta}$  et  $\phi$  par le processus itératif décrit ci-haut.

### 1.3.3. Test du modèle généralisé

Une fois que nous avons trouvé une solution stable pour  $\phi$  et  $\boldsymbol{\beta}$ , et si  $\hat{\phi} \neq 1$ , nous sommes intéressés à savoir si le modèle généralisé de régression de Poisson

est plus approprié que le modèle de Poisson ordinaire. Pour ce faire, nous testons les hypothèses suivantes:

$$H_0 : \phi = 1 \text{ vs } H_1 : \phi \neq 1.$$

Rejeter l'hypothèse  $H_0$  implique que le modèle de Poisson devrait être rejeté. Le test statistique est basé sur (voir Consul et Famoye, 1992)

$$\begin{aligned} L &= -2 \log \left( \frac{L(\widehat{\alpha}, \widehat{\beta}, \phi = 1)}{L(\widehat{\alpha}, \widehat{\beta}, \widehat{\phi})} \right) \\ &= 2 \left( l(\widehat{\alpha}, \widehat{\beta}, \widehat{\phi}) - l(\widehat{\alpha}, \widehat{\beta}, \phi = 1) \right) \\ &= 2 \sum_{i=1}^n \left\{ \log(\widehat{\mu}_i / \widehat{\mu}_i) + (y_i - 1) \log \left( \frac{\widehat{\mu}_i + (\widehat{\phi} - 1)y_i}{\widehat{\mu}_i} \right) \right. \\ &\quad \left. - y_i \log \widehat{\phi} - \frac{\widehat{\mu}_i + (\widehat{\phi} - 1)y_i}{\widehat{\phi}} + \widehat{\mu}_i \right\}, \end{aligned}$$

où  $\widehat{\mu}_i = \exp\{\widehat{\alpha} + \widehat{\beta}x_i\}$  et  $\widehat{\phi}$  sont les estimateurs du maximum de vraisemblance des  $\mu_i$  et de  $\phi$  sous le modèle généralisé, sans aucune valeur spécifique des paramètres, et  $\widehat{\mu}_i = \exp\{\widehat{\alpha} + \widehat{\beta}x_i\}$  sont les estimateurs du maximum de vraisemblance des  $\mu_i$  sous le modèle de Poisson ordinaire, c'est-à-dire lorsque  $H_0$  est vraie sous le modèle généralisé ou encore, lorsque  $\phi = 1$ . La valeur de  $L$  est comparée à la valeur critique d'une loi du khi-deux à un degré de liberté.

#### 1.3.4. Test sur $\beta$

À la section précédente, nous avons vu comment choisir le modèle le plus approprié. Si nous rejetons le modèle de Poisson ordinaire, nous désirons tout de même tester s'il y a une tendance apparente dans le modèle généralisé. Nous testons donc les mêmes hypothèses qu'à la section 1.1.3 et le test du rapport de

vraisemblance nous donne comme statistique

$$\begin{aligned}
 l_2 &= -2 \log \left( \frac{L(\widehat{\alpha}, \beta_0 = 0, \widehat{\phi})}{L(\widehat{\alpha}, \widehat{\beta}, \widehat{\phi})} \right) \\
 &= 2 \sum_{i=1}^n \left\{ \log(\widehat{\mu}_i / \widehat{\mu}_i) + (y_i - 1) \log \left( \frac{\widehat{\mu}_i + (\widehat{\phi} - 1) y_i}{\widehat{\mu}_i + (\widehat{\phi} - 1) y_i} \right) - y_i \log(\widehat{\phi} / \widehat{\phi}) \right. \\
 &\quad \left. + \frac{\widehat{\mu}_i + (\widehat{\phi} - 1) y_i}{\widehat{\phi}} - \frac{\widehat{\mu}_i + (\widehat{\phi} - 1) y_i}{\widehat{\phi}} \right\},
 \end{aligned}$$

où  $\widehat{\mu}_i = \exp\{\widehat{\alpha} + \widehat{\beta}x_i\}$  et  $\widehat{\phi}$  sont les estimateurs du maximum de vraisemblance des  $\mu_i$  et de  $\phi$  sous le modèle généralisé, sans aucune spécification des paramètres, et  $\widehat{\mu}_i = \exp\{\widehat{\alpha}\}$  et  $\widehat{\phi}$  sont les estimateurs du maximum de vraisemblance des  $\mu_i$  et de  $\phi$ , toujours sous le modèle généralisé, lorsque  $H_0$  est vraie. Ici encore, nous comparons la valeur de  $l_2$  à la valeur critique d'un khi-deux où le nombre de degrés de liberté est égale à la longueur de  $\beta_{(1)}$ .

#### 1.4. EXEMPLE NUMÉRIQUE

Pour terminer le premier chapitre, nous appliquons les techniques vues à un exemple tiré de Lindley (1965, p.190). Le jeu de données est composé des ensembles de triplets nés en Norvège entre 1911 et 1940. Nous retrouvons les estimations de  $\alpha$  et  $\beta$ , pour le modèle ordinaire et le modèle généralisé, dans le tableau 1.4.1.

TABLEAU 1.4.1. Valeurs des estimateurs de  $\alpha$  et  $\beta$  pour le modèle standard et le modèle généralisé des ensembles de triplets

	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\phi}$
Modèle standard	4,249	-0,205	-
Modèle généralisé	4,261	-0,209	0,500



Dans le tableau 1.4.2, nous retrouvons les observations, les valeurs de la covariable, ainsi que les estimations des  $\mu_i$  pour les deux modèles. Le test d'hypothèse concernant le paramètre  $\beta$  du modèle ordinaire rejette l'hypothèse que ce dernier soit égale à zéro, avec une valeur-p de  $4,01 \times 10^{-7}$ . Le test d'hypothèses rejette à 5%, le modèle ordinaire, avec une valeur-p de  $3,708 \times 10^{-2}$ . Nous conservons donc le modèle généralisé, où il y a sous-dispersion à travers les données. Concernant le test par rapport au paramètre  $\beta$  pour le modèle avec sous-dispersion, nous rejetons encore une fois l'hypothèse  $H_0$  avec une valeur-p de  $2,03 \times 10^{-5}$ .

De plus, nous avons introduit une valeur aberrante dans le jeu de données pour vérifier que l'estimateur du modèle ordinaire converge vraiment vers ce que nous pensions. Nous avons tout simplement fait varier la valeur de  $K$  de 1 à 20, qui multiplie la deuxième observation du jeu de données. Dans le tableau 1.4.3 et les figures 1.4.1 et 1.4.2, nous retrouvons la vraie valeur de l'estimation, ainsi que celle vers laquelle nous croyons qu'elle va converger pour différents  $K$ . Pour le modèle généralisé, nous présentons les différentes estimations pour  $\alpha$ ,  $\beta$  et  $\phi$ , lorsque  $K$  varie, dans le tableau 1.4.4, ainsi que les graphiques correspondants aux figures 1.4.3, 1.4.4 et 1.4.5. Par ces derniers, nous remarquons que  $\hat{\phi}$  croît de façon linéaire lorsque  $K$  augmente, tandis que  $\hat{\alpha}$  et  $\hat{\beta}$  croient de façon logarithmique. La croissance linéaire étant plus grande que la croissance logarithmique, la présence d'une valeur aberrante influence plus l'estimation de la variance. Il est possible d'ajuster le modèle mais plus la valeur sera aberrante, plus l'erreur sera grande.

TABLEAU 1.4.2. Ensembles de triplets, nés en Norvège, ainsi que les estimations des  $\mu_i$ 

Période	1911-15	1916-20	1921-25	1926-30	1931-35	1936-40
Séries de nombre $x_i$ des ensembles	1	2	3	4	5	6
Ensembles de triplets $y_i$	52	52	40	30	24	20
$\hat{\mu}_i$ du modèle ordinaire	57,080	46,507	37,892	30,873	25,154	20,495
$\hat{\mu}_i$ du modèle généralisé	57,533	46,681	37,876	30,732	24,936	20,233

Dans ce chapitre, nous avons présenté l'approche classique de la régression poissonnienne pour deux modèles, soient le modèle standard et le modèle généralisé. Dans le prochain chapitre, nous développons une autre approche. Celle-ci se base sur des concepts bayésiens qui sont appliqués à l'approximation normale pour notre modèle.

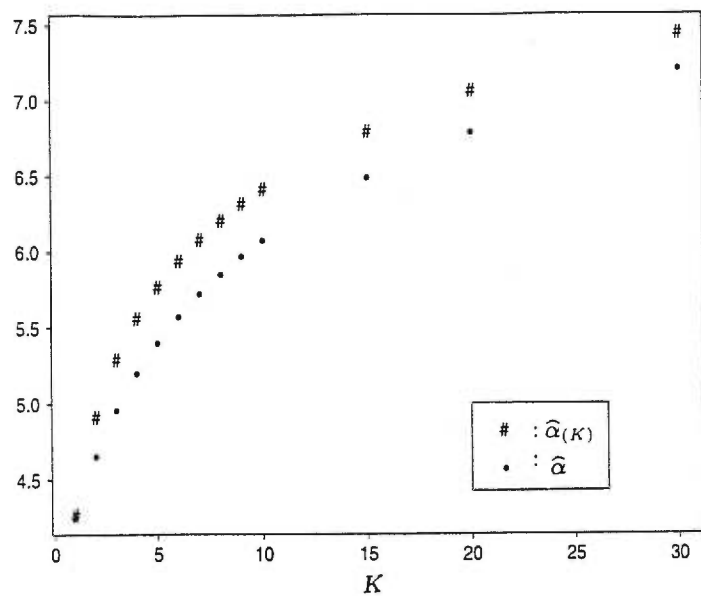


FIGURE 1.4.1. Graphique de  $\hat{\alpha}$  et de  $\hat{\alpha}(K)$  en fonction de  $K$  pour le modèle ordinaire des ensembles de triplets

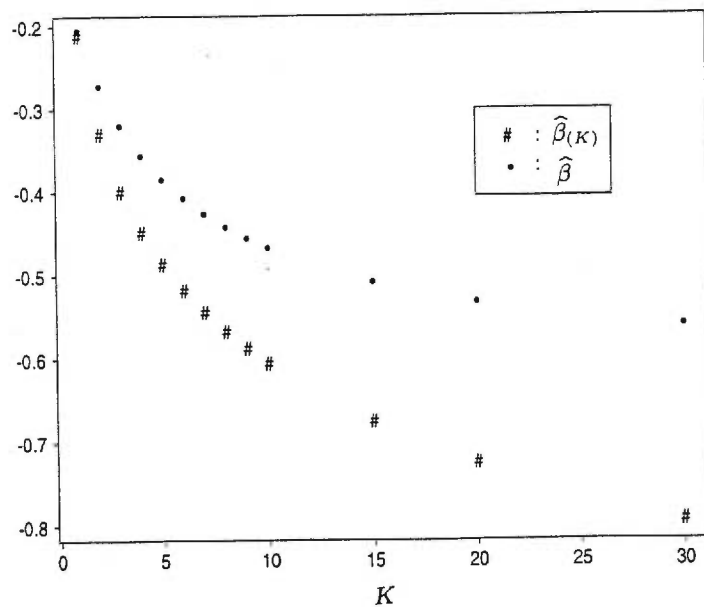


FIGURE 1.4.2. Graphique de  $\hat{\beta}$  et de  $\hat{\beta}(K)$  en fonction de  $K$  pour le modèle ordinaire des ensembles de triplets

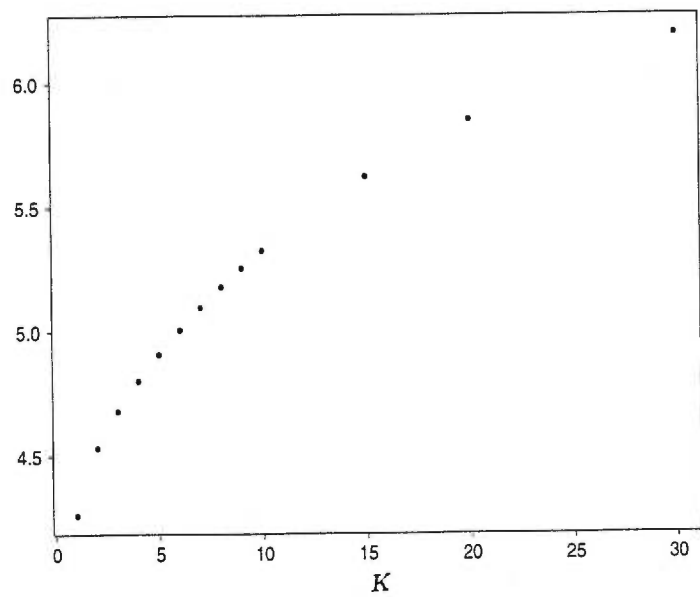


FIGURE 1.4.3. Graphique de  $\hat{\alpha}$  en fonction de  $K$  pour le modèle généralisé des ensembles de triplets

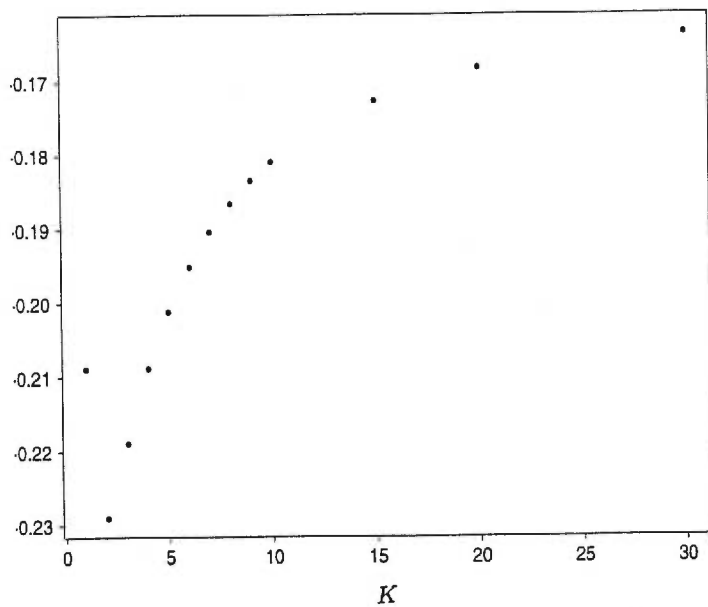


FIGURE 1.4.4. Graphique de  $\hat{\beta}$  en fonction de  $K$  pour le modèle généralisé des ensembles de triplets

TABLEAU 1.4.3. Valeurs des estimateurs du modèle ordinaire pour différents K des ensembles de triplets

$K$	$\hat{\alpha}_{(K)} = (\hat{\alpha}_0 + \log K \hat{\alpha}_*)$	$\hat{\beta}_{(K)} = (\hat{\beta}_0 + \log K \hat{\beta}_*)$	$\hat{\alpha}$	$\hat{\beta}$
1	4,266	-0,211	4,249	-0,205
2	4,913	-0,330	4,653	-0,272
3	5,292	-0,399	4,958	-0,320
4	5,560	-0,449	5,201	-0,356
5	5,768	-0,487	5,402	-0,384
6	5,939	-0,518	5,573	-0,407
7	6,082	-0,545	5,722	-0,426
8	6,207	-0,567	5,853	-0,442
9	6,317	-0,5887	5,970	-0,456
10	6,415	-0,606	6,076	-0,467
15	6,794	-0,675	6,489	-0,508
20	7,062	-0,725	6,785	-0,532
30	7,441	-0,794	7,203	-0,560

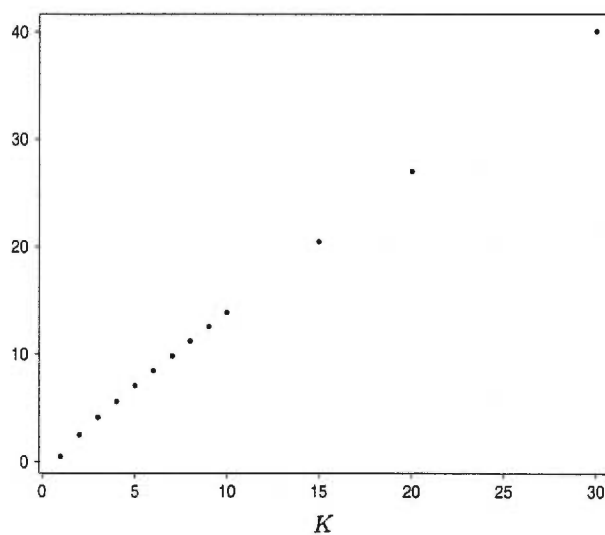


FIGURE 1.4.5. Graphique de  $\hat{\phi}$  en fonction de  $K$  pour le modèle généralisé des ensembles de triplets

TABLEAU 1.4.4. Valeurs des estimateurs du modèle général pour différents K des ensembles de triplets

K	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\phi}$
1	4,261	-0,209	0,500
2	4,533	-0,229	2,504
3	4,680	-0,219	4,146
4	4,801	-0,209	5,601
5	4,909	-0,201	7,074
6	5,006	-0,195	8,466
7	5,095	-0,190	9,835
8	5,178	-0,187	11,189
9	5,254	-0,184	12,531
10	5,325	-0,181	13,865
15	5,622	-0,173	20,469
20	5,851	-0,168	27,017
30	6,196	-0,164	40,066

## Chapitre 2

---

### APPROCHE BAYÉSIENNE DE LA RÉGRESSION POISSONNIENNE AVEC L'APPROXIMATION NORMALE

Jusqu'à maintenant, nous avons appliqué des méthodes classiques à la régression de Poisson. Nous nous intéressons maintenant aux méthodes bayésiennes, afin de faire de l'inférence sur les paramètres inconnus. Nous faisons d'abord un survol de quelques concepts bayésiens qui nous seront nécessaires dans le présent chapitre et le suivant. Par la suite, nous approximations notre modèle par un modèle normal afin d'obtenir facilement la forme des estimateurs de Bayes et de faire des tests. Nous présentons ensuite un estimateur de Bayes hiérarchique robuste par rapport aux valeurs aberrantes et nous faisons un test sur le paramètre  $\beta$ . Finalement, nous appliquons ces notions à un exemple numérique. Notons que tout ce chapitre se limite au modèle simple, c'est-à-dire avec deux régresseurs seulement car le modèle général se traite de façon similaire.

## 2.1. INTRODUCTION

Dans cette première section, nous introduisons quelques concepts nécessaires à l'estimation bayésienne. En admettant qu'une grande partie de l'inférence statistique conduit à une décision finale, il convient de pouvoir comparer les procédures décisionnelles à l'aide d'un critère d'évaluation qui décrit les conséquences de chaque décision en fonction des paramètres du modèle (Robert, 1992, section 2.1). Ceci est fait à l'aide des fonctions de perte souvent appelées coûts.

**Définition 2.1.** *Soit  $\mathcal{D}$  l'espace des décisions. Une fonction de perte est une fonction  $L$  telle que  $L(\theta, \delta) : \Theta \times \mathcal{D} \rightarrow [0, \infty)$ .*

Dans le cadre de l'estimation, la fonction de perte évalue la pénalité résultant de l'estimation d'un paramètre  $\theta$  à l'aide d'un estimateur  $\delta$ . Soit un modèle statistique,  $f(y|\theta)$ ,  $\theta \in \Theta$ . Étant donné un sous-ensemble d'intérêt  $\Theta_0$  de  $\Theta$ , nous cherchons à déterminer si  $\theta$ , le vrai paramètre, appartient à  $\Theta_0$ , c'est-à-dire à tester l'hypothèse nulle  $H_0 : \theta \in \Theta_0$ , par rapport à l'alternative  $H_1 : \theta \in \Theta_1$ , où  $\Theta_1 \cap \Theta_0 = \emptyset$ . Il existe plusieurs fonctions de perte mais nous en présentons seulement deux ainsi que les règles de Bayes correspondantes dans les propositions suivantes.

**Proposition 2.1.** *La règle de Bayes  $\delta^\pi$  de  $g(\theta)$ , associée à la loi a priori  $\pi$  et à la fonction de perte '0 - 1', c'est-à-dire*

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



où  $\Theta = \Theta_0 \cup \Theta_1$  et  $\Theta_0 \cap \Theta_1 = \emptyset$  est donnée par

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

**Proposition 2.2.** La règle de Bayes  $\delta^\pi$  de  $g(\theta)$ , associée à la loi a priori  $\pi$  et à la fonction de perte quadratique

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

est la moyenne a posteriori, c'est-à-dire

$$\begin{aligned} \delta^\pi(x) &= \mathbb{E}^{\pi(\theta|x)}[g(\theta)] \\ &= \frac{\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) d\theta}{\int_{\Theta} f(x|\theta) \pi(\theta) d\theta}. \end{aligned}$$

Nous utiliserons la perte quadratique pour toutes les estimations bayésiennes ponctuelles qui sont faites. Dans ce cas, la règle de Bayes est aussi appelée estimateur de Bayes. Une détermination subjective des lois *a priori* s'avère impossible lorsque l'information *a priori* sur le modèle est trop vague ou peu fiable. C'est pourquoi nous faisons appel aux lois *a priori* conjuguées et aux lois non informatives.

**Définition 2.2.** Une famille  $\mathcal{F}$  de lois sur  $\Theta$  est dite conjuguée (ou fermée par échantillonnage) si, pour tout  $\pi \in \mathcal{F}$ , la loi a posteriori  $\pi(\theta|x)$ , appartient également à  $\mathcal{F}$ .

L'approche bayésienne, pour effectuer des tests statistiques, utilise fréquemment les cotes de Bayes, souvent appelées facteurs de Bayes dont nous donnons la définition.

**Définition 2.3.** *La cote de Bayes pour confronter les hypothèses  $H_0$  et  $H_1$ , est donnée par le rapport suivant.*

$$B^\pi(y) = \frac{\mathbb{P}^{\pi(\theta|y)}(\theta \in \Theta_0)}{\mathbb{P}^{\pi(\theta|y)}(\theta \in \Theta_1)} \bigg/ \frac{\mathbb{P}^{\pi(\theta)}(\theta \in \Theta_0)}{\mathbb{P}^{\pi(\theta)}(\theta \in \Theta_1)}.$$

Ce rapport mesure donc le changement des vraisemblances relatives des deux hypothèses causé par les observations. Si la cote est inférieure à 0,1, nous rejetons l'hypothèse nulle tandis que si la cote est supérieure à 10, nous l'acceptons. Si la cote se retrouve entre 0,1 et 10, nous essayons d'avoir plus d'informations ou de raffiner notre jugement *a priori*.

## 2.2. APPROXIMATION NORMALE

Lindley (1965) a montré que la plupart des lois *a posteriori*, qui se conduisent raisonnablement bien dans des grands échantillons, peuvent être approximées par la distribution normale multidimensionnelle. Nous estimons, de façon approximative, les moyennes et la matrice de variance-covariance par les estimateurs du maximum de vraisemblance pour les paramètres concernés et la matrice d'information, évaluée aux estimateurs. Quoique la loi *a posteriori* de  $\alpha$ , sous une loi *a priori* judicieusement choisie, peut être évaluée de façon analytique, celle de  $\beta$  n'est pas facile à manipuler théoriquement, bien qu'une tabulation exacte peut être faite pour n'importe quel cas particulier (voir El-Sayyad, 1973). C'est pourquoi nous suggérons l'approximation normale pour la distribution *a posteriori* de  $\alpha$  et  $\beta$ , ce qui est facile à calculer. Dans le but de justifier la validité de cette approximation, nous examinons l'ordre de l'erreur d'approximation si nous remplaçons la vraie densité par une loi normale (voir Bartlett et Kendall, 1946).

