Université de Montréal

INFÉRENCE POUR LA FAMILLE DE SUNDT

par

Arié Haziza

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

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maitrise ès Sciences (Msc)
en statistique

Août 1997

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

INFÉRENCE POUR LA FAMILLE DE SUNDT

présenté par

Arié Haziza

a été évalué par un jury composé des personnes suivantes :

Urs Maag
(président-rapporteur)

Louis G. Doray
(directeur de recherche)

Narayan Giri
(membre du jury)

Mémoire accepté le :

17.12.1997

REMERCIEMENTS

Il m'a fallu des feuilles blanches, des crayons et puis du temps, de l'obstination et de la chance aussi. Il m'a fallu des rencontres, beaucoup de rencontres, des instants de soleil et des coins d'ombre quelquefois.

J'aimerais exprimer mes remerciements sincères à Monsieur Louis.G.Doray pour m'avoir introduit à un domaine fort intéressant de la recherche actuarielle, pour m'avoir soutenu certaines fois alors que le découragement me gagnait ainsi que pour son aide financière.

Qu'il me soit permis de remercier particulièrement Monsieur Farhat pour son aide précieuse tout au long de la rédaction. Ce mémoire salue la naissance de ses deux filles jumelles.

Ma reconnaissance va aussi à Monsieur Duchesne pour nos discussions informelles qui, sans doute, ont été fort bénéfiques pour moi et pour lui, je le souhaite.

Je ne peux tourner cette page sans saluer ceux qui, tout au long de ma scolarité, m'ont donné un peu d'eux-même. Ainsi, ma reconnaissance va à mes professeurs et en particulier à Madame Frigon, à Monsieur Yatracos et à Monsieur Angers, qui a accepté de travailler avec moi durant tout un été.

Enfin, j'exprime toute ma grattitude aux parents de ma mère, à mes parents, à mes frères ainsi qu'à mes oncles.

Je dédie ce mémoire aux parents de mon père qui auraient été si fiers de moi.

SOMMAIRE

Les familles de distributions discrètes définies de façon récursive jouent un rôle de premier plan dans la modélisation en assurance. Au chapitre 1, nous présentons trois de ces familles dont la dernière, la famille de Sundt, représente une généralisation des deux précédantes. Au chapitre 2, nous définissons les familles obtenues en tronquant ces trois familles. Nous considérons le problème d'estimation, par la méthode de distance quadratique minimale, pour ces familles tronquées. Nous énonçons et démontrons les propriétés des estimateurs obtenus. Puis, nous élargissons certains des résultats obtenus au cas des familles non-tronquées. Au chapitre 3, nous développons des tests d'hypothèses pour les familles tronquées. La méthode de distance quadratique minimale permet d'aborder l'inférence statistique de façon unifiée. Au chapitre 4, nous présentons un exemple illustrant les méthodes développées dans ce mémoire. Nous discutons aussi les avantages et certaines limites de ces méthodes.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

CHAPITRE 1

FAMILLES DISCRÈTES DÉFINIES DE FAÇON RÉCURSIVE

Ces dernières années, une partie de la recherche en théorie du risque, s'est concentrée sur le modèle d'assurance collectif pour une période finie.

Essentiellement, ce modèle suppose que, durant la période considérée, le montant total des réclamations des assurés d'une certaine compagnie d'assurance, noté S, peut s'écrire comme une somme aléatoire $S = \sum_{j=1}^{N} X_j$, X_j étant la variable aléatoire du montant de la j^{ieme} réclamation et N la variable aléatoire du nombre de réclamations survenues durant la période. N peut ainsi prendre les valeurs $0, 1, 2, 3, \ldots$. Une hypothèse de base de ce modèle consiste à supposer les X_j indépendantes, identiquement distribuées et indépendantes de N. Cette hypothèse permet de calculer la variance de S. De plus, une application du théorème limite centrale permet, sous certaines conditions de régularité, d'établir la prime suffisante pour avoir une grande probabilité de couvrir la perte.

La littérature classique s'est attardée en long et en large sur le cas où N est Poisson, ce qui a abouti au modèle fort connu des actuaires (voir Bowers et al. (1986)), le modèle Poisson composé. Une introduction intuitive à ce modèle est donnée par Adelson (1966). Une implication importante de ce premier modèle tient au fait que l'on peut établir la fonction de probabilité de S en utilisant une méthode récursive, lorsque les montants des réclamations sont des entiers non nuls. D'autres distributions discrètes ont ensuite été considérées comme alternatives à l'hypothèse Poissonienne, restrictive puisqu'elle suppose l'égalité de la moyenne et de la variance. Ainsi, la loi binomiale négative assouplit cette hypothèse mais on peut aussi citer les travaux de Willmot (1987) sur la distribution Poisson-inverse gaussienne.

Des familles plus vastes de lois discrètes qui contiennent comme cas particuliers certaines des distributions citées plus haut ont ensuite été définies. De cette manière, on assouplit l'hypothèse sur la distribution de N à considérer. Ainsi, à la section (1.1), on présente les deux premières familles, introduites par Panjer (1981) et par Sundt et Jewell (1981). De la même façon qu'avec la distribution de Poisson, on peut établir la loi de S en utilisant une méthode récursive. A ce sujet, Panjer a noté que la méthode habituelle pour évaluer la fonction de distribution de S requiert le calcul d'un grand nombre de convolutions de la distribution conditionelle du montant des réclamations, étant donné qu'une réclamation est survenue. Lorsque le nombre espéré de ces réclamations est élevé, les calculs peuvent devenir ardus, même avec les facilités actuelles. Cependant, lorsque le montant des réclamations est discret, on peut dériver la fonction de distribution de S sans utiliser de convolutions. Cela peut avoir comme effet de réduire de façon très significative les calculs, particulièrement pour des portefeuilles d'assurance de grande taille. Sundt et Jewell (1981) s'attardent aux propriétés de ces familles et présentent les distributions qui y appartiennent.

Schröter (1990) présente une première généralisation de la famille de Panjer. Il donne une caractérisation des distributions de cette famille ainsi que certaines propositions sur l'espace des paramètres car il n'a pu le décrire dans sa totalité. La famille de Schröter est présentée à la section (1.2). Puis Sundt (1992) définit une famille qui généralise les familles de Panjer et de Schröter. Une discussion exhaustive de cette famille est présentée à la section (1.3). La formule récursive pour S a aussi été étendue au cas où N est élément de la famille de Sundt.

1.1. Les familles $(a, b, 0, \infty)$ et $(a, b, 0, w)$

1.1.1. Généralités

Nous présentons d'abord la famille introduite par Panjer (1981).

Définition 1.1.1. *Soit N, une variable aléatoire discrète, prenant ses valeurs dans $\{0,1,2,3,.....\}$. N appartient à la famille $(a, b, 0, \infty)$, qu'on note $N \in \mathcal{F}\ (a, b, 0, \infty)$, si elle satisfait l'équation récursive suivante:*

$$p_n = (a + \frac{b}{n})p_{n-1}, \;\; n = 1, 2, 3... \tag{1.1.1}$$

où $p_n = P[N = n]$ et $(a, b)'$ est tel que (1.1.1) définit bien une loi de probabilité.

Sundt et Jewell (1981) démontrent que seules les lois de Poisson, binomiale et binomiale négative appartiennent à $\mathcal{F}\ (a, b, 0, \infty)$, avec les valeurs de a et b admissibles contenues au tableau (1.1).

Distribution	Valeur de a	Valeur de b
Poisson	0	$\lambda > 0$
Binomiale	$-\frac{p}{1-p}$, $p \in [0,1]$	$(m+1)\frac{p}{1-p}$, $m = 1,2,3,...$
Binomiale négative	$\frac{\beta}{1+\beta}$, $\beta > 0$	$(r-1)\frac{\beta}{1+\beta}$, r un réel positif

TABLEAU 1.1. Les distributions de la famille de Panjer

Pour $N \in \mathcal{F}(a,b,0,\infty)$, on vérifie que

$$\begin{aligned}
E(N) &= \mu = \sum_{n=0}^{\infty} np_n = \frac{a+b}{1-a}\ , \\
Var(N) &= \sigma^2 = \sum_{n=0}^{\infty} n^2p_n - \mu^2 = \frac{a+b}{(1-a)^2} \\
\text{et } \gamma_1 &= \frac{E(N-\mu)^3}{(\sigma^2)^{\frac{3}{2}}} = \frac{\sum_{n=0}^{\infty} n^3p_n - 3\mu\sigma^2 - \mu^3}{(\sigma^2)^{\frac{3}{2}}} \\
&= \frac{1+a}{(a+b)^{\frac{1}{2}}}\ ,
\end{aligned}$$

où γ_1 est le coefficient d'assymétrie de la variable N.

Parallèlement à Panjer (1981), Sundt et Jewell (1981) ont considéré une famille de lois discrètes où le domaine infini de la famille de Panjer est tronqué.

DÉFINITION 1.1.2. *Soit N, une variable aléatoire discrète, prenant ses valeurs dans $\{0,1,2,3,....,w\}$. $N \in \mathcal{F}(a,b,0,w)$, si elle satisfait l'équation récursive suivante:*

$$p_n = (a + \frac{b}{n})p_{n-1}, \quad n=1,2,3...,w. \tag{1.1.2}$$

où $p_n = P[N = n]$ et $(a,b)'$ est tel que (1.1.2) définit bien une loi de probabilité.

Nous illustrons ici la famille $(a,b,0,w)$ avec l'exemple suivant.

EXEMPLE 1.1.1. *Considérons une variable aléatoire N qui suit une loi binomiale négative(r,β) telle que définie plus haut. Soit N^* définie comme suit:*

$$\begin{aligned}
p[N^*=n] &= p_n^* = \left[\sum_{j=0}^{w} \frac{\Gamma(r+j)}{\Gamma(r)j!}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^j\right]^{-1} \frac{\Gamma(r+n)}{\Gamma(r)n!}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^n \\
&= \lambda\frac{\Gamma(r+n)}{\Gamma(r)n!}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^n, \ n = 0,1,2,\ldots..,w.
\end{aligned}$$

On peut montrer que la relation récursive suivante tient:

$$\begin{aligned}
p_n^* &= (a+\frac{b}{n})p_{n-1}^*, \ n=1,2,3\ldots\ ,w, \text{ avec} \\
a &= \frac{\beta}{1+\beta} \text{ et } b=(r-1)a.
\end{aligned}$$

Ainsi, on voit qu'en général, si $N \in \mathcal{F}(a,b,0,\infty)$, il est possible de définir une variable $N^* \in \mathcal{F}(a,b,0,w)$, en tronquant les valeurs de N supérieures à k.

1.2. La famille de Schröter

1.2.1. Généralités

Schröter (1990) publie ses résultats sur une première généralisation de la famille $(a,b,0,\infty)$.

DÉFINITION 1.2.1. *Soit N, une variable aléatoire discrète, prenant ses valeurs dans $\{0,1,2,3,\ldots..\}$. N appartient à la famille de Schröter, si elle satisfait l'équation récursive suivante:*

$$p_n = (a+\frac{b}{n})p_{n-1} + \frac{c}{n}p_{n-2}, p_{-1}=0, n=1,2,3\ldots \qquad (1.2.1)$$

où $a < 1$ et $(a,b,c)'$ est tel que (1.2.1) définit une loi de probabilité valide.

On note par $R2(a,b,c)$ la distribution de N. La famille $(a,b,0,\infty)$ est incluse dans la famille de Schröter en posant $c=0$. Dans le cas où $a+b+c=0$, (1.2.1) est réduite à la loi dégénérée à 0.

1.2.2. Quelques propriétés de la famille de Schröter

Soit $\Psi(s)$ la fonction génératrice des probabilités de N:

$$\Psi(s) \quad = \quad E[s^n] = \sum_{i=1}^{\infty} s^n p_n,\ s \in [0,1]. \tag{1.2.2}$$

Rappelons qu'il y a une relation biunivoque entre Ψ(s) et la distribution de N. De plus, Ψ(s) permet de trouver la forme explicite de la fonction de probabilité de N, en se servant du fait que

$$p_n \quad = \quad \frac{1}{n!}\frac{d^n}{ds^n}\Psi(s)|_{s=0}. \tag{1.2.3}$$

On montre que

$$\begin{aligned} \Psi^{'}(s) &= as\Psi^{'}(s) + (a+b+cs)\Psi(s);\ \text{d'où} && (1.2.4) \\ \frac{d}{ds}ln\Psi(s) &= \frac{a+b+cs}{1-as} && (1.2.5) \end{aligned}$$

Une condition nécessaire et suffisante pour que N appartienne à la famille de Schröter est que $\frac{d}{ds}ln\Psi(s)$ puisse s'écrire comme le rapport de deux polynômes en s, celui du numérateur avec un degré d'au plus 1 et celui du dénominateur avec un degré d'au plus 1 et un terme constant égal à 1.

A la prochaine section, nous aurons un résultat analogue à celui-ci pour une famille plus vaste, la famille de Sundt. Pour trouver la forme explicite de p_n, en se servant de (1.2.3), il faut résoudre l'équation différentielle (1.2.5) en distinguant les cas où $a=0$ et $a\neq 0$.

Pour a $\neq 0$, nous avons

$$\Psi(s) = \left(\frac{1-a}{1-as}\right)^{\frac{a(a+b)+c}{a^2}}$$

et pour $a = 0$, on obtient

$$\Psi(s) = e^{\frac{c}{2}(s^2-1)+b(s-1)}.$$

Pour une variable aléatoire N dans cette famille, les expressions pour la moyenne et la variance sont

$$E(N) = \frac{a+b+c}{1-a}$$
$$Var(N) = \frac{a+b+(2-a)c}{(1-a)^2}$$

Schröter (1990) a étudié l'espace possible des paramètres a, b et c. Il ne l'a pas décrit entièrement mais il propose un certain nombre de caractéristiques. A la prochaine section, nous verrons que Sundt (1992) qui a introduit une famille plus vaste, ne s'est pas aventuré non plus à étudier les caractéristiques de l'espace paramétrique. Comme membre de la famille de Schröter, on peut citer la convolution d'une Poisson avec les éléments de $\mathcal{F}$(a,b,0,∞) . En particulier, la convolution d'une loi de Poisson et d'une loi binomiale négative, connue sous le nom de distribution de Delaporte est traitée par Willmot et Sundt (1989). A la prochaine section, nous obtiendrons des résultats plus élaborés en nous situant dans le contexte d'une famille plus générale que celle-ci.

1.3. La famille de Sundt

1.3.1. Généralités

Dans cette section, on présente une généralisation des familles des sections précédentes, due à Sundt (1992).

Définition 1.3.1. *Soit N, une variable aléatoire discrète, prenant ses valeurs dans {0,1,2,3,...,}. $N \in \mathcal{R}_k$, si elle satisfait l'équation récursive suivante:*

$$p_n = \sum_{i=1}^{k}(a_i + \frac{b_i}{n})p_{n-i}, \ k \geq 1, \ p_{-n} = 0 \text{ pour } n > 0, n = 1, 2, 3... \quad (1.3.1)$$

où $a = (a_1, a_2, ..., a_k)'$ et $b = (b_1, b_2, ..., b_k)'$ sont tels que (1.3.1) définit bien une loi de probabilité.

Notons par $\mathcal{R}_k[a, b]$ la distribution de N. On a clairement que $\mathcal{R}_{k-1} \subset \mathcal{R}_k$. La famille $(a, b, 0, \infty)$ est équivalente à $\mathcal{R}_1$. En effet, il s'agira, pour une distribution $N \in \mathcal{F}(a, b, 0, \infty)$, de poser $a_i = b_i = 0$ pour $i = 2, 3, ..., k$. La famille de Schröter est équivalente à $\mathcal{R}_2$ en posant $a_2 = a_i = b_i = 0$ pour $i = 3, 4, ..., k$. Dans le cas où $a_i = b_i = 0$ pour $i = 1, 2, , ..., k$, (1.3.1) se réduit à la loi dégénérée en 0. Aussi, la représentation (1.3.1) implique $p_0 > 0$.

1.3.2. Quelques propriétés de $\mathcal{R}_k$

Soit $N \in \mathcal{R}_k$ et soit Ψ(s) sa fonction génératrice des probabilités.

$$\text{Définissons } \rho(s) = \frac{d}{ds}ln\Psi(s) = \frac{\Psi'(s)}{\Psi(s)} = \frac{\sum_{i=1}^{k}(ia_i + b_i)s^{i-1}}{1 - \sum_{i=1}^{k} a_i s^i}. \quad (1.3.2)$$

D'où le théorème suivant:

THÉORÈME 1.3.2.1. *$N \in \mathcal{R}_k$ si et seulement si $\rho(s)$ s'exprime comme un rapport de deux polynômes, celui du numérateur avec un degré d'au plus k-1 et celui du dénominateur avec un degré d'au plus k et un terme constant égale à 1.*

En multipliant (1.3.2) par $(1+qs)$ où q est un nombre quelconque, on peut écrire:

$$\rho(s) = \frac{\sum_{i=1}^{k+1}(ic_i + d_i)s^{i-1}}{1 - \sum_{i=1}^{k+1} c_i s^i},$$

avec

$$\begin{aligned} c_i &= a_i + q a_{i-1}, \\ d_i &= b_i + q(b_{i-1} - a_{i-1}),\ i = 1, 2, 3..., k+1\ , \\ a_0 &= -1,\ b_0 = b_{k+1} = a_{k+1} = 0. \end{aligned}$$

D'où $\mathcal{R}_k[a,b] \equiv \mathcal{R}_{k+1}[c,d]$, $c = (c_1, ..., c_{k+1})'$, $d = (d_1, ..., d_{k+1})'$.