**Théorème 2.1.** Soit  $\mu$  ayant une distribution *Gamma*  $(y, 1)$ , c'est-à-dire

$$f(\mu) = \frac{1}{\Gamma(y)} \mu^{y-1} e^{-\mu} \mathbb{I}_{(0, \infty)}(\mu).$$

Alors, pour  $y$  grand,  $\log \mu$  sera approximativement normal avec moyenne  $\log y$  et variance  $1/y$ .

**Démonstration 2.1.** Nous faisons le changement de variables suivant:

$$t = \log \mu \Rightarrow \mu = e^t. \text{ D'où } dt = \frac{d\mu}{\mu} \Rightarrow d\mu = e^t dt.$$

La distribution devient donc

$$f(t) = \frac{e^{-(e^t - yt)}}{\Gamma(y)}.$$

Soit  $g(t) = e^t - yt$ . Nous développons  $g(t)$  en série de Taylor autour de  $\log y$ .

Ainsi

$$g(t) \simeq y(1 - \log y) + \frac{y}{2}(t - \log y)^2 + \frac{y}{3!}(t - \log y)^3 + \frac{g'''(\xi)}{4!}(t - \log y)^4,$$

où  $\xi \in (t, \log y)$ .

Nous substituons la dernière équation dans  $f(t)$  et nous obtenons,

$$\begin{aligned} f(t) &\simeq \frac{e^{-y(1 - \log y)}}{\Gamma(y)} e^{-[\frac{y}{6}(t - \log y)^3 + \frac{y'''(\xi)}{24}(t - \log y)^4]} e^{-\frac{y}{2}(t - \log y)^2} \\ &= K h(t) e^{-\frac{y}{2}(t - \log y)^2}, \end{aligned}$$

$$\text{où } K = \frac{e^{-y(1 - \log y)}}{\Gamma(y)} \text{ et } h(t) = e^{-[\frac{y}{6}(t - \log y)^3 + \frac{y'''(\xi)}{24}(t - \log y)^4]}.$$

Maintenant, en développant  $h(t)$  en série de Taylor autour de  $\log y$ , nous obtenons

$$h(t) \simeq 1 - \frac{y}{3!}(t - \log y)^3 + \frac{h'''(\xi)}{4!}(t - \log y)^4.$$

À nouveau, nous substituons cette équation dans  $f(t)$ , et nous obtenons

$$f(t) \simeq \frac{\sqrt{2\pi}}{\sqrt{y}} K \left[ 1 - \frac{y}{6}(t - \log y)^3 + \frac{h'''(\xi)}{24}(t - \log y)^4 \right] \varphi(t),$$

où  $\varphi(t) = \frac{\sqrt{y}}{\sqrt{2\pi}} e^{-\frac{y}{2}(t - \log y)^2}$ .

Nous voulons calculer la norme euclidienne de l'erreur qui est donnée par

$$\| f - \varphi \| = \sqrt{\int_0^\infty |f(t) - \varphi(t)|^2 dt}.$$

Nous calculons d'abord

$$\begin{aligned} \int_0^\infty |f(t) - \varphi(t)|^2 dt &= \int_0^\infty c \left\{ a^2 \left[ 1 - \frac{2y}{6}(t - \log y)^3 + \frac{2h'''(\xi)}{24}(t - \log y)^4 \right. \right. \\ &\quad \left. \left. + \frac{y^2}{36}(t - \log y)^6 - \frac{2yh'''(\xi)}{6 \times 24}(t - \log y)^7 + \left[ \frac{h'''(\xi)}{24} \right]^2 (t - \log y)^8 \right] \right. \\ &\quad \left. - 2a \left[ 1 - \frac{y}{6}(t - \log y)^3 + \frac{h'''(\xi)}{24}(t - \log y)^4 \right] + 1 \right\} \varphi^*(t) dt, \end{aligned}$$

où  $a = \frac{\sqrt{2\pi} e^{-y(1-\log y)}}{\sqrt{y} \Gamma(y)}$ ,  $c = \frac{\sqrt{y}}{2\sqrt{\pi}}$  et  $\varphi^*(t) = \frac{\sqrt{y}}{\sqrt{\pi}} e^{-y(t - \log y)^2}$ . Or, nous savons que si  $X \sim N(0, \sigma^2)$ , alors

$$\mathbb{E}[X^{2k+1}] = 0,$$

$$\mathbb{E}[X^{2k}] = \frac{(2k)! \sigma^{2k}}{2^k k!}.$$

Puisque  $t \sim N(\log y, 1/(2y))$ , alors  $(t - \log y) \sim N(0, 1/(2y))$  et nous obtenons

$$\begin{aligned} \int_0^\infty |f(t) - \varphi(t)|^2 dt &= c \left\{ a^2 \left[ 1 + \frac{y^2 a_3}{36 \times (2y)^6} + \frac{a_2}{12 \times (2y)^4} + \frac{a_4}{(24)^2 \times (2y)^8} \right] \right. \\ &\quad \left. - 2a - \frac{a a_2}{12 \times (2y)^4} + 1 \right\} \\ &= c \left\{ (a - 1) \left[ (a - 1) + \frac{a b_2}{y^4} \right] + \frac{a^2}{y^4} \left( b_3 + \frac{b_4}{y^4} \right) \right\}, \end{aligned}$$

avec  $a_k = \frac{(2k)!}{2^k k!}$ ,  $b_2 = \frac{a_2}{12 \times 2^4}$ ,  $b_3 = \frac{a_3}{36 \times 2^6}$  et  $b_4 = \frac{a_4}{(24)^2 \times 2^8}$ . En utilisant l'approximation de Stirling, nous avons  $\Gamma(y) \simeq y^{y-1/2} e^{-y} \sqrt{2\pi} \left(1 + \frac{1}{12y}\right)$  et alors  $a = \frac{12y}{1+12y}$ . En substituant  $a$  dans l'intégrale, nous obtenons finalement

$$\int_0^\infty |f(t) - \varphi(t)|^2 dt \simeq \frac{\sqrt{y}}{2\sqrt{\pi}} \left\{ \frac{1}{144y^2} - \frac{b_2}{12y^5} + \frac{b_3}{y^4} + \frac{b_4}{y^8} \right\} \approx \mathcal{O}(y^{-3/2}).$$

L'erreur que nous cherchions est donnée par

$$\|f - \varphi\| \approx \mathcal{O}\left(\frac{1}{y^{3/4}}\right).$$

Nous pouvons donc voir que l'approximation normale est valable pour notre modèle pour de grandes valeurs de  $y$ .  $\square$

### 2.2.1. Estimateurs de Bayes

Premièrement, considérons comme loi *a priori* sur  $\mu$ , la loi gamma suivante:

$$f(\mu) = \frac{1}{\Gamma(y)} \mu^{y-1} e^{-\mu} d\mu.$$

Nous savons que si une variable aléatoire a une telle distribution, alors son logarithme est approximativement de loi normale. Et donc, pour  $y$  assez grand,  $\log \mu \simeq N(\log y, 1/y)$ . Nous obtenons pour la fonction de vraisemblance des observations,

$$L(\boldsymbol{\mu}) = \prod_{i=1}^n \exp \left\{ -\frac{1}{2} y_i (\log \mu_i - \log y_i)^2 \right\}$$

$$L(\alpha, \beta) = \exp \left\{ -\frac{1}{2} \sum_{i=1}^n y_i (\alpha + \beta x_i - \log y_i)^2 \right\}.$$

Si nous choisissons la loi *a priori* de  $\alpha$  et  $\beta$  comme étant proportionnelle à  $d\alpha d\beta$ , alors, à condition que les  $y_i$  soient assez grands, la loi *a posteriori* de  $\alpha$  et  $\beta$  est

approximativement

$$\pi(\alpha, \beta | \mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n y_i (\alpha + \beta x_i - \log y_i)^2 \right\},$$

qui est une distribution binormale pour  $\alpha$  et  $\beta$ . Si nous posons

$$\mathbf{t} \equiv (\sqrt{y_1} \log y_1, \dots, \sqrt{y_n} \log y_n)'$$

et

$$Z' \equiv \begin{pmatrix} \sqrt{y_1} & \dots & \sqrt{y_n} \\ x_1 \sqrt{y_1} & \dots & x_n \sqrt{y_n} \end{pmatrix},$$

nous obtenons le théorème suivant.

**Théorème 2.2.** *Le vecteur des moyennes a posteriori fourni par les estimateurs de Bayes, est donné par*

$$\begin{pmatrix} \mathbb{E}[\alpha] \\ \mathbb{E}[\beta] \end{pmatrix} = (Z'Z)^{-1}Z'\mathbf{t},$$

et la matrice de variance-covariance de  $\alpha$  et  $\beta$  est donnée par

$$\begin{pmatrix} V(\alpha) & Cov(\alpha, \beta) \\ Cov(\alpha, \beta) & V(\beta) \end{pmatrix} = (Z'Z)^{-1} = \begin{pmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 y_i \end{pmatrix}^{-1},$$

qui est la même que celle que nous aurions obtenue en utilisant l'approximation bayésienne de Lindley (1965), avec les estimateurs du maximum de vraisemblance.

Si  $z_i = \log y_i$ , nous obtenons alors,

$$\mu_\alpha = \mathbb{E}[\alpha] = \frac{\left( \sum_{i=1}^n x_i^2 y_i \right) \left( \sum_{i=1}^n y_i z_i \right) - \left( \sum_{i=1}^n x_i y_i \right) \left( \sum_{i=1}^n x_i y_i z_i \right)}{\left( \sum_{i=1}^n y_i \right) \left( \sum_{i=1}^n x_i^2 y_i \right) - \left( \sum_{i=1}^n x_i y_i \right)^2},$$

$$\begin{aligned}\mu_\beta = \mathbb{E}[\beta] &= \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i y_i z_i\right) - \left(\sum_{i=1}^n x_i y_i\right)\left(\sum_{i=1}^n y_i z_i\right)}{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i^2 y_i\right) - \left(\sum_{i=1}^n x_i y_i\right)^2}, \\ \sigma_\alpha^2 = \text{Var}(\alpha) &= \frac{\sum_{i=1}^n x_i^2 y_i}{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i^2 y_i\right) - \left(\sum_{i=1}^n x_i y_i\right)^2}, \\ \sigma_\beta^2 = \text{Var}(\beta) &= \frac{\sum_{i=1}^n y_i}{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i^2 y_i\right) - \left(\sum_{i=1}^n x_i y_i\right)^2}, \\ \rho = \text{Corr}(\alpha, \beta) &= \frac{-\sum_{i=1}^n x_i y_i}{\sqrt{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i^2 y_i\right)}}.\end{aligned}$$

### 2.2.2. Valeur aberrante

Dans cette section, nous utilisons l'approche bayésienne hiérarchique, afin d'estimer les paramètres de régression, en supposant une erreur multinormale. Nous développons les grandes lignes de l'article de Angers et MacGibbon (1996). Nous pouvons appliquer cet article, à condition que la matrice de variance-covariance soit diagonale, et nous trouvons un estimateur de Bayes hiérarchique qui est partiellement insensible à la présence de valeurs aberrantes. Posons  $\boldsymbol{\varphi} = (\alpha, \beta)'$  et  $\boldsymbol{\mu} = e^{X\boldsymbol{\varphi}}$ . En utilisant l'approximation par la loi normale discutée à la section 2.2.1, nous avons  $\log \mathbf{Y} = X\boldsymbol{\varphi} + \boldsymbol{\varepsilon}$ , où  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \Sigma_M)$  et  $\Sigma_M = \text{diag}\left(\frac{1}{y_1}, \dots, \frac{1}{y_n}\right)$ . En faisant le changement de variables  $T_i = \sqrt{y_i} \log y_i$  et alors  $\mathbf{T} = (\sqrt{y_1} \log y_1, \dots, \sqrt{y_n} \log y_n)'$ . Nous avons donc le modèle suivant.

$$\mathbf{T} = Z\boldsymbol{\varphi} + \boldsymbol{\varepsilon}^*, \text{ où } \boldsymbol{\varepsilon}^* \sim N(\mathbf{0}, I) \text{ et } Z = \begin{pmatrix} \sqrt{y_1} & x_1 \sqrt{y_1} \\ \vdots & \vdots \\ \sqrt{y_n} & x_n \sqrt{y_n} \end{pmatrix}.$$

La densité *a priori* de premier niveau pour  $\boldsymbol{\varphi}$  est donnée par  $\boldsymbol{\varphi} \sim N_2(\mathbf{0}, S)$ , où  $S = \text{diag}(s_1^2, s_2^2)$  et celle pour  $S$  est  $\pi_2(S) \propto |I_p + SZ'Z|^{-l}$  où  $l$  est choisi assez grand pour que  $\pi_2(S)$  soit propre. La fonction de vraisemblance est

$$L(\boldsymbol{\varphi}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{T} - Z\boldsymbol{\varphi})' (\mathbf{T} - Z\boldsymbol{\varphi}) \right\}. \quad (2.2.1)$$

**Théorème 2.3.** *La densité a posteriori de  $\boldsymbol{\varphi}$ , si  $S$  est connue, est*

$$\boldsymbol{\varphi} | \mathbf{T}, S \sim N_2 \left( S(S + \Sigma)^{-1} \hat{\boldsymbol{\varphi}}_{LS}, S(S + \Sigma)^{-1} \Sigma \right),$$

où  $\Sigma = (Z'Z)^{-1}$  et  $\hat{\boldsymbol{\varphi}}_{LS} = \Sigma Z' \mathbf{T}$ .

Les densités a posteriori sont données par

$$\pi_1(\boldsymbol{\varphi} | \mathbf{T}, S) \propto e^{-\frac{1}{2}(\boldsymbol{\varphi} - S(S + \Sigma)^{-1} \hat{\boldsymbol{\varphi}}_{LS})' A (\boldsymbol{\varphi} - S(S + \Sigma)^{-1} \hat{\boldsymbol{\varphi}}_{LS})},$$

$$\pi_2(S | \mathbf{T}) \propto \frac{1}{|S + \Sigma|^{l+1/2}} e^{-\frac{1}{2} \hat{\boldsymbol{\varphi}}_{LS}' (S + \Sigma)^{-1} \hat{\boldsymbol{\varphi}}_{LS}},$$

où  $A = \Sigma^{-1} + S^{-1}$ .

Conséquemment, sous une perte quadratique, l'estimateur de Bayes de premier niveau est donné par

$$\hat{\boldsymbol{\varphi}} | \mathbf{z}, S = S(S + \Sigma)^{-1} \hat{\boldsymbol{\varphi}}_{LS}.$$

**Théorème 2.4.** *L'estimateur de Bayes hiérarchique robuste est donné par*

$$\hat{\boldsymbol{\varphi}}^H = (I_p - AG_l A') \hat{\boldsymbol{\varphi}}_{LS},$$

où  $G_l = \text{diag}(g_l(z_1^2/2), \dots, g_l(z_p^2/2))$ , avec

$$g_l(x) = \begin{cases} \frac{l-1/2}{l+1/2} & \text{si } x = 0, \\ \frac{\gamma(l+1/2, x)}{x \gamma(l-1/2, x)} & \text{sinon,} \end{cases}$$



$$\mathbf{z}' = (z_1, \dots, z_p) = Q^{-1} \widehat{\varphi}_{LS},$$

et où  $Q$  représente la matrice obtenue en faisant la décomposition de Cholesky de  $(Z'Z)^{-1}$ , c'est-à-dire que  $Q$  doit satisfaire  $QQ' = (Z'Z)^{-1}$ .

La démonstration de ces deux théorèmes se retrouve dans Angers et MacGibbon (1996).

Afin d'assurer la robustesse de l'estimateur obtenu, nous portons une attention particulière aux matrices de design de type suivant. Nous écrivons  $Z$  comme

$$\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix},$$

où  $Z_{11}$  est une matrice  $k \times m$ ,  $Z_{12}$  une  $k \times (p - m)$ ,  $Z_{21}$  une  $(n - k) \times m$  et  $Z_{22}$  une  $(n - k) \times (p - m)$  pour des entiers  $k$  et  $m$  tels que  $0 \leq k \leq n$ ,  $0 \leq m \leq p$  et  $m \leq k$ .

**Définition 2.4.** *Nous supposons que  $Z \in \mathcal{D}$ , la classe des matrices de design  $Z$  qui satisfont les conditions suivantes.*

1. la première colonne de  $Z_{11}$  contient seulement des zéros,
2.  $Z_{12}$  est orthogonale à la première colonne de  $Z_{11}(Z'_{11}Z_{11} + Z'_{21}Z_{21})^{-1}Z'_{11}$ ,
3.  $Z_{21}$  est orthogonale à la première colonne de  $(Z'_{11}Z_{11} + Z'_{21}Z_{21})^{-1}Z'_{11}$ .

Nous utilisons une décomposition analogue pour le vecteur  $\mathbf{T}$  qui est,  $\mathbf{T} = (\mathbf{T}'_{(1)}\mathbf{T}'_{(2)})'$ , où  $\mathbf{T}_{(1)}$  est un vecteur  $k \times 1$ , contenant la composante  $T_1$  et  $\mathbf{T}_{(2)}$  est  $(n - k) \times 1$ .

**Théorème 2.5.** Soit  $\mathbf{T} = Z\boldsymbol{\varphi} + \boldsymbol{\varepsilon}^*$  défini tel que précédemment. Soit  $Q$  la matrice définie par la décomposition de Cholesky de  $(Z'Z)^{-1}$  qui est,

$$\begin{pmatrix} Q_{11} & 0 \\ Q_{21} & Q_{22} \end{pmatrix},$$

où  $Q_{11}$  est une matrice triangulaire inférieure  $m \times m$ ,  $Q_{21}$  une matrice  $(p-m) \times m$  et  $Q_{22}$  une matrice triangulaire inférieure  $(p-m) \times (p-m)$  telle que  $QQ' = (Z'Z)^{-1}$ .

Si  $|T_1| \rightarrow \infty$  et si la matrice de design  $Z$  est un membre de la classe  $\mathcal{D}$ , alors

$$\hat{\boldsymbol{\varphi}}^H = \begin{pmatrix} \hat{\boldsymbol{\varphi}}_{LS}^{(1)} \\ (I_{p-m} - Q_{22}G_{l(2)}Q_{22}^{-1})\hat{\boldsymbol{\varphi}}_{LS}^{(2)} + Q_{22}G_{l(2)}Q_{22}^{-1}Q_{11}^{-1}\hat{\boldsymbol{\varphi}}_{LS}^{(1)} \end{pmatrix},$$

où  $\hat{\boldsymbol{\varphi}}_{LS}^{(1)}$  correspond aux  $m$  premières coordonnées de  $\hat{\boldsymbol{\varphi}}_{LS}$ ,  $\hat{\boldsymbol{\varphi}}_{LS}^{(2)}$ , aux  $p-m$  restantes et  $G_{l(2)} = \text{diag}(g_l(z_{m+1}^2/2), \dots, g_l(z_p/2))$ .

Dans la dernière section, nous appliquons cet estimateur à un exemple dans lequel nous introduisons une valeur aberrante. Notons que  $\alpha$  et  $\beta$  sont affectés par la présence de valeurs aberrantes, car la matrice de "design" ne satisfait pas les conditions énoncées à la section 2.2.2.

### 2.2.3. Test sur $\beta$

Nous souhaitons tester s'il y a une tendance apparente et cela se fait en testant les hypothèses

$$H_0 : \beta = 0 \text{ vs } H_1 : \beta \neq 0.$$

Pour ce faire, nous utilisons la cote de Bayes, qui est donnée par

$$B^\pi(\mathbf{y}) = \frac{\mathbb{P}^{\pi(\alpha, \beta | \mathbf{y})}((\alpha, \beta) \in \Theta_0)}{\mathbb{P}^{\pi(\alpha, \beta | \mathbf{y})}((\alpha, \beta) \in \Theta_1)} \bigg/ \frac{\mathbb{P}^{\pi(\alpha, \beta)}((\alpha, \beta) \in \Theta_0)}{\mathbb{P}^{\pi(\alpha, \beta)}((\alpha, \beta) \in \Theta_1)}, \quad (2.2.2)$$

où  $\Theta_0 = \mathbb{R} \times \{0\}$  et  $\Theta_1 = \mathbb{R}^2$ .

Pour calculer la cote de Bayes, nous avons besoin d'une loi *a priori* qui soit propre (Robert, 1992, section 5.1). Nous prenons comme loi *a priori*, la loi conjuguée, qui est entièrement déterminée par la fonction de vraisemblance. En posant  $\boldsymbol{\varphi} = (\alpha, \beta)'$ , nous prenons comme loi *a priori*, la loi conjuguée  $\boldsymbol{\varphi} \sim N_2(\mathbf{0}, S)$ , où  $S = \text{diag}(s_1^2, s_2^2)$  est connue. La fonction de vraisemblance est donnée par l'équation (2.2.1) que nous pouvons écrire sous la forme suivante:

$$L(\boldsymbol{\varphi}) \propto e^{-\frac{1}{2}\mathbf{r}'\mathbf{r}} e^{-\frac{1}{2}(\boldsymbol{\varphi}-\widehat{\boldsymbol{\varphi}}_{LS})'\Sigma^{-1}(\boldsymbol{\varphi}-\widehat{\boldsymbol{\varphi}}_{LS})},$$

où  $\mathbf{r} = \mathbf{T} - Z\widehat{\boldsymbol{\varphi}}_{LS}$ .

Au théorème 2.3, nous avons vu que la densité *a posteriori* de  $\boldsymbol{\varphi}$  est donnée par

$$\boldsymbol{\varphi} \mid \mathbf{T}, S \sim N_2(S(S + \Sigma)^{-1}\widehat{\boldsymbol{\varphi}}_{LS}, A^{-1}),$$

avec  $A^{-1} = S(S + \Sigma)^{-1}\Sigma$ . Nous retrouvons la cote de Bayes dans le théorème suivant.

**Théorème 2.6.** *Soient  $\pi_0$  la probabilité a priori que  $H_0$  soit vraie et  $g_i$ , la densité a priori de  $(\alpha, \beta)$  sous l'hypothèse  $H_i$ ,  $i = 0, 1$ . Alors, la cote de Bayes est donnée par*

$$\begin{aligned} B^\pi(\mathbf{y}) &= \frac{\int_{-\infty}^{\infty} f(\mathbf{y} \mid \alpha, \beta_0) g_0(\alpha, \beta_0) d\alpha}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{y} \mid \alpha, \beta) g_1(\alpha, \beta) d\alpha d\beta} \\ &= \frac{m_0(\mathbf{y})}{m_1(\mathbf{y})} \\ &= \frac{|AS|^{1/2}}{\sqrt{2\pi} |A_\alpha S_\alpha|^{1/2}} e^{-\frac{1}{2}\{\mathbf{r}'_\alpha \mathbf{r}_\alpha - \mathbf{r}'\mathbf{r} + \widehat{\alpha}_{LS}^2 (S_\alpha + \Sigma_\alpha)^{-1} - \widehat{\boldsymbol{\varphi}}'_{LS} (S + \Sigma)^{-1} \widehat{\boldsymbol{\varphi}}_{LS}\}}, \end{aligned}$$

où  $\Sigma$ ,  $\mathbf{r}$ ,  $S$ ,  $A$  et  $\widehat{\boldsymbol{\varphi}}_{LS}$  sont tels que définis précédemment et  $\mathbf{r}_\alpha = \mathbf{T} - W\widehat{\alpha}_{LS}$ ,  $S_\alpha = s_1^2$ ,  $\Sigma_\alpha = (W'W)^{-1}$ ,  $\widehat{\alpha}_{LS} = \Sigma_\alpha W'\mathbf{T}$  et  $W = (\sqrt{y_1}, \dots, \sqrt{y_n})'$ .