On montre aussi que pour $N \in \mathcal{R}_k$, on a :

$$E(N) = \frac{\sum_{i=1}^{k}(ia_i + b_i)}{1 - \sum_{i=1}^{k} a_i} \tag{1.3.3}$$

$$Var(N) = \frac{\sum_{i=1}^{k} i\left[(i + E(N)a_i + b_i)\right]}{1 - \sum_{i=1}^{k} a_i} \tag{1.3.4}$$

On voit que (1.3.3) et (1.3.4) généralisent les formules des moyennes et variances données aux sections antérieures.

1.3.3. Un exemple

Nous illustrons la section 1.2.2 dans le cas où N suit une loi binomiale négative(r,β) telle que définie à la section 1.1. On a déjà montré qu'on a la relation récursive suivante pour cette distribution :

$$p_n = \left(\frac{\beta}{1+\beta} + \frac{(r-1)}{n}\frac{\beta}{1+\beta}\right) p_{n-1} \ , \ n = 1, 2, 3, ...,$$

D'où $N \in \mathcal{R}_1$. Notons cette distribution par $\mathcal{R}_1 \left[a = \frac{\beta}{1+\beta}, b = (r-1)\frac{\beta}{1+\beta}\right]$.

Une autre façon de dériver ce résultat consiste à considérer $\rho(s)$. Ainsi, on a:

$$\begin{aligned}
\Psi(s) &= \sum_{n=0}^{\infty} \frac{\Gamma(r+n)}{\Gamma(r)n!} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta s}{1+\beta}\right)^n \\
&= \sum_{n=0}^{\infty} \binom{-r}{n} \left(\frac{1}{1+\beta}\right)^r \left(\frac{-\beta s}{1+\beta}\right)^n \\
&= \{1 - \beta(s-1)\}^{-r}, |s| < \frac{1+\beta}{\beta}.
\end{aligned}$$

D'où

$$\rho(s) = \frac{\beta r}{1 - \beta(s-1)} \ , \ |s| < \frac{1+\beta}{\beta} \ ,$$

qu'on peut écrire comme

$$\rho(s) = \frac{\beta r - \beta + \beta}{1 - \beta(s-1)} = \frac{(r-1)\beta + \beta}{1 - \beta(s-1)}$$

$$= \frac{\frac{(r-1)\beta+\beta}{1+\beta}}{\frac{1+\beta}{1+\beta} - \frac{\beta s}{1+\beta}} = \frac{\frac{\beta}{1+\beta} + (r-1)\frac{\beta}{1+\beta}}{1 - \frac{\beta}{1+\beta}s} \qquad (1.3.5)$$

D'aprés le théorème (1.3.2.1), il vient:

$$a_1 = \frac{\beta}{1+\beta}, \ b_1 = (r-1)\frac{\beta}{1+\beta}.$$

$N \in \mathcal{R}_1$ et on note sa distribution par $\mathcal{R}_1 \left[a = \frac{\beta}{1+\beta}, b = (r-1)\frac{\beta}{1+\beta}\right]$.

Aussi, multipliant (1.3.5) par $(1+qs)$, pour tout q, on obtient:

$$\rho(s) = \frac{\frac{\beta}{1+\beta} + (r-1)\frac{\beta}{1+\beta} + \frac{\beta}{1+\beta}qs + (r-1)\frac{\beta}{1+\beta}qs}{1 - \left(\frac{\beta}{1+\beta} + q\right)s - \frac{\beta}{1+\beta}qs^2}.$$

On peut ainsi conclure que $N \in \mathcal{R}_2$ et que sa distribution est

$$\mathcal{R}_2 \left[a = (\frac{\beta}{1+\beta} + q, \frac{\beta}{1+\beta}q)', b = \left((r-1)\frac{\beta}{1+\beta} - q, r\frac{\beta}{1+\beta}q - 2(\frac{\beta}{1+\beta})q\right)'\right].$$

1.3.4. Les convolutions

Soient N et M , $N \in \mathcal{R}_k$ et $M \in \mathcal{R}_l$ ayant pour distributions respectives $\mathcal{R}_k[a,b]$ et $\mathcal{R}_l[c,d]$ ainsi que leurs fonctions génératrices des probabilités $\Psi_1(s)$ et $\Psi_2(s)$. On supposera de plus que N et M sont des variables aléatoires indépendantes.

On est intéressé ici à la distribution de la variable $O = N + M$. Posons $\Psi(s)$ sa fonction génératrice des probabilités. On a alors

$$\begin{aligned}
\rho(s) &= \frac{d}{ds}ln\Psi(s) = \frac{d}{ds}ln[\Psi_1(s)\Psi_2(s)] = \frac{d}{ds}ln\Psi_1(s) + \frac{d}{ds}ln\Psi_2(s) \\
&= \frac{\sum_{i=1}^{k}(ia_i + b_i)s^{i-1}}{1 - \sum_{i=1}^{k} a_i s^i} + \frac{\sum_{i=1}^{l}(ic_i + d_i)s^{i-1}}{1 - \sum_{i=1}^{l} c_i s^i} \\
&= \frac{\left(\sum_{i=1}^{k}(ia_i + b_i)s^{i-1}\right)\left(1 - \sum_{i=1}^{l} c_i s^i\right) + \left(\sum_{i=1}^{l}(ic_i + d_i)s^{i-1}\right)\left(1 - \sum_{i=1}^{k} a_i s^i\right)}{\left(1 - \sum_{i=1}^{k} a_i s^i\right)\left(1 - \sum_{i=1}^{l} c_i s^i\right)}.
\end{aligned}$$

Le degré du polynôme du numérateur est d'au plus $k+l-1$ et celui du dénominateur d'au plus $k+l$. Par le théorème (1.3.2.1), on tire le corollaire suivant:

COROLLAIRE 1.3.4.1. *La convolution de $\mathcal{R}_k[a,b]$ et de $\mathcal{R}_l[c,d]$ appartient à $\mathcal{R}_{k+l}$.*

Remarquons que si les polynômes du dénominateur ont un facteur en commun de degré q, la convolution de $\mathcal{R}_k[a,b]$ et de $\mathcal{R}_l[c,d]$ appartiendra à $\mathcal{R}_{k+l-q}$.

Dans le cas où $l=k$ et $c=a$, on obtient:

$$\begin{aligned}\rho(s) &= \frac{d}{ds} ln\Psi(s) = \frac{\sum_{i=1}^{k}(ia_i+b_i)s^{i-1}}{1-\sum_{i=1}^{k}a_is^i} + \frac{\sum_{i=1}^{k}(ia_i+d_i)s^{i-1}}{1-\sum_{i=1}^{k}a_is^i} \\ &= \frac{\sum_{i=1}^{k}\left\{ia_i+(ia_i+b_i+d_i)\right\}s^{i-1}}{1-\sum_{i=1}^{k}a_is^i}.\end{aligned}$$

Encore par le théorème (1.3.2.1), on dérive le second corollaire:

COROLLAIRE 1.3.4.2. *La convolution de $\mathcal{R}_k[a,b]$ et de $\mathcal{R}_k[a,d]$ est une distribution $\mathcal{R}_k[a,e]$, avec $a=(a_1,a_2,...,a_k)'$ et $e=(e_1,...,e_k)'$ où $e_i=ia_i+b_i+d_i$, i=1,2,....,k.*

Sundt (1992) généralise le corollaire (1.3.4.2) au cas où on considère la convolution de $\mathcal{R}_k[a,b^{(1)}], \mathcal{R}_k[a,b^{(2)}], ..., \mathcal{R}_k[a,b^{(m)}]$, qu'on suppose indépendantes, où les $b^{(j)}$ sont des vecteurs de dimension k. Il démontre que cette convolution a pour distribution $\mathcal{R}_k[c,\beta]$ avec $a=(a_1,a_2,...,a_k)'$ et $\beta=(\beta_1,\beta_2,...,\beta_k)'$ où $\beta_i=(m-1)ia_i+\sum_{j=1}^{m}b_i^{(j)}$, $i=1,2,....,k$.

Ainsi la distribution de la m^{ieme} convolution de $\mathcal{R}_k[a,b]$ est $\mathcal{R}_k[a,\beta]$ avec $a = (a_1, a_2, ..., a_k)'$ et $\beta = (\beta_1, \beta_2, ..., \beta_k)$ où $\beta_i = (m-1)ia_i + mb_i$, $i = 1, 2, ..., k$. Aussi, la convolution de $\mathcal{R}_1[a, b^{(1)}], \mathcal{R}_1[a, b^{(2)}], ..., \mathcal{R}_1[a, b^{(m)}]$ est $\mathcal{R}_1[a,\beta]$ où $\beta = (m-1)a + \sum_{j=1}^{m} b_j$.

Nous pouvons vérifier ici deux résultats classiques. Le premier stipule que la convolution de m variables aléatoires Poisson(λ_j), $j = 1, 2, ..., m$, indépendantes, suit une loi de Poisson($\sum_{j=1}^{m} \lambda_j$). En effet, nous posons N_j et sa distribution $\mathcal{R}_1[a = 0, b = \lambda_j]$, $j = 1, 2, ..., m$, qui est la distribution d'une loi de Poisson(λ_j). On a alors, par ce qui est écrit plus haut, que $\sum_{j=1}^{m} N_j$ a une distribution $\mathcal{R}_1[a = 0, b = (m-1).0 + \sum_{j=1}^{m} \lambda_j]$ qui est la distribution d'une loi de Poisson(λ_j). Le second concerne la convolution de m variables aléatoires binomiale négative (r_j, β), $j = 1, 2, ..., m$. De la même façon que plus haut, prenons N_j et sa distribution $\mathcal{R}_1[a = \frac{\beta}{1+\beta}, b = (r-1)\frac{\beta}{1+\beta}]$, $j = 1, 2, ..., m$. On a alors que $\sum_{j=1}^{m} N_j$ a pour distribution $\mathcal{R}_1[a = \frac{\beta}{1+\beta}, b = (m-1)\frac{\beta}{1+\beta} + \frac{\beta}{1+\beta}(\sum_{j=1}^{m} r_j - m)]$ ou $\mathcal{R}_1[a = \frac{\beta}{1+\beta}, b = (\sum_{j=1}^{m} r_j - 1)\frac{\beta}{1+\beta}]$ qui est la distribution d'une loi binomiale négative($\sum_{j=1}^{m} r_j, \beta$).

Jusqu'ici, nous n'avons présenté que les convolutions dans le cas où a est identique pour les m distributions de $\mathcal{R}_k$. Le théorème suivant, démontré par Sundt (1992), généralise la situation, dans le cas où les m distributions appartiennent à $\mathcal{R}_1$.

Théorème 1.3.4.1. *La convolution de* $\mathcal{R}_1[a_1,b_1], \mathcal{R}_1[a_2,b_2], ..., \mathcal{R}_1[a_m,b_m]$ *est* $\mathcal{R}_1[\alpha,\beta]$, *où*

$$\begin{aligned}
\alpha_i &= (-1)^{i+1} \sum_{1\le j_1<j_2<...<j_i\le m} \prod_{k=1}^{i} a_{j_k}, i = 1,2,...,m, \\
\beta_i &= (-1)^{i+1} \sum_{r=1}^{m} b_r \sum_{\substack{1\le j_1<j_2<...<j_{i-1}\le m \\ j_t\neq r(t=1,...,i-1)}} \prod_{k=1}^{i-1} a_{j_k}, i = 2,...,m \\
\beta_1 &= \sum_{j=1}^{m} b_j.
\end{aligned}$$

Il découle du théorème (1.3.4.1) que la convolution de $\mathcal{R}_1[a_1,b_1]$ et de $\mathcal{R}_1[a_2,b_2]$ est $\mathcal{R}_2[\alpha = (a_1+a_2, a_1a_2)', \beta = (b_1+b_2, -(a_1b_2+a_2b_1))']$.

Exemple 1.3.1. *La convolution d'une loi binomiale(r,p) et d'une loi binomiale négative(s,q) est une distribution* $\mathcal{R}_2[\alpha,\beta]$, *où*

$$\begin{aligned}
\alpha &= \left(\frac{-p}{1-p} + \frac{q}{1+q}, \frac{pq}{(1-p)(1+q)}\right)', \\
\beta &= \left(\frac{(r+1)p}{1-p} + \frac{(s-1)q}{1+q}, \frac{-pq(r-s+2)}{(1-p)(1+q)}\right)'.
\end{aligned}$$

Exemple 1.3.2. *La convolution d'une loi binomiale(r,p) et d'une loi de Poisson(λ) est une distribution*

$$\mathcal{R}_2\left[\alpha = (\frac{-p}{1-p}, 0)', \beta = \left(\frac{(r+1)p}{1-p} + \lambda, \frac{p\lambda}{1-p}\right)'\right].$$

Exemple 1.3.3. *La convolution d'une loi de Poisson(λ) et d'une loi binomiale négative(s,q) est une distribution*

$$\mathcal{R}_2\left[\alpha = (\frac{q}{1+q}, 0)', \beta = \left(\frac{(s-1)q}{1+q} + \lambda, \frac{-q\lambda}{1+q}\right)'\right].$$

On obtient aussi du théorème (1.3.4.1) que la convolution de $\mathcal{R}_1[a_1, b_1]$, $\mathcal{R}_1[a_2, b_2]$ et de $\mathcal{R}_1[a_3, b_3]$ est $\mathcal{R}_3[\alpha, \beta]$, avec

$$\begin{aligned} \alpha &= (a_1 + a_2 + a_3, -(a_1a_2 + a_1a_3 + a_2a_3), a_1a_2a_3)', \\ \beta &= (b_1 + b_2 + b_3, -(a_1b_2b_3 + a_2b_1b_3 + a_3b_1b_2), a_1a_2b_3 + a_1a_3b_2 + a_2a_3b_1)'. \end{aligned}$$

EXEMPLE 1.3.4. *La convolution d'une loi de Poisson(λ), d'une loi binomiale(r,p) et d'une Binomiale négative(s,q) est une distribution $\mathcal{R}_3[\alpha, \beta]$, où:*

$$\begin{aligned} \alpha &= \left(\frac{-p}{1-p} + \frac{q}{1+q}, \frac{pq}{(1-p)(1+q)}, 0\right)', \\ \beta &= \left(\lambda + \frac{(r+1)p}{1-p} + \frac{(s-1)q}{1+q}, \frac{-pq\lambda(r-s+2)}{(1-p)(1+q)}, \frac{-pq\lambda}{(1-p)(1+q)}\right)'. \end{aligned}$$

1.3.5. Les mixtures

Sundt (1992) présente les mixtures de deux distributions $\mathcal{R}_k[a, b]$ et $\mathcal{R}_\ell[c, d]$. Nous allons ici élargir le contexte aux combinaisons convexes de m distributions. Soient N_1, N_2,..., N_m, leur distribution respective $\mathcal{R}_{k_1}[a_1, b_1], \mathcal{R}_{k_2}[a_2, b_2]$,... et $\mathcal{R}_{k_m}[a_m, b_m]$, ainsi que leur fonction génératrice des probabilités $\Psi_1(s), \Psi_2(s), \ldots.$, $\Psi_m(s)$. Nous pouvons, comme cas particulier, prendre $k_1 = k_2 = \ldots.. = k_m$. Considérons la mixture de ces distributions:

$$\sum_{i=1}^{m} \alpha_i \mathcal{R}_{k_i}[a_i, b_i] \text{ où } \sum_{i=1}^{m} \alpha_i = 1,\ \alpha_i \in]0, 1[.$$

On obtient une variable aléatoire discrète. Posons P_n la mixture du haut et $\Psi(s)$ sa fonction génératrice des probabilités. On montre aisément que

$$\begin{aligned} \Psi(s) &= \sum_{i=1}^{m} \alpha_i \Psi_i(s), \text{ d'où} \\ \rho(s) &= \frac{\sum_{i=1}^{m} \alpha_i \frac{d}{ds}\Psi_i(s)}{\sum_{i=1}^{m} \alpha_i \Psi_i(s)}. \end{aligned}$$

Maintenant, à l'aide du théorème (1.3.2.1), nous pouvons vérifier si P_n est une distribution de $\mathcal{R}_{k_*}$, pour un k_* fini. En effet, il est intéressant de noter ici qu'une mixture d'éléments de la famille de Sundt ne fait pas forcément partie de cette famille. L'exemple le plus probant consiste en la mixture de m lois de Poisson(λ_i), $i = 1, 2, 3..., m$ où les λ_i sont différents. On a alors que $\Psi_i(s) = e^{\lambda_i(s-1)}$ et $\frac{d}{ds}\Psi_i(s)$ $= \lambda_i e^{\lambda_i(s-1)}$.

D'où

$$\rho(s) = \frac{\sum_{i=1}^{m} \alpha_i \lambda_i e^{\lambda_i(s-1)}}{\sum_{i=1}^{m} \alpha_i e^{\lambda_i(s-1)}},$$

qu'on ne peut écrire comme un rapport de deux polynômes. Ainsi, cette mixture n'appartient pas à la famille de Sundt. Une autre situation consiste à considérer deux Binomiales négatives (r_1, β) et (r_2, β) avec $r_1 < r_2$. On peut montrer que si $r_2 - r_1$ est un nombre entier, la mixture des distributions du haut appartient à $\mathcal{R}_{r_2 - r_1 + 1}$ alors que si $r_2 - r_1$ n'est pas entier, cette mixture n'appartient pas à la famille de Sundt.

1.4. Autre formes récursives

Dans ce mémoire, nous sommes intéréssés par les distributions qui admettent la représentation (1.3.1). Nous avons vu que cette représentation implique $p_0 > 0$.

Sundt et Jewell (1981) ont introduit une modification de la famille de distributions discrètes tronquée à 0 $\mathcal{F}(a, b, 1, \infty)$ qui admet la forme récursive:

$$p_n \quad = \quad (a + \frac{b}{n})p_{n-1} \ , \ p_0 = 0 \ , \ p_1 > 0 \ , \ \text{n=2,3,4,...}$$

Ils ont montré que les lois de Poisson tronquée à 0, binomiale tronquée à 0, binomiale négative tronquée à 0, ainsi que la distribution logarithmique, appartiennent à cette famille. Willmot (1987) ajoute à cette famille la distribution ETNB (loi binomiale négative tronquée à 0 avec $-1 < r < 0$) . Seules ces distributions y font partie. Une première généralisation consiste à considérer la forme récursive

$$p_n^* \quad = \quad (a + \frac{b}{n})p_{n-1}^* \ , \ p_0^* = 0 \ , \ p_1^* > 0 \ , \ n = 2, 3, 4, ..., w.$$

La famille formée des distributions qui admettent cette forme récursive est appelée $\mathcal{F}(a, b, 1, w)$. En général, on peut introduire la famille des distributions qui admettent la forme récursive:

$$p_n^* \quad = \quad (a + \frac{b}{n})p_{n-1}^* \ , \ p_0^* = 0 \ , \ p_1^* = 0 \ ,...., \ p_{c-2}^* = 0 \ , \ p_{c-1}^* > 0 \ , \ n = c, c+1, ..., w \ , \ c > 2,$$

qu'on nomme $\mathcal{F}(a, b, c-1, w)$.

On peut donner comme exemple de distribution qui appartient à $\mathcal{F}$(a,b,c-1,w) la loi de Poisson tronquée(λ), $\lambda > 0$:

$$p_n^* \quad = \quad \left[\sum_{j=c}^{w} \frac{\lambda^j e^{-\lambda}}{j!}\right]^{-1} \frac{\lambda^n e^{-\lambda}}{n!} \ , \ n = c, c+1, ..., w.$$

Au chapitre 2, nous introduisons de nouvelles familles définies récursivement, qui vont servir comme base de travail. On présente une méthode d'estimation, par la distance quadratique minimale, pour estimer les paramètres de N, qu'on suppose appartenir aux familles citées plus haut. Les trois familles présentées dans ce premier chapitre ont une caractéristique commune très importante puisqu'on trouve là l'essence même de la méthode d'estimation présentée: elles admettent toutes une représentation récursive.

Au chapitre 3, nous traitons le problème des tests d'hypothèse dans le cadre de ces familles.

Au chapitre 4, nous illustrons, à l'aide d'un exemple, les méthodes présentées dans ce mémoire.

CHAPITRE 2

ESTIMATION PAR LA MÉTHODE DE DISTANCE QUADRATIQUE MINIMALE

Au chapitre précédent, nous avons présenté des familles de lois discrètes définies de façon récursive, dont la famille de Sundt représente une généralisation.

Luong et Garrido (1993) ont proposé une méthode d'estimation pour la famille $\mathcal{F}(a, b, 0, w)$ et élargi les résultats à $\mathcal{F}(a, b, 0, \infty)$. A la section 2.1, nous exposons la démarche de leur recherche. A la section 2.2, nous généralisons cette méthode au cas de la famille de Sundt. Nous débutons d'abord par la famille $\mathcal{R}_2$ tronquée que nous définissons. Nous verrons qu'à partir de $\mathcal{R}_2$, il est facile de définir des distributions de $\mathcal{R}_2$ tronquées. Nous démontrons que les estimateurs obtenus sont asymptotiquement normaux, convergents en probabilité et asymptotiquement efficaces. La robustesse de l'estimateur est aussi explorée. Certaines difficultés sont relevées quant à la nature de l'espace paramétrique. Nous élargissons ensuite les résultats au cas de la famille $\mathcal{R}_2$ qui généralise la famille de Schröter.

Enfin, nous considérons le cas général de la famille $\mathcal{R}_k$ où k est un entier positif fini quelconque. Nous voyons que les résultats pour la famille $\mathcal{R}_2$ se généralisent à la famille $\mathcal{R}_k$.

2.1. Estimation pour la famille $\mathcal{F}(a, b, 0, w)$

Nous débutons tout d'abord par la famille intoduite par Sundt et Jewell (1981). On suppose que $N \in \mathcal{F}(a, b, 0, w)$. Par (1.1.2), on a

$$\begin{aligned}
p_1^* &= (a + \frac{b}{1})p_0^* \\
p_2^* &= (a + \frac{b}{2})p_1^* = p_0^* \prod_{n=1}^{2}(a + \frac{b}{n}) \\
\ldots. &= \ldots. \\
p_w^* &= (a + \frac{b}{w})p_{w-1}^* = p_0^* \prod_{n=1}^{w}(a + \frac{b}{n})
\end{aligned}$$

Comme $\sum_{n=0}^{w} p_n^* = 1$, on obtient une expression pour p_0^*:

$$p_0^* = \left[1 + \sum_{n=1}^{w} \prod_{j=1}^{n}(a + \frac{b}{j})\right]^{-1}.$$

Cela nous permet de dériver une expression pour p_n^*, en fonction de a et b:

$$p_n^* = \left[\prod_{j=1}^{n}(a + \frac{b}{j})\right]\left[1 + \sum_{n=1}^{w} \prod_{j=1}^{n}(a + \frac{b}{j})\right]^{-1}, \; n = 1, 2, 3..., w. \qquad (2.1.1)$$

Soient $N_1, N_2, ..., N_m$, des variables aléatoires i.i.d provenant de la famille $\mathcal{R}_k$.

Pour estimer $\theta = (a, b)'$ par la méthode du maximum de vraisemblance, il faut maximiser

$$l(a, b) = \sum_{n=0}^{w} f_n^* ln(p_n^*)$$

où f_n^* est la fréquence observée de $n = 0, 1, 2, 3..., w$. L'estimateur du maximum de vraisemblance (EMV) de θ, dénoté $\hat{\theta} = (\hat{a}, \hat{b})$, est la solution du système d'équations suivant:

$$\frac{\partial l(a,b)}{\partial a}|_{\theta=\hat{\theta}} =$$

$$\sum_{n=0}^{w} f_n^* \left\{ \sum_{j=1}^{n} (\hat{a} + \frac{\hat{b}}{j})^{-1} - \left[1 + \sum_{h=1}^{w} \prod_{j=1}^{h} (\hat{a} + \frac{\hat{b}}{j}) \right]^{-1} \left[\sum_{h=1}^{w} (\sum_{j=1}^{h} (\hat{a} + \frac{\hat{b}}{j}))(\prod_{j=1}^{h} (\hat{a} + \frac{\hat{b}}{j})) \right] \right\} = 0,$$

$$\frac{\partial l(a,b)}{\partial b}|_{\theta=\hat{\theta}} =$$

$$\sum_{n=0}^{w} f_n^* \left\{ \sum_{j=1}^{n} (\hat{a}j + \hat{b})^{-1} - \left[1 + \sum_{h=1}^{w} \prod_{j=1}^{h} (\hat{a} + \frac{\hat{b}}{j}) \right]^{-1} \left[\sum_{n=1}^{w} (\sum_{h=1}^{n} (\hat{a}j + \hat{b})^{-1})(\prod_{j=1}^{n} (\hat{a} + \frac{\hat{b}}{j})) \right] \right\} = 0.$$

On constate que l'EMV est difficile à calculer, car on doit trouver les racines de polynômes de degré élevé. On doit aussi être en mesure de distinguer les maximums locaux du maximum global. C'est dans cet esprit que Luong et Garrido ont proposé une méthode alternative d'estimation.

A partir de l'échantillon observé $n_1, n_2, ..., n_m$, on estime les probabilités p_n^* à l'aide des fréquences observées

$$\hat{p}_n^* = \frac{f_n^*}{m} , n = 0, 1, 2, ..., w.$$

Les équations récursives, linéaires en $(a, b)'$, suggèrent le modèle de régression linéaire suivant:

$$\hat{p}_n^* = \hat{p}_{n-1}^* a + (\frac{\hat{p}_{n-1}^*}{n}) b + \varepsilon_n , n = 1, 2, ..., w, \tag{2.1.2}$$

où ε_n est une erreur aléatoire de moyenne 0.

Posons:

$$\hat{Y} = (\hat{p}_1^*, \hat{p}_2^*, ..., \hat{p}_w^*)'$$

$$\hat{X} = \begin{pmatrix} \hat{p}_0^* & \hat{p}_0^* \\ \hat{p}_1^* & \hat{p}_1^*/2 \\ \hat{p}_2^* & \hat{p}_2^*/3 \\ ..... & .... \\ \hat{p}_{w-1}^* & \hat{p}_{w-1}^*/w \end{pmatrix}$$

$$\Xi = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_w)'.$$

Sous forme matricielle, le modèle (2.1.2) s'écrit comme $\hat{Y} = \hat{X}\theta + \Xi$.

Si $m \to \infty$, $\hat{X}$ est une matrice de plein rang avec probabilité 1, et elle tend en probabilité vers la matrice X (qu'on note $\hat{X} \xrightarrow{p} X$) définie comme

$$X = \begin{pmatrix} p_0^* & p_0^* \\ p_1^* & p_1^*/2 \\ p_2^* & p_2^*/3 \\ ..... & .... \\ p_{w-1}^* & p_{w-1}^*/w \end{pmatrix}$$

En effet, $\hat{X}[i,j] \xrightarrow{p} X[i,j]$, pour $i = 1, 2, ..., w$ et $j = 1, 2$. Il suit (voir Serfling (1980), p.6) que $\hat{X} \xrightarrow{p} X$. Et comme $E(\hat{p}_i^*) = p_i^*$, $E(\hat{X}) = X$.

Huard (1995) a montré que

$$\begin{aligned} E(\varepsilon_n) &= 0 \ , \ n = 1, 2, ..., w \\ Var(\varepsilon_n) &= \frac{1}{m}\left(\frac{p_n^* p_{n-1}^* + p_n^{*2}}{p_{n-1}^*}\right) \ , \ n = 1, 2, ..., w \\ Cov(\varepsilon_n, \varepsilon_{n+1}) &= -\frac{1}{m} p_{n+1}^* \ , \ n = 1, 2, ..., w \\ Cov(\varepsilon_n, \varepsilon_{n+\ell}) &= 0 \ , \ \ell > 1, \ n = 1, 2, ..., w, \end{aligned}$$

et que la matrice de variance-covariance de Ξ est égale à

$$
\begin{aligned}
Var(\Xi) = \textstyle\sum_\theta(\Xi) &= \frac{1}{m}\begin{pmatrix}
\frac{p_0^*p_1^*+p_1^{*2}}{p_0^*} & -p_2^* & 0 & .... & 0 \\
-p_2^* & \frac{p_1^*p_2^*+p_2^{*2}}{p_1^*} & -p_3^* & .... & 0 \\
.... & .... & .... & .... & .... \\
0 & .... & .... & -p_w^* & \frac{p_{w-1}^*p_w^*+p_w^{*2}}{p_{w-1}^*}
\end{pmatrix} \\
&= \frac{1}{m}\textstyle\sum_\theta^*(\Xi), \text{ où } \sum_\theta^*(\Xi) \text{ est la matrice ci-haut.}
\end{aligned}
$$

Par (2.1.1), on voit que $\sum_\theta^*(\Xi)$ est fonction de $\theta = (a,b)'$.

Dans un modèle standard de régression, on considère le problème de minimisation de la somme des carrés des résidus pour obtenir l'estimateur des moindres carrés du vecteur des paramètres. Dans notre contexte, nous devons tenir compte de la structure particulière de $\sum_\theta(\Xi)$. On choisit de minimiser la distance $\Xi'[\sum_\theta^*(\Xi)]^{-1}\Xi$ par rapport à θ. Comme $\sum_\theta(\Xi)$ est une matrice définie positive, $\Xi'[\sum_\theta^*(\Xi)]^{-1}\Xi$ est un scalaire positif.

DÉFINITION 2.1.1. *L'estimateur de distance quadratique minimale (EDQM) de* $\theta = (a,b)'$, $\hat\theta = (\hat a, \hat b)'$, *minimise l'expression* $\Xi'[\sum_\theta^*(\Xi)]^{-1}\Xi$.

Pour obtenir l'EDQM de θ, il suffit de minimiser l'expression

$$\phi = (\hat Y - \hat X\theta)[\textstyle\sum_\theta^*(\Xi)]^{-1}(\hat Y - \hat X\theta)',$$

par rapport à θ. Après dérivation, il vient

$$\hat\theta = (\hat X'[\textstyle\sum_\theta^*(\Xi)]^{-1}\hat X)^{-1}\hat X'[\textstyle\sum_\theta^*(\Xi)]^{-1}\hat Y,$$

qui n'est pas un estimateur dans le sens usuel car $\sum_\theta^*(\Xi)$ est fonction de θ, et on doit estimer $p_0^*, p_1^*, ..., p_w^*$. Pour cela, Huard (1995, p.64) a utilisé une méthode des moindres carrés pondérés itérés.

On obtient ainsi deux suites $\{\hat{\theta}_1, \hat{\theta}_2, ...\}$ et

$\{[\hat{\sum}_{\hat{\theta}_1}(\Xi)]^{-1}, [\hat{\sum}_{\hat{\theta}_2}(\Xi)]^{-1}, ...\}$. Luong et Garrido (1993) montrent que $\hat{\theta}_i \xrightarrow{p} \theta$, i=1,2,..., et que $[\hat{\sum}_{\hat{\theta}_i}(\Xi)]^{-1} \xrightarrow{p} [\sum_{\theta}^*(\Xi)]^{-1}$, i=1,2,... .

Luong et Garrido (1993) montrent que $\sqrt{m}\hat{\theta} \xrightarrow{\mathcal{L}} N_2(\theta, (\hat{X}'[\sum_{\theta}^*(\Xi)]^{-1}\hat{X})^{-1})$ et que $\hat{\theta}$ est un estimateur asymptotiquement efficace de θ. La robustesse de l'estimateur $\hat{\theta}$ au sens de Hampel (1974, 1986) est aussi démontrée.

A la prochaine section, nous démontrons ces propriétés dans le cadre général de la famille de Sundt.

2.2. Estimation pour la famille de Sundt $\mathcal{R}_2$

2.2.1. La famille $\mathcal{R}_2$ tronquée

Nous définissons d'abord la famille $\mathcal{R}_2$ tronquée.

DÉFINITION 2.2.1. *Soit N, une variable aléatoire discrète, avec domaine 0,1,2,...,w. N appartient à la famille tronquée de Sundt d'ordre 2, si elle admet la forme récursive suivante:*

$$p_n^* = \sum_{i=1}^{2}(a_i + \frac{b_i}{n})p_{n-i}^*,\ p_{-1}^* = 0,\ n = 1, 2, ..., w, \tag{2.2.1}$$

où a_1, b_1, a_2, b_2 sont tels que (2.2.1) définit bien une fonction de probabilité valide.

Notre but est d'estimer le paramètre $\theta = (a_1, b_1, a_2, b_2)'$. La méthode du maximum de vraisemblance est difficilement applicable car on ne peut déterminer une forme analytique des probabilités théoriques en fonction de θ. On pourrait utiliser des algorithmes numériques pour trouver le maximum global de la fonction de vraisemblance mais cette alternative semble difficile à mettre en pratique. N'oublions pas aussi que, dans le cas de la famille de Sundt $\mathcal{R}_k$, nous devrons estimer $\theta = (a_1, b_1, a_2, b_2, ..., a_w, b_w)'$, où w peut être assez grand. Appelons Θ l'espace des paramètres. Au chapitre précédent, on a remarqué que Sundt (1992) ne s'est pas aventuré dans l'étude de cet espace. Nous définissons Θ de la façon suivante:

$$\Theta \quad = \quad \{(a_1, b_1, a_2, b_2)' \in R^4 : 0 < p_i^* < 1, i = 0, 1, 2, ..., w, \sum_{i=0}^{w} p_i^* = 1\}.$$

Plus loin dans cette section, on démontre que la méthode de distance quadratique minimale produit un estimateur convergent en probabilité vers le vecteur des paramètres. La méthode que nous présentons ne tient pas compte des contraintes de l'espace paramétrique. Une autre approche serait de considérer un problème de régression sous contraintes.

Soit $n_1, n_2, ..., n_m$ l'échantillon observé de $N_1, N_2, ..., N_m$, qui satisfait (2.2.1).

Les équations récursives, linéaires en $(a_1, b_1, a_2, b_2)'$, suggèrent le modèle de régression linéaire suivant:

$$\hat{p}_n^* \quad = \quad \sum_{i=1}^{2} (a_i + \frac{b_i}{n}) \hat{p}_{n-i}^* + \varepsilon_n \ , \ \hat{p}_{-1}^* = 0, \ n = 1, 2, ..., w \qquad (2.2.2)$$

Ceci définit un modèle de régresion linéaire en $\theta = (a_1, b_1, a_2, b_2)$'. Posons:

$$\hat{Y} \quad = \quad (\hat{p}_1^*, \hat{p}_2^*, ..., \hat{p}_w^*)'$$

$$\hat{X} = \begin{pmatrix} \hat{p}_0^* & \hat{p}_0^* & 0 & 0 \\ \hat{p}_1^* & \hat{p}_1^*/2 & \hat{p}_0^* & \hat{p}_0^*/2 \\ \hat{p}_2^* & \hat{p}_2^*/3 & \hat{p}_1^* & \hat{p}_1^*/3 \\ \cdots & \cdots & \cdots & \cdots \\ \hat{p}_{w-1}^* & \hat{p}_{w-1}^*/w & \hat{p}_{w-2}^* & \hat{p}_{w-2}^*/w \end{pmatrix}$$

$$\Xi = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_w)'.$$

Sous forme matricielle, le modèle (2.2.2) s'écrit comme $\hat{Y} = \hat{X}\theta + \Xi$.