**Démonstration 2.2.** Sous  $H_1$ , nous avons

$$\begin{aligned}
m_1(\mathbf{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\mathbf{r}'\mathbf{r}} e^{-\frac{1}{2}(\boldsymbol{\varphi}-\widehat{\boldsymbol{\varphi}}_{LS})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}-\widehat{\boldsymbol{\varphi}}_{LS})} \frac{1}{|S|^{1/2}} e^{-\frac{1}{2}\boldsymbol{\varphi}'S^{-1}\boldsymbol{\varphi}} d\boldsymbol{\varphi} \\
m_1(\mathbf{y}) &= 2\pi \frac{|A^{-1}|^{1/2}}{|S|^{1/2}} e^{-\frac{1}{2}\mathbf{r}'\mathbf{r}} e^{-\frac{1}{2}\widehat{\boldsymbol{\varphi}}'_{LS}(S+\boldsymbol{\Sigma})^{-1}\widehat{\boldsymbol{\varphi}}_{LS}} \\
&\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi |A^{-1}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{\varphi}-S(S+\boldsymbol{\Sigma})^{-1}\widehat{\boldsymbol{\varphi}}_{LS})'A(\boldsymbol{\varphi}-S(S+\boldsymbol{\Sigma})^{-1}\widehat{\boldsymbol{\varphi}}_{LS})} d\boldsymbol{\varphi} \\
&= \frac{2\pi}{|AS|^{1/2}} e^{-\frac{1}{2}\{\mathbf{r}'\mathbf{r}+\widehat{\boldsymbol{\varphi}}'_{LS}(S+\boldsymbol{\Sigma})^{-1}\widehat{\boldsymbol{\varphi}}_{LS}\}}.
\end{aligned}$$

Sous  $H_0$ , nous avons le modèle suivant:

$$\mathbf{T} = W\boldsymbol{\alpha} + \boldsymbol{\varepsilon}^*, \text{ avec } W = (\sqrt{y_1}, \dots, \sqrt{y_n})' \text{ et } S_\alpha = s_1^2.$$

Alors,

$$\begin{aligned}
m_0(\mathbf{y}) &= \int_{-\infty}^{\infty} e^{-\frac{1}{2}\mathbf{r}'_\alpha \mathbf{r}_\alpha} e^{-\frac{1}{2}(\alpha-\widehat{\alpha}_{LS})^2 \boldsymbol{\Sigma}_\alpha^{-1}} \frac{1}{|S_\alpha|^{1/2}} e^{-\frac{1}{2}\alpha^2 S_\alpha^{-1}} d\alpha \\
&= \sqrt{2\pi} \frac{|A_\alpha^{-1}|^{1/2}}{|S_\alpha|^{1/2}} e^{-\frac{1}{2}\mathbf{r}'_\alpha \mathbf{r}_\alpha} e^{-\frac{1}{2}\widehat{\alpha}^2_{LS}(S_\alpha+\boldsymbol{\Sigma}_\alpha)^{-1}} \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} |A_\alpha^{-1}|^{1/2}} e^{-\frac{1}{2}(\alpha-S_\alpha(S_\alpha+\boldsymbol{\Sigma}_\alpha)^{-1}\widehat{\alpha}_{LS})^2 A_\alpha} d\alpha \\
&= \frac{\sqrt{2\pi}}{|A_\alpha S_\alpha|^2} e^{-\frac{1}{2}\{\mathbf{r}'_\alpha \mathbf{r}_\alpha + \widehat{\alpha}^2_{LS}(S_\alpha+\boldsymbol{\Sigma}_\alpha)^{-1}\}}.
\end{aligned}$$

Nous retrouvons la cote de Bayes en faisant le rapport des deux densités marginales.  $\square$

### 2.3. EXEMPLE NUMÉRIQUE

Pour clore ce chapitre, nous appliquons les méthodes au même exemple vu à la section 1.4. Pour l'approximation normale, nous obtenons les estimateurs

suivants:

$$\hat{\alpha} = \mathbb{E}[\alpha] = 4,256,$$

$$\hat{\beta} = \mathbb{E}[\beta] = -0,206,$$

$$Var(\alpha) = 1,967 \times 10^{-2},$$

$$Var(\beta) = 1,771 \times 10^{-3}$$

$$\text{et } Cov(\alpha, \beta) = -5,168 \times 10^{-3}.$$

Nous retrouvons, dans le tableau 2.3.1, les  $\hat{\mu}_i$ , ainsi que les observations et les valeurs de la covariable. De la même façon que dans le chapitre précédent, nous avons introduit une valeur aberrante dans le jeu de données. Les valeurs des estimateurs de Bayes avec l'approximation normale et des estimateurs de Bayes hiérarchiques que nous avons développé dans les sections 2.2.1 et 2.2.2, avec  $l = \frac{1}{2} + \frac{1}{n^2}$  sont données dans le tableau 2.3.2 pour différentes valeurs de  $K$ . De plus, nous retrouvons le graphique des erreurs absolues entre les estimateurs de Bayes avec l'approximation normale et les estimateurs de Bayes hiérarchiques dans la figure 2.3.1. Nous remarquons que l'estimateur hiérarchique n'est pas robuste aux valeurs aberrantes car la matrice de "design" ne satisfait pas les conditions que nous avons mentionnées. Concernant le test pour  $\beta = 0$ , nous obtenons une cote de Bayes de  $2,13 \times 10^{-4}$ . Tout comme avec le test classique, nous rejetons l'hypothèse  $H_0$  et conservons le paramètre  $\beta$  dans le modèle.

Dans ce chapitre, nous avons étudié le modèle sous l'approximation normale. Nous avons utilisé des concepts bayésiens afin de calculer les estimateurs des paramètres  $\alpha$  et  $\beta$ . De plus, nous avons développé un estimateur hiérarchique aux valeurs aberrantes. Dans le prochain chapitre, nous nous concentrons sur l'étude bayésienne du modèle en considérant la vraie distribution des paramètres.

TABLEAU 2.3.1. Ensembles de triplets, nés en Norvège, ainsi que les estimations des  $\mu_i$ 

Période	1911-15	1916-20	1921-25	1926-30	1931-35	1936-40
Séries de nombre $x_i$ des ensembles	1	2	3	4	5	6
Ensembles de triplets $y_i$	52	52	40	30	24	20
$\hat{\mu}_i$ du modèle simple	57,374	46,694	38,00118	30,927	25,170	20,484

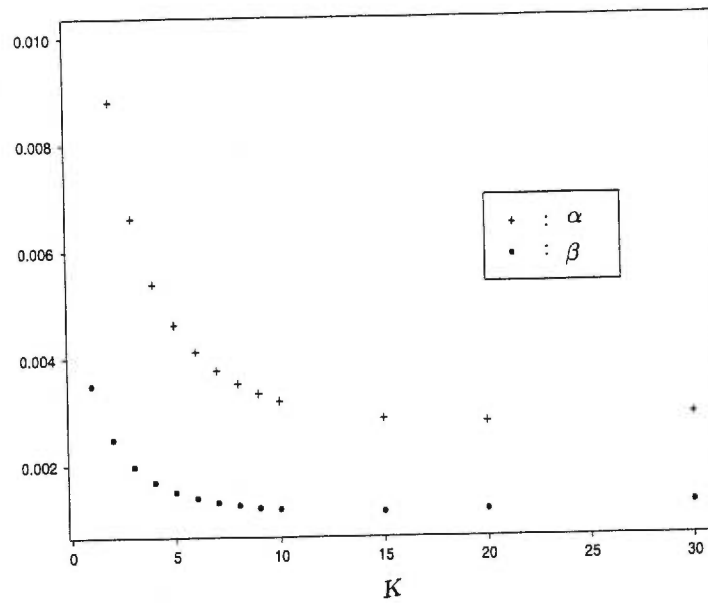
FIGURE 2.3.1. Graphique des erreurs absolues des estimateurs de  $\alpha$  et  $\beta$  en fonction de  $K$  pour le modèle ordinaire des ensembles de triplets

TABLEAU 2.3.2. Valeurs des estimateurs de Bayes avec l'approximation normale et des estimateurs de Bayes hiérarchiques pour différents K

K	$\hat{\alpha}_N$	$\hat{\beta}_N$	$\hat{\alpha}_H$	$\hat{\beta}_H$
1	4,256	-0,206	4,242	-0,203
2	4,806	-0,302	4,797	-0,299
3	5,299	-0,388	5,293	-0,386
4	5,717	-0,461	5,711	-0,459
5	6,072	-0,522	6,067	-0,521
6	6,378	-0,576	6,374	-0,574
7	6,645	-0,622	6,641	-0,621
8	6,881	-0,663	6,878	-0,662
9	7,093	-0,700	7,089	-0,699
10	7,283	-0,733	7,280	-0,732
15	8,023	-0,862	8,020	-0,861
20	8,546	-0,953	8,543	-0,952
30	9,273	-1,080	9,270	-1,079

## Chapitre 3

---

# APPROCHE BAYÉSIENNE DE LA RÉGRESSION POISSONNIENNE AVEC LA DISTRIBUTION EXACTE

Dans le chapitre précédent, nous avons approximé la distribution *a posteriori* de  $\alpha$  et  $\beta$  par une loi normale. Dans le présent chapitre, nous étudions la distribution exacte mais nous devons utiliser des méthodes numériques pour approximer les estimateurs. Nous optons pour la méthode de Naylor-Smith et l'approximation analytique de Laplace. Dans ce chapitre, nous nous concentrons sur le modèle log-linéaire simple. Pour ce modèle, nous donnons la forme des estimateurs. De plus, nous observons l'effet d'une valeur aberrante sur ces derniers. Dans la dernière section, nous développons une loi conjuguée pour le modèle simple et nous faisons des tests sur le paramètre  $\beta$ . Nous présentons un exemple pour chacune de ces sections.

### 3.1. FORME DES ESTIMATEURS DE BAYES

Nous nous souvenons que pour le modèle log-linéaire simple, la fonction de vraisemblance est donnée par



$$\begin{aligned}
f(\mathbf{y} \mid \alpha, \beta, \mathbf{x}) &= \prod_{i=1}^n f(y_i \mid \alpha, \beta, x_i) \\
&= \frac{e^{\alpha \sum_{i=1}^n y_i} e^{\beta \sum_{i=1}^n x_i y_i} e^{-e^{\alpha} \sum_{i=1}^n e^{\beta x_i}}}{\prod_{i=1}^n y_i!}.
\end{aligned} \tag{3.1.1}$$

C'est cette fonction de vraisemblance exacte, que nous utiliserons tout au long de ce chapitre.

Premièrement, nous devons spécifier la distribution *a priori* pour les paramètres. Puisque  $\alpha$  et  $\beta$  varient de  $-\infty$  à  $+\infty$ , une loi *a priori* vague, représentant notre ignorance est

$$\pi(\alpha, \beta) \propto d\alpha d\beta, \tag{3.1.2}$$

(voir Jeffreys, 1948, section 3.0). En combinant ceci avec la fonction de vraisemblance donnée par l'équation (3.1.1), nous obtenons la loi conjointe *a posteriori* de  $\alpha$  et  $\beta$ ,

$$\pi(\alpha, \beta \mid \mathbf{y}) \propto e^{\alpha \sum_{i=1}^n y_i + \beta \sum_{i=1}^n x_i y_i} e^{-e^{\alpha} \sum_{i=1}^n e^{\beta x_i}}. \tag{3.1.3}$$

La loi marginale de  $\beta$  peut être obtenue en intégrant l'équation (3.1.3) par rapport à  $\alpha$ . Pour ce faire, nous posons  $u = e^{\alpha}$  et  $A = \sum_{i=1}^n e^{\beta x_i}$ . Nous obtenons alors,

$$\begin{aligned}
\pi(\beta \mid \mathbf{y}) &\propto e^{\beta \sum_{i=1}^n x_i y_i} \int_0^{+\infty} u^{\sum_{i=1}^n y_i - 1} e^{-Au} du \\
&= \frac{e^{\beta \sum_{i=1}^n x_i y_i}}{(\sum_{i=1}^n e^{\beta x_i})^{\sum_{i=1}^n y_i}}.
\end{aligned}$$

Quoique cette loi *a posteriori* ne soit pas facile à manipuler théoriquement, une tabulation exacte, pour n'importe quel cas particulier, peut facilement être construite, de laquelle la moyenne et la variance de  $\beta$  peuvent être évaluées et n'importe quelle inférence peut être faite.

Nous cherchons à estimer les  $\mu_i = e^{\alpha + \beta x_i}$ . Afin de faciliter le calcul des intégrales, nous faisons le changement de variables suivant:  $u = e^\alpha$  et  $v = e^\beta$ . Nous sommes donc intéressés à estimer les moments de  $u$  et  $v$ . De façon générale, nous cherchons à intégrer

$$I_{\{k,l\}} = \int_0^{+\infty} \int_0^{+\infty} u^k v^l f(\mathbf{y} | u, v) \pi(u, v) |J_{T^{-1}}(u, v)| dudv,$$

où  $k = 0, 1, 2$  et  $l = 0, 1, 2, x_1, \dots, x_n$ . Le déterminant du jacobien de la transformation inverse est donné par  $|J_{T^{-1}}| = \frac{1}{uv}$ . Alors, nous retrouvons

$$\begin{aligned} I_{\{k,l\}} &= \int_0^{+\infty} \int_0^{+\infty} \frac{u^{\sum y_i + k - 1} v^{\sum x_i y_i + l - 1} e^{-u \sum v^{x_i}}}{\prod_{i=1}^n y_i!} dudv \\ &= \frac{\Gamma(\sum_{i=1}^n y_i + k)}{\prod_{i=1}^n y_i!} \int_0^{+\infty} \frac{v^{\sum x_i y_i + l - 1}}{(\sum_{i=1}^n v^{x_i})^{\sum y_i + k}} dv. \end{aligned} \quad (3.1.4)$$

Nous pouvons maintenant écrire l'estimateur bayésien de  $\mu_i$  comme étant

$$\delta_{\mu_i}^\pi(\mathbf{y}) = I_{\{1, x_i\}} / I_{\{0, 0\}}.$$

Nous voyons, dans la littérature, que les gens estiment plutôt  $\alpha$  et  $\beta$  à la place des  $\mu_i$ . Dans ce chapitre, nous procéderons comme suit pour l'estimation des  $\mu_i$ . Premièrement, nous calculons les estimateurs bayésiens de  $u$  et  $v$ , qui sont donnés par  $\delta_u^\pi(\mathbf{y}) = I_{\{1, 0\}} / I_{\{0, 0\}}$  et  $\delta_v^\pi(\mathbf{y}) = I_{\{0, 1\}} / I_{\{0, 0\}}$ . Afin de retrouver les estimateurs de  $\alpha$  et  $\beta$ , nous utilisons le théorème suivant.

**Théorème 3.1.** *Soit la transformation  $Y = g(X)$ . Alors,*

$$\mathbb{E}[Y] = \mu_y \approx g(\mu_x) + \frac{1}{2} \sigma_x^2 g''(\mu_x)$$

et

$$\text{Var}(Y) = \sigma_y^2 \approx \sigma_x^2 [g'(\mu_x)]^2.$$

La perte *a posteriori* est définie par

$$\begin{aligned}\rho(\pi, \delta \mid \mathbf{y}) &= \mathbb{E}^{\pi(\theta \mid \mathbf{y})} [L(\theta, \delta)] \\ &= \int_{\Theta} L(\theta, \delta) \pi(\theta \mid \mathbf{y}) d\theta.\end{aligned}$$

Pour une perte quadratique, la perte *a posteriori* est donc égale à la variance *a posteriori*. Par le théorème précédent, nous définissons des estimateurs pseudo-bayésiens pour  $\alpha$  et  $\beta$  par

$$\begin{aligned}\delta_{\alpha}^{\pi}(\mathbf{y}) &= \log \left( \frac{I_{\{1,0\}}}{I_{\{0,0\}}} \right) - \frac{1}{2} \left( \frac{I_{\{0,0\}}}{I_{\{1,0\}}^2} \right) \left( I_{\{2,0\}} - \frac{I_{\{1,0\}}^2}{I_{\{0,0\}}} \right), \\ \delta_{\beta}^{\pi}(\mathbf{y}) &= \log \left( \frac{I_{\{0,1\}}}{I_{\{0,0\}}} \right) - \frac{1}{2} \left( \frac{I_{\{0,0\}}}{I_{\{0,1\}}^2} \right) \left( I_{\{0,2\}} - \frac{I_{\{0,1\}}^2}{I_{\{0,0\}}} \right).\end{aligned}$$

Les pertes *a posteriori* correspondantes sont données par

$$\begin{aligned}\rho(\pi, \delta_{\alpha}^{\pi} \mid \mathbf{y}) &= \left( \frac{I_{\{2,0\}}}{I_{\{1,0\}}} - \left( \frac{I_{\{1,0\}}}{I_{\{0,0\}}} \right)^2 \right) \left( \frac{I_{\{0,0\}}}{I_{\{1,0\}}} \right)^2, \\ \rho(\pi, \delta_{\beta}^{\pi} \mid \mathbf{y}) &= \left( \frac{I_{\{0,2\}}}{I_{\{0,1\}}} - \left( \frac{I_{\{0,1\}}}{I_{\{0,0\}}} \right)^2 \right) \left( \frac{I_{\{0,0\}}}{I_{\{1,0\}}} \right)^2.\end{aligned}$$

Dans les deux prochaines sections, nous utilisons la méthode de Naylor-Smith et l'approximation de Laplace, afin de calculer les estimateurs de Bayes.

### 3.1.1. Méthode de Naylor-Smith

Nous cherchons à calculer les estimateurs de Bayes pour  $\alpha$  et  $\beta$ , qui sont donnés par les espérances *a posteriori*. Puisque nous connaissons déjà la densité *a posteriori* de  $\beta$ , nous développons en premier lieu, l'estimateur de  $\beta$ , pour ensuite exprimer celui de  $\alpha$  en fonction de ce dernier. Dans Naylor et Smith (1982), nous voyons que les intégrales de la forme  $\int_{-\infty}^{+\infty} e^{-t^2} f(t) dt$ , peuvent être approximées par la formule de type gaussien  $\sum_{i=1}^m \omega_i f(t_i)$ , où  $\omega_i = \frac{2^{m-1} m! \sqrt{\pi}}{m^2 [H_{m-1}(t_i)]^2}$  sont les poids

d'Hermite d'ordre  $m$  et  $t_i$  est le  $i$ -ième zéro du polynôme d'Hermite de degré  $m$ , noté  $H_m(t)$  et qui satisfait la relation de récurrence suivante:  $H_{m+1}(t) = 2tH_m(t) - 2mH_{m-1}(t)$ , avec  $H_0 = 1$  et  $H_1 = 2t$ . L'erreur d'approximation est donnée par  $\frac{m! \sqrt{\pi}}{2^m (2m)!} f^{(2m)}(\xi)$ , pour un certain  $\xi$ , alors si  $f(t)$  est en réalité un polynôme de degré  $2m - 1$ , le reste sera de zéro et l'approximation sera exacte. Nous voulons utiliser la méthode d'intégration numérique gaussienne pour intégrer une fonction  $g(t)$  quelconque.

**Théorème 3.2.** *L'approximation par la méthode numérique gaussienne pour des intégrales de la forme  $\int_{-\infty}^{+\infty} g(t)dt$  est donnée par*

$$\sum_{i=1}^m m_i g(r_i), \quad (3.1.5)$$

où  $m_i = \omega_i e^{t_i^2} \sqrt{2\sigma^2}$  et  $r_i = \mu + \sqrt{2\sigma^2} t_i$ .

**Démonstration 3.1.** Posons  $h(t)$  telle que

$$g(t) = \frac{h(t)}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

En posant  $r = \frac{t-\mu}{\sqrt{2\sigma^2}}$ , nous retrouvons

$$\int_{-\infty}^{+\infty} g(t)dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} h(\mu + \sqrt{2\sigma^2}r) e^{-r^2} dr.$$

En utilisant la formule de Gauss-Hermite, nous obtenons

$$\begin{aligned} \int_{-\infty}^{+\infty} g(t)dt &\simeq \sum_{i=1}^m \frac{\omega_i}{\sqrt{\pi}} h(\mu + \sqrt{2\sigma^2}t_i) \\ &= \sum_{i=1}^m \omega_i \sqrt{2\sigma^2} e^{t_i^2} g(\mu + \sqrt{2\sigma^2}t_i). \end{aligned}$$

En posant  $m_i = \omega_i e^{t_i^2} \sqrt{2\sigma^2}$  et  $r_i = \mu + \sqrt{2\sigma^2}t_i$ , nous obtenons le résultat voulu.

□

Plus la fonction  $h(t)$  qui sert à approcher  $g(t)$  se comportera comme un polynôme, plus précise sera l'approximation. Afin d'appliquer cette méthode, nous devons aussi calculer la moyenne  $\mu$  et la variance  $\sigma^2$  de la variable aléatoire  $T$  de loi  $g(t)$ , c'est-à-dire

$$\mu = \frac{\int_{-\infty}^{+\infty} t g(t) dt}{\int_{-\infty}^{+\infty} g(t) dt},$$

$$\sigma^2 = \frac{\int_{-\infty}^{+\infty} t^2 g(t) dt}{\int_{-\infty}^{+\infty} g(t) dt} - \mu^2.$$

Puisque ces quantités sont presque toujours inconnues, nous les estimons à l'aide de la méthode d'intégration gaussienne (voir théorème 3.2). Les formules itératives pour calculer ces quantités sont données par

$$\mu_{j+1} = \frac{\sum_{i_j=1}^{n_j^*} r_{i_j} m_{i_j} g(r_{i_j})}{\sum_{i_j=1}^{n_j^*} m_{i_j} g(r_{i_j})},$$

$$\sigma_{j+1}^2 = \frac{\sum_{i_j=1}^{n_j^*} r_{i_j}^2 m_{i_j} g(r_{i_j})}{\sum_{i_j=1}^{n_j^*} m_{i_j} g(r_{i_j})} - \mu_j^2,$$

où  $m_{i_j} = \omega_{i_j} e^{t_{i_j}^2} \sqrt{2\sigma_j^2}$  et  $r_{i_j} = \mu_j + \sqrt{2\sigma_j^2} t_{i_j}$ . Nous construisons une suite d'itérations successives  $(\mu_j, \sigma_j^2)$  pour  $j = 0, 1, \dots$ , en prenant  $\mu_0 = 0$  et  $\sigma_0^2 = 0,01$  comme point de départ. Nous disons que la convergence est atteinte lorsque la différence relative entre deux itérations est plus petite qu'une certaine précision. Lorsque nous avons trouvé le point de convergence  $(\mu_*, \sigma_*^2)$ , nous le remplaçons dans l'équation (3.1.5) afin de calculer l'intégrale recherchée. Puisque cette méthode dépend du nombre de termes utilisés dans la série, nous débutons avec peu de termes tout en augmentant le nombre, jusqu'à ce que la stabilité soit atteinte.

Nous utilisons donc cette méthode afin de calculer les moments de  $v$ , pour la loi de densité  $g(v)$ , qui correspond à la densité *a posteriori*

$$\pi(v \mid \mathbf{y}) \propto \frac{v^{\sum_{i=1}^n x_i y_i - 1}}{(\sum_{i=1}^n v^{x_i})^{\sum_{i=1}^n y_i}}.$$

Puisque  $v$  varie de 0 à l'infini, nous devons refaire un changement de variable afin d'appliquer la méthode. Après avoir trouvé l'estimateur de  $v$ , nous pouvons maintenant calculer celui de  $u$  à l'aide de l'équation (3.1.4) en posant  $l = 0$  et  $k = 1$  dans cette équation. Ensuite, pour retrouver les estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$ , nous utilisons le théorème 3.1. Nous appliquons cette méthode à un exemple numérique dans la section 3.1.3.

### 3.1.2. Approximation analytique de Laplace

Lorsque la fonction à intégrer est suffisamment régulière, la méthode de Laplace, qui n'est valide qu'asymptotiquement, permet d'éviter les simulations. Nous voulons évaluer l'espérance *a posteriori*

$$\mathbb{E}^{\pi(\theta|\mathbf{y})} [g(\theta)] = \frac{\int_{\Theta} g(\theta) f(\mathbf{y} \mid \theta) \pi(\theta) d\theta}{\int_{\Theta} f(\mathbf{y} \mid \theta) \pi(\theta) d\theta},$$

que nous pouvons écrire sous la forme

$$\mathbb{E}^{\pi(\theta|\mathbf{y})} [g(\theta)] = \frac{\int_{\Theta} b_N(\theta) \exp\{-nh_N(\theta)\} d\theta}{\int_{\Theta} b_D(\theta) \exp\{-nh_D(\theta)\} d\theta}, \quad (3.1.6)$$

où  $n$  correspond en général à la taille échantillonnale. Nous disons que l'équation (3.1.6) est écrite sous forme standard si  $h_N(\theta) = h_D(\theta)$  ce qui implique que  $b_N = gb_D$ , et sous forme totalement exponentielle si  $b_N(\theta) = b_D(\theta)$  et donc  $h_N = h_D - (1/n) \log g$ . Le développement de Laplace nous donne une approximation analytique de l'espérance *a posteriori* d'une fonction quelconque des paramètres.

**Théorème 3.3.** *Le développement de Laplace est donné par*

$$\int_{-\infty}^{+\infty} b(\theta) e^{-nh(\theta)} d\theta = \sqrt{2\pi\sigma^2} e^{-n\hat{h}} \left\{ \hat{b} + \frac{1}{2n} \left[ \sigma^2 \hat{b}'' - \sigma^4 \hat{b}' \hat{h}''' \right. \right. \\ \left. \left. + \frac{5}{12} \hat{b}(\hat{h}''')^2 \sigma^6 - \frac{1}{4} \hat{b} \hat{h}^{(4)} \sigma^4 \right] \right\} + \mathcal{O}(n^{-2}), \quad (3.1.7)$$

où  $b$ ,  $h$  et leurs dérivées sont évaluées en  $\hat{\theta}$ , argument du maximum et  $\sigma^2 = [h''(\hat{\theta})]^{-1}$ .

La démonstration est donnée dans Tierney, Kass et Kadane (1989). Notons que pour appliquer le développement de Laplace, il est nécessaire que  $\Theta = \mathbb{R}$  puisque nous approximations à l'aide d'une loi normale. Si nous supposons que  $h_N$  et  $h_D$  vérifient  $\hat{h}_N - \hat{h}_D = \mathcal{O}(n^{-1}), \dots, \hat{h}_N^{(4)} - \hat{h}_D^{(4)} = \mathcal{O}(n^{-1})$  (ce qui est automatiquement vérifié pour la forme standard), nous avons le théorème suivant.

**Théorème 3.4.** *L'approximation de Laplace de  $\mathbb{E}^{\pi(\theta|\mathbf{y})} [g(\theta)]$  est donnée par*

$$\mathbb{E}^{\pi(\theta|\mathbf{y})} [g(\theta)] = \hat{g} + \frac{\hat{\sigma}^2 \hat{b}'_D \hat{g}'}{n \hat{b}_D} + \frac{\hat{\sigma}^2 \hat{g}''}{2n} - \frac{\hat{\sigma}^4 \hat{h}''' \hat{g}'}{2n} + \mathcal{O}(n^{-2}),$$

pour l'écriture sous forme standard et elle sera donnée par

$$\mathbb{E}^{\pi(\theta|\mathbf{y})} [g(\theta)] = \frac{\hat{b}_N}{\hat{b}_D} \sqrt{\frac{\hat{\sigma}_N^2}{\hat{\sigma}_D^2}} e^{-n(\hat{h}_N - \hat{h}_D)} + \mathcal{O}(n^{-2}),$$

pour la forme totalement exponentielle si  $g$  est positive et  $g(\hat{\theta}_D)$  est uniformément bornée inférieurement en  $n$ .

La densité *a priori* sur  $\alpha$  et  $\beta$  est donnée par l'équation (3.1.2). Sous la perte quadratique, nous savons que l'estimateur de Bayes  $\delta^\pi$  est donné par l'espérance *a posteriori*. En faisant le même changement de variables qu'à la section précédente,

nous retrouvons les estimateurs de Bayes de  $u$  et  $v$ .

$$\delta^\pi(\mathbf{y}) = \frac{\Gamma\left(\sum y_i + k\right) \int_0^{+\infty} \frac{v^{\sum x_i y_i + l - 1}}{(\sum v^{x_i})^{\sum y_i + k}} dv}{\Gamma\left(\sum y_i\right) \int_0^{+\infty} \frac{v^{\sum x_i y_i - 1}}{(\sum v^{x_i})^{\sum y_i}} dv},$$

où  $l = 0, 1$  et  $v = 0, x_1, \dots, x_n$ .

Nous cherchons à approximer, à l'aide de l'approximation analytique de Laplace, la forme suivante:

$$J_{kl} = \frac{\int_0^{+\infty} e^{-nh_N(v)} dv}{\int_0^{+\infty} e^{-nh_D(v)} dv},$$

où

$$h_N(v) = \frac{1}{n} \left( \sum_{i=1}^n y_i + k \right) \log \left( \sum_{i=1}^n v^{x_i} \right) - \frac{1}{n} \left( \sum_{i=1}^n x_i y_i + l - 1 \right) \log v,$$

$$h_D(v) = \bar{y} \log \left( \sum_{i=1}^n v^{x_i} \right) - \frac{1}{n} \left( \sum_{i=1}^n x_i y_i - 1 \right) \log v.$$

Afin de simplifier la notation, nous posons

$$A_l = \frac{1}{n} \left( \sum_{i=1}^n x_i y_i + l - 1 \right),$$

$$B_k = \frac{1}{n} \left( \sum_{i=1}^n y_i + k \right).$$

Alors, nous pouvons écrire

$$h_N(v) = B_k \log \left( \sum_{i=1}^n v^{x_i} \right) - A_l \log v,$$

$$h_D(v) = B_0 \log \left( \sum_{i=1}^n v^{x_i} \right) - A_0 \log v.$$



Après avoir calculé les dérivées premières et secondes de  $h_N$  et  $h_D$ , le développement de Laplace devient

$$J_{kl} = \left( \frac{\hat{v}_N^2 \left[ \sum_{i=1}^n x_i y_i - 1 + \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n \hat{v}_D^{x_i}} \left( \sum_{i=1}^n x_i (x_i - 1) \hat{v}_D^{x_i} - \frac{(\sum_{i=1}^n x_i \hat{v}_D^{x_i})^2}{\sum_{i=1}^n \hat{v}_D^{x_i}} \right) \right]}{\hat{v}_D^2 \left[ \sum_{i=1}^n x_i y_i + l - 1 + \frac{(\sum_{i=1}^n y_i + k)}{\sum_{i=1}^n \hat{v}_N^{x_i}} \left( \sum_{i=1}^n x_i (x_i - 1) \hat{v}_N^{x_i} - \frac{(\sum_{i=1}^n x_i \hat{v}_N^{x_i})^2}{\sum_{i=1}^n \hat{v}_N^{x_i}} \right) \right]} \right)^{1/2} \\ \times e^{-n(\hat{h}_N - \hat{h}_D)} + \mathcal{O}(n^{-2}),$$

où  $\hat{v}_N$  est tel que  $\sum_{i=1}^n (B_k x_i - A_l) \hat{v}_N^{x_i} = 0$  et  $\hat{v}_D$  tel que  $\sum_{i=1}^n (B_0 x_i - A_0) \hat{v}_D^{x_i} = 0$ .

Nous procédons comme à la section précédente, c'est-à-dire que nous calculons l'estimateur de Bayes des  $\mu_i$  ainsi que les estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$  et les pertes *a posteriori* correspondantes. Nous retrouvons les estimations suivantes:

$$\delta_{\mu_i} = J_{1x_i} \sum_{i=1}^n y_i, \\ \delta_{\alpha}^{\pi}(\mathbf{y}) = \log \left( J_{10} \sum_{i=1}^n y_i \right) - \frac{1}{2} \left[ \frac{(\sum_{i=1}^n y_i) (\sum_{i=1}^n y_i + 1) J_{20} - (J_{10} \sum_{i=1}^n y_i)^2}{[J_{10} (\sum_{i=1}^n y_i)]^2} \right], \\ \delta_{\beta}^{\pi}(\mathbf{y}) = \log (J_{01}) - \frac{1}{2} \left[ \frac{J_{02} - (J_{01})^2}{[(J_{01})]^2} \right], \\ \rho(\pi, \delta_{\alpha}^{\pi} | \mathbf{y}) = \frac{(\sum_{i=1}^n y_i) (\sum_{i=1}^n y_i + 1) J_{20} - (J_{10} \sum_{i=1}^n y_i)^2}{[J_{10} (\sum_{i=1}^n y_i)]^2} \text{ et} \\ \rho(\pi, \delta_{\beta}^{\pi} | \mathbf{y}) = \frac{J_{02} - (J_{01})^2}{[(J_{01})]^2}.$$

Dans la section suivante, nous appliquons les notions vues dans les sections 3.1.1 et 3.1.2 à un exemple.

### 3.1.3. Exemple numérique

Afin de comparer les résultats des chapitres précédents avec ceux du présent chapitre, nous donnons les estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$  obtenus par la méthode de Naylor-Smith et l'approximation de Laplace pour le même exemple. Nous retrouvons ces estimateurs dans le tableau 3.1.1, ainsi que les pertes *a posteriori* correspondantes. Nous comparons les estimateurs bayésiens  $\delta_{\mu_i}^\pi = \mathbb{E}[uv^{x_i}]$  aux estimateurs pseudo-bayésiens  $\hat{\mu}_i = e^{\hat{\alpha} + \hat{\beta}x_i}$ , où  $\hat{\alpha} = \delta_\alpha^\pi$  et  $\hat{\beta} = \delta_\beta^\pi$  dans le tableau 3.1.2 pour les deux méthodes.

TABLEAU 3.1.1. Estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$  ainsi que des pertes *a posteriori* correspondantes

	$\delta_\alpha^\pi$	$\delta_\beta^\pi$	$\rho_\alpha^\pi$	$\rho_\beta^\pi$
Méthode de Naylor-Smith	4,247	-0,206	$1,910 \times 10^{-2}$	$1,702 \times 10^{-3}$
Approximation de Laplace	4,247	-0,206	$1,910 \times 10^{-2}$	$1,701 \times 10^{-3}$

TABLEAU 3.1.2. Ensembles de triplets, nés en Norvège, ainsi que les estimations des  $\mu_i$ 

Période	1911-15	1916-20	1921-25	1926-30	1931-35	1936-40
Séries de nombre $x_i$ des ensembles	1	2	3	4	5	6
Ensembles de triplets $y_i$	52	52	40	30	24	20
$\hat{\mu}_i$ par la méthode de Naylor-Smith	56,884	46,318	37,715	30,710	25,006	20,361
$\delta_{\mu_i}^{\pi}$ par la méthode de Naylor-Smith	57,194	46,459	37,803	30,812	25,157	20,574
$\hat{\mu}_i$ par l'approximation de Laplace	56,884	46,318	37,715	30,710	25,006	20,361
$\delta_{\mu_i}^{\pi}$ par l'approximation de Laplace	57,195	46,460	37,804	30,812	25,157	20,574

### 3.2. LOIS CONJUGUÉES

L'utilisation de lois conjuguées est attrayante puisqu'elle permet de manipuler explicitement les lois *a posteriori*. Raiffa et Schlaifer (1961) justifient l'utilisation des lois conjuguées par un "principe d'invariance". Ce principe dit que la modification apportée par les observations doit rester de dimension finie, c'est-à-dire

que lorsque l'observation de  $x \sim f(x | \theta)$  vient modifier  $\pi(\theta)$  en  $\pi(\theta | x)$ , l'information apportée par  $x$  sur  $\theta$  est limitée, donc ne doit pas conduire à une remise en cause de la forme de  $\pi(\theta)$  mais seulement de ses paramètres (Robert, 1992, section 3.2). Dans cette dernière section, nous déterminons une famille de lois *a priori* conjuguées, pour ensuite trouver les estimateurs de Bayes sous la fonction de perte quadratique. De plus, nous retrouvons la cote de Bayes associée au test fait sur le paramètre  $\beta$ , ainsi que le comportement de celle-ci en présence de valeurs aberrantes.

### 3.2.1. Détermination de la loi *a priori* conjuguée

Une famille de lois conjuguées est déterminée uniquement par la fonction de vraisemblance. Nous avons, pour le modèle log-linéaire simple, la fonction de vraisemblance suivante:

$$f(\mathbf{y} | \alpha, \beta, \mathbf{x}) \propto e^{\alpha \sum y_i} e^{\beta \sum x_i y_i} e^{-e^\alpha \sum e^{\beta x_i}}.$$

Donc, la loi *a priori* conjuguée sera de la forme

$$\pi(\alpha, \beta) \propto e^{a\alpha} e^{b\beta} e^{-ce^\alpha e^{d\beta}} e^{-fe^\alpha}.$$

Posons  $u = e^\alpha$  et  $v = e^\beta$ . Nous avons déjà vu que le jacobien de la transformation inverse est donné par  $1/(uv)$ . Nous cherchons à intégrer

$$L_{kl} = \int_0^{+\infty} \int_0^{+\infty} u^{k+a-1} v^{l+b-1} e^{-(cv^d+f)u} dudv.$$

Nous intégrons premièrement par rapport à  $u$  et nous retrouvons

$$\begin{aligned} L_{kl} &= \int_0^{+\infty} \frac{v^{l+b-1} \Gamma(k+a)}{(cv^d+f)^{k+a}} dv \\ &= \frac{\Gamma(k+a)}{f^{k+a}} \int_0^{+\infty} \frac{v^{(l+b)-1}}{(\frac{c}{f}v^d+1)^{k+a}} dv. \end{aligned}$$

Or, si  $p > 0$  et  $0 < \mu < Up$ , nous avons

$$\int_0^{+\infty} x^{\mu-1} (1 + \beta x^p)^{-U} dx = \frac{1}{p} \frac{1}{\beta^{\mu/p}} Be\left(\frac{\mu}{p}, U - \frac{\mu}{p}\right)$$

(Gradshteyn et Ryzhik, 1980, section 3.251). Nous obtenons donc,

$$L_{kl} = \frac{1}{d f^{k+a}} \left(\frac{f}{c}\right)^{\frac{l+b}{d}} \Gamma\left(\frac{l+b}{d}\right) \Gamma\left(k+a - \left(\frac{l+b}{d}\right)\right), \quad (3.2.1)$$

si  $d > 0$  et  $0 < l+b < d(k+a)$ .

Nous pouvons maintenant calculer les moments *a priori* de  $u$  et  $v$ . Ces derniers sont donnés par

$$\mathbb{E}[u] = \frac{a - b/d}{f},$$

$$Var(u) = \frac{a - b/d}{f^2}, \quad \text{où } d > 0, 0 < b < ad,$$

$$\mathbb{E}[v] = \left(\frac{f}{c}\right)^{1/d} \frac{\Gamma\left(\frac{b+1}{d}\right) \Gamma\left(a - \frac{b+1}{d}\right)}{\Gamma\left(\frac{b}{d}\right) \Gamma\left(a - b/d\right)},$$

$$Var(v) = \frac{(f/c)^{2/d}}{\Gamma\left(\frac{b}{d}\right) \Gamma\left(a - b/d\right)} \left\{ \Gamma\left(\frac{b+2}{d}\right) \Gamma\left(a - \frac{b+2}{d}\right) - \frac{[\Gamma\left(\frac{b+1}{d}\right) \Gamma\left(a - \frac{b+1}{d}\right)]^2}{\Gamma\left(\frac{b}{d}\right) \Gamma\left(a - b/d\right)} \right\},$$

où  $d > 0$ ,  $f > 0$ ,  $0 < b < ad - 1$ .

Nous cherchons une densité *a priori* sur  $u$  et  $v$  qui soit la plus vague possible. Pour ce faire, nous cherchons à rendre les variances très grandes. En posant  $k = 0$  et  $l = 0$  dans l'équation (3.2.1), nous retrouvons la constante de normalisation  $C = \frac{1}{d} f^{N-a} \Gamma(N) \Gamma(m/d)$ , où  $N = b/d$ ,  $c = 1$  et  $a = N + m/d$ . Pour faciliter notre problème, c'est-à-dire maximiser la variance *a priori* de  $u$  et  $v$ , nous fixons quelques inconnues, de sorte à obtenir ce que nous cherchons. Nous posons

$$N = b/d = 1 \Rightarrow b = d,$$

$$a = N + m/d = b/d + \delta = \delta + 1.$$

Alors, nous pouvons réécrire les variances de  $u$  et  $v$  comme suit:

$$\begin{aligned} \text{Var}(u) &= \frac{a - b/d}{f^2} = \frac{a - N}{f^2} = \frac{m}{df^2} = \delta/f^2, \\ \text{Var}(v) &= \frac{f^{2/d}}{d\Gamma(\delta)} \left\{ 2\Gamma(2/d)\Gamma(\delta - 2/d) - \frac{[\Gamma(1/d)\Gamma(\delta - 1/d)]^2}{d\Gamma(\delta)} \right\}, \end{aligned}$$

où  $d \leq \delta/2$ . Plus  $d$  s'approche de  $\delta/2$ , plus la loi *a priori* sera vague. Nous prenons donc  $d = \delta/2$  et nous obtenons la loi *a priori* suivante:

$$\pi(\alpha, \beta) \propto e^{(\delta+1)\alpha} e^{\frac{\delta}{2}\beta} e^{-e^\alpha e^{\frac{\delta}{2}\beta}} e^{-fe^\alpha} \quad (3.2.2)$$

et la loi *a posteriori* est donnée par

$$\pi(\alpha, \beta \mid \mathbf{y}) \propto e^{(\delta+1+\sum y_i)\alpha} e^{(\frac{\delta}{2}+\sum x_i y_i)\beta} e^{-e^\alpha (e^{\frac{\delta}{2}\beta} + f + \sum e^{\beta x_i})}.$$

Il nous reste deux inconnues à fixer, soient  $\delta$  et  $f$ . Nous remarquons que pour retrouver la densité *a posteriori* donnée par l'équation (3.1.3), nous n'avons qu'à substituer

$$\begin{aligned} \sum_{i=1}^n y_i &\text{ par } \delta + \sum_{i=0}^n y_i, \\ \sum_{i=1}^n x_i y_i &\text{ par } \sum_{i=0}^n x_i y_i \\ \text{et } \sum_{i=1}^n v^{x_i} &\text{ par } f + \sum_{i=0}^n v^{x_i} \end{aligned}$$

dans l'équations précédente, où  $x_0 = \delta/2$  et  $y_0 = 1$ . Tout d'abord, nous déterminons  $\delta$  en ajustant la droite de régression du logarithme des observations et nous cherchons la valeur de  $x_0$  pour laquelle  $y_0 = 1$ . Nous choisissons  $\delta$  égale à  $2x_0$ . Selon les propriétés de la loi Gamma, si  $\frac{\Gamma(4/\delta)\Gamma(\delta-4/\delta)}{\delta\Gamma(\delta)} < 0$ , nous choisissons  $\delta \pm 1$ , de façon à ce que la variance de  $v$  soit positive. Pour fixer la valeur de  $f$ , nous

posons l'espérance *a priori* de  $u$  égale à l'estimateur des moindres carrés. Nous obtenons alors l'équation suivante.

$$f = \frac{\delta}{e^{\hat{\alpha}_{MC}}}.$$

Nous pouvons maintenant trouver les estimateurs de Bayes et faire toute l'étude que nous avons vue dans la section précédente.

### 3.2.2. Forme des estimateurs de Bayes par l'approximation analytique de Laplace

Pour calculer les estimateurs de Bayes ainsi que leurs pertes *a posteriori* correspondantes, nous n'avons qu'à faire les mêmes substitutions qu'à la section précédente dans les équations de la section 3.1.2. Les estimateurs de  $u$  et  $v$  ont la forme suivante:

$$\delta^\pi(\mathbf{y}) = \frac{\Gamma\left(\delta + \sum_{i=0}^n y_i + k\right) \int_0^{+\infty} \frac{v^{\sum_{i=0}^n x_i y_i + l - 1}}{(f + \sum_{i=0}^n v^{x_i})^{\delta + \sum_{i=0}^n y_i + k}} dv}{\Gamma\left(\delta + \sum_{i=0}^n y_i\right) \int_0^{+\infty} \frac{v^{\sum_{i=0}^n x_i y_i - 1}}{(f + \sum_{i=0}^n v^{x_i})^{\delta + \sum_{i=0}^n y_i}} dv},$$

où  $l = 0, 1$  et  $v = 0, x_1, \dots, x_n$ . Afin de simplifier la notation, nous posons

$$A_l = \frac{1}{n} \left( \sum_{i=0}^n x_i y_i + l - 1 \right), \quad B_k = \frac{1}{n} \left( \sum_{i=0}^n y_i + \delta + k \right),$$

$$E_N = \sum_{i=0}^n \hat{v}_N^{x_i}, \quad E_D = \sum_{i=0}^n \hat{v}_D^{x_i},$$

$$F_N = \sum_{i=0}^n x_i \hat{v}_N^{x_i}, \quad F_D = \sum_{i=0}^n x_i \hat{v}_D^{x_i},$$

$$G_N = \sum_{i=0}^n x_i (x_i - 1) \hat{v}_N^{x_i}, \quad G_D = \sum_{i=0}^n x_i (x_i - 1) \hat{v}_D^{x_i}.$$

Nous obtenons alors,

$$J_{kl} = \left( \frac{\hat{v}_N^2 \left[ A_0 + \frac{B_0}{f+E_D} \left( G_D - \frac{F_D^2}{f+E_D} \right) \right]}{\hat{v}_D^2 \left[ A_l + \frac{B_k}{f+E_N} \left( G_N - \frac{F_N^2}{f+E_N} \right) \right]} \right)^{1/2} \times e^{-n(\hat{h}_N - \hat{h}_D)} + \mathcal{O}(n^{-2}). \quad (3.2.3)$$

L'estimateur de Bayes des  $\mu_i$  est donné par  $\delta_{\mu_i}^\pi = nB_0J_{1x_i}$ , où  $J_{1x_i}$  est donnée par l'équation (3.2.3), et nous retrouvons les estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$ , ainsi que leurs pertes *a posteriori*, dans les deux théorèmes suivants.