Si $m \to \infty$, $\hat{X}$ est une matrice de plein rang avec probabilité 1, et elle tend en probabilité vers la matrice X où

$$X = \begin{pmatrix} p_0^* & p_0^* & 0 & 0 \\ p_1^* & p_1^*/2 & p_0^* & p_0^*/2 \\ p_2^* & p_2^*/3 & p_1^* & p_1^*/3 \\ \cdots & \cdots & \cdots & \cdots \\ p_{w-1}^* & p_{w-1}^*/w & p_{w-2}^* & p_{w-2}^*/w \end{pmatrix}.$$

Aussi $\mathrm{E}(\hat{X}) = X$.

Intéressons-nous maintenant à Ξ et à la structure de la matrice de variance-covariance des erreurs, appelée $\sum_\theta(\Xi)$. On montre la proposition suivante. La preuve de cette proposition est donnée à l'appendice A.

Proposition 2.2.1.

(a) $E(\varepsilon_n) = 0$, $n= 1,2,\ldots\ldots,w$

(b) $Var(\varepsilon_1) = \frac{1}{m} p_1^* (1 + c_{1,1})$

(c) $Var(\varepsilon_n) = \frac{1}{m}\left[p_n^*\{1 + \sum_{j=1}^{2} c_{j,n}\} - (p_{n-1}^* + p_{n-2}^*)\prod_{j=1}^{2} c_{j,n}\right]$, $n= 2,3,4,\ldots,w$

(d) $Cov(\varepsilon_n, \varepsilon_{n+1}) = \frac{1}{m}\left[c_{1,n}p_{n+1}^* - c_{1,n+1}p_n^*\{1 + c_{1,n}\}\right]$, $n= 1, 2,3, \ldots,w\text{-}1$

(e) $Cov(\varepsilon_n, \varepsilon_{n+2}) = \frac{-1}{m} c_{2,n+2} p_n^*$, $n= 1, 2,3, \ldots,w\text{-}2$

(f) $Cov(\varepsilon_n, \varepsilon_{n+j}) = 0, j \geq 3$, $n= 1, 2,3, \ldots,w\text{-}j$, *où* $c_{1,n} = (a_1 + \frac{b_1}{n}), c_{2,n} = (a_2 + \frac{b_2}{n})$, *n= 1, 2, ...,w.*

On pose $Var(\Xi) = \sum_\theta(\Xi) = \frac{1}{m}\sum_\theta^*(\Xi)$.

DÉFINITION 2.2.2. *L'EDQM de θ, $\hat{\theta} = (\hat{a_1}, \hat{b_1}, \hat{a_2}, \hat{b_2})'$ minimise l'expression* $\Xi'[\sum_\theta^*(\Xi)]^{-1}\Xi$.

La solution est donnée par

$$\hat{\theta} \;=\; (\hat{X}'[\textstyle\sum_\theta^*(\Xi)]^{-1}\hat{X})^{-1}\hat{X}'[\textstyle\sum_\theta^*(\Xi)]^{-1}\hat{Y}.$$

Comme $\sum_\theta^*(\Xi)$ est fonction de θ, on doit utiliser le procédé itératif qui suit:

- Poser $[\sum_\theta^*(\Xi)]^{-1} = I_w$ et calculer $\hat{\theta}_1 = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{Y}$, où I_w est la matrice identité de dimension w.

- Avec $\hat{\theta}_1$, estimer $[\sum_\theta^*(\Xi)]^{-1}$ par $[\hat{\sum}_{\hat{\theta}_1}(\Xi)]^{-1}$

- Avec $[\hat{\sum}_{\hat{\theta}_1}(\Xi)]^{-1}$, calculer

$$\hat{\theta}_2 \;=\; (\hat{X}'[\textstyle\hat{\sum}_{\hat{\theta}_1}(\Xi)]^{-1}\hat{X})^{-1}\hat{X}'[\textstyle\hat{\sum}_{\hat{\theta}_1}(\Xi)]^{-1}\hat{Y}$$

.....

Dans les prochaines pages, nous énonçons et démontrons trois propositions sur les propriétés de l'EDQM de θ.

Proposition 2.2.2. *Le vecteur Ξ est asymptotiquement normal, de moyenne 0 et de matrice de variance-covariance $\sum_\theta(\Xi)$.*

Preuve : Ecrivons

$$\begin{aligned}
\varepsilon_1 &= \hat{p}_1^* - (a_1 + \frac{b_1}{1})\hat{p}_0^* \\
&= \frac{1}{m}\sum_{i=1}^{m}\left\{Y_{1,i} - (a_1 + \frac{b_1}{1})Y_{0,i}\right\} \\
&= \frac{1}{m}\sum_{i=1}^{m} X_i^{(1)},
\end{aligned}$$

où $X_i^{(1)} = Y_{1,i} - (a_1 + \frac{b_1}{1})Y_{0,i}$, $Y_{1,i}$ est une loi binomiale$(1, p_1)$, $Y_{0,i}$ est une loi binomiale$(1, p_0)$.

Pour $\ell = 2, 3, ..., w$, écrivons $\varepsilon_\ell = \frac{1}{m}\sum_{i=1}^{m} X_i^{(\ell)}$, où

$$X_i^{(\ell)} = Y_{\ell,i} - (a_1 + \frac{b_1}{\ell})Y_{\ell-1,i} - (a_2 + \frac{b_2}{\ell})Y_{\ell-2,i}$$

où $Y_{\ell,i}$ est une loi binomiale$(1, p_\ell)$, $Y_{\ell-1,i}$ est une loi binomiale$(1, p_{\ell-1})$ et $Y_{\ell-2,i}$ est une loi binomiale$(1, p_{\ell-2})$.

Ξ peut s'écrire comme

$$\Xi = \frac{1}{m}\begin{pmatrix} X_1^{(1)} + X_2^{(1)} + ... + X_m^{(1)} \\ X_1^{(2)} + X_2^{(2)} + ... + X_m^{(2)} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ X_1^{(w)} + X_2^{(w)} + ... + X_m^{(w)} \end{pmatrix}$$

Posons $T_i = (X_i^{(1)}, X_i^{(2)}, ..., X_i^{(w)})'$, $i = 1, 2, ..., m$. Ecrivons $\Xi = \frac{1}{m}\sum_{i=1}^{m} T_i$. Les T_i sont identiquement distribués de moyenne 0 et de matrice de variance-covariance $\sum_\theta^*(\Xi)$. Par le théorème limite centrale multivarié (voir Serfling (1980), p.28), on a donc que $\sqrt{m}\Xi \xrightarrow{\mathcal{L}} N_w(0, \sum_\theta^*(\Xi))$. $\square$

De la proposition (2.2.2), on conclut que $\sqrt{m}\hat{\theta} \xrightarrow{\mathcal{L}} N_4\left(\theta, \left(X'\left[\sum_\theta^*(\Xi)\right]^{-1} X\right)^{-1}\right)$.

<u>**Proposition**</u> 2.2.3. *Si* $\| \left(X' \left[\sum_\theta^*(\Xi)\right]^{-1} X\right)^{-1} \| < \infty$, $\hat{\theta}$ *est un estimateur convergent de* θ. *On utilise ici la norme euclidienne classique.*

Preuve : Du fait que $\sqrt{m}\hat{\theta} \xrightarrow{\mathcal{L}} N_4\left(\theta, \left(X' \left[\sum_\theta^*(\Xi)\right]^{-1} X\right)^{-1}\right)$, le résultat suit si $\| \left(X' \left[\sum_\theta^*(\Xi)\right]^{-1} X\right)^{-1} \| < \infty$. $\square$

<u>**Proposition**</u> 2.2.4. $\hat{\theta}$ *est un estimateur asymptotiquement efficace de* θ.

Preuve : Appelons $I(\theta)$ la matrice d'information de Fisher, qui est la matrice de variance-covariance du vecteur score $\partial logp_n(\theta)/\partial\theta$. Il est connu (voir Gouriéroux (1989)) que $I(\theta) = V_\theta\left[\partial logp_n(\theta)/\partial\theta\right]$. Pour démontrer que $\hat{\theta}$ est un estimateur asymptotiquement efficace, il faut montrer (voir Serfling (1980), p.142) que $\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta, [I(\theta)]^{-1}\right)$. Or, l'on sait que $AsVar(\hat{\theta}) = \left(X' \left[\sum_\theta^*(\Xi)\right]^{-1} X\right)^{-1}$. Il s'agit donc de montrer que

$$I(\theta) \quad = \quad X' \left[\textstyle\sum_\theta^*(\Xi)\right]^{-1} X.$$

Considérons $T_1 = (X_1^{(1)}, X_1^{(2)}, ..., X_1^{(w)})'$ tel que défini dans la preuve de la proposition 2.2.3. Soit $\sum$, la matrice de variance-covariance du vecteur $-X' \left[\sum_\theta^*(\Xi)\right]^{-1} T_1$:

$$\begin{aligned}
\sum &= X' \left[\textstyle\sum_\theta^*(\Xi)\right]^{-1} Var(T_1) \left[\textstyle\sum_\theta^*(\Xi)\right]^{-1} X \\
&= X' \left[\textstyle\sum_\theta^*(\Xi)\right]^{-1} \textstyle\sum_\theta^* \left[\textstyle\sum_\theta^*(\Xi)\right]^{-1} X \\
&= X' \left[\textstyle\sum_\theta^*(\Xi)\right]^{-1} X.
\end{aligned}$$

Il reste donc à démontrer que $-X' \left[\sum_\theta^*(\Xi)\right]^{-1} T_1$ est égal à $\partial logp_n(\theta)/\partial\theta$;

Posons $\mathcal{S}(n,\theta)$ le vecteur score et

$$\begin{aligned}
\mathcal{S}_{a_1}(n,\theta) &= \partial\mathcal{S}(n,\theta)/\partial a_1,\\
\mathcal{S}_{b_1}(n,\theta) &= \partial\mathcal{S}(n,\theta)/\partial b_1,\\
\mathcal{S}_{a_2}(n,\theta) &= \partial\mathcal{S}(n,\theta)/\partial a_2,\\
\mathcal{S}_{b_2}(n,\theta) &= \partial\mathcal{S}(n,\theta)/\partial b_2.
\end{aligned}$$

$$\mathcal{S}(n,\theta) = (\mathcal{S}_{a_1}(n,\theta), \mathcal{S}_{b_1}(n,\theta), \mathcal{S}_{a_2}(n,\theta), \mathcal{S}_{b_2}(n,\theta))'.$$

Soient

$$\begin{aligned}
\mathcal{T}_{a_1}(n,\theta) &= \sum_{i=1}^{w} \alpha_{1i} X_1^{(i)},\\
\mathcal{T}_{b_1}(n,\theta) &= \sum_{i=1}^{w} \beta_{1i} X_1^{(i)},\\
\mathcal{T}_{a_2}(n,\theta) &= \sum_{i=1}^{w} \alpha_{2i} X_1^{(i)} et\\
\mathcal{T}_{b_2}(n,\theta) &= \sum_{i=1}^{w} \beta_{2i} X_1^{(i)},
\end{aligned}$$

où $\alpha_{1i}, \beta_{1i}, \alpha_{2i} et \beta_{2i}$ sont des constantes, et $X_1^{(i)}$ sont définis dans la preuve de la proposition (2.2.2).

Considérons d'abord le problème de choisir $\alpha_{11}, \alpha_{12}, ..., \alpha_{1w}$ qui minimisent

$$A_{a_1}(\alpha_{11}, \alpha_{12}, ..., \alpha_{1w}) = \sum_{n=0}^{w} p_n^* \left[\mathcal{S}_{a_1}(n,\theta) - \mathcal{T}_{a_1}(n,\theta)\right]^2, \ n = 0, 1, 2, ..., w.$$

Appelons la solution $\alpha^*_{a_1} = (\alpha^*_{11}, \alpha^*_{12}, ..., \alpha^*_{1w})'$. On a :

$$
\begin{aligned}
\mathcal{T}_{a_1}(w,\theta) &= a_w, \\
\mathcal{T}_{a_1}(w-1,\theta) &= a_{w-1} - a_w(a_1 + \frac{b_1}{w}), \\
\mathcal{T}_{a_1}(w-2,\theta) &= a_{w-2} - a_{w-1}(a_1 + \frac{b_1}{w-1}) - a_w(a_2 + \frac{b_2}{w}), \\
\mathcal{T}_{a_1}(w-3,\theta) &= a_{w-3} - a_{w-2}(a_1 + \frac{b_1}{w-2}) - a_{w-1}(a_2 + \frac{b_2}{w-1}), \\
&\dots \\
\mathcal{T}_{a_1}(1,\theta) &= a_1 - a_2(a_1 + \frac{b_1}{2}) - a_3(a_2 + \frac{b_2}{3}), \\
\mathcal{T}_{a_1}(0,\theta) &= -a_1(a_1 + \frac{b_1}{1}).
\end{aligned}
$$

Et, $\rho_w = \mathcal{S}_{a_1}(w,\theta) - \mathcal{T}_{a_1}(w,\theta) = \mathcal{S}_{a_1}(w,\theta) - a_w$. D'où

$$
\begin{aligned}
\rho_{w-1} &= \mathcal{S}_{a_1}(w-1,\theta) - \mathcal{T}_{a_1}(w-1,\theta) \\
&= \mathcal{S}_{a_1}(w-1,\theta) - a_{w-1} + a_w(a_1 + \frac{b_1}{w}) \\
&= \mathcal{S}_{a_1}(w-1,\theta) - a_{w-1} + (\mathcal{S}_{a_1}(w,\theta) - \rho_w)(a_1 + \frac{b_1}{w}).
\end{aligned}
$$

Pour évaluer $A_{a_1}(a_{11}, a_{12}, ..., a_{1w})$, il faut donc donner une valeur initiale à ρ_w. Posons $\rho_w = 0$, ce qui implique que $a_w = \mathcal{S}_{a_1}(w,\theta)$. Notre problème revient donc à déterminer $\alpha_{11}, \alpha_{12}, ..., \alpha_{1w}$ qui minimisent

$$
\begin{aligned}
B_{a_1}(\alpha_{11}, \alpha_{12}, ..., \alpha_{1w}) &= B_{a_1} \\
&= \sum_{n=0}^{w-1} p^*_n \left[\mathcal{S}_{a_1}(n,\theta) - \mathcal{T}_{a_1}(n,\theta)\right]^2
\end{aligned}
$$

Pour ce faire, résolvons le système suivant:

$$\partial B_{a_1}/\partial\alpha^*_{11} = \partial B_{a_1}/\partial\alpha^*_{12} = ... = \partial B_{a_1}/\partial\alpha^*_{1w} = 0.$$

Pour $j = 1, 2, ..., w$, on a

$$\begin{aligned}
\partial B_{a_1}/\partial \alpha_{1j} = 0 \iff & \sum_{n=0}^{w-1} p_n^* \left[\mathcal{S}_{a_1}(n,\theta) - \mathcal{T}_{a_1}(n,\theta)\right] X_1^{(j)} = 0 \\
\iff & \sum_{n=0}^{w} p_n^* \mathcal{S}_{a_1}(n,\theta) X_1^{(j)} = \sum_{n=0}^{w} p_n^* \mathcal{T}_{a_1}(n,\theta) X_1^{(j)} + p_w^* \mathcal{S}_{a_1}(w,\theta) X_1^{(j)} \\
& - p_w^* \mathcal{T}_{a_1}(w,\theta) X_1^{(j)} \\
\iff & \sum_{i=1}^{w} \alpha_{1i}^* E\left[X_1^{(i)} X_1^{(j)}\right] = E\left[\mathcal{S}_{a_1} X_1^{(j)}\right]
\end{aligned}$$

Sous forme matricielle, on peut écrire:

$$\Sigma_\theta^*(\Xi)\alpha_{a_1}^* = \begin{pmatrix} E\left[\mathcal{S}_{a_1} X_1^{(1)}\right] \\ E\left[\mathcal{S}_{a_1} X_1^{(2)}\right] \\ \cdots\cdots \\ E\left[\mathcal{S}_{a_1} X_1^{(w)}\right] \end{pmatrix} = - \begin{pmatrix} E\left[\partial X_1^{(1)}/\partial a_1\right] \\ E\left[\partial X_1^{(2)}/\partial a_1\right] \\ \cdots\cdots \\ E\left[\partial X_1^{(w)}/\partial a_1\right] \end{pmatrix} = -E\left[\partial \mathcal{T}_1/\partial a_1\right]$$

Pour voir la seconde égalité ci-haut, remarquons que pour $j = 1, 2, ..., w$, on a $\partial E\left[X_1^{(j)}\right]/\partial a_1 = 0$.

Ainsi,

$$\begin{aligned}
0 &= \frac{\partial \sum_{n=0}^{w} X_1^{(j)}(n) p_n^*}{\partial a_1} \\
&= \sum_{n=0}^{w} \frac{\partial X_1^{(j)}(n)}{\partial a_1} p_n^* + X_1^{(j)}(n) \frac{\partial p_n^*/\partial a_1}{p_n^*} p_n^*.
\end{aligned}$$

D'où $E\left[\partial X_1^{(j)}/\partial a_1\right] = -E\left[\mathcal{S}_{a_1} X_1^{(j)}\right]$.

On obtient ainsi $\alpha_{a_1}^* = -\left[\Sigma_\theta^*(\Xi)\right]^{-1} E\left[\partial \mathcal{T}_1/\partial a_1\right]$.

De la même façon, on peut considérer les trois problèmes qui suivent:

- Trouver $\beta_{11}, \beta_{12}, ..., \beta_{1w}$ qui minimisent

$$B_{b_1}(\beta_{11}, \beta_{12}, ..., \beta_{1w}) = \sum_{n=0}^{w-1} p_n^* \left[\mathcal{S}_{b_1}(n,\theta) - \mathcal{T}_{b_1}(n,\theta)\right]^2.$$

Appelons la solution du système $\beta_{b_1}^* = (\beta_{11}^*, \beta_{12}^*, ..., \beta_{1w}^*)'$.

On trouve $\beta^*_{b_1} = -\left[\Sigma^*_\theta(\Xi)\right]^{-1} E\left[\partial T_1/\partial b_1\right].$

• Trouver $\alpha_{21}, \alpha_{22}, ..., \alpha_{2w}$ qui minimisent

$$B_{a_2}(\alpha_{21}, \alpha_{22}, ..., \alpha_{2w}) \quad = \quad \sum_{n=0}^{w-1} p^*_n \left[\mathcal{S}_{a_2}(n,\theta) - \mathcal{T}_{a_2}(n,\theta)\right]^2$$

Appelons la solution du système $\alpha^*_{a_2} = (\alpha^*_{21}, \alpha^*_{22}, ..., \alpha^*_{2w})'$.

On trouve $\alpha^*_{a_2} = -\left[\Sigma^*_\theta(\Xi)\right]^{-1} E\left[\partial T_1/\partial a_2\right].$

• Trouver $\beta_{21}, \beta_{22}, ..., \beta_{2w}$ qui minimisent

$$B_{b_2}(\beta_{21}, \beta_{22}, ..., \beta_{2w}) \quad = \quad \sum_{n=0}^{w-1} p^*_n \left[\mathcal{S}_{b_2}(n,\theta) - \mathcal{T}_{b_2}(n,\theta)\right]^2.$$

Appelons la solution du système $\beta^*_{b_2} = (\beta^*_{21}, \beta^*_{22}, ..., \beta^*_{2w})'$.

On trouve $\beta^*_{b_2} = -\left[\Sigma^*_\theta(\Xi)\right]^{-1} E\left[\partial T_1/\partial b_2\right].$

$$\begin{aligned} \text{Posons } \mathcal{T}^* &= (\mathcal{T}^*_{a_1}, \mathcal{T}^*_{b_1}, \mathcal{T}^*_{a_2}, \mathcal{T}^*_{b_2})' \\ &= ((\alpha^*_{a_1})'T_1, (\beta^*_{b_1})'T_1, (\alpha^*_{a_2})'T_1, (\beta^*_{b_2})'T_1)' \\ &= -\left(E\left[\partial T_1'/\partial a_1\right], E\left[\partial T_1'/\partial b_1\right], E\left[\partial T_1'/\partial a_2\right], E\left[\partial T_1'/\partial b_2\right]\right)' \left[\Sigma^*_\theta(\Xi)\right]^{-1} T_1 \\ &= -X'[\Sigma^*_\theta(\Xi)]^{-1} T_1. \end{aligned}$$

On voit que $\mathcal{T}^*$ est la projection de $\mathcal{S}(n,\theta)$ sur l'espace engendré par $\left\{X_1^{(1)}, X_1^{(2)}, ..., X_1^{(w)}\right\}$. $\mathcal{T}^*$ appartient à cet espace, d'où $\mathcal{T}^* = \mathcal{S}(n,\theta)$. $\square$

2.2.1.1. *La robustesse de l'EDQM de θ dans la famille $\mathcal{R}_2$ tronquée*

Pour une présentation exhaustive de la théorie de la robustesse, nous renvoyons à Hampel (1974) pour le traitement intuitif des courbes d'influence, Hampel (1986).

Soit F_θ la fonction de répartition théorique de $N_1, N_2, ..., N_m$. Supposons que nous ayons un estimateur de θ, et que nous puissions l'exprimer comme une fonctionnelle de la fonction de répartition échantillonale F_m. Duchesne (1995) fait remarquer que la notion de fonctionnelle permet de ne plus tenir compte de la taille de l'échantillon qu'à travers F_m seulement. On fait subir une perturbation infinitésimale au modèle théorique F_θ en y introduisant une valeur contaminée, disons n_* en considérant l'expression qui suit:

$$F_\epsilon = (1-\epsilon)F_\theta + \epsilon\delta_{n_*}$$

où δ_{n_*} est la fonction de répartition d'une loi dégénérée à n_* et $\epsilon \in [0,1]$. Notons que $n_* \in \{0, 1, ..., w\}$.

Définition 2.2.3. *La fonction d'influence (IF) de la statistique $\hat{\theta}$ évaluée à $(1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}$ est donnée par*

$$\begin{aligned} IF(n_*, \hat{\theta}, (1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left[\hat{\theta}\left((1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}\right) - \hat{\theta}(F_\theta)\right] \\ &= \frac{\partial}{\partial\varepsilon}\hat{\theta}\left[(1-\varepsilon)F_\theta + \varepsilon\partial_{n_*}\right]|_{\varepsilon=0} \end{aligned}$$

Ainsi, on pourra conclure que $\hat{\theta}$ est robuste, au sens de Hampel, si la fonction d'influence reste bornée pour tout n_* considéré.

Dans le contexte de la famille $\mathcal{R}_2$ tronquée, nous savons que $\hat{\theta} = (\hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2)'$ est la solution du système :

$$\begin{aligned}
\frac{\partial}{\partial a_1}\Xi'\left[\Sigma_\theta^*(\Xi)\right]^{-1}\Xi &= 0\\
\frac{\partial}{\partial b_1}\Xi'\left[\Sigma_\theta^*(\Xi)\right]^{-1}\Xi &= 0\\
\frac{\partial}{\partial a_2}\Xi'\left[\Sigma_\theta^*(\Xi)\right]^{-1}\Xi &= 0\\
\frac{\partial}{\partial b_2}\Xi'\left[\Sigma_\theta^*(\Xi)\right]^{-1}\Xi &= 0,
\end{aligned}$$

où $\Xi = \hat{Y} - \hat{X}\theta$.

Définissons les fonctions indicatrices $I_j\left[x \in [j, j+1)\right] = 1$ si $x \in [j, j+1)$ et 0 sinon ($j = 0, 1, ..., w$).

On peut écrire alors

$$\begin{aligned}
\Xi' &= \Big(\int [I_1 - (a_1 + b_1)I_0]dF_m, \int [I_2 - (a_1 + b_1/2)I_1 - (a_2 + b_2/2)I_0]dF_m,\\
&\quad ..., \int [I_w - (a_1 + b_1/w)I_{w-1} - (a_2 + b_2/w)I_{w-2}]dF_m\Big)'.
\end{aligned}$$

$\hat{\theta}$ peut s'écrire comme $\hat{\theta} = \hat{\theta}(F_m)$ et donc comme fonctionnelle de la fonction de répartition échantillonale.

Soit maintenant la mixture $F_\varepsilon = (1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}$, $0 < \varepsilon < 1$, telle que définie précédemment. L'estimateur $\hat{\theta}$ de θ sera alors donné, sous cette perturbation, par la solution du système suivant:

$$\begin{aligned}
L_{a_1} &= \frac{\partial}{\partial a_1}\Delta'[\Sigma_\theta^*(\Delta)]^{-1}\Delta = 0\\
L_{a_2} &= \frac{\partial}{\partial b_1}\Delta'[\Sigma_\theta^*(\Delta)]^{-1}\Delta = 0\\
L_{b_1} &= \frac{\partial}{\partial a_2}\Delta'[\Sigma_\theta^*(\Delta)]^{-1}\Delta = 0\\
L_{b_2} &= \frac{\partial}{\partial b_2}\Delta'[\Sigma_\theta^*(\Delta)]^{-1}\Delta = 0,
\end{aligned}$$

où

$$\begin{aligned}
\Delta' &= \Big(\int [I_1 - (a_1+b_1)I_0]dF_\varepsilon, \int [I_2 - (a_1+b_1/2)I_1 - (a_2+b_2/2)I_0]dF_\varepsilon,\\
&\quad ..., \int [I_w - (a_1+b_1/w)I_{w-1} - (a_2+b_2/w)I_{w-2}]dF_\varepsilon\Big)'.
\end{aligned}$$

Pour la suite, nous allons utiliser une forme explicite pour Δ':

$$\begin{aligned}
\Delta' &= ((1-\varepsilon)p_1^* - (a_1+b_1)(1-\varepsilon)p_0^*,\\
&\quad (1-\varepsilon)p_2^* - (a_1+\frac{b_1}{2})(1-\varepsilon)p_1^* - (a_2+\frac{b_2}{2})(1-\varepsilon)p_0^*, ...,\\
&\quad (1-\varepsilon)p_{n_*-1}^* - (a_1+\frac{b_1}{n_*-1})(1-\varepsilon)p_{n_*-2}^* - (a_2+\frac{b_2}{n_*-1})(1-\varepsilon)p_{n_*-3}^*,\\
&\quad (1-\varepsilon)p_{n_*}^* + \varepsilon - (a_1+\frac{b_1}{n_*})(1-\varepsilon)p_{n_*-1}^* - (a_2+\frac{b_2}{n_*})(1-\varepsilon)p_{n_*-2}^*,\\
&\quad (1-\varepsilon)p_{n_*+1}^* - (a_1+\frac{b_1}{n_*+1})[(1-\varepsilon)p_{n_*}^* + \varepsilon] - (a_2+\frac{b_2}{n_*+1})(1-\varepsilon)p_{n_*-1}^*,\\
&\quad (1-\varepsilon)p_{n_*+2}^* - (a_1+\frac{b_1}{n_*+2})(1-\varepsilon)p_{n_*+1}^* - (a_2+\frac{b_2}{n_*+2})(1-\varepsilon)[(1-\varepsilon)p_{n_*}^* + \varepsilon], ...,\\
&\quad (1-\varepsilon)p_w^* - (a_1+\frac{b_1}{w})(1-\varepsilon)p_{w-1}^* - (a_2+\frac{b_2}{w})(1-\varepsilon)p_{w-2}^*)'.
\end{aligned}$$

On a aussi

$$\begin{aligned}
\frac{\partial L_{a_i}}{\partial \varepsilon} &= \sum_{j=1}^{2}\left[\frac{\partial L_{a_i}}{\partial a_j}\frac{\partial a_j}{\partial \varepsilon} + \frac{\partial L_{a_i}}{\partial b_j}\frac{\partial b_j}{\partial \varepsilon}\right], \qquad i = 1,2\\
\frac{\partial L_{b_i}}{\partial \varepsilon} &= \sum_{j=1}^{2}\left[\frac{\partial L_{b_i}}{\partial a_j}\frac{\partial a_j}{\partial \varepsilon} + \frac{\partial L_{b_i}}{\partial b_j}\frac{\partial b_j}{\partial \varepsilon}\right], \qquad i = 1,2.
\end{aligned}$$

La fonction d'influence est donc donnée par

$$IF(n_*, \hat{\theta}, (1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}) = \frac{\partial}{\partial\varepsilon}\hat{\theta}\left[(1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}\right]|_{\varepsilon=0}$$

$$= \begin{pmatrix} \frac{\partial L_{a_1}}{\partial a_1} & \frac{\partial L_{a_1}}{\partial b_1} & \frac{\partial L_{a_1}}{\partial a_2} & \frac{\partial L_{a_1}}{\partial b_2} \\ \frac{\partial L_{b_1}}{\partial a_1} & \frac{\partial L_{b_1}}{\partial b_1} & \frac{\partial L_{b_1}}{\partial a_2} & \frac{\partial L_{b_1}}{\partial b_2} \\ \frac{\partial L_{a_2}}{\partial a_1} & \frac{\partial L_{a_2}}{\partial b_1} & \frac{\partial L_{a_2}}{\partial a_2} & \frac{\partial L_{a_2}}{\partial b_2} \\ \frac{\partial L_{b_2}}{\partial a_1} & \frac{\partial L_{b_2}}{\partial b_1} & \frac{\partial L_{b_2}}{\partial a_2} & \frac{\partial L_{b_2}}{\partial b_2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial L_{a_1}}{\partial\varepsilon} \\ \frac{\partial L_{b_1}}{\partial\varepsilon} \\ \frac{\partial L_{a_2}}{\partial\varepsilon} \\ \frac{\partial L_{b_2}}{\partial\varepsilon} \end{pmatrix} |_{\varepsilon=0}. \qquad (2.2.3)$$

Proposition 2.2.5.

$$IF(n_*, \hat{\theta}, (1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}) = \left(X'[\Sigma_\theta^*(\Xi)]^{-1}X\right)^{-1} X'[\Sigma_\theta^*(\Xi)]^{-1}V$$
$$\text{où } V = \left(0, 0, .., 0, -1, a_1 + \frac{b_1}{n_*+1}, a_2 + \frac{b_2}{n_*+2}, 0, 0, .., 0\right)'.$$

Ainsi, la fonction d'influence est bornée pour tout n_* et donc, l'EDQM de θ est robuste au sens de Hampel.

Preuve : D'abord, nous démontrons que

$$\begin{aligned} &\left(\frac{\partial L_{a_1}}{\partial a_1}, \frac{\partial L_{a_1}}{\partial b_1}, \frac{\partial L_{a_1}}{\partial a_2}, \frac{\partial L_{a_1}}{\partial b_2}\right)|_{\varepsilon=0} \\ = \;& 2\left(\frac{\partial}{\partial a_1}\Delta' U \frac{\partial}{\partial a_1}\Delta, \frac{\partial}{\partial a_1}\Delta' U \frac{\partial}{\partial b_1}\Delta, \frac{\partial}{\partial a_1}\Delta' U \frac{\partial}{\partial a_2}\Delta, \frac{\partial}{\partial a_1}\Delta' U \frac{\partial}{\partial b_2}\Delta\right)|_{\varepsilon=0}, \end{aligned}$$

où on pose U $= [\sum_\theta^*(\Delta)]^{-1}$. Les calculs sont similaires pour les trois autres vecteurs-lignes de la matrice de (2.2.3).

$$\bullet\frac{\partial L_{a_1}}{\partial a_1} = \frac{\partial}{\partial a_1}\left[\frac{\partial}{\partial a_1}(\Delta' U\Delta)\right] = \frac{\partial}{\partial a_1}\left[\frac{\partial}{\partial a_1}(\Delta')U\Delta + \Delta' U\frac{\partial}{\partial a_1}(\Delta)\right]$$
$$= \frac{\partial^2}{\partial a_1^2}(\Delta')U\Delta + \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) + \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) + \Delta' U\frac{\partial^2}{\partial a_1^2}(\Delta).$$

Comme $\Delta|_{\varepsilon=0} = 0$, on obtient

$$\frac{\partial L_{a_1}}{\partial a_1}|_{\varepsilon=0} = 2\frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial a_1}(\Delta)|_{\varepsilon=0}.$$

$$\bullet\frac{\partial L_{a_1}}{\partial b_1} = \frac{\partial}{\partial b_1}\left[\frac{\partial}{\partial a_1}(\Delta')U\Delta + \Delta' U\frac{\partial}{\partial a_1}(\Delta)\right]$$

$$= \frac{\partial^2}{\partial b_1\partial a_1}(\Delta')U\Delta + \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial b_1}(\Delta) + \frac{\partial}{\partial b_1}(\Delta')U\frac{\partial}{\partial a_1}(\Delta)$$
$$+ (\Delta')U\frac{\partial^2}{\partial b_1\partial a_1}(\Delta).$$

Ainsi, $\frac{\partial L_{a_1}}{\partial b_1}|_{\varepsilon=0} = 2\frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial b_1}(\Delta)|_{\varepsilon=0}$.

- De la même façon, on obtient

$$\frac{\partial L_{a_1}}{\partial a_2} = 2\frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial a_2}(\Delta)|_{\varepsilon=0} \text{ et}$$
$$\frac{\partial L_{a_1}}{\partial b_2} = 2\frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial b_2}(\Delta)|_{\varepsilon=0}.$$

Sous forme matricielle, on peut écrire ainsi la matrice de (2.2.3) comme

$$\frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) & \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial b_1}(\Delta) & \dots & \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial b_2}(\Delta) \\ \frac{\partial}{\partial b_1}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) & \frac{\partial}{\partial b_1}(\Delta')U\frac{\partial}{\partial b_1}(\Delta) & \dots & \frac{\partial}{\partial b_1}(\Delta')U\frac{\partial}{\partial b_2}(\Delta) \\ \frac{\partial}{\partial a_2}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) & \frac{\partial}{\partial a_2}(\Delta')U\frac{\partial}{\partial b_1}(\Delta) & \dots & \frac{\partial}{\partial a_2}(\Delta')U\frac{\partial}{\partial b_2}(\Delta) \\ \frac{\partial}{\partial b_2}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) & \frac{\partial}{\partial b_2}(\Delta')U\frac{\partial}{\partial b_1}(\Delta) & \dots & \frac{\partial}{\partial b_2}(\Delta')U\frac{\partial}{\partial b_2}(\Delta) \end{pmatrix}^{-1}\Bigg|_{\varepsilon=0}$$

$$= \frac{1}{2}\left[\begin{pmatrix} \frac{\partial}{\partial a_1}(\Delta') \\ \frac{\partial}{\partial b_1}(\Delta') \\ \frac{\partial}{\partial a_2}(\Delta') \\ \frac{\partial}{\partial b_2}(\Delta') \end{pmatrix} U\left(\frac{\partial}{\partial a_1}(\Delta), \frac{\partial}{\partial b_1}(\Delta), \frac{\partial}{\partial a_2}(\Delta), \frac{\partial}{\partial b_2}(\Delta)\right)\right]^{-1}\Bigg|_{\varepsilon=0}$$

On montre facilement que

$$\begin{aligned} \frac{\partial}{\partial a_1}(\Delta')|_{\varepsilon=0} &= -(p_0^*, p_1^*, ..., p_{w-1}^*)' \\ \frac{\partial}{\partial b_1}(\Delta')|_{\varepsilon=0} &= -(p_0^*, p_1^*/2, p_2^*/3, ..., p_{w-1}^*/w)' \\ \frac{\partial}{\partial a_2}(\Delta')|_{\varepsilon=0} &= -(0, p_0^*, p_1^*, ..., p_{w-2}^*)' \\ \frac{\partial}{\partial b_2}(\Delta')|_{\varepsilon=0} &= -(0, p_0^*/2, p_1^*/3, ..., p_{w-2}^*/w)'. \end{aligned}$$

La matrice se réduit ainsi à $\frac{1}{2}\left[X' \left[\sum_{\theta}^*(\Xi)\right]^{-1} X\right]^{-1}$.