**Théorème 3.5.** *L'estimateur pseudo-bayésien pour  $\alpha$  est donné par*

$$\delta_\alpha^\pi(\mathbf{y}) = \log(nDJ_{10}) - \frac{1}{2} \frac{\rho(\pi, \delta_u^\pi | \mathbf{y})}{[nB_0J_{10}]^2},$$

où  $\rho(\pi, \delta_u^\pi | \mathbf{y}) = (nB_0)(nB_0 + 1)J_{20} - (nB_0J_{10})^2$ ,  $J_{10}$  et  $J_{20}$  sont données par l'équation (3.2.3) et où

$$\begin{aligned} \hat{v}_N \text{ satisfait } \sum_{i=0}^n (B_1x_i - A_0) \hat{v}_N^{x_i} &= A_0f, \\ \hat{v}_D \text{ satisfait } \sum_{i=0}^n (B_0x_i - A_0) \hat{v}_D^{x_i} &= A_0f, \\ \hat{h}_N &= B_1 \log \left( f + \sum_{i=0}^n \hat{v}_N^{x_i} \right) - A_0 \log \hat{v}_N, \\ \hat{h}_D &= B_0 \log \left( f + \sum_{i=0}^n \hat{v}_D^{x_i} \right) - A_0 \log \hat{v}_D. \end{aligned}$$

**Théorème 3.6.** *L'estimateur pseudo-bayésien pour  $\beta$  est donné par*

$$\delta_\beta^\pi(\mathbf{y}) = \log(J_{01}) - \frac{1}{2} \frac{\rho(\pi, \delta_v^\pi | \mathbf{y})}{[J_{01}]^2},$$



où  $\rho(\pi, \delta_v^\pi | \mathbf{y}) = J_{02} - (J_{01})^2$ ,  $J_{01}$  et  $J_{02}$  sont données par l'équation (3.2.3) et où

$$\hat{v}_N \text{ satisfait } \sum_{i=0}^n (B_0 x_i - A_1) \hat{v}_N^{x_i} = A_1 f,$$

$$\hat{v}_D \text{ satisfait } \sum_{i=0}^n (B_0 x_i - A_0) \hat{v}_D^{x_i} = A_0 f,$$

$$\hat{h}_N = B_0 \log \left( f + \sum_{i=0}^n \hat{v}_N^{x_i} \right) - A_1 \log \hat{v}_N,$$

$$\hat{h}_D = B_0 \log \left( f + \sum_{i=0}^n \hat{v}_D^{x_i} \right) - A_0 \log \hat{v}_D.$$

### 3.2.3. Test sur $\beta$

Comme pour le modèle approximé par la distribution normale, nous voulons savoir s'il y a une tendance apparente. Puisque pour utiliser la cote de Bayes, donnée par l'équation (2.2.2), nous avons besoin d'une loi *a priori* qui soit propre, nous prenons la loi *a priori* conjuguée que nous avons trouvée dans la section 3.2.1. Avec cette densité, nous pouvons faire le test sur  $\beta$  pour le modèle simple. Encore une fois, nous voulons tester les hypothèses suivantes:

$$H_0 : \beta = 0 \text{ vs } H_1 : \beta \neq 0.$$

Nous avons comme loi *a priori*, la loi conjuguée donnée par l'équation (3.2.2). Afin de calculer la cote de Bayes, nous nous référons au théorème 2.6, ainsi qu'aux mêmes variables qu'à la section précédente et nous obtenons le résultat suivant:

#### **Théorème 3.7.**

$$B^\pi(\mathbf{y}) = \frac{1}{I_\beta (f + n + 1)^{nB_0}},$$

où

$$I_\beta = \sqrt{2\pi\hat{\sigma}^2} e^{-n\hat{h}} \left\{ 1 + \frac{1}{2n} \left[ \frac{5}{12} (\hat{h}''')^2 \hat{\sigma}^6 - \frac{1}{4} \hat{h}^{(4)} \hat{\sigma}^4 \right] \right\} + \mathcal{O}(n^{-2}), \quad (3.2.4)$$

où  $\hat{\sigma}^2 = [\hat{h}''(\beta)]^{-1}$ ,  $\hat{h}''$ ,  $\hat{h}'''$  et  $\hat{h}^{(4)}$  sont évaluées en  $\hat{\beta}$ , où  $\hat{\beta}$  est la valeur de  $\beta$  qui maximise  $h(\beta)$ , c'est-à-dire que  $\hat{\beta}$  est tel que  $\sum_{i=0}^n (x_i \bar{y}_* - \bar{z}_*) e^{\hat{\beta} x_i} = \bar{z}_* f$ ,  $B_0 = \frac{1}{n} \sum_{i=0}^{n+1} y_i$  et où  $\bar{y}_* = \frac{1}{n} \sum_{i=0}^{n+1} y_i$  et  $\bar{z}_* = \frac{1}{n} \sum_{i=0}^n x_i y_i$ .

**Démonstration 3.2.** En posant  $u = e^\alpha$ , nous retrouvons

$$\begin{aligned} I_0 &= \int_{-\infty}^{\infty} f(\mathbf{y} \mid \alpha, 0) g_0(\alpha, 0) d\alpha \\ &= \int_0^{\infty} \frac{u^{(\delta+1+\sum_{i=1}^n y_i)-1} e^{-u(f+n+1)}}{\prod y_i!} du \\ &= \frac{\Gamma(nB_0)}{\prod y_i! (f+n+1)^{nD}} \end{aligned}$$

et

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{y} \mid \alpha, \beta) g_1(\alpha, \beta) d\alpha d\beta \\ &= \int_{-\infty}^{\infty} \frac{e^{(\frac{\delta}{2} + \sum_{i=1}^n x_i y_i)\beta}}{\prod y_i!} \int_0^{\infty} u^{(\delta+1+\sum_{i=1}^n y_i)-1} e^{-u(e^{\frac{\delta}{2}\beta} + f + \sum_{i=1}^n e^{\beta x_i})} du d\beta \\ &= \frac{\Gamma(nB_0)}{\prod y_i!} I_\beta, \end{aligned}$$

où  $g_0(\alpha, 0) = \frac{\pi(\alpha, \beta)}{\pi(\Theta_0)} \mathbb{I}_{\Theta_0}(\alpha, \beta)$ ,  $g_1(\alpha, \beta) = \frac{\pi(\alpha, \beta)}{\pi(\Theta_1)} \mathbb{I}_{\Theta_1}(\alpha, \beta)$ ,  $\Theta_0 = \mathbb{R} \times \{0\}$  et  $\Theta_1 = \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ .

Afin de faciliter la notation, nous posons  $y_{n+1} = \delta$ . Nous obtenons alors,

$$\begin{aligned} I_\beta &= \int_{-\infty}^{\infty} \frac{e^{\beta \sum_{i=0}^n x_i y_i}}{(f + \sum_{i=0}^n e^{\beta x_i})^{\sum_{i=0}^{n+1} y_i}} d\beta \\ &= \int_{-\infty}^{\infty} e^{-nh(\beta)} d\beta, \end{aligned}$$

avec

$$h(\beta) = \bar{y}_* \log \left( f + \sum_{i=0}^n e^{\beta x_i} \right) - \beta \bar{z}_*. \quad (3.2.5)$$

Nous utilisons l'approximation analytique de Laplace pour calculer  $I_\beta$  et nous retrouvons l'équation (3.2.4). Après quelques calculs, nous retrouvons les dérivées deuxième, troisième et quatrième suivantes:

$$\begin{aligned} h''(\beta) &= \bar{y}_* \left\{ \frac{\sum_{i=0}^n x_i^2 e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} - \left( \frac{\sum_{i=0}^n x_i e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} \right)^2 \right\} \\ &= \bar{y}_* K_2, \end{aligned} \quad (3.2.6)$$

$$\begin{aligned} h'''(\beta) &= \bar{y}_* \left\{ \frac{\sum_{i=0}^n x_i^3 e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} - 3 \frac{(\sum_{i=0}^n x_i^2 e^{\beta x_i}) (\sum_{i=0}^n x_i e^{\beta x_i})}{(f + \sum_{i=0}^n e^{\beta x_i})^2} \right. \\ &\quad \left. + 2 \left( \frac{\sum_{i=0}^n x_i e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} \right)^3 \right\} \\ &= \bar{y}_* K_3 \end{aligned} \quad (3.2.7)$$

$$\begin{aligned} \text{et } h^{(4)}(\beta) &= \bar{y}_* \left\{ \frac{\sum_{i=0}^n x_i^4 e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} - 4 \frac{(\sum_{i=0}^n x_i^3 e^{\beta x_i}) (\sum_{i=0}^n x_i e^{\beta x_i})}{(f + \sum_{i=0}^n e^{\beta x_i})^2} \right. \\ &\quad - 3 \left( \frac{\sum_{i=0}^n x_i^2 e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} \right)^2 + 12 \frac{(\sum_{i=0}^n x_i^2 e^{\beta x_i}) (\sum_{i=0}^n x_i e^{\beta x_i})^2}{(f + \sum_{i=0}^n e^{\beta x_i})^3} \\ &\quad \left. - 6 \left( \frac{\sum_{i=0}^n x_i e^{\beta x_i}}{f + \sum_{i=0}^n e^{\beta x_i}} \right)^4 \right\} \\ &= \bar{y}_* K_4. \end{aligned} \quad (3.2.8)$$

À noter que nous pouvons écrire  $\bar{y}_* = \left(\frac{\delta+1}{n}\right) + \bar{y}$ ,  $\bar{z}_* = \frac{\delta}{2n} + \bar{z}$  et  $\sum_{i=0}^n x_i e^{\beta x_i} = \frac{\delta}{2} e^{\frac{\delta}{2}\beta} + \sum_{i=1}^n x_i e^{\beta x_i}$ .  $\square$

### 3.2.4. Valeur aberrante

Nous voulons voir comment se comporte la cote de Bayes lorsqu'une des observations est très grande. Nous supposons que  $y_I \rightarrow +\infty$  et que  $x_I y_I \gg x_i y_i$ ,  $\forall i \neq I$ . Alors,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{y_I}{n} + \frac{(n-1)}{n} \bar{y}_{(n)}, \text{ où } \bar{y}_{(n)} = \frac{1}{n-1} \sum_{i \neq I}^n y_i$$

et

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{x_I y_I}{n} + \frac{(n-1)}{n} \bar{z}_{(n)}, \text{ où } \bar{z}_{(n)} = \frac{1}{n-1} \sum_{i \neq I}^n x_i y_i.$$

Dans la section précédente, nous avons vu que la cote de Bayes dépend de la fonction  $h(\beta)$ , donnée par l'équation (3.2.4) et de ses dérivées. Regardons ce qui se passe avec la première dérivée, lorsque  $y_I \rightarrow +\infty$ .

**Lemme 3.1.**

$$\begin{aligned} h'(\beta) = 0 &\iff \sum_{i=0}^n (x_i \bar{y}_* - \bar{z}_*) e^{\hat{\beta} x_i} - \bar{z}_* f = 0 \\ &\iff \sum_{i=0}^n e^{\hat{\beta} x_i} \left[ \left( x_i - x_I - \frac{x_I f}{(n+1) e^{\hat{\beta} x_i}} \right) \right. \\ &\quad \left. + \frac{1}{y_I} \left( (\delta+1) x_i - \delta/2 + (n-1)(x_i \bar{y}_{(n)} - \bar{z}_{(n)}) \right) \right. \\ &\quad \left. - \frac{(\delta/2 + (n-1)\bar{z}_{(n)})f}{(n+1) e^{\hat{\beta} x_i}} \right] = 0. \end{aligned}$$

Nous remarquons que le dernier terme placé entre parenthèses est un  $\mathcal{O}(y_I^{-1})$ . Donc,  $h'(\beta)$  ne dépend donc pas des  $y_i$  si  $|y_i| \ll |y_I|$  et la solution de  $h'(\beta) = 0$  ne dépend pas de  $y_I$ . Regardons maintenant pour les dérivées deuxième, troisième et quatrième.

**Lemme 3.2.**

$$\begin{aligned}
 h''(\beta) &= \bar{y}_* K_2 \\
 &= \left( 1 + \frac{\delta + 1 + (n-1)\bar{y}_{(n)}}{y_I} \right) \frac{y_I}{n} K_2, \\
 h'''(\beta) &= \bar{y}_* K_3 \\
 &= \left( 1 + \frac{\delta + 1 + (n-1)\bar{y}_{(n)}}{y_I} \right) \frac{y_I}{n} K_3, \\
 h^{(4)}(\beta) &= \bar{y}_* K_4 \\
 &= \left( 1 + \frac{\delta + 1 + (n-1)\bar{y}_{(n)}}{y_I} \right) \frac{y_I}{n} K_4,
 \end{aligned}$$

où les constantes  $K_2$ ,  $K_3$  et  $K_4$  sont données dans les équations (3.2.6), (3.2.7) et (3.2.8).

Alors,

$$I_\beta \propto \sqrt{2\pi\hat{\sigma}^2} e^{-n\hat{h}} \quad \text{si } y_I \rightarrow +\infty. \quad (3.2.9)$$

Nous pouvons écrire l'équation (3.2.5) de la façon suivante.

$$\begin{aligned}
 h(\beta) &= \log \left( f + \sum_{i=0}^n e^{\beta x_i} \right) \left[ \frac{\delta + 1}{n} + \frac{y_I}{n} + \frac{(n-1)}{n} \bar{y}_{(n)} \right] \\
 &\quad - \beta \left( \frac{\delta}{2n} + \frac{x_I y_I}{n} + \frac{(n-1)}{n} \bar{z}_{(n)} \right) \\
 &= \left[ \log \left( f + \sum_{i=0}^n e^{\beta x_i} \right) - \beta x_I \right] \frac{y_I}{n} \\
 &\quad - \frac{n-1}{n} \left[ \left( \frac{\delta + 1}{n-1} + \bar{y}_{(n)} \right) \log \left( f + \sum_{i=0}^n e^{\beta x_i} \right) - \beta \left( \frac{\delta}{2(n-1)} + \bar{z}_{(n)} \right) \right].
 \end{aligned}$$

Nous remplaçons le premier terme de l'équation précédente dans l'équation (3.2.9) et nous obtenons,

$$\begin{aligned} I_\beta &\propto \sqrt{\frac{2\pi n}{y_I K_2}} e^{-n[\log(f + \sum e^{\beta x_i}) - \beta x_I] \frac{y_I}{n}} \\ &\propto \frac{cte}{\sqrt{y_I}} \left( \frac{e^{\beta x_I}}{f + \sum e^{\beta x_i}} \right)^{y_I}. \end{aligned}$$

La cote de Bayes devient donc,

$$\begin{aligned} B^\pi(\mathbf{y}) &= \frac{1}{I_\beta (f + n + 1)^{nB_0}} \\ &\propto \sqrt{y_I} \left( \frac{f + \sum e^{\beta x_i}}{e^{\beta x_I} (f + n + 1)} \right)^{y_I}. \end{aligned} \quad (3.2.10)$$

Si le terme à l'intérieur de la parenthèse est plus petit que un, alors la cote tendra vers zéro pour de grandes valeurs de  $y_I$ . Sinon, elle tendra vers l'infini.

### 3.2.5. Exemple numérique

Reprenons, pour la dernière fois, notre même exemple sur les triplets. Nous avons vu dans la section 3.1.3 que la méthode de Naylor-Smith et l'approximation de Laplace donnaient des résultats similaires. Pour trouver les estimateurs avec la loi *a priori* conjuguée, nous utilisons seulement l'approximation de Laplace. Premièrement, nous calculons  $\delta$  et  $f$  et nous trouvons 40,439 et 0,568 comme valeurs. Les estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$  sont 4,261 et  $-0,210$  respectivement. Les pertes *a posteriori* correspondantes sont de  $9,100 \times 10^{-3}$  et  $8,172 \times 10^{-4}$ . Les estimateurs bayésiens  $\delta_{\mu_i}^\pi$  et pseudo-bayésiens  $\hat{\mu}_i$  des  $\mu_i$  sont donnés dans le tableau 3.2.1. Concernant le test sur le paramètre  $\beta$ , nous obtenons une cote de Bayes de  $2,330 \times 10^{-22}$  et nous rejetons donc l'hypothèse nulle et concervons ce paramètre dans le modèle. Le comportement de la cote de Bayes en présence d'une valeur aberrante est sensiblement le même que celui de l'approximation que nous retrouvons dans l'équation (3.2.10).

TABLEAU 3.2.1. Ensembles de triplets, nés en Norvège, ainsi que les estimations des  $\mu_i$ 

Période	1911-15	1916-20	1921-25	1926-30	1931-35	1936-40
Séries de nombre $x_i$ des ensembles	1	2	3	4	5	6
Ensembles de triplets $y_i$	52	52	40	30	24	20
$\hat{\mu}_i$ par l'approximation de Laplace	57,441	46,557	37,736	30,586	24,791	20,094
$\delta_{\mu_i}^\pi$ par l'approximation de Laplace	57,607	46,653	37,813	30,673	24,901	20,232

Dans ce chapitre, nous avons étudié le modèle de régression poissonienne en considérant la distribution exacte des données. Nous avons introduit deux méthodes numériques afin de calculer les estimateurs bayésiens et pseudo-bayésiens de  $\alpha$  et  $\beta$ , pour se rendre compte que les résultats sont similaires. Finalement, nous avons développé une loi conjuguée comme loi *a priori* afin de détecter si le modèle avait une tendance. Nous arrivons aux mêmes conclusions qu'avec l'approximation normale et les méthodes classiques. Dans le prochain chapitre, nous faisons l'analyse d'un exemple à l'aide de toutes les méthodes vues dans ce mémoire.

## Chapitre 4

---

### EXEMPLE NUMÉRIQUE

Dans ce dernier chapitre, nous illustrons toutes les méthodes présentées dans les chapitres précédents, à l'aide d'un exemple pratique. Pour ce faire, nous utilisons le jeu de données que Barbour et Brown (1974) ont déjà considéré, sur les diverses espèces de poissons provenant de 70 lacs de différentes parties du monde. Stein et Juritz (1988) ont appliqué le modèle de régression de Poisson avec la distribution gaussienne inverse à ce jeu de données. Notons que jusqu'à maintenant, nous avons étudié un exemple qui comportait seulement 6 observations, tandis que celui-ci en comporte 70.

La variable réponse  $Y$  est le nombre d'espèces de poissons et la variable explicative  $x$  est le log de l'aire du lac. Pour cet exemple, nous avons le même modèle que celui que nous avons étudié jusqu'à maintenant, c'est-à-dire

$$\mathbb{E}[Y_i | x_i] = \mu_i = \exp(\alpha + \beta x_i).$$

Pour avoir une petite idée du jeu de données, nous retrouvons une représentation graphique de ce dernier dans la figure 4.0.1.



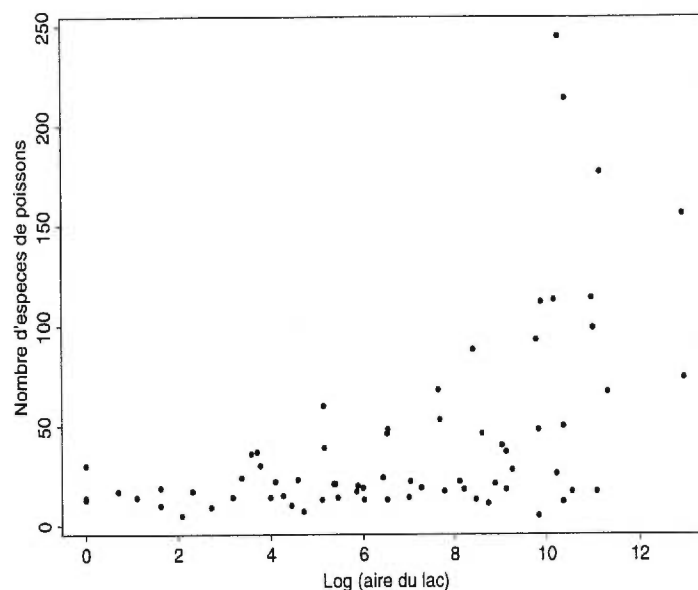


FIGURE 4.0.1. *Graphique des observations par rapport à la covariable*

#### 4.1. ANALYSE À L'AIDE DES MÉTHODES CLASSIQUES

Dans cette section, nous présentons les résultats obtenus à l'aide des méthodes classiques vues au premier chapitre. Nous retrouvons les estimateurs du maximum de vraisemblance, pour le modèle standard et le modèle généralisé, dans le tableau 4.1.1. En ce qui concerne le test classique sur le paramètre  $\beta$  du modèle standard, la statistique du khi-deux nous suggère de conserver ce paramètre, la valeur de la statistique étant très grande, c'est-à-dire de 1110,47. La valeur de l'estimateur de  $\phi$  étant de 5,50, nous nous doutons que les données soient surdispersées. Nous nous en assurons en testant l'hypothèse  $H_0 : \phi = 1$  contre  $H_1 : \phi \neq 1$ . La valeur de la statistique du khi-deux pour ce test est de 1263,95, ce qui est très significatif, la valeur-p étant très près de 0. Nous croyons donc que le modèle le plus approprié est le modèle généralisé de régression de Poisson. Nous rejetons l'hypothèse que

le paramètre  $\beta$  soit égale à zéro pour le modèle généralisé puisque la valeur-p est de  $2,25 \times 10^{-4}$ . Finalement, nous étudions le comportement des estimateurs pour les deux modèles, en présence d'une valeur aberrante que nous avons placée sur la plus grande observation, c'est-à-dire la 16-ième. Nous retrouvons les résultats dans les tableaux 4.1.2 et 4.1.3. Nous voyons, par les figures 4.1.1 et 4.1.2, que les estimateurs du modèle standard  $\hat{\alpha}$  et  $\hat{\beta}$  varient plus drastiquement en fonction de  $K$  que les estimateurs  $\hat{\alpha}_{(K)}$  et  $\hat{\beta}_{(K)}$ , mais les deux séries d'estimateurs ont le même comportement. Pour le modèle généralisé, l'estimateur de  $\alpha$  augmente avec  $K$  tandis que celui de  $\beta$  diminue lorsque  $K$  augmente. Nous remarquons que la relation entre  $K$  et  $\phi$  pour le modèle généralisé est à peu près linéaire.