Il reste à calculer maintenant $(\frac{\partial L_{a_1}}{\partial \varepsilon}, \frac{\partial L_{b_1}}{\partial \varepsilon}, \frac{\partial L_{a_2}}{\partial \varepsilon}, \frac{\partial L_{b_2}}{\partial \varepsilon})'|_{\varepsilon=0}$:

- $\frac{\partial L_{a_1}}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon}\left[\frac{\partial}{\partial a_1}(\Delta')U\Delta + \Delta' U\frac{\partial}{\partial a_1}(\Delta)\right]$

$$\begin{aligned} &= \frac{\partial^2}{\partial\varepsilon\partial a_1}(\Delta')U\Delta + \frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial \varepsilon}(\Delta) + \frac{\partial}{\partial \varepsilon}(\Delta')U\frac{\partial}{\partial a_1}(\Delta) \\ &+ (\Delta')U\frac{\partial^2}{\partial\varepsilon\partial a_1}(\Delta). \end{aligned}$$

Il vient ainsi $\frac{\partial L_{a_1}}{\partial \varepsilon}|_{\varepsilon=0} = 2\frac{\partial}{\partial a_1}(\Delta')U\frac{\partial}{\partial \varepsilon}(\Delta)|_{\varepsilon=0}$.

• De la même façon, on obtient

$$\begin{aligned}
\frac{\partial L_{b_1}}{\partial \varepsilon}|_{\varepsilon=0} &= 2\frac{\partial}{\partial b_1}(\Delta')U\frac{\partial}{\partial \varepsilon}(\Delta)|_{\varepsilon=0} \\
\frac{\partial L_{a_2}}{\partial \varepsilon}|_{\varepsilon=0} &= 2\frac{\partial}{\partial a_2}(\Delta')U\frac{\partial}{\partial \varepsilon}(\Delta)|_{\varepsilon=0} \\
\frac{\partial L_{b_2}}{\partial \varepsilon}|_{\varepsilon=0} &= 2\frac{\partial}{\partial b_2}(\Delta')U\frac{\partial}{\partial \varepsilon}(\Delta)|_{\varepsilon=0}.
\end{aligned}$$

Et, $\frac{\partial}{\partial \varepsilon}(\Delta)'|_{\varepsilon=0} = \Big(0, 0, .., 0, 1, -(a_1 + \frac{b_1}{n_*+1}), -(a_2 + \frac{b_2}{n_*+2}), 0, 0, .., 0\Big)'$.

Ainsi, $(\frac{\partial L_{a_1}}{\partial \varepsilon}, \frac{\partial L_{b_1}}{\partial \varepsilon}, \frac{\partial L_{a_2}}{\partial \varepsilon}, \frac{\partial L_{b_2}}{\partial \varepsilon})'|_{\varepsilon=0} =$

$2X'U\Big(0, 0, .., 0, -1, a_1 + \frac{b_1}{n_*+1}, a_2 + \frac{b_2}{n_*+2}, 0, 0, .., 0\Big)'$.

Finalement, on a que

$$IF(n_*, \hat{\theta}, (1-\varepsilon)F_\theta + \varepsilon\delta_{n_*}) = \Big(X'\textstyle\sum_\theta^*(\Xi)X\Big)^{-1} X'\textstyle\sum_\theta^*(\Xi)V$$

$$\text{où } V = \left(0, 0, .., 0, -1, a_1 + \frac{b_1}{n_*+1}, a_2 + \frac{b_2}{n_*+2}, 0, 0, .., 0\right)',$$

qui est une combinaison linéaire de V, un vecteur borné pour $n_* \in \{0, 1, 2, ..., w\}$. □

2.2.2. La famille $\mathcal{R}_2$

Jusqu'ici, on a résolu le problème d'estimation dans le cadre de la famille $\mathcal{R}_2$ tronquée. Si on suppose que notre échantillon est issu de la famille $\mathcal{R}_2$, c'est à dire si on laisse tendre w vers l'infini, la convergence en probabilité, la normalité asymptotique et l'efficacité asymptotique de $\hat{\theta}$ restent valides. En pratique, on peut fixer $w = A$, pour$A > 0$ et supposer que toute valeur de w supérieure à A est abérrante et est rejetée. Dans ce cas, l'efficacité asymptotique de $\hat{\theta}$ ne tient plus mais la convergence et la normalité asymptotique de $\hat{\theta}$ sont des propriétés encore valides. Il s'agira de choisir A assez grand pour avoir la plus grande efficacité asymptotique possible de $\hat{\theta}$.

Un point fort de cette méthode réside dans la facilité de sa mise en oeuvre. Cependant, la taille de l'échantillon considéré doit être assez grande, ce qui peut se justifier dans certains domaines comme l'actuariat comme nous le verrons au chapitre 4. Pour ce qui est du procédé itératif, il apparait que, souvent, quelques itérations suffisent à la convergence de la méthode, pourvu que la taille de l'échantillon soit assez grande (voir Luong et Doray (1996), Doray et Luong (1995, 1997)).

2.3. Les familles $\mathcal{R}_k$ tronquée et $\mathcal{R}_k$

Nous présentons le problème d'estimation dans le cadre général des familles $\mathcal{R}_k$ tronquée et $\mathcal{R}_k$, pour un k fixé. Certaines difficultés rencontrées sont discutées. On commence d'abord par la famille $\mathcal{R}_k$ tronquée, définie comme suit :

Définition 2.3.1. *Soit N, une variable aléatoire discrète, prenant ses valeurs dans 0, 1, ..., w. N appartient à la famille tronquée de Sundt d'ordre k fini, si elle satisfait l'équation récursive suivante :*

$$p_n^* \;=\; \sum_{i=1}^{k}(a_i+\frac{b_i}{n})p_{n-i}^*,\;\; k\;\geq 1,\;\; p_{-1}^*=...=p_{-(k-1)}^*=0,\; n=1,...w(2.3.1)$$

où $(a_1,a_2,...,a_k)'$ *et* $(b_1,b_2,...,b_k)'$ *sont tels que (2.3.1) définit une fonction de probabilité valide.*

Soit un échantillon $n_1, n_2, ..., n_m$. On est intéressé par l'estimation des paramètres de $\theta = (\ a_1, b_1, ..., a_k, b_k)$'. Dans le cadre général, on peut dériver un modèle de régression linéaire en θ

$$\hat{p}_n^* \ = \ \sum_{i=1}^{k}(a_i + \frac{b_i}{n})\hat{p}_{n-1}^* + \varepsilon_n, \hat{p}_{-1}^* = ... = \hat{p}_{-(k-1)}^* = 0,\ n = 1, 2, 3...w. (2.3.2)$$

On pose $\hat{Y} = (\hat{p_1}^*, \hat{p_2}^*, ..., \hat{p_w}^*)'$,

$$\hat{X} \ = \ \begin{pmatrix}
\hat{p}_0^* & \hat{p}_0^* & 0 & 0 & 0 & 0 & ... & 0 & 0 & 0 & 0 \\
\hat{p}_1^* & \frac{\hat{p}_1^*}{2} & \hat{p}_0^* & \frac{\hat{p}_0^*}{2} & 0 & 0 & ... & 0 & 0 & 0 & 0 \\
\hat{p}_2^* & \frac{\hat{p}_2^*}{3} & \hat{p}_1^* & \frac{\hat{p}_1^*}{3} & \hat{p}_0^* & \frac{\hat{p}_0^*}{3} & ... & 0 & 0 & 0 & 0 \\
\hat{p}_3^* & \frac{\hat{p}_3^*}{4} & \hat{p}_2^* & \frac{\hat{p}_2^*}{4} & \hat{p}_1^* & \frac{\hat{p}_1^*}{4} & ... & 0 & 0 & 0 & 0 \\
. & . & . & . & . & . & ... & . & . & . & . \\
\hat{p}_{k-2}^* & \frac{\hat{p}_{k-2}^*}{k-1} & \hat{p}_{k-3}^* & \frac{\hat{p}_{k-3}^*}{k-1} & \hat{p}_{k-4}^* & \frac{\hat{p}_{k-4}^*}{k-1} & ... & \hat{p}_0^* & \frac{\hat{p}_0^*}{k-1} & 0 & 0 \\
\hat{p}_{k-1}^* & \frac{\hat{p}_{k-1}^*}{k} & \hat{p}_{k-2}^* & \frac{\hat{p}_{k-2}^*}{k} & \hat{p}_{k-3}^* & \frac{\hat{p}_{k-3}^*}{k} & ... & \hat{p}_1^* & \frac{\hat{p}_1^*}{k} & \hat{p}_0^* & \frac{\hat{p}_0^*}{k} \\
\hat{p}_k^* & \frac{\hat{p}_k^*}{k+1} & \hat{p}_{k-1}^* & \frac{\hat{p}_{k-1}^*}{k+1} & \hat{p}_{k-2}^* & \frac{\hat{p}_{k-2}^*}{k+1} & ... & \hat{p}_2^* & \frac{\hat{p}_2^*}{k+1} & \hat{p}_1^* & \frac{\hat{p}_1^*}{k+1} \\
. & . & . & . & . & . & ... & . & . & . & . \\
\hat{p}_{w-1}^* & \frac{\hat{p}_{w-1}^*}{w} & \hat{p}_{w-2}^* & \frac{\hat{p}_{w-2}^*}{w} & \hat{p}_{w-3}^* & \frac{\hat{p}_{w-3}^*}{w} & ... & \hat{p}_{w-k+1}^* & \frac{\hat{p}_{w-k+1}^*}{w} & \hat{p}_{w-k}^* & \frac{\hat{p}_{w-k}^*}{w}
\end{pmatrix}$$

$$\Xi = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_w)'.$$

Sous forme matricielle, le modèle (2.3.2) s'écrit comme $\hat{Y} = \hat{X}\theta + \Xi$. Si $w > k$, la matrice $\hat{X}$ est de plein rang avec probabilité 1, si $m \to \infty$, et elle tend en probabilité vers son homologue théorique, dénoté X. Aussi, $\mathrm{E}(\hat{X}) = X$.

Nous définissons l'espace des paramètres ainsi:

$$\Theta \quad = \quad \{(a1, b1, ..., a_k, b_k)' \in R^{2k} : 0 < p_i^* < 1, i = 0, 1, 2, ..., w, \sum_{i=0}^{w} p_i^* = 1\}.$$

Les propositions (2.2.2) à (2.2.5) se généralisent. Du fait que l'écriture est plus lourde en supposant un k quelconque, nous avons opté pour la clarté en démontrant les résultats dans le cas où $k = 2$.

Posons $\sum_\theta(\Xi)$, la matrice de variance-covariance des termes d'erreur. Définissons $\sum_\theta(\Xi) = \frac{1}{m} \sum_\theta^*(\Xi)$, ce que l'on peut faire car nous travaillons avec des proportions binomiales.

DÉFINITION 2.3.2. *L'EDQM de* θ*,* $\hat{\theta} = (a_1, b_1, a_2, b_2, ..., a_k, b_k)'$ *minimise l'expression* $\Xi' \left[\sum_\theta^*(\Xi)\right]^{-1} \Xi$.

On obtient

$$\hat{\theta} \quad = \quad \left(\hat{X}' \left[\sum_\theta^*(\Xi)\right]^{-1} \hat{X}\right)^{-1} \hat{X} \left[\sum_\theta^*(\Xi)\right]^{-1} \hat{Y}.$$

Encore une fois, comme $\sum_\theta^*(\Xi)$ est une matrice constante qui est fonction des paramètres inconnus, un procédé itératif est nécessaire.

2.3.1. La matrice de variance-covariance

Les résultats de la section (2.3) supposent que la forme de la matrice de variance-covariance est connue pour un k quelconque. Nous avons réalisé la difficulté à faire ressortir une forme générale pour les termes de $\sum_\theta(\Xi)$. En fait, pour déterminer la forme de la matrice de variance-covariance pour la famille de Sundt d'ordre k, on peut se servir de la forme de la matrice de variance-covariance pour la famille de Sundt d'ordre $k-1$. Pour illustrer cela, nous présentons l'exemple suivant.

EXEMPLE 2.3.1. *Situons-nous dans la famille de Sundt d'ordre* 3. *On est intéressé par le calcul de la variance de ε_1, de ε_2 et de ε_3. Les variances de ε_1 et ε_2 dans la famille $\mathcal{R}_3$ demeurent les mêmes que celles trouvées pour la famille $\mathcal{R}_2$ dans la proposition (2.2.2). Pour ε_3, on a*

$$\varepsilon_3 = p_3^* - (a_1 + \frac{b_1}{3})p_2^* - (a_2 + \frac{b_2}{3})p_1^* - (a_3 + \frac{b_3}{3})p_0^*.$$

Et donc

$$\begin{aligned} Var(\varepsilon_3) &= Var\left[p_3^* - (a_1 + \frac{b_1}{3})p_2^* - (a_2 + \frac{b_2}{3})p_1^*\right] + Var\left[(a_3 + \frac{b_3}{3})p_0^*\right] \\ &- 2(a_3 + \frac{b_3}{3})Cov\left[p_3^* - (a_1 + \frac{b_1}{3})p_2^* - (a_2 + \frac{b_2}{3})p_1^*, p_0^*\right]. \end{aligned}$$

On peut maintenant se sevir des expressions pour la variance de la proposition (2.2.2) pour calculer les deux termes de variance.

On pourrait se contenter de l'estimateur qui ne tient pas compte de la matrice $\sum_\theta^*(\Xi)$:

$$\hat{\theta} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}\hat{Y},$$

mais l'efficacité asymptotique de cet estimateur sera inférieure à celui qui prend en considération la structure de $\sum_\theta^*(\Xi)$. On pourrait aussi se contenter d'une matrice approximative où l'on tient compte des termes de la diagonale. Une fois encore, la convergence en probabilité de l'estimateur obtenu tient mais son efficacité sera inférieure à celui qui utilise toute l'information à notre disposition, sur les dépendances des résidus du modèle.

2.3.2. La famille $\mathcal{R}_k$

La discussion est similaire à celle de la section (2.2.2), et donc nous référons à cette section pour les résultats qui restent valides pour le cas non-tronqué.

CHAPITRE 3

TESTS D'HYPOTHÈSES

Au chapitre 2, nous avons présenté une méthode d'estimation qui s'avère être une excellente alternative aux méthodes usuelles. Huard (1996) a travaillé sur des tests de discrimination entre la loi de poisson et les autres éléments de la famille de Panjer.

Dans un premier temps, nous élargissons le contexte de sa recherche en supposant que l'échantillon observé provient de la famille $\mathcal{R}_2$. Nous testons l'hypothèse nulle qui stipule que les observations sont issues d'une loi de poisson. On traite les cas où le paramètre est connu et celui, plus habituel, où il faut l'estimer. Dans le cas où on conclut, à un certain niveau de confiance, que la loi de Poisson n'est pas appropriée pour modéliser les données, nous proposons de tester l'hypothèse nulle qui stipule que les données observées proviennent d'un autre membre de la famille de Panjer. Et si on rejette encore cette dernière hypothèse, on peut tester si l'échantillon observé provient d'un membre de la famille de Schröter. L'idée derrière cette série de tests réside dans le fait que la loi de poisson est incluse dans la famille de Panjer elle-même incluse dans la famille de Schröter qui, elle, est incluse dans la famille $\mathcal{R}_2$. Remarquons que l'on peut aussi effectuer l'un des tests, au choix, mais le fait que l'on ait une telle construction pour nos familles imbriquées, permet d'aborder de façon élégante le problème des tests d'hypothèses.

Dans un second temps, en nous situant dans la famille $\mathcal{R}_k$, où k est un entier non-nul fini et spécifié, nous testons l'hypothèse nulle selon laquelle l'échantillon observé est issu de la famille $\mathcal{R}_l$, où l est un entier non-nul strictement inférieur à k et $l > 1$.

3.1. Tests d'hypothèses dans la famille $\mathcal{R}_2$

Dans cette section, nous supposons que l'échantillon observé $n_1, n_2, ..., n_m$ provient de la famille $\mathcal{R}_2$. Cette famille, nous l'avons vu au chapitre 2, suggère la représentation suivante:

$$\hat{p}_n = \sum_{i=1}^{2}(a_i + \frac{b_i}{n})\hat{p}_{n-i} + \varepsilon_n,\ \hat{p}_{-1} = 0,$$

où $\hat{p}_n = \frac{f_n}{m}$, f_n dénote la fréquence des observations dans l'échantillon prenant la valeur n, ε_n est la variable aléatoire des erreurs et m est la taille de l'échantillon. Soit $\hat{Y} = (\hat{p}_1, \hat{p}_2, ..., \hat{p}_w)'$ où w est une valeur fixée; lorsque la taille de l'échantillon tend vers l'infini, ces fréquences sont non-nulles avec probabilité 1. Soit aussi $I_j\,[n \in [j, j+1)]$, $j = 0, 1, 2, ..., w$, définie à la section 2.2. On peut écrire

$$\hat{Y} = (f_1/m, f_2/m, ..., f_w/m)' = \left(\int_{I_1} dF_m, \int_{I_2} dF_m, ..., \int_{I_w} dF_m\right)',$$

où F_m est la fonction de répartition empirique de l'échantillon observé.

Asymptotiquement, on sait que

$$E(\hat{Y}) = X\theta \quad = \begin{pmatrix} \int_{I_0} dF_\theta & \int_{I_0} dF_\theta & 0 & 0 \\ \int_{I_1} dF_\theta & \frac{1}{2}\int_{I_1} dF_\theta & \int_{I_0} dF_\theta & \frac{1}{2}\int_{I_0} dF_\theta \\ \int_{I_2} dF_\theta & \frac{1}{3}\int_{I_2} dF_\theta & \int_{I_1} dF_\theta & \frac{1}{3}\int_{I_1} dF_\theta \\ \cdot & \cdot & \cdot & \cdot \\ \int_{I_{w-1}} dF_\theta & \frac{1}{3}\int_{I_{w-1}} dF_\theta & \int_{I_{w-2}} dF_\theta & \frac{1}{3}\int_{I_{w-2}} dF_\theta \end{pmatrix} \theta,$$

où F_θ est la fonction de répartition théorique.

Considérons la mesure de distance suivante, entre F_m et F_θ, traitée par Luong et Thompson (1987):

$$d(F_m, F_\theta) \quad = \quad (\hat{Y} - X\theta)' \left[\sum_\theta(\Xi)\right]^{-1} (\hat{Y} - X\theta).$$

Une telle mesure de distance a été utilisée par Doray et Luong (1995), Huard (1996).

Tout au long de ce chapitre, nous utilisons le théorème suivant qui concerne la distribution d'une forme quadratique. Pour une discussion de ce dernier, nous renvoyons à Moore (1977).

THÉORÈME 3.1.0.1. *Supposons que* $Y \sim N_p(0, \sum)$ *et soit* C *une matrice symétrique définie positive de dimension* $(p \times p)$*. Si* $\sum C$ *est idempotente et si* $trace(\sum C) = \nu$*, alors la forme quadratique* $Y'CY$ *suit une loi* χ^2_ν*. Si on remplace* C *par* $\hat{C}$*, un estimateur convergent de* C*, alors* $Y'\hat{C}Y \xrightarrow{\mathcal{L}} \chi^2_\nu$*.*

3.1.1. Test d'hypothèse pour H_0 : $a_1 = 0, b_1 = b_{10}, a_2 = 0, b_2 = 0$

Nous testons ici l'hypothèse nulle selon laquelle les données proviennent d'une loi de poisson de moyenne b_{10} positive connue. Soit $\theta_0 = (0, b_{10}, 0, 0)'$. Pour ce faire, considérons

$$
\begin{aligned}
d(F_m, F_{\theta_0}) &= (\hat{Y} - X\theta_0)' \left[\sum\nolimits_{\theta_0}(\Xi)\right]^{-1} (\hat{Y} - X\theta_0) \\
&= m(\hat{Y} - X\theta_0)' \left[\sum\nolimits_{\theta_0}^{*}(\Xi)\right]^{-1} (\hat{Y} - X\theta_0).
\end{aligned}
$$

Sous H_0, nous avons $\sqrt{m}(\hat{Y} - X\theta_0) \xrightarrow{\mathcal{L}} N(0, \sum_1^*)$,

où $\sum_1^* = mVar(\hat{Y} - X\theta_0) = mVar(\Xi) = \sum_{\theta_0}^*$.

D'où, $\sum_1^* \left[\sum_{\theta_0}^*(\Xi)\right]^{-1} = I_w$ est une matrice idempotente dont la trace est égale à w. Par le théorème (3.1.0.1), il suit que $d(F_m, F_{\theta_0}) \xrightarrow{\mathcal{L}} \chi_w^2$.

Remarquons que $d(F_m, F_{\theta_0})$ dépend de X qui est inconnue. On utilise un estimateur convergent de X et comme $\hat{d}(F_m, F_{\theta_0}) \xrightarrow{p} d(F_m, F_{\theta_0})$, où $\hat{d}(F_m, F_{\theta_0})$ est la distance calculée en remplaçant X par $\hat{X}$, il découle que $\hat{d}(F_m, F_{\theta_0}) \xrightarrow{\mathcal{L}} \chi_w^2$.

Le test consiste donc à rejeter H_0 au niveau approximatif (1-α) si

$$
\hat{d}(F_m, F_{\theta_0}) > \chi^2_{w;1-\alpha}
$$

où $\chi^2_{w;1-\alpha}$ est le quantile d'une loi χ_w^2 d'ordre $(1 - \alpha)$.

3.1.2. Test d'hypothèse pour H_0 : $a_1 = 0, a_2 = 0, b_2 = 0$

Dans un contexte pratique, les vraies valeurs des paramètres sont habituellement inconnues. Il faut donc les estimer. Nous testons alors l'hypothèse nulle selon laquelle les données proviennent d'une loi de poisson de moyenne inconnue b_1.

Posons $\tilde{\theta} = (0, \tilde{b_1}, 0, 0)'$, l'EDQM de θ calculé en posant au départ $a_1 = a_2 = b_2 = 0$. Ajoutons que $\tilde{b_1}$ doit être un nombre positif, de façon à avoir une fonction de probabilité valide.

Soit l'expression de distance suivante :

$$\begin{aligned} d(F_m, F_{\tilde{\theta}}) &= (\hat{Y} - X\tilde{\theta})' \left[\textstyle\sum_{\tilde{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\tilde{\theta}) \\ &= m(\hat{Y} - X\tilde{\theta})' \left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\tilde{\theta}). \end{aligned}$$

Sous H_0, $\sqrt{m}(\hat{Y} - X\tilde{\theta}) \xrightarrow{\mathcal{L}} N(0, \sum^*_2)$, où

$$\begin{aligned} \textstyle\sum^*_2 &= mVar(\hat{Y} - X\tilde{\theta}) \\ &= mVar(\hat{Y} - X^{(2)}\tilde{b}_1) \text{ , et } X^{(2)} \text{ est la seconde colonne de } X. \end{aligned}$$

On trouve

$$\begin{aligned} \textstyle\sum^*_2 &= mVar\left(\hat{Y} - X^{(2)}\left[X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}\hat{Y}\right) \\ &= mVar\left(\left[I - X^{(2)}\left[X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}\right]\hat{Y}\right) \\ &= \left[I - X^{(2)}\left[X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}\right]\textstyle\sum^*_{\tilde{\theta}}(\Xi) \\ &\times \left[I - \left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}X^{(2)}\left[X^{(2)'}\left[\textstyle\sum^*_{\tilde{\theta}}(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\right]. \end{aligned}$$

D' où

$$\begin{aligned}\Sigma_2^*\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1} &= \left[I - X^{(2)}\left[X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}\right] \\ &\times \left[I - X^{(2)}\left[X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}\right] \\ &= I - X^{(2)}\left[X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1},\end{aligned}$$

qui est une une matrice idempotente avec trace

$$tr\left(\Sigma_2^*\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}\right) = tr\left[I - X^{(2)}\left[X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}X^{(2)}\right]^{-1}X^{(2)'}\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}\right]$$
$$= w - tr(I_1) = w - 1.$$

On conclut ainsi que $d(F_m, F_{\tilde{\theta}}) \xrightarrow{\mathcal{L}} \chi^2_{w-1}$. Le test consiste donc à rejeter H_0 au niveau approximatif (1-α) si

$$\hat{d}(F_m, F_{\tilde{\theta}}) \quad > \quad \chi^2_{w-1;1-\alpha}\text{ , où}$$

$\chi^2_{w-1;1-\alpha}$ est le quantile d'une loi χ^2_{w-1} d'ordre $(1-\alpha)$, et $\hat{d}(F_m, F_{\tilde{\theta}})$ est la distance calculée en utilisant $\hat{X}$.

3.1.3. Test d'hypothèse pour H_0 : $a_1 = 0, a_2 = 0, b_2 = 0$

Une autre approche pour tester si l'échantillon provient d'une loi Poisson consiste à utiliser la distribution asymptotique de $\hat{\theta}$. En effet, nous savons que

$$\sqrt{m}\hat{\theta} \xrightarrow{\mathcal{L}} N\left(0, \left(X'\left[\Sigma_{\tilde{\theta}}^*(\Xi)\right]^{-1}X\right)^{-1}\right).$$

Soit $\left(X'\left[\Sigma_{\hat{\theta}}(\Xi)\right]^{-1}X\right)^{-1}[j,j]$, $j = 1, 3, 4$, le j^{ieme} élément de la diagonale de la matrice $\left(X'\left[\Sigma_{\hat{\theta}}(\Xi)\right]^{-1}X\right)^{-1}$. On peut calculer les trois intervalles de confiance de niveau approximatif (1-α) suivants:

$$
\begin{aligned}
a_1 &\in \left(\hat{a}_1 \pm z_{1-\alpha/2}\sqrt{\left(X' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} X\right)^{-1}[1,1]}\right), \\
a_2 &\in \left(\hat{a}_2 \pm z_{1-\alpha/2}\sqrt{\left(X' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} X\right)^{-1}[3,3]}\right) \textit{ et} \\
b_2 &\in \left(\hat{b}_2 \pm z_{1-\alpha/2}\sqrt{\left(X' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} X\right)^{-1}[4,4]}\right)
\end{aligned}
$$

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

Il est bien connu (voir Seber (1977), p.125) que le niveau de confiance simultané n'est pas de $(1-\alpha)$. Nous utiliserons ici l'approche de Bonferroni pour avoir un niveau de confiance global près de $(1-\alpha)$. On veut déterminer les intervalles de confiance pour a_1, a_2 et b_2.

Nous pouvons utiliser un niveau de confiance individuel de $\alpha/3$ de façon à avoir un niveau de confiance global $(1-\alpha)$ approximativement.

Les trois intervalles corrigés sont:

$$
\begin{aligned}
a_1 &\in \left(\hat{a}_1 \pm z_{1-\alpha/6}\sqrt{\left(X' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} X\right)^{-1}[1,1]}\right), \\
a_2 &\in \left(\hat{a}_2 \pm z_{1-\alpha/6}\sqrt{\left(X' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} X\right)^{-1}[3,3]}\right) \textit{ et} \\
b_2 &\in \left(\hat{b}_2 \pm z_{1-\alpha/6}\sqrt{\left(X' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} X\right)^{-1}[4,4]}\right)
\end{aligned}
$$

où $z_{1-\alpha/6}$ est le quantile d'ordre $(1-\alpha/6)$ d'une loi $N(0,1)$.

Si 0 appartient aux trois intervalles ci-haut, on pourra conclure, avec un niveau de confiance approximatif de $(1-\alpha)$ que les données proviennent bien d'une loi de Poisson.

3.1.4. Test d'hypothèse pour H_0 : $a_2 = 0, b_2 = 0$

On veut tester l'hypothèse nulle selon laquelle l'échantillon observé est issu de la famille de Panjer. On suppose que a_1 et b_1 sont des paramètres inconnus.

Soit $\bar{\theta} = (\bar{a}_1, \bar{b}_1, 0, 0)$, l'EDQM de θ obtenu en posant $a_2 = b_2 = 0$ au départ.

On considère la distance

$$\begin{aligned} d(F_m, F_{\bar{\theta}}) &= (\hat{Y} - X\bar{\theta})' \left[\Sigma_{\bar{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\bar{\theta}) \\ &= m(\hat{Y} - X\bar{\theta})' \left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\bar{\theta}). \end{aligned}$$

Sous H_0, $\sqrt{m}(\hat{Y} - X\bar{\theta}) \xrightarrow{\mathcal{L}} N(0, \Sigma^*_3)$, où

$$\begin{aligned} \Sigma^*_3 &= mVar(\hat{Y} - X\bar{\theta}) \\ &= mVar\left(\hat{Y} - X^{(1,2)}(\bar{a}_1, \bar{b}_1)'\right) \text{ et} \end{aligned}$$

$X^{(1,2)}$ est la matrice formée des deux premières colonnes de X.

On calcule

$$\begin{aligned} \Sigma^*_3 &= mVar\left(\hat{Y} - X^{(1,2)}\left[X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}X^{(1,2)}\right]^{-1}X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}\hat{Y}\right) \\ &= mVar\left(\left[I - X^{(1,2)}\left[X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}X^{(1,2)}\right]^{-1}X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}\right]\hat{Y}\right) \\ &= \left[I - X^{(1,2)}\left[X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}X^{(1,2)}\right]^{-1}X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}\right]\Sigma^*_{\bar{\theta}}(\Xi) \\ &\times \left[I - \left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}X^{(1,2)}\left[X^{(1,2)'}\left[\Sigma^*_{\bar{\theta}}(\Xi)\right]^{-1}X^{(1,2)}\right]^{-1}X^{(1,2)'}\right]. \end{aligned}$$

Ainsi,

$$
\begin{aligned}
\Sigma_3^* \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1} &= \left[I - X^{(1,2)} \left[X^{(1,2)'} \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1} X^{(1,2)}\right]^{-1} X^{(1,2)'} \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1}\right] \\
&\times \left[I - X^{(1,2)} \left[X^{(1,2)'} \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1} X^{(1,2)}\right]^{-1} X^{(1,2)'} \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1}\right] \\
&= I - X^{(1,2)} \left[X^{(1,2)'} \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1} X^{(1,2)}\right]^{-1} X^{(1,2)'} \left[\Sigma_{\bar{\theta}}^*(\Xi)\right]^{-1}
\end{aligned}
$$

qui est une une matrice idempotente dont la trace est égale à $w - 2$.

On a ainsi $d(F_m, F_{\bar{\theta}}) \xrightarrow{\mathcal{L}} \chi^2_{w-2}$. On rejette donc H_0 au niveau de confiance approximatif (1-α) si

$$
\hat{d}(F_m, F_{\bar{\theta}}) \quad > \quad \chi^2_{w-2;1-\alpha},
$$

où $\chi^2_{w-2;1-\alpha}$ est le quantile d'une loi χ^2_{w-2} d'ordre $(1 - \alpha)$, et $\hat{d}(F_m, F_{\bar{\theta}})$ est la distance calculée en utilisant $\hat{X}$. On pourrait aussi utiliser l'approche présentée à la section 3.1.3.

3.1.5. Test d'hypothèse pour H_0 : $a_2 = 0$,

Si on rejette les hypotèses nulles testées précédemment, on peut vouloir tester si notre échantillon provient de la famille de Schröter. Soient $\ddot{\theta} = (\ddot{a}_1, \ddot{b}_1, 0, \ddot{b}_2)'$ l'EDQM de θ que l'on obtient en posant $a_2 = 0$ et la distance

$$
\begin{aligned}
d(F_m, F_{\ddot{\theta}}) &= (\hat{Y} - X\ddot{\theta})' \left[\Sigma_{\ddot{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\ddot{\theta}) \\
&= m(\hat{Y} - X\ddot{\theta})' \left[\Sigma_{\ddot{\theta}}^*(\Xi)\right]^{-1} (\hat{Y} - X\ddot{\theta}).
\end{aligned}
$$

Sous H_0, $\sqrt{m}(\hat{Y} - X\ddot{\theta}) \xrightarrow{\mathcal{L}} N(0, \Sigma_4^*)$, où

$$
\Sigma_4^* \quad = \quad mVar(\hat{Y} - X\ddot{\theta}).
$$

On obtient

$$\Sigma_4^* \left[\Sigma_{\ddot{\theta}}^*(\Xi)\right]^{-1} \; = \; I - X^{(1,2,4)} \left[X^{(1,2,4)'} \left[\Sigma_{\ddot{\theta}}^*(\Xi)\right]^{-1} X^{(1,2,4)}\right]^{-1} X^{(1,2,4)'} \left[\Sigma_{\ddot{\theta}}^*(\Xi)\right]^{-1}$$

où $X^{(1,2,4)}$ est la matrice X sans sa troisième colonne. La matrice $\Sigma_4^* \left[\Sigma_{\ddot{\theta}}^*(\Xi)\right]^{-1}$ est idempotente et la trace est égale à $w - 3$.

Il vient que $d(F_m, F_{\ddot{\theta}}) \xrightarrow{\mathcal{L}} \chi^2_{w-3}$. Le test consiste donc à rejeter H_0 au niveau approximatif (1-α) si

$$\hat{d}(F_m, F_{\ddot{\theta}}) \; > \; \chi^2_{w-3;1-\alpha},$$

où $\chi^2_{w-3;1-\alpha}$ est le quantile d'une loi χ^2_{w-3} d'ordre $(1-\alpha)$. Une autre approche permet de construire un test basé sur la distribution asymptotique de $\hat{a}_2$. On testera l'hypotèse $H_0 : a_2 = 0$ au niveau approximatif α. On rejetera cette dernière au niveau de confiance approximatif (1-α) lorsque

$$|\hat{a}_2| / \sqrt{\left(X' \left[\Sigma_{\hat{\theta}}(\Xi)\right]^{-1} X\right)^{-1} [3,3]} \; > \; z_{1-\alpha/2} \text{ , où}$$

$\left(X' \left[\Sigma_{\hat{\theta}}(\Xi)\right]^{-1} X\right)^{-1} [3,3]$ est le troisième élément de la diagonale de la matrice $\left(X' \left[\Sigma_{\hat{\theta}}(\Xi)\right]^{-1} X\right)^{-1}$ et $z_{1-\alpha/2}$ le quantile d'ordre $(1-\alpha/2)$ d'une loi N(0,1).