TABLEAU 4.1.1. Valeurs des estimateurs de  $\alpha$  et  $\beta$  pour le modèle standard et le modèle généralisé des espèces de poissons

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\phi}$
Modèle standard	2,140	0,205	-
Modèle généralisé	3,014	0,099	5,498

TABLEAU 4.1.2. Valeurs des estimateurs du modèle ordinaire pour différents  $K$  des espèces de poissons

$K$	$\hat{\alpha}_{(K)} = (\hat{\alpha}_0 + \log K \hat{\alpha}_*)$	$\hat{\beta}_{(K)} = (\hat{\beta}_0 + \log K \hat{\beta}_*)$	$\hat{\alpha}$	$\hat{\beta}$
1	2,339	0,144	2,140	0,205
2	2,316	0,150	2,082	0,221
3	2,302	0,154	2,033	0,235
4	2,293	0,156	1,990	0,248
5	2,285	0,158	1,954	0,259
6	2,279	0,160	1,923	0,270
7	2,274	0,162	1,896	0,278
8	2,269	0,163	1,873	0,287
9	2,266	0,164	1,853	0,294
10	2,262	0,165	1,836	0,301
15	2,248	0,169	1,783	0,330
20	2,239	0,171	1,766	0,350
30	2,225	0,175	1,784	0,377

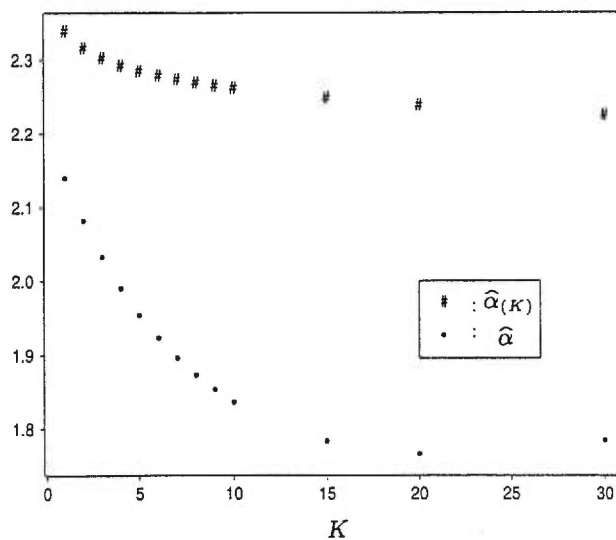


FIGURE 4.1.1. Graphique de  $\hat{\alpha}$  et de  $\hat{\alpha}_{(K)}$  en fonction de  $K$  pour le modèle ordinaire des espèces de poissons

TABLEAU 4.1.3. Valeurs des estimateurs du modèle général pour différents  $K$  des espèces de poissons

$K$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\phi}$
1	3,014	0,099	5,498
2	3,136	0,094	6,327
3	3,244	0,089	7,134
4	3,339	0,086	7,921
5	3,424	0,083	8,692
6	3,502	0,081	9,456
7	3,576	0,079	10,220
8	3,641	0,077	10,985
9	3,699	0,076	11,758
10	3,758	0,075	12,443
15	4,003	0,071	16,111
20	4,198	0,069	19,742
30	4,499	0,066	26,958

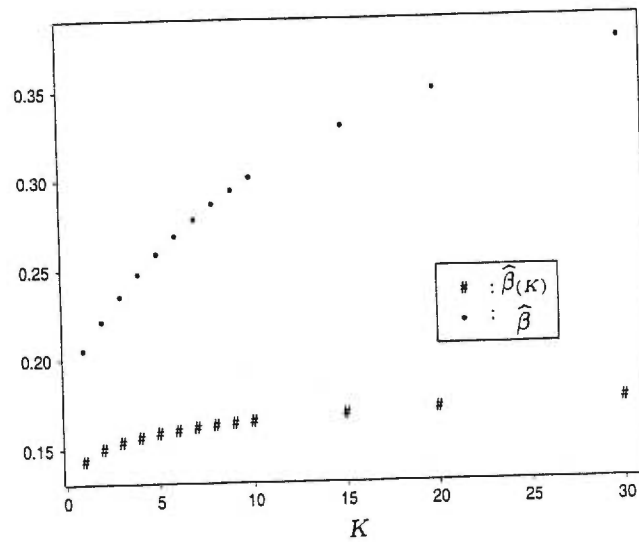


FIGURE 4.1.2. Graphique de  $\hat{\beta}$  et de  $\hat{\beta}_{(K)}$  en fonction de  $K$  pour le modèle ordinaire des espèces de poissons

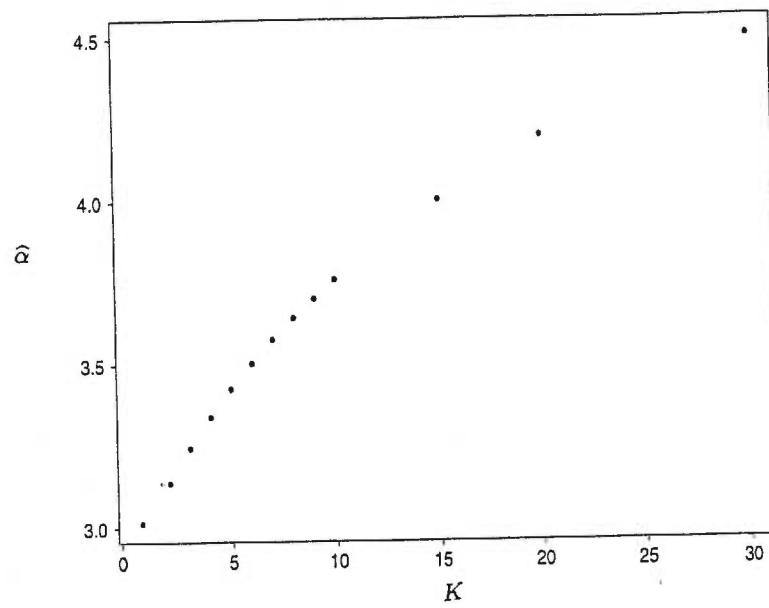


FIGURE 4.1.3. Graphique de  $\hat{\alpha}$  en fonction de  $K$  pour le modèle généralisé des espèces de poissons

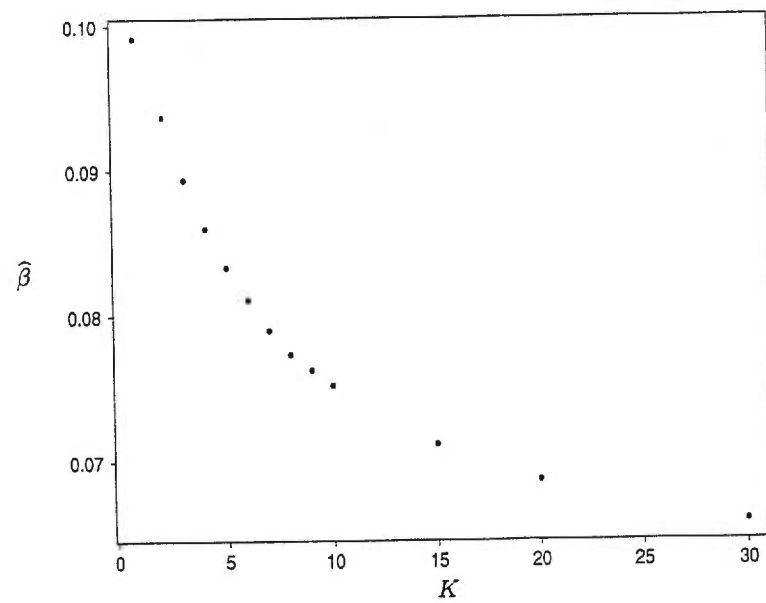


FIGURE 4.1.4. Graphique de  $\hat{\beta}$  en fonction de  $K$  pour le modèle généralisé des espèces de poissons

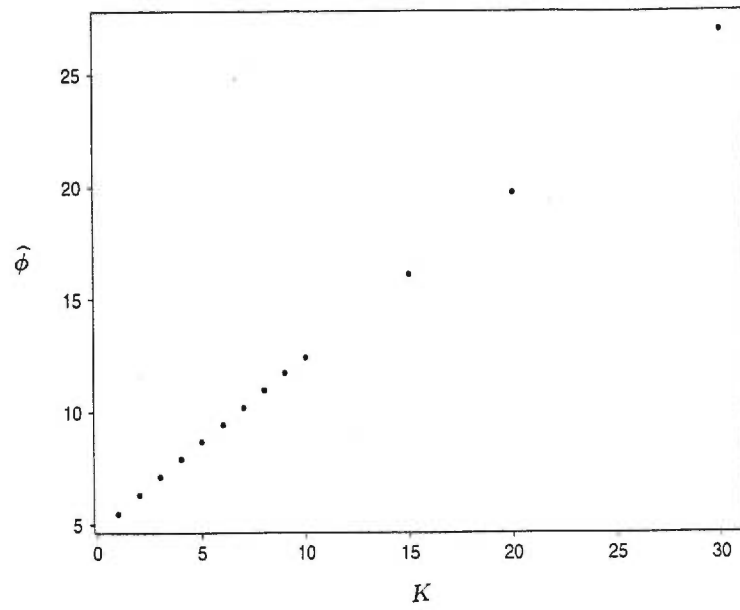


FIGURE 4.1.5. Graphique de  $\hat{\phi}$  en fonction de  $K$  pour le modèle généralisé des espèces de poissons

## 4.2. ANALYSE À L'AIDE DES MÉTHODES BAYÉSIENNES AVEC L'APPROXIMATION NORMALE

Dans le deuxième chapitre, nous avons étudié le modèle de régression poissonienne à l'aide de méthodes bayésiennes, en approchant la densité *a posteriori* par la loi normale. Nous appliquons maintenant ces notions à notre exemple de 70 observations. Les estimateurs de Bayes pour  $\alpha$  et  $\beta$  sont  $\mathbb{E}[\alpha] = 2,302$  et  $\mathbb{E}[\beta] = 0,217$ . Leurs variances sont respectivement de  $3,145 \times 10^{-3}$  et de  $3,722 \times 10^{-5}$ . Le test concernant le paramètre  $\beta$  nous donne une cote de Bayes de  $3,47 \times 10^{-276}$ , ce qui est très significatif. Nous conservons donc ce paramètre dans le modèle, tout comme nous l'avons fait avec la méthode classique. Nous introduisons une valeur aberrante sur la même donnée que dans la section précédente et nous étudions le comportement des estimateurs de Bayes avec l'approximation normale et des estimateurs de Bayes hiérarchiques de  $\alpha$  et  $\beta$ . Nous retrouvons les valeurs de ces derniers pour différents  $K$  dans le tableau 4.2.1. Par la figure 4.2.1, nous remarquons que les erreurs absolues entre les estimateurs de Bayes par l'approximation normale et les estimateurs de Bayes hiérarchiques sont très petites et ont un comportement similaire en fonction de  $K$ .

TABLEAU 4.2.1. Valeurs des estimateurs de Bayes avec l'approximation normale et des estimateurs de Bayes hiérarchiques pour différents  $K$  des espèces de poissons

$K$	$\hat{\alpha}_N$	$\hat{\beta}_N$	$\hat{\alpha}_H$	$\hat{\beta}_H$
1	2,302	0,217	2,301	0,217
2	2,214	0,248	2,213	0,248
3	2,124	0,279	2,123	0,279
4	2,037	0,309	2,037	0,309
5	1,957	0,338	1,956	0,338
6	1,882	0,364	1,881	0,364
7	1,813	0,388	1,812	0,388
8	1,748	0,411	1,747	0,411
9	1,689	0,431	1,688	0,431
10	1,633	0,451	1,632	0,451
15	1,405	0,531	1,403	0,531
20	1,234	0,590	1,232	0,591
30	0,990	0,676	0,986	0,676

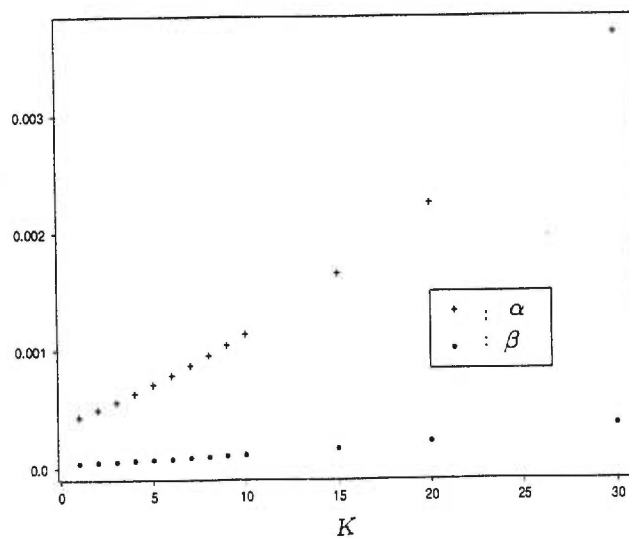


FIGURE 4.2.1. Graphique des erreurs absolues des estimateurs de  $\alpha$  et  $\beta$  en fonction de  $K$  pour le modèle ordinaire des espèces de poissons



### 4.3. ANALYSE À L'AIDE DES MÉTHODES BAYÉSIENNES AVEC LA DISTRIBUTION EXACTE

Dans le chapitre 3, nous avons présenté deux méthodes d'approximation pour le calcul des intégrales, afin d'estimer les paramètres. Dans l'exemple numérique de la section 3.1.3, nous avons remarqué que les deux méthodes apportaient des résultats similaires. De plus, les estimateurs du modèle avec ou sans la loi conjuguée sont pratiquement les mêmes. Pour ces raisons, nous appliquons seulement les notions vues dans la section 3.2 à notre jeu de données. Nous considérons donc la loi conjuguée et utilisons l'approximation analytique de Laplace pour le calcul des estimateurs bayésiens et pseudo-bayésiens. Nous retrouvons  $-32,586$  et  $-3,142$  comme valeurs de  $\delta$  et  $f$ . Les estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$ , ainsi que les pertes *a posteriori*, sont donnés dans le tableau 4.3.1. En ce qui concerne les estimateurs bayésiens des  $\mu_i$ , nous retrouvons un tableau synthèse dans la prochaine section. Encore une fois, nous conservons le paramètre  $\beta$  dans notre modèle, avec une cote de Bayes de  $8,377 \times 10^{-234}$ . Nous arrivons aux mêmes conclusions qu'au chapitre 3 pour ce qui est du comportement de la cote de Bayes en présence d'une valeur aberrante.

TABLEAU 4.3.1. Estimateurs pseudo-bayésiens de  $\alpha$  et  $\beta$  ainsi que des pertes *a posteriori* correspondantes des espèces de poissons

	$\delta_{\alpha}^{\pi}$	$\delta_{\beta}^{\pi}$	$\rho_{\alpha}^{\pi}$	$\rho_{\beta}^{\pi}$
Approximation de Laplace	2,12	0,21	$4,00 \times 10^{-3}$	$4,75 \times 10^{-5}$

#### 4.4. ANALYSE DES RÉSULTATS

Afin de bien voir les différences entre les méthodes que nous avons vues, nous retrouvons les observations, les valeurs de la covariable ainsi que les estimateurs des  $\mu_i$  dans les tableaux 4.4.1, 4.4.2 et 4.4.3. Les  $\hat{\mu}_{st}$  correspondent aux estimateurs classiques du modèle standard, les  $\hat{\mu}_{gen}$  correspondent aux estimateurs classiques du modèle généralisé, les  $\hat{\mu}_{norm}$  correspondent aux estimateurs bayésiens de l'approximation normale, les  $\hat{\mu}_{ps.B}$  et  $\hat{\mu}_B$  correspondent aux estimateurs pseudo-bayésiens et bayésiens avec la loi conjuguée. Nous nous doutons que les estimateurs classiques du modèle standard et les estimateurs pseudo-bayésiens seront semblables, puisque les estimations de  $\alpha$  et  $\beta$  ne diffèrent pas beaucoup les unes des autres. Nous ne pouvons nous prononcer en ce qui concerne les estimateurs de Bayes, étant donné que nous ne pouvons obtenir les estimations de  $\alpha$  et  $\beta$ . Par contre, en se basant sur les résultats de la section 3.1.3, nous nous attendons à ce qu'ils soient similaires aux estimateurs pseudo-bayésiens. Puisque notre étude est axée sur les méthodes bayésiennes, nous retrouvons les graphiques comparant les estimateurs de Bayes des  $\mu_i$  aux autres estimateurs dans les figures 4.4.1, 4.4.2, 4.4.3 et 4.4.4. Nous remarquons que les estimateurs classiques du modèle généralisé sont en moyenne, inférieurs aux estimateurs bayésiens, tandis que ceux avec l'approximation normale sont supérieurs.

Dans ce dernier chapitre, nous avons appliqué toutes les notions étudiées dans ce mémoire à un exemple numérique comportant 70 observations sur les diverses espèces de poissons qui se retrouvent dans différents lacs. D'une méthode à l'autre, nous nous rendons compte que les résultats sont similaires, mis à part le modèle généralisé, puisque celui-ci considère la notion de surdispersion.

TABLEAU 4.4.1. Estimateurs des  $\mu_i$  pour toutes les méthodes avec les observations correspondantes ainsi que les valeurs de la covariable pour les espèces de poissons

$x_i$	$y_i$	$\hat{\mu}_{st}$	$\hat{\mu}_{gen}$	$\hat{\mu}_{norm}$	$\hat{\mu}_{ps.B}$	$\hat{\mu}_B$
1,609	10	11,823	23,884	14,166	11,658	11,674
3,714	37	18,214	29,422	22,362	18,023	18,037
5,142	60	24,421	33,896	30,484	24,223	24,235
10,155	113	68,378	55,712	90,465	68,390	68,405
10,995	99	81,258	60,551	108,560	81,386	81,410
0,000	13	8,495	20,362	9,990	8,354	8,371
3,784	30	18,480	29,629	22,707	18,288	18,302
10,968	114	80,811	60,390	107,929	80,935	80,958
9,877	112	64,583	54,198	85,170	64,565	64,578
2,303	17	13,632	25,582	16,465	13,457	13,473
4,443	10	21,155	31,627	26,194	20,959	20,972
0,000	14	8,495	20,362	9,990	8,354	8,371
5,159	39	24,509	33,955	30,600	24,310	24,322
1,099	14	10,645	22,705	12,679	10,488	10,505
3,989	14	19,273	30,237	23,739	19,080	19,094
11,320	67	86,852	62,528	116,471	87,036	87,065
3,584	36	17,733	29,046	21,740	17,544	17,558
0,000	30	8,495	20,362	9,990	8,354	8,371
1,609	19	11,823	23,884	14,166	11,658	11,674
8,584	46	49,523	47,679	64,337	49,402	49,411

TABLEAU 4.4.2. Estimateurs des  $\mu_i$  pour toutes les méthodes avec les observations correspondantes ainsi que les valeurs de la covariable pour les espèces de poissons

(suite)

$x_i$	$y_i$	$\hat{\mu}_{st}$	$\hat{\mu}_{gen}$	$\hat{\mu}_{norm}$	$\hat{\mu}_{ps.B}$	$\hat{\mu}_B$
7,636	68	40,763	43,404	52,377	40,600	40,608
9,770	93	63,179	53,626	83,215	63,150	63,162
6,512	13	32,357	38,826	41,037	32,167	32,177
7,673	53	41,074	43,563	52,799	40,912	40,920
7,771	17	41,904	43,986	53,927	41,745	41,753
10,257	245	69,830	56,280	92,496	69,854	69,870
8,392	88	47,610	46,782	61,714	47,479	47,488
3,367	24	16,963	28,430	20,743	16,776	16,790
9,112	37	55,196	50,241	72,147	55,110	55,120
8,102	22	44,857	45,456	57,950	44,712	44,720
8,196	18	45,728	45,880	59,139	45,587	45,595
10,401	214	71,922	57,087	95,426	71,964	71,981
11,149	177	83,861	61,480	112,237	84,015	84,041
11,074	17	82,589	61,028	110,439	82,730	82,755
10,358	50	71,285	56,843	94,534	71,322	71,339
9,826	5	63,904	53,922	84,224	63,881	63,893
7,026	22	35,958	40,855	45,877	35,778	35,786
12,956	156	121,554	73,541	166,131	122,143	122,215
12,985	74	122,283	73,754	167,183	122,881	122,955
5,106	13	24,243	33,776	30,249	24,044	24,056
8,733	11	51,063	48,390	66,453	50,952	50,961
9,820	48	63,833	53,893	84,125	63,809	63,822
3,178	14	16,317	27,901	19,909	16,131	16,146
9,244	28	56,708	50,901	74,237	56,632	56,642
0,693	17	9,795	21,810	11,612	9,644	9,660

TABLEAU 4.4.3. Estimateurs des  $\mu_i$  pour toutes les méthodes avec les observations correspondantes ainsi que les valeurs de la covariable pour les espèces de poissons (suite et fin)

$x_i$	$y_i$	$\hat{\mu}_{st}$	$\hat{\mu}_{gen}$	$\hat{\mu}_{norm}$	$\hat{\mu}_{ps.B}$	$\hat{\mu}_B$
10,545	17	74,085	57,910	98,461	74,146	74,165
5,398	21	25,742	34,769	32,229	25,544	25,555
8,445	13	48,124	47,025	62,419	47,996	48,005
5,442	14	25,977	34,922	32,540	25,779	25,790
8,875	21	52,576	49,076	68,535	52,474	52,483
6,423	24	31,774	38,487	40,257	31,583	31,593
10,347	12	71,123	56,781	94,307	71,158	71,175
10,211	26	69,166	56,021	91,567	69,185	69,200
6,006	13	29,167	36,929	36,775	28,971	28,982
5,989	19	29,063	36,866	36,637	28,867	28,878
7,262	19	37,747	41,823	48,291	37,572	37,581
4,094	22	19,695	30,554	24,288	19,500	19,514
4,263	15	20,388	31,068	25,191	20,192	20,206
2,708	9	14,815	26,632	17,979	14,635	14,651
4,585	23	21,783	32,077	27,016	21,586	21,599
6,528	48	32,465	38,889	41,182	32,275	32,285
5,357	21	25,523	34,626	31,940	25,325	25,336
6,516	46	32,387	38,843	41,077	32,197	32,206
6,985	14	35,658	40,690	45,472	35,476	35,485
4,710	7	22,347	32,475	27,756	22,150	22,162
2,079	5	13,021	25,023	15,686	12,850	12,865
9,020	40	54,157	49,783	70,713	54,064	54,074
9,112	18	55,196	50,241	72,147	55,110	55,120
5,878	20	28,407	36,462	35,763	28,210	28,221
5,849	17	28,241	36,359	35,543	28,045	28,055

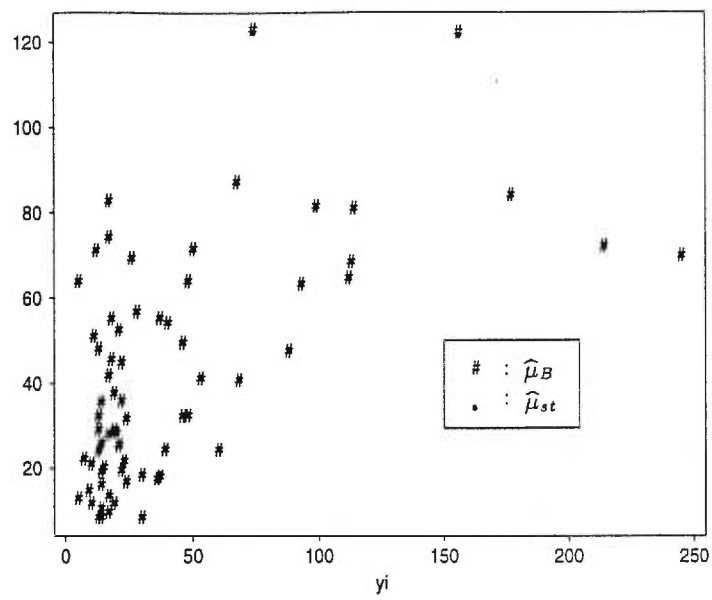


FIGURE 4.4.1. Graphique des  $\hat{\mu}_{st}$  et des  $\hat{\mu}_B$  pour les espèces de poissons

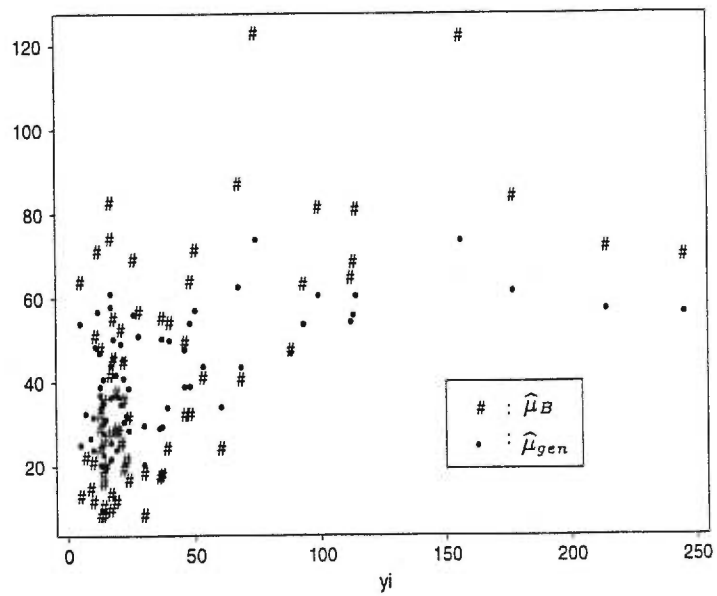


FIGURE 4.4.2. Graphique des  $\hat{\mu}_{gen}$  et des  $\hat{\mu}_B$  pour les espèces de poissons

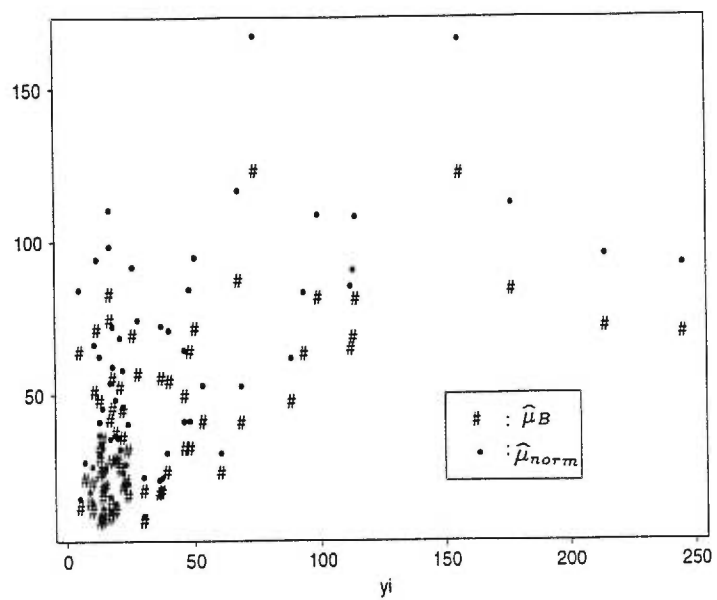


FIGURE 4.4.3. Graphique des  $\hat{\mu}_{norm}$  et des  $\hat{\mu}_B$  pour les espèces de poissons

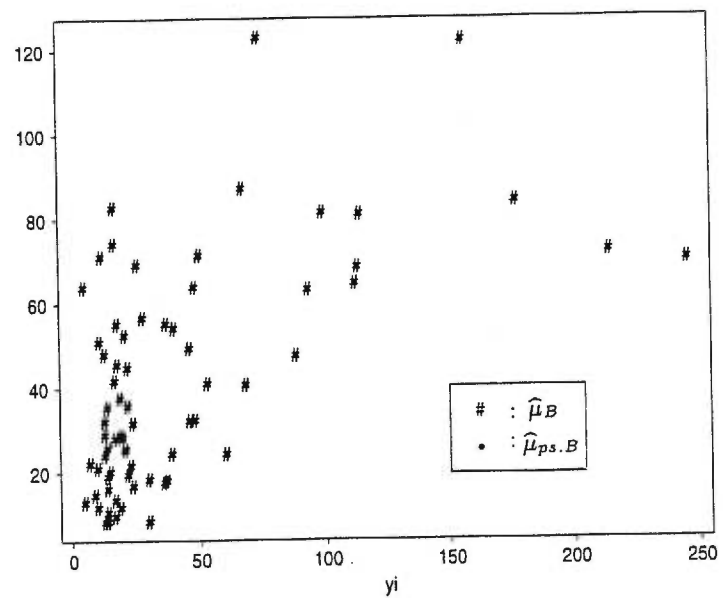


FIGURE 4.4.4. Graphique des  $\hat{\mu}_{ps.B}$  et des  $\hat{\mu}_B$  pour les espèces de poissons

## CONCLUSION

---

Tout au long de ce mémoire, nous avons été en mesure de nous familiariser avec un type de régression peu connu qui est la régression poissonnienne.

Après avoir introduit le modèle ordinaire de régression de Poisson au premier chapitre, nous avons présenté un modèle moins populaire qui prend en considération le niveau de dispersion des données. C'est le modèle généralisé de régression poissonnienne qui utilise la distribution généralisée de Poisson. Pour chacun de ces modèles, nous avons calculé les estimateurs classiques du maximum de vraisemblance à l'aide de la méthode de Newton-Raphson. De plus, nous avons donné les statistiques concernant les tests détectant lequel des deux modèles est le plus approprié et s'il y a une tendance apparente pour chacun d'eux. Aussi, nous avons étudié le comportement des estimateurs en présence d'une valeur aberrante. Finalement, nous avons illustré toutes ces notions à l'aide d'un exemple numérique dans la dernière section de ce chapitre.

Au deuxième chapitre, nous avons présenté les éléments de la théorie bayésienne qui nous étaient nécessaires pour faire l'étude du modèle de régression poissonnienne standard. À partir de celui-ci, nous avons approximé la distribution *a posteriori* des paramètres par une loi normale et avons vérifié sa validité. Par la suite, nous avons calculé les estimateurs de Bayes de ce modèle pour ensuite développer un estimateur hiérarchique robuste par rapport à la présence de valeurs aberrantes. Afin de tester s'il y avait une tendance dans le modèle, nous avons utilisé la cote de Bayes.



Nous avons repris, au troisième chapitre, le même modèle qu'au chapitre précédent mais en considérant cette fois-ci la distribution *a posteriori* exacte des paramètres. Pour ce modèle, nous avons calculé les estimateurs bayésiens et pseudo-bayésiens des paramètres à l'aide de la méthode de Naylor-Smith et de l'approximation analytique de Laplace, les intégrales n'étant pas calculables analytiquement. Par l'exemple numérique appliqué pour ce modèle, nous nous sommes rendus compte que les deux méthodes menaient à des résultats similaires. De plus, afin de calculer la cote de Bayes pour tester la tendance du modèle, nous avons déterminé une loi *a priori* conjuguée pour les paramètres. Nous avons donné les estimateurs pour ce nouveau modèle et avons étudié le comportement de la cote de Bayes en présence d'une valeur aberrante. Par l'exemple numérique donné dans la dernière section de ce chapitre, nous avons remarqué que les estimateurs bayésiens et pseudo-bayésiens des  $\mu_i$  étaient semblables.

Enfin, le dernier chapitre présentait une étude d'un exemple numérique, afin de comparer toutes les techniques vues tout au long des chapitres précédents. Nous avons noté que les estimateurs classiques du modèles standard étaient similaires aux estimateurs bayésiens et pseudo-bayésiens avec la distribution exacte des paramètres. Nous avons aussi vu que les estimateurs classiques du modèle généralisé et les estimateurs bayésiens avec l'approximation normale étaient un peu différents des estimateurs de Bayes. Notons que les estimateurs bayésiens et pseudo-bayésiens étaient ceux calculés à partir du modèle avec la loi *a priori* conjuguée.

Finalement, bien que nous ayons limité notre étude au modèle simple à deux régresseurs, il est possible de généraliser au cas d'un modèle de régression poissonienne multiple à l'aide de la théorie présentée dans ce mémoire.

## Annexe A

---

### PROGRAMMES

```

# *****
# * PROGRAMMES CONCERNANT LE CHAPITRE 1*
# *   Tous faits en S-Plus   *
# *****
#
# *****
#   Definition des donnees et calcul de l'estimateur
#   des moindres carres
# *****
#
#
#   n_6
#   xi1_matrix(c(1,1,1,1,1,1),nrow=6,ncol=1)
#   xi2_matrix(c(1,2,3,4,5,6),nrow=6,ncol=1)
#   xi_cbind(xi1,xi2)
#   yi_matrix(c(52,52,40,30,24,20),nrow=6,ncol=1)
#   X_xi
#   Z_log(yi)
#   BO_solve(t(X)%*%X)%*%t(X)%*%Z
#
#
# *****
#   Calcul de l'estimateur du maximum de vraisemblance pour
#   le modele standard, a l'aide de la methode de Newton-Raphson
# *****
#
#
#   bet_NULL
#   ui_NULL

```

```

q_NULL
H_NULL
new.bet_B0

  for (i in 1:10)
    {

bet_matrix(new.bet, ncol=1, nrow=2)
ui_exp(xi%*%bet)

q_c(sum((yi-ui)*xi1),sum((yi-ui)*xi2))
q_matrix(q,nrow=2,ncol=1)
H_c(-sum(ui*(xi1)^2),-sum(ui*xi1*xi2),
      -sum(ui*xi1*xi2),-sum(ui*(xi2)^2))
H_matrix(H,nrow=2,ncol=2)
Hi_solve(H)
new.bet_bet-Hi%*%q
      if (abs(sqrt(sum(new.bet-bet)^2)/sqrt(sum(bet^2)))
          < 0.00000000001) break

mu.st_ exp(xi%*%new.bet)

    }

#
#
# *****
#   Calcul de la statistique pour tester l'hypothese d'une
#   tendance apparente pour le modele ordinaire
# *****
#
#   *****
#   On estime alpha lorsque beta = 0
#   sous le modele ordinaire
#   *****
#
#
alph0_solve(t(xi1)%*%xi1)%*%t(xi1)%*%Z
alp.0_NULL
ui0_NULL

```

```

q0_NULL
H0_NULL
new.alp_alph0

  for (i in 1:10)
    {

alp.0_matrix(new.alp, ncol=1, nrow=1)
ui0_exp(xi1%%alp.0)

q0_c(sum((yi-ui0)*xi1))
q0_matrix(q0,nrow=1,ncol=1)
H0_c(-sum(ui0*(xi1)^2))
H0_matrix(H0,nrow=1,ncol=1)
Hi0_solve(H0)
new.alp_alp.0-Hi0%%q0
      if (abs(sqrt(sum(new.alp-alp.0)^2)/sqrt(sum(alp.0^2)))
        < 0.00000000001) break

mu.st0_ exp(xi1%%new.alp)

    }

lo_2*(sum(yi*log(mu.st/mu.st0)-mu.st+mu.st0))
1-pchisq(lo,1)
#
#
# *****
#   Calcul de l'estimateur du maximum de vraisemblance pour
#   le modele generalise, a l'aide de la methode de
#   Newton-Raphson "amortie"
# *****
#
#   *****
#   On prend le beta du modele standard et on calcule
#   le phi et le nouveau beta
#   *****
#
#
#

```

```

k_3
Ui_exp(xi%%new.bet)
phitild_sqrt((sum((yi-ui)^2/ui))/(n-k))
Bet_NULL
Ui_NULL
q_NULL
H_NULL
new.Bet_new.bet

for (i in 1:10)
{

  Bet_matrix(new.Bet, ncol=1, nrow=2)
  Ui_exp(xi%%Bet)

  q_c(sum(((1-Ui/phitild+(yi-1)*Ui/(Ui+(phitild-1)*yi))*xi1)),
        sum(((1-Ui/(phitild)+(yi-1)*Ui/(Ui+(phitild-1)*yi))*xi2)))
  q_matrix(q,nrow=2,ncol=1)
  H_c(sum((yi-1)*(phitild-1)*yi/((Ui+(phitild-1)*yi)^2)-1/phitild
          *Ui*xi1^2),
        sum(((yi-1)*(phitild-1)*yi/((Ui+(phitild-1)*yi)^2)-1/phitild
          *Ui*xi1*xi2),
        sum(((yi-1)*(phitild-1)*yi/((Ui+(phitild-1)*yi)^2)-1/phitild
          *Ui*xi1*xi2),
        sum(((yi-1)*(phitild-1)*yi/((Ui+(phitild-1)*yi)^2)-1/phitild
          *Ui*xi2^2))

  H_matrix(H,nrow=2,ncol=2)
  Hi_solve(H)
  new.Bet_Bet-Hi%%q
  if (abs(sqrt(sum(new.Bet-Bet)^2)/sqrt(sum(Bet^2)))
      < 0.00000000001) break

}

#
# *****
#   On ajuste le modele generalise a l'aide
#   de la methode de Newton-Raphson
#           "amortie"
# *****

```

```

tet_NULL
n.phi_NULL
n.bet_NULL
nui_NULL
qt_NULL
Ht_NULL
new.tet_c(new.Bet,phitild)

for (i in 1:20)
{

tet_matrix(new.tet, ncol=1, nrow=3)
n.phi_tet[3]
n.bet_tet[c(1:2)]

nui_exp(xi%*%n.bet)

qt_c(sum(((1-nui/n.phi+(yi-1)*nui/
          (nui+(n.phi-1)*yi))*xi1)),
      sum(((1-nui/(n.phi)+(yi-1)*nui/
          (nui+(n.phi-1)*yi))*xi2)),
      sum(yi*(yi-1)/(nui+(n.phi-1)*yi)-yi/
          n.phi+(nui-yi)/(n.phi^2)))
qt_ matrix(qt,nrow=3,ncol=1)
Ht_c(sum(((yi-1)*(n.phi-1)*yi/
          ((nui+(n.phi-1)*yi)^2)-1/n.phi)*nui*xi1^2),
      sum(((yi-1)*(n.phi-1)*yi/
          ((nui+(n.phi-1)*yi)^2)-1/n.phi)*nui*xi1*xi2),
      sum((1/(n.phi^2)-(yi-1)*yi/
          ((nui+(n.phi-1)*yi)^2))*nui*xi1),
      sum(((yi-1)*(n.phi-1)*yi/
          ((nui+(n.phi-1)*yi)^2)-1/n.phi)*nui*xi1*xi2),
      sum(((yi-1)*(n.phi-1)*yi/
          ((nui+(n.phi-1)*yi)^2)-1/n.phi)*nui*xi2^2),
      sum((1/(n.phi^2)-(yi-1)*yi/
          ((nui+(n.phi-1)*yi)^2))*nui*xi2),
      sum((1/(n.phi^2)-(yi-1)*yi/
          ((nui+(n.phi-1)*yi)^2))*nui*xi1),
      sum((1/(n.phi^2)-(yi-1)*yi/

```

```

        ((nui+(n.phi-1)*yi)^2))*nui*xi2),
        sum(yi/(n.phi^2)-yi^2*(yi-1)/
        ((nui+(n.phi-1)*yi)^2)-2*(nui-yi)/(n.phi^3)))
Ht_matrix(Ht,nrow=3,ncol=3)
Hti_solve(Ht)
incr_Hti%%qt

for(i2 in 1:10)
{
    new.tet.tmp_tet-incr
    ulchap_exp(xi%%new.tet[c(1:2)])
    phi.max_max(1-ulchap/yi)
    print(c(i,i2,new.tet.tmp[3],phi.max))
    if(new.tet.tmp[3]>phi.max) break
    incr_incr/2
}
new.tet_new.tet.tmp
    if (new.tet[3] < 0.5) {new.tet[3]_0.5}
phi_new.tet[3]
    if (max (abs ((new.tet-tet)/tet))<0.01) break
mu.gen_exp(xi%%new.tet[c(1:2)])
}

#
# *****
#          Calcul de la statistique pour le
#          test du modele generalise
# *****
#
#
# L_2*(sum(log(mu.gen/mu.st)+(yi-1)*log((mu.gen+(phi-1)*yi)/mu.st)
#       -yi*log(phi)-(mu.gen+(phi-1)*yi)/phi+mu.st))
# 1-pchisq(L,1)
#
#
# *****
#          Calcul de la statistique pour tester l'hypothese d'une
#          tendance apparente pour le modele generalise
# *****
#

```

```

# *****
#      On estime alpha et phi lorsque beta = 0
#      sous le modele generalise
# *****
#
#
k0_2
Ui.H0_exp(xi1%*%new.alp)
phitild.H0_sqrt((sum((yi-ui.H0)^2/ui.H0))/(n-k0))
Alp.0_NULL
Ui.0_NULL
qo_NULL
Ho_NULL
new.Alp_new.alp

for (i in 1:10)
{

Alp.0_matrix(new.Alp, ncol=1, nrow=1)
Ui.0_exp(xi1%*%Alp.0)

qo_c(sum((1-Ui.0/phitild.H0+(yi-1)*Ui.0/
          (Ui.0+(phitild.H0-1)*yi))*xi1))
qo_matrix(qo,nrow=1,ncol=1)
Ho_c(sum((yi-1)*(phitild.H0-1)*yi/
          ((Ui.0+(phitild.H0-1)*yi)^2)-1/phitild.H0)*Ui.0*xi1^2))
Ho_matrix(Ho,nrow=1,ncol=1)
Hio_solve(Ho)
new.Alp_Alp.0-Hio%*%qo
  if (abs(sqrt(sum(new.Alp-Alp.0)^2)/sqrt(sum(Alp.0^2)))
      < 0.00000000001) break

}

tet0_NULL
n.phi0_NULL
n.alp_NULL
nui0_NULL
qt0_NULL

```



```

Ht0_NULL
new.tet0_c(new.Alp,phitild.H0)

for (i in 1:30)
{

tet0_matrix(new.tet0, ncol=1, nrow=2)
n.phi0_tet0[2]
n.alp_tet0[1]
n.alp_matrix(n.alp, ncol=1, nrow=1)

nui0_exp(xi1%%n.alp)

qt0_c(sum((1-nui0/n.phi0+(yi-1)*nui0/
          (nui0+(n.phi0-1)*yi))*xi1),
       sum(yi*(yi-1)/(nui0+(n.phi0-1)*yi)-yi/
          n.phi0+(nui0-yi)/(n.phi0^2)))
qt0_ matrix(qt0,nrow=2,ncol=1)
Ht0_c(sum(((yi-1)*(n.phi0-1)*yi/
          ((nui0+(n.phi0-1)*yi)^2)-1/n.phi0)*nui0*xi1^2),
       sum((1/(n.phi0^2)-(yi-1)*yi/
          ((nui0+(n.phi0-1)*yi)^2))*nui0*xi1),
       sum((1/(n.phi0^2)-(yi-1)*yi/
          ((nui0+(n.phi0-1)*yi)^2))*nui0*xi1),
       sum(yi/(n.phi0^2)-yi^2*(yi-1)/((nui0+
          (n.phi0-1)*yi)^2)-2*(nui0-yi)/(n.phi0^3)))
Ht0_matrix(Ht0,nrow=2,ncol=2)
Hti0_solve(Ht0)
incr0_Hti0%%qt0

for(i2 in 1:10)
{
new.tet.tmp_tet0-incr0
mu.gen0_exp(xi1%%new.tet0[1])
phi.max_max(1-mu.gen0/yi)
print(c(i,i2,new.tet.tmp[2],phi.max))
if(new.tet.tmp[2]>phi.max) break
incr0_incr0/2
}

```

```

new.tet0_new.tet.tmp
phi0_new.tet0[2]
  if (max (abs ((new.tet0-tet0)/tet0))<0.01) break
mu.gen0_exp(xi1%*%new.tet0[1])
}

12_2*sum(log(mu.gen/mu.gen0)+(yi-1)*log((mu.gen+(phi-1)*yi)/
(mu.gen0+(phi0-1)*yi))-yi*log(phi/phi0)+
(mu.gen0+(phi0-1)*yi)/phi0 -(mu.gen+(phi-1)*yi)/phi)
1-pchisq(12,1)
#
#
# *****
#      Calcul de l'approximation de l'estimateur du modele
#      standard en presence d'une valeur aberrante
# *****
#
#      *****
#      On calcule ce qui suit pour differentes
#      valeurs de K
#      *****
#
#
# yi_matrix(c(52,K*52,40,30,24,20),nrow=6,ncol=1)
# Z_log(yi)
# xI_matrix(c(1,2), ncol=1, nrow=2)
# B0_solve(t(X)%*%X)%*%t(X)%*%Z
# betoil_solve(t(X)%*%X)%*%xI
# betk_B0+log(K)*betoil
#
#
# *****
#      Calcul de l'estimateur du maximum de vraisemblance pour le
#      modele standard et generalise en presence d'une valeur
#      aberrante, a l'aide de la methode de Newton-Raphson
# *****
#      Refaire les programmes deja donnees avec les elements suivants,
#      en faisant varier le K
# *****

```

```

n_6
xi1_matrix(c(1,1,1,1,1,1),nrow=6,ncol=1)
xi2_matrix(c(1,2,3,4,5,6),nrow=6,ncol=1)
xi_cbind(xi1,xi2)
yi_matrix(c(52,K*52,40,30,24,20),nrow=6,ncol=1)
X_xi
Z_log(yi)
B0_solve(t(X)%*%X)%*%t(X)%*%Z
#
#
#          *****
#          * PROGRAMMES CONCERNANT LE CHAPITRE 2*
#          *      Tous faits en S-Plus      *
#          *****
#
# *****
# Calcul des estimateurs de Bayes a l'aide de l'esperance et de
# la matrice de variance-covariance pour l'approximation
# normale de la distribution a posteriori
# des parametres
# *****
#
#
xi_c(1,2,3,4,5,6)
yi_c(52,52,40,30,24,20)
zi_log(yi)
zi1_sqrt(yi)
zi2_xi*zi1
Z_cbind(zi1,zi2)
t_matrix(zi1*zi,ncol=1,nrow=6)

vari_solve(t(Z)%*%Z)
esp_vari)%*%t(Z)%*%t
mu.al_esp[1]
mu.bet_esp[2]
sig.al_vari[1,1]
sig.bet_vari[2,2]
rh.o_vari[1,2]/(sqrt(sig.al*sig.bet))
mu.i_exp(mu.al+mu.bet*xi)

```

```

# *****
#           Calcul de la cote de Bayes afin de
#           faire le test de beta = 0
# *****
#
#
det_function(x) prod(eigen(x)$values)

cote_function(cte)
{
  Sig_vari
  S_matrix(c(Sig[1,1],0,0,Sig[2,2]),ncol=2,nrow=2)
  S_S+diag(c(cte,cte))
  mc_esp
  r_t-Z%%mc
  A_solve(Sig)+solve(S)
  A.S_A%%S
  term1_(t(r)%%r+t(mc)%%solve(Sig+S)%%mc)

  W_matrix(zil,nrow=6, ncol=1)
  Sig0_solve(t(W)%%W)
  S0_Sig0+cte
  mc0_Sig0%%t(W)%%t
  r0_t-W%%mc0
  A0_solve(Sig0)+solve(S0)
  A0.S0_A0%%S0
  term0_(t(r0)%%r0+t(mc0)%%solve(Sig0+S0)%%mc0)

  cote_(sqrt(det(A.S)/(2*pi*det(A0.S0))))
  *exp(-0.5*(t(r0)%%r0-t(r)%%r+mc0^2*solve(S0+Sig0)
  -t(mc)%%solve(S+Sig)%%mc))
  cote
}

cote(4.5)
#
#
#
#

```

```

# *****
#      Calcul de l'estimateur hierarchique de Bayes en presence
#      d'une valeur aberrante, en faisant varier le K
# *****
#
#
yi_matrix(c(52,K*52,40,30,24,20),nrow=6,ncol=1)
zi1_sqrt(yi)
zi2_xi*zi1
t_matrix(sqrt(yi)*log(yi),ncol=1,nrow=6)
Z_cbind(zi1,zi2)

Sig_solve(t(Z)%*%Z)
Sig_(Sig+t(Sig))/2
mc_Sig%*%t(Z)%*%t
dec_chol(Sig)
Q_t(matrix(dec,nrow=2,ncol=2))
vl.pr_eigen(t(Z)%*%Z)$values
C_matrix(c(vl.pr[1],0,0,vl.pr[2]),nrow=2,ncol=2)
A_Q%*%C^.5
Ip_matrix(c(1,0,0,1),nrow=2,ncol=2)

z_solve(Q)%*%mc
z_matrix(z,nrow=2,ncol=1)
z1_z^2/2
l_1/2+1/(n^2)
gl_gamma(1+1/2)*pgamma(z1,1+1/2)/
      (z1*gamma(1-1/2)*pgamma(z1,1-1/2))
G1_matrix(c(gl[1],0,0,gl[2]),nrow=2,ncol=2)
est.H_(Ip-A%*%G1%*%t(A))%*%mc
#
#
# *****
#      Calcul des estimateurs de Bayes, en presence d'une valeur
#      aberrante, a l'aide de l'esperance et de
#      la matrice de variance-covariance pour l'approximation
#      normale de la distribution a posteriori
#      des parametres
# *****

```

```

xi_c(1,2,3,4,5,6)
yi_c(52,K*52,40,30,24,20)
zi1_sqrt(yi)
zi2_xi*zi1
t_matrix(zi1*log(yi),ncol=1,nrow=6)
Z_cbind(zi1,zi2)

esp_solve(t(Z)%*%Z)%*%t(Z)%*%t
vari_solve(t(Z)%*%Z)

#
#
#
# *****
# * PROGRAMMES CONCERNANT LE CHAPITRE 3*
# * Faits en S-Plus et en Mathematica *
# *****
#
# *****
# Calcul du point de convergence de l'esperance et de la variance
# avec la vraie distribution, a l'aide de la methode
# numerique gaussienne ( En S-Plus )
# *****
#
#
xi_c(1,2,3,4,5,6)
yi_c(52,52,40,30,24,20)
my_mean(yi)
s.y_sum(yi)
s.xy_sum(xi*yi)
u0_0
sig0_0.01
t_c(-4.30444857,-3.46265693,-2.74847072,-2.09518326,-1.47668273,
-.87871379,-.29174551,.29174551,.87871379,1.47668273,
2.09518326,2.74847072,3.46265693,4.30444857)
w_c(.00000001,.00000472,.00035509,.00785005,.06850553,.27310561,
.53640591,.53640591,.27310561,.06850553,.00785005,.00035509,
.00000472,.00000001)

g_function(b)

```

```

    {
      exp(b*(s.xy-1))/(sum(exp(b*xi))^(s.y))
    }

u_NULL
sig_NULL
new.u_u0
new.sig_sig0

for (i in 1:20)
{
  m_NULL
  v_NULL
  gt_NULL

  for ( j in 1:14 )
  {
    m_sqrt(2*new.sig)*exp(t^2)*w
    v_new.u+sqrt(2*new.sig)*t
    gt[j]_g(v[j])
  }

  u[i]_sum( v * m * gt)/sum(m*gt)
  sig[i]_sum(m*(v^2)*gt)/sum(m*gt)-(new.u)^2
  new.u_u[i]
  new.sig_sig[i]
}

#
# *****
#   Calcul de l'estimateur de Bayes des mu_i ainsi que
#   des estimateurs pseudo-bayesiens de alpha et
#   beta par la methode de Naylor-Smith
# *****
#
# *****
#   On applique la boucle suivante pour
#   differentes combinaisons de k et l
#   afin de calculer les moments
# *****

```

```

mi_sqrt(2*new.sig)*exp(t^2)*w
vi_new.u+sqrt(2*new.sig)*t

I.kl_NULL
den0_NULL
gt0_NULL

for ( j in 1:14 )
  {
    den0[j]_sum(exp(vi[j]*xi))

    g0_function(b)
    {
      exp(b*(s.xy+1))/(den0[j]^(s.y+k))
    }
    gt0[j]_g0(vi[j])
  }

I.kl_sum(mi*gt0)

est.u_sum(yi)*I.10/I.00
est.v_I.01/I.00
pert.u_sum(yi)*(sum(yi)+1)*I.20/I.00-est.u^2
pert.v_I.02/I.00-est.v^2
est.al_log(est.u)-0.5*(1/est.u^2)*pert.u
est.bet_log(est.v)-0.5*(1/est.v^2)*pert.v
pert.al_pert.u/(est.u^2)
pert.bet_pert.v/(est.v^2)
mu.chap_exp(est.al+est.bet*xi)

I.1xi_NULL
for (i in 1:6)
{
  den_NULL
  gt1_NULL

  for ( j in 1:14 )
    {
      den[j]_sum(exp(vi[j]*xi))
    }
}

```



```

        g1_function(b)
        {
            exp(b*(s.xy+xi[i]))/(den[j]^(s.y+1))
        }
        gt1[j]_g1(vi[j])
    }

    I.1xi[i]_sum(mi*gt1)
}
mu.bayes_sum(yi)*I.1xi/I.00
#
#
# *****
# Calcul des estimateurs de Bayes de u et v, ainsi que des pertes
# a posteriori correspondantes, avec la vraie distribution,
# a l'aide de l'approximation analytique de Laplace
# ( En Mathematica )
# *****
#
#
xi = {1,2,3,4,5,6}
yi = {52,52,40,30,24,20}
n = 6
c = (Apply[Plus,xi*yi]-1)/n
d = (Apply[Plus,yi])/n
FindRoot[d*Apply[Plus,xi*vd^xi]-c*Apply[Plus,vd^xi] == 0, {vd,1}]
vD = vd /.%
sox = Apply[Plus,xi]
soy = Apply[Plus,yi]
soxy = Apply[Plus,xi*yi]
sovd = Apply[Plus,vd^xi]
soxvd = Apply[Plus,xi*vD^xi]
sox2vd = Apply[Plus,(xi^2)*(vD^xi)]
#
# *****
# Nous appliquons la fonction suivante
# pour chaque cas
# *****

```

```

a = (Apply[Plus,xi*yi]+l-1)/n
b = (Apply[Plus,yi]+k)/n
FindRoot[b*Apply[Plus,xi*vn^xi]-a*Apply[Plus,vn^xi] == 0, {vn,1}]
vN = vn /.%
sovn = Apply[Plus,vN^xi]
soxvn = Apply[Plus,xi*vN^xi]
sox2vn = Apply[Plus,(xi^2)*(vN^xi)]
num = (vN^2)*(soxy-1+soy*(sox2vd-soxvd-soxvd^2/sovd)/sovd)
den = (vD^2)*(soxy-1+l+(soy+k)*(sox2vn-soxvn-soxvn^2/sovn)/sovn)
hn = -a*Log[vN]+b*Log[sovn]
hd = -c*Log[vD]+d*Log[sovd]
#
# *****
# Cas 1: L'esperance a posteriori de u
# *****
l = 0
k = 1
# Nous trouvons:
#
in1u = Sqrt[num/den]*Exp[-n*(hn-hd)]
espu = soy*in1u
#
# *****
# Cas 2: L'esperance a posteriori de v
# *****
l = 1
k = 0
# Nous trouvons:
#
in1v = Sqrt[num/den]*Exp[-n*(hn-hd)]
espv = in1v
#
# *****
# Cas 3: La variance a posteriori de u
# *****
l = 0
k = 2
# Nous trouvons:
#

```

```

n2u = Sqrt[num/den]*Exp[-n*(hn-hd)]
rou = soy*(soy+1)*in2u-(espu)^2
#
# *****
# Cas 4: La variance a posteriori de v
# *****
l = 2
k = 0
# Nous trouvons:
#
in2v = Sqrt[num/den]*Exp[-n*(hn-hd)]
rov = in2v-(espv)^2
#
#
# *****
# Programme qui retourne les estimateurs pseudo-bayesiens
# de alpha et beta, ainsi que les pertes
# a posteriori correspondantes
# *****
#
#
delal = Log[espu]-rou/(2*espu^2)
delbet = Log[espv]-rov/(2*espv^2)
roal = rou/(espu^2)
robet = rov/(espv^2)
mui = Exp[delal+xi*delbet]
#
#
# *****
# Programme qui retourne les estimateurs de
# Bayes des mu_i
# *****
#
#
xi = {1,2,3,4,5,6}
yi = {52,52,40,30,24,20}
n = 6
k = 1
l = xi

```

```

FindRoot[d*Apply[Plus,xi*vd^xi]-c*Apply[Plus,vd^xi] == 0, {vd,1}]
vD = vd /.%
sox = Apply[Plus,xi]
soy = Apply[Plus,yi]
soxy = Apply[Plus,xi*yi]
sovd = Apply[Plus,vd^xi]
soxvd = Apply[Plus,xi*vd^xi]
sox2vd = Apply[Plus,(xi^2)*(vd^xi)]
hd = -c*Log[vD]+d*Log[sovd]

a = Array[0,Length[yi]]
vN1 = Array[0,Length[yi]]
Do[a[[i]] = (Apply[Plus,xi*yi]+1[[i]]-1)/n;
    b = (Apply[Plus,yi]+k)/n;
    c = (Apply[Plus,xi*yi]-1)/n;
    d = (Apply[Plus,yi])/n;
vN1[[i]] = FindRoot[b*Apply[Plus,xi*vN^xi]-a[[i]]*Apply[Plus,vN^xi]
    == 0, {vN,1}],
{i,1,Length[yi]}]
vN1
vN = vN /.%

sovn = Array[0,Length[vN]]
soxvn = Array[0,Length[vN]]
sox2vn = Array[0,Length[vN]]
num = Array[0,Length[vN]]
den = Array[0,Length[vN]]
hn = Array[0,Length[vN]]
Do[sovn[[i]] = Apply[Plus,vN[[i]]^xi];
    soxvn[[i]] = Apply[Plus,xi*vN[[i]]^xi];
    sox2vn[[i]] = Apply[Plus,(xi^2)*(vN[[i]]^xi)];
    num[[i]] = (vN[[i]]^2)*(soxy-1+soy*(sox2vd-
        soxvd-soxvd^2/sovd)/sovd);
    den[[i]] = (vD^2)*(soxy-1+1[[i]]+(soy+k)*(sox2vn[[i]]-
        soxvn[[i]]-soxvn[[i]]^2/sovn[[i]))/sovn[[i]];
    hn[[i]] = -a[[i]]*Log[vN[[i]]]+b*Log[sovn[[i]]],
{i,1,Length[vN]}]
in1xi = Sqrt[num/den]*Exp[-n*(hn-hd)]
mubayes = soy*in1xi

```



```

# *****
#           Nous appliquons la fonction suivante
#           pour chaque cas
# *****
#
#
a = (Apply[Plus,xi*yi]+1-1)/n
b = (Apply[Plus,yi]+del+k)/n
FindRoot[b*Apply[Plus,xi*vn^xi]-a*Apply[Plus,vn^xi]-a*f==0,{vn,1}]
vN = vn /.%
sovn = f + Apply[Plus,vN^xi]
soxvn = Apply[Plus,xi*vN^xi]
sox2vn = Apply[Plus,(xi^2)*(vN^xi)]
num = (vN^2)*(soxy-1+soy*(sox2vd-soxvd-soxvd^2/sovd)/sovd)
den = (vD^2)*(soxy-1+1+(soy+k)*(sox2vn-soxvn-soxvn^2/sovn)/sovn)
hn = -a*Log[vN]+b*Log[sovn]
hd = -c*Log[vD]+d*Log[sovd]
#
#
# *****
# Cas 1: L'esperance a posteriori de u
# *****
l = 0
k = 1
# Nous trouvons:
#
in1u = Sqrt[num/den]*Exp[-n*(hn-hd)]
esp1u = soy*in1u
#
# *****
# Cas 2: L'esperance a posteriori de v
# *****
l = 1
k = 0
# Nous trouvons:
#
in1v = Sqrt[num/den]*Exp[-n*(hn-hd)]
esp1v = in1v
#

```

```

# *****
# Cas 3: La variance a posteriori de u
# *****
  l = 0
  k = 2
# Nous trouvons:
#
  n2u = Sqrt[num/den]*Exp[-n*(hn-hd)]
  rou = soy*(soy+1)*in2u-(espu)^2
#
# *****
# Cas 4: La variance a posteriori de v
# *****
  l = 2
  k = 0
# Nous trouvons:
#
  in2v = Sqrt[num/den]*Exp[-n*(hn-hd)]
  rov = in2v-(espv)^2
#
#
# *****
#       Programme qui retourne les estimateurs pseudo-bayesiens
#       de alpha et beta, ainsi que les pertes
#       a posteriori correspondantes
# *****
#
#
  delal = Log[espu]-rou/(2*espu^2)
  delbet = Log[espv]-rov/(2*espv^2)
  roal = rou/(espu^2)
  robet = rov/(espv^2)
  mui = Exp[delal+xi*delbet]
#
#
# *****
#       Programme qui retourne les estimateurs de
#       Bayes des mu_i
# *****

```

```

xi = {del/2,1,2,3,4,5,6}
yi = {1,52,52,40,30,24,20}
n = 6
k = 1
l = xi
FindRoot[d*Apply[Plus,xi*vd^xi]-c*Apply[Plus,vd^xi]-c*f==0,{vd,1}]
vD = vd /.%
soy = del + Apply[Plus,yi]
soxy = Apply[Plus,xi*yi]
sovd = f + Apply[Plus,vD^xi]
soxvd = Apply[Plus,xi*vD^xi]
sox2vd = Apply[Plus,(xi^2)*(vD^xi)]
hd = -c*Log[vD]+d*Log[sovd]

a = Array[0,Length[yi]]
vN1 = Array[0,Length[yi]]
Do[a[[i]] = (Apply[Plus,xi*yi]+1[[i]]-1)/n;
    b = (Apply[Plus,yi]+k+del)/n;
    c = (Apply[Plus,xi*yi]-1)/n;
    d = (Apply[Plus,yi]+del)/n;
vN1[[i]] = FindRoot[b*Apply[Plus,xi*vn^xi]-a[[i]]*Apply[Plus,vn^xi]
    -a[[i]]*f == 0, {vn,1}],
{i,1,Length[yi]}]
vN1
vN = vn /.%

sovn = Array[0,Length[vN]]
soxvn = Array[0,Length[vN]]
sox2vn = Array[0,Length[vN]]
num = Array[0,Length[vN]]
den = Array[0,Length[vN]]
hn = Array[0,Length[vN]]
Do[sovn[[i]] = f + Apply[Plus,vN[[i]]^xi];
    soxvn[[i]] = Apply[Plus,xi*vN[[i]]^xi];
    sox2vn[[i]] = Apply[Plus,(xi^2)*(vN[[i]]^xi)];
    num[[i]] = (vN[[i]]^2)*(soxy-1+soy*(sox2vd-
        soxvd-soxvd^2/sovd)/sovd);
    den[[i]] = (vD^2)*(soxy-1+1[[i]]+(soy+k)*(sox2vn[[i]]-
        soxvn[[i]]-soxvn[[i]]^2/sovn[[i]])/sovn[[i]]);

```



```

hn[[i]] = -a[[i]]*Log[vN[[i]]]+b*Log[sovn[[i]]],
{i,1,Length[vN]}

in1xi = Sqrt[num/den]*Exp[-n*(hn-hd)]
in1xi = in1xi[{{2,3,4,5,6,7}}]
mubayes = soy*in1xi

#
#
# *****
# Programme qui retourne la cote de Bayes
# *****
#
#
del = 40.439
f = 0.567558
xi = {del/2,1,2,3,4,5,6}
yi = {1,52,52,40,30,24,20}
yi2 = {1,52,52,40,30,24,20,del}
n = 6
zi = xi*yi
yb = (Apply[Plus,yi2])/n
zb = (Apply[Plus,zi])/n

FindRoot[Apply[Plus,(yb*xi-zb)*Exp[b*xi]]- zb*f == 0, {b,0}]
bet = b /.%

a = f + Apply[Plus,Exp[bet*xi]]
b = Apply[Plus,xi*Exp[bet*xi]]
c = Apply[Plus,xi^2*Exp[bet*xi]]
d = Apply[Plus,xi^3*Exp[bet*xi]]
e = Apply[Plus,xi^4*Exp[bet*xi]]
h = yb*Log[a]-bet*zb
h2 = yb*(c/a - (b/a)^2)
sig2 = 1/h2
h3 = yb*(d/a - 3*c*b/(a^2) + 2*((b/a)^3))
h4 = yb*(e/a - 4*d*b/(a^2) -3*((c/a)^2) + 12*c*(b^2)/(a^3)
      - 6*((b/a)^4))
pi = N[Pi]

```

```

ibet = Sqrt[2*pi*sig2]*Exp[-n*h]*(1+(5*(h3^2)*(sig2^3)/
      12-h4*(sig2^2)/4)/(2*n))
ds = (Apply[Plus,yi2])/n
de = ibet*((f+n+1)^(n*ds))
cote = 1/de
#
#
#          *****
#          * PROGRAMMES CONCERNANT LE CHAPITRE 4*
#          *   Faits en S-Plus et en Mathematica   *
#          *****
#          *****
#          Refaire les programmes precedents en
#          changeant le nombre d'observations (n=70)
#          avec les donnees suivantes
#          *****
#
n_70
xi1_matrix(c(rep(1,70)),nrow=70,ncol=1)
xi2_log(matrix(c(5,41,171,25719,59596,1,44,58016,19477,10,85,1,
      174,3,54,82414,36,1,5,5346,2072,17500,673,2150,
      2370,28490,4413,29,9065,3302,3626,32893,69484,
      64500,31500,18500,1125,423488,436000,165,6206,
      18400,24,10340,2,38000,221,4650,231,7154,616,
      31153,27195,406,399,1425,60,71,15,98,684,212,
      676,1080,111,8,8264,9065,357,347),
      nrow=70,ncol=1))
xi_cbind(xi1,xi2)
yi_matrix(c(10,37,60,113,99,13,30,114,112,17,
      10,14,39,14,14,67,36,30,19,46,
      68,93,13,53,17,245,88,24,37,22,
      18,214,177,17,50,5,22,156,74,13,
      11,48, 14,28,17,17,21,13,14,21,
      24,12,26,13,19,19,22,15,9,23,
      48,21,46,14,7,5,40,18,20,17),
      nrow=70,ncol=1)
#
#
*****

```

## BIBLIOGRAPHIE

---

- [1] Angers, J.-F. et MacGibbon, B. (1996). Protection against outliers in bayesian linear models with econometric applications. *Advances in Econometrics*, **11**, PartA, 211-229.
- [2] Barbour, C.D. et Brown, J.H. (1974). Fish species diversity in lakes. *The American Naturalist*, **108**, 473-489.
- [3] Bartlett, M.S. et Kendall, D.G. (1946). The statistical analysis of variance-heterogeneity and logarithmic transformation. *Journal of the Royal Statistical Society Suppl.*, **8**, 128.
- [4] Bernardinelli, L. et Montomoli, C. (1992). Empirical Bayes Versus Fully Bayesian Analysis of Geographical Variation in Disease Risk. *Statistics in Medicine*, **11**, 983-1007.
- [5] Breslow, N. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**, 38-44.
- [6] Christiansen, Cindy L. et Morris, Carl N. (1997). Hierarchical Poisson Regression Modeling. *Journal of the American Statistical Association*, **92**, No 438, 618-632.
- [7] Clayton, D. et Kaldor, J. (1987). Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics*, **43**, 671-681.
- [8] Consul, P.C. et Famoye, Felix (1992). Generalized Poisson Regression Model. *Communication in Statistics, A-Theory and Methods*, **21**, 89-109.
- [9] Conte, S.D. et de Boor, C. (1980). *Elementary numerical analysis, an algorithmic approach, 3thd edition*. McGraw-Hill Book Company, New York.
- [10] El-Sayyad, G.M. (1973). Bayesian and classical analysis of Poisson regression. *Journal of the Royal Statistical Society, Series B*, **35**, 445-451.
- [11] Gradshteyn, I.S. et Ryzhik, I.M. (1980). *Table of integrals, series and products*, Academic Press.
- [12] Jeffreys, H. (1948). *Theory of Probability, 2nd edition*. Oxford: Clarendon Press.
- [13] Keeler, E.B. et Rolph, J.E. (1988). The Demand for Episodes of Treatment in the Health Insurance Experiment. *Journal of the Health Economics*, **7**, 337-367.

- [14] Lawless, J.F. (1987). Negative Binomial and Mixed Poisson Regression. *Canadian Journal of Statistics*, **15**, 209-225.
- [15] Lindley, D.V. (1965). *Introduction to Probability and Statistics, Vol.2*. Cambridge University Press.
- [16] McCullagh, P. et Nelder, J.A. (1989). *Generalized Linear Models, 2nd edition*. Chapman and Hall, London.
- [17] Naylor, J.C. et Smith, A.F.M. (1982). Application of a method for the efficient computation of posterior distribution. *Journal of the Royal Statistical Society, Series C*, **31**, 214-225.
- [18] Raiffa, H. et Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University.
- [19] Robert, C. (1992). *L'analyse statistique bayésienne*. Economica, Paris.
- [20] Stein, G.Z. et Juritz, J.M. (1988). Linear models with an inverse Gaussian-Poisson error distribution. *Communication in Statistics, A-Theory and Methods*, **17**, No 2, 557-571.
- [21] Stoer, J. et Burlirsch, R. (1980). *Introduction to numerical analysis*. Springer-Verlag, New York.
- [22] Tierney, L., Kass, R.E. et Kadane, J.B. (1989). Fully exponential Laplace approximations. *Journal of the American Statistical Association*, **84**, 710-716.