3.2. TESTS D'HYPOTHÈSE DANS $\mathcal{R}_k$

Nous supposons dans cette section que notre échantillon observé $n_1, n_2, ..., n_m$ provient de la famille $\mathcal{R}_k$. Cette famille suggère la représentation suivante:

$$\hat{p}_n \; = \; \sum_{i=1}^{k} (a_i + \frac{b_i}{n}) \hat{p}_{n-i} + \varepsilon_n, \; \hat{p}_{-1} = \hat{p}_{-2} = ... = \hat{p}_{-(k-1)} = 0$$

Semblablement à la démarche de la section 3.1, nous en venons à considérer la mesure de distance suivante entre F_m et F_θ:

$$\begin{aligned} d(F_m, F_\theta) &= (\hat{Y} - X\theta)' \left[\Sigma_\theta(\Xi)\right]^{-1} (\hat{Y} - X\theta) \\ &= m(\hat{Y} - X\theta)' \left[\Sigma^*_\theta(\Xi)\right]^{-1} (\hat{Y} - X\theta), \text{ où} \end{aligned}$$

$\theta = (a_1, b_1, a_2, b_2, ..., a_k, b_k)'$. Nous désirons tester l'hypothèse nulle stipulant que les données proviennent de la famille $\mathcal{R}_l$, où $l < k$ et l est un entier non nul. Soit $H_0 = (a_1, b_1, a_2, b_2, ..., a_l, b_l, a_{l+1} = 0, ..., a_k = 0, b_k = 0)'$, où les paramètres de l'hypothèse nulle sont inconnus.

Soient $\tilde{\theta} = (\tilde{a}_1, \tilde{b}_1, \tilde{a}_2, \tilde{b}_2, ..., \tilde{a}_l, \tilde{b}_l, 0, 0, ...0)','$ l' EDQM de θ obtenu en posant $a_{l+1} = b_{l+1} = ... = a_k = b_k = 0$ au départ et

$$\begin{aligned} d(F_m, F_{\tilde{\theta}}) &= (\hat{Y} - X\tilde{\theta})' \left[\Sigma_{\tilde{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\tilde{\theta}) \\ &= m(\hat{Y} - X\tilde{\theta})' \left[\Sigma^*_{\tilde{\theta}}(\Xi)\right]^{-1} (\hat{Y} - X\tilde{\theta}). \end{aligned}$$

Sous H_0, $\sqrt{m}(\hat{Y} - X\tilde{\theta}) \xrightarrow{\mathcal{L}} N(0, \Sigma^*_5)$, où

$$\begin{aligned} \Sigma^*_5 &= mVar(\hat{Y} - X\tilde{\theta}). \\ &= mVar(\hat{Y} - X^{(1,2,...,2l-1,2l)}(a_1, b_1, ..., a_l, b_l)'), \end{aligned}$$

où $X^{(1,2,...,2l-1,2l)}$ est la matrice formée des $2l$ premières colonnes de X. Pour alléger

l'écriture, convenons de poser $X^{(1,2,...,2l-1,2l)} = X^{(2l)}$. Il résulte

$$\Sigma^*_5 \left[\Sigma^*_{\tilde{\theta}}(\Xi)\right]^{-1} = I - X^{(2l)} \left[X^{(2l)} \left[\Sigma^*_{\tilde{\theta}}(\Xi)\right]^{-1} X^{(2l)}\right]^{-1} X^{(2l)} \left[\Sigma^*_{\tilde{\theta}}(\Xi)\right]^{-1}$$

qui est une matrice idempotente de trace égale à $w - l$.

Ainsi, $d(F_m, F_{\tilde{\theta}}) \xrightarrow{\mathcal{L}} \chi^2_{w-l}$. Le test consiste donc à rejeter H_0 au niveau approximatif (1-α) si

$$\hat{d}(F_m, F_{\tilde{\theta}}) \quad > \quad \chi^2_{w-l;1-\alpha},$$

où $\chi^2_{w-l;1-\alpha}$ est le quantile d'une loi χ^2_{w-l} d'ordre $(1-\alpha)$ et $\hat{d}(F_m, F_{\tilde{\theta}})$ est telle que définie tout au long de ce chapitre.

3.3. Discussion

Tout d'abord, nous avons supposé implicitement que les estimateurs obtenus pour mettre en oeuvre les tests présentés dans ce chapitre appartiennent à l'espace paramétrique. Par exemple, pour les tests qui concernent la loi de poisson, nous imposons à l'estimateur de b_1 d' être un nombre réel strictement positif. Dans le cadre général de la famille $\mathcal{R}_k$, nous devons vérifier que les probabilités estimées, en utilisant les valeurs des paramètres estimés, sont comprises entre 0 et 1.

Le choix de la valeur de w est arbitraire lorsqu'on tronque le domaine infini de notre variable aléatoire. Il faut donc tenter d'utiliser le maximum d'information à notre disposition pour faire ce choix. On devrait prendre le plus grand w possible pour des fins d'efficacité de l'estimateur.

On a vu que les diverses statistiques des tests présentés dans ce chapitre convergent en loi vers une variable aléatoire χ^2. Dans un travail ultérieur, il faudrait faire des simulations pour voir comment se comporte le niveau du test, par exemple en fonction de la taille de l'échantillon ou encore des valeurs des paramètres. On pourrait aussi étudier, par simulation, la puissance des tests développés en choisissant des hypothèses alternatives appropriées.

Aussi, dans le cadre des tests qui supposent $\mathcal{R}_2$ comme famille de référence, on a obtenu la représentation complète de la matrice de la variance-covariance. Mais, si on veut utiliser les résultats de la section (3.2) où k peut être supérieur à 2, on doit trouver la forme explicite de la matrice de variance-covariance ou utiliser un estimateur convergent de la matrice de variance-covariance comme à la section (2.2.2). Dans un travail subséquent, il faudrait étudier l'effet de l'utilisation de cette estimation sur les tests.

CHAPITRE 4

UN EXEMPLE NUMÉRIQUE

A la section (4.1), nous illustrons, à l'aide d'un exemple, les méthodes développées dans ce mémoire. A la section (4.2), nous discutons certaines difficultés rencontrées en pratique et nous concluons le mémoire.

4.1. Exemple

Nous considérons la loi obtenue en tronquant la convolution d'une loi de poisson avec $\lambda = 2$ et d'une loi binomiale négative, avec $s = 2$ et $q = 2$. Cette convolution, nous l'avons vu au chapitre 1, appartient à la famille de Schröter. Le domaine de la loi tronquée est l'ensemble $\{0,1,2,...,8\}$. Pour obtenir les probabilités de la loi tronquée, p_0^*, p_1^*, ..., p_8^*, nous multiplions les probabilités de la loi non-tronquée, p_0, p_1, ..., p_8, par $(\sum_{n=0}^{8} p_n)^{-1}$. Les probabilités théoriques de la loi tronquée sont présentés au tableau (4.1). La méthode de la fonction de répartition est utilisée pour générer l'échantillon qui va servir pour la suite. Le programme S-Plus écrit à cette fin est présenté à l'appendice B. La taille de notre échantillon est de 15 000. Cette taille est très raisonnable dans le contexte de l'actuariat où la variable N représente le nombre d'accidents des assurés d'une compagnie d'assurance, durant une cetaine période. On peut citer les travaux de Lemaire (1985) qui a étudié un portefeuille de 106 965 assurés pour étudier le phénomène décrit ci-haut. Les fréquences observées, pour l'échantillon utilisé pour la suite, sont présentées au tableau (4.1).

Considérons le modèle de régression suivant:

$$\hat{p}_n^* = (a_1 + \frac{b_1}{n})\hat{p}_{n-1}^* + \frac{b_2}{n}\hat{p}_{n-2}^* + \varepsilon_n, \; \hat{p}_{-1}^* = 0, \; n = 1, 2, 3, ..., 8.$$

Sous forme matricielle, le modèle s'écrit:

$$\hat{Y} = \hat{X}\theta + \Xi,$$

où

$$\hat{Y} = (0.0657, 0.1127, 0.1486, 0.1605, 0.1571, 0.1315, 0.1153, 0.0901)',$$

$$\hat{X} = \begin{pmatrix} 0.0186 & 0.0186 & 0 \\ 0.0657 & 0.0329 & 0.0093 \\ 0.1127 & 0.0376 & 0.0219 \\ 0.1486 & 0.0372 & 0.0282 \\ 0.1605 & 0.0321 & 0.0297 \\ 0.1571 & 0.0262 & 0.0267 \\ 0.1315 & 0.0188 & 0.0224 \\ 0.1153 & 0.0144 & 0.0164 \end{pmatrix},$$

$$\theta = (a_1, b_1, b_2)'.$$

Nous calculons d'abord une première valeur de θ en posant $\sum_\theta = I_8$. On obtient $\ddot{\theta} = (0.5561, 2.6469, -0.6766)'$. Avec cette valeur de $\ddot{\theta}$, on peut évaluer numériquement $\sum_\theta$. Les programmes S-Plus qui font ce calcul sont présentés à l'appendice B. Avec cette matrice $\hat{\sum}_{\ddot{\theta}}$, on peut calculer une seconde valeur pour θ. On répète le procédé itératif jusqu'à obtenir la précision désirée pour les trois valeurs. Dans notre example, on a obtenu la convergence des trois valeurs, avec une précision de 4 chiffres après le point décimal, après 5 itérations.

La solution finale est $\ddot{\theta} = (0.6411, 2.6235, -1.1267)'$, et la matrice de variance-covariance estimée des paramètres est:

$$\hat{Var}(\ddot{\theta}) = \begin{pmatrix} 0.0041 & 0.0021 & -0.0258 \\ 0.0021 & 0.0012 & -0.0135 \\ -0.0258 & -0.0135 & 0.1607 \end{pmatrix}.$$

Pour voir si le paramètre b_2 est significatif dans le modèle, on calcule l'intervalle de confiance au niveau de approximatif de 95%. On obtient:

$$b_2 \in (-1.7858, -0.4675).$$

On conclut donc, au niveau de confiance approximatif de 95%, que le paramètre b_2 est significativement différent de 0.

On peut maintenant utiliser les résultats de la section (3.1.5) pour tester la qualité de l'ajustement en calculant la distance présentée. Dans notre exemple, on obtient $\hat{d}(F_{15000}, F_{\ddot{\theta}}) = 6.0114$. Comme $6.0114 < \chi^2_{5;0.95} = 11.07$, on conclut que la famille considérée dans cette illustration est appropriée pour modéliser cet échantillon. Au tableau (4.1), on présente les fréquences estimées par le modèle.

n	Probabilités théoriqes de la loi tronquée	Fréquences observées de l'échantillon	Fréquences estimées par le modèle
0	0.0191	279	299
1	0.0636	986	911
2	0.1145	1691	1769
3	0.1499	2229	2192
4	0.1616	2408	2415
5	0.1540	2357	2304
6	0.1352	1973	2090
7	0.1123	1730	1625
8	0.0898	1350	1398

TABLEAU 4.1. Résultats de la simulation

4.2. Discussion et conclusion

L'exemple de la section (4.1) fait ressortir la simplicité de la mise en oeuvre pratique de la méthode d'estimation présentée au chapitre 2. On obtient la convergence du procédé itératif après 5 itérations, ce qui est arrivé fréquemment dans d'autres applications. Cependant, il faut souligner que, dans certains cas, le procédé itératif ne converge pas. On a remarqué qu'en pratique, s'il y a convergence du procédé itératif, celle-ci a lieu en moins de 10 itérations. Aussi, les variances estimées des estimateurs $\hat{a}_1$, $\hat{b}_1$ et $\hat{b}_2$ sont toutes les trois positives. Mais, en pratique, des cas sont survenus où une ou plusieurs des variances estimées étaient négatives. On rencontre cette situation dans d'autre domaines, comme en théorie du sondage, par exemple. L' élaboration d' estimateurs à variance non négative est un problème qui mérite une attention particulière.

Dans ce mémoire, nous avons présenté une méthode d'estimation qui s'avère être une alternative intéressante à la méthode du maximum de vraisemblance, qui est difficile à mettre en oeuvre dans le contexte de la famille de Sundt. Les propriétés asymptotiques permettent d'obtenir des résultats fiables, en présence d'échantillons de grande taille. Nous avons tout de même fait ressortir certaines difficultés d'ordre pratique. Au chapitre 3, des tests d'hypothèses ont été proposés et il faudrait étudier, au moyen de simulations, les comportements pratiques de ces tests. Finalement, la méthode de distance quadratique minimale permet un traitement unifié du problème de l'estimation et du problème des tests d'hypothèses. Des travaux de recherche se poursuivent pour généraliser les méthodes de ce mémoire à d'autres familles définies de façon récursive.

BIBLIOGRAPHIE

ADELSON, R.M. (1966), Compound poisson distributions, *Operational Research Quarterly,* 17, 73-75.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A., NESBIT, C.J. (1986), *Actuarial Mathematics,* Society of Actuaries.

DORAY, L.G., LUONG, A. (1995), Quadratic distance estimators for the zeta family, *Insurance: Mathematics and Economics,* 16, 255-260.

DORAY, L.G., LUONG, A. (1997), Efficient estimators for the Good family, *Communications in Statistics,* 26, 1075-1088.

DUCHESNE, P. (1996), Quelques aspects de la robustesse dans les modèles de régression et pour le cas de l'estimation de la position et de l'échelle, *Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal.*

GOURIÉROUX, C. , MONFORT, A. (1989), *Statistique et Modèles Econométriques, Economica.*

HAMPEL, F.R. (1974), The influence curve and its role in robust estimation, *Journal of the American Statistical Association,* 69, 346, 383-393.

HAMPEL, F.R, RONCHETTI, E.M, ROUSSEAU, P.J., STAHEL, W.A (1986), *Robust Statistics: The Approach Based on Influence Functions, Wiley: New York.*

HUARD, L. (1995), Nouveaux Tests d'Ajustement pour la Loi de Poisson et Généralisation, *Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal.*

LEMAIRE, J. (1985), *Automobile Insurance Actuarial Models, Hueboner International, Series on Risk, Insurance, and Economics Security.*

LUONG, A., DORAY, L.G. (1996), Goodness of fit test statistics for the zeta family, *Insurance: Mathematics and Economics,* 19, 45-53.

LUONG, A., GARRIDO, J. (1993) Minimum quadratic distance estimation for a parametric family of discrete distributions defined recursively, *Australian Journal of Statistic,* 35, 59-67.

LUONG, A., THOMPSON, M.E. (1987), Minimum-distance methods based on quadratic distances for transforms, *The Canadian Journal of Statistics,* 15, No.3, 239-251.

MOORE, D.S. (1977), Generalized inverses, Wald's method and the construction of chi-squared tests of fit, *Journal of the American Statistical Association,* 72, 131-137.

PANJER, H.H. (1981), Recursive evaluation of a family of compound distributions, *Astin Bulletin,* 12, 22-26.

SCHRÖTER, K.J. (1990), On a family of counting distributions and recursions for related compound distributions, *Scandinavian Actuarial Journal,* 161-175.

SEBER, G.A.F. (1977), *Linear Regression Analysis, Wiley: New York.*

SERFLING, R.J. (1980), *Approximation Theorems of Mathematical Statistics, Wiley: New York.*

SUNDT, B. (1992), On some extensions of Panjer's class of counting distributions, *Astin Bulletin,* 22, 61-80.

SUNDT, B., JEWELL, W.S. (1981), Further results on recursive evaluation of compound distributions, *Astin Bulletin,* 12, 27-39.

WILLMOT, G.E. (1987), The poisson-inverse gaussian distribution as an alternative to the negative binomial, *Scandinavian Actuarial Journal,* 113-127.

WILLMOT, E., SUNDT, B. (1989), On evaluation of the Delaporte distribution and related distributions, *Scandinavian Actuarial Journal,* 101-113.

APPENDICE A

PREUVE DE LA PROPOSITION (2.2.1)

Comme $\hat{p}_n^* = \frac{f_n^*}{m}$, que f_n^* suit une loi binomiale(m, p_n^*) et que (f_i^*, f_j^*), $i \neq j$, suit une loi trinomiale (m, p_i^*, p_j^*), nous obtenons les résultats suivants:

• D'abord, pour $n = 1, 2, 3, ..., w$, on a:

$E(\varepsilon_n) = E[\hat{p}_n^* - \sum_{i=1}^{2}(a_i + \frac{b_i}{n})\hat{p}_{n-i}^*] =$

$p_n - \sum_{i=1}^{2}(a_i + \frac{b_i}{n})p_{n-i} = 0,\ n = 1, 2, 3, ..., w.$

•$Var(\varepsilon_1) = Var[\hat{p}_1^* - (a_1 + \frac{b_1}{1})\hat{p}_0^*] =$

$Var(\hat{p}_1^*) + (a_1 + \frac{b_1}{1})^2 Var(\hat{p}_0^*) - 2(a_1 + \frac{b_1}{1})Cov(\hat{p}_1^*, \hat{p}_0^*) =$

$\frac{1}{m}\{p_1^*(1 - p_1^*) + (a_1 + \frac{b_1}{1})^2 p_0^*(1 - p_0^*) + 2(a_1 + \frac{b_1}{1})p_1^* p_0^*\} =$

$\frac{1}{m}\{p_1^* - (p_1^*)^2 + (a_1 + \frac{b_1}{1})p_1^*(1 - p_0^*) + 2(p_1^*)^2\} = \frac{1}{m}\{\frac{p_1^* p_0^*}{p_0^*} + \frac{(a_1 + \frac{b_1}{1})p_1^* p_0^*}{p_0^*}\} =$

$\frac{1}{m}\{\frac{p_1^* p_0^*}{p_0^*} + (p_1^*)^2 p_0^*\}.$

• Pour $n = 2, 3, 4, ..., w$, on a:

$$Var(\varepsilon_n) = Var[\hat{p}_n^* - (a_1 + \tfrac{b_1}{n})\hat{p}_{n-1}^* - (a_2 + \tfrac{b_2}{n})\hat{p}_{n-2}^*] =$$

$$\tfrac{1}{m}\{p_n^*(1 - p_n^*) + (a_1 + \tfrac{b_1}{n})^2 p_{n-1}^*(1 - p_{n-1}^*) + (a_2 + \tfrac{b_2}{n})^2 p_{n-2}^*(1 - p_{n-2}^*)$$

$$+2(a_1 + \tfrac{b_1}{n})p_n^* p_{n-1}^* + 2(a_2 + \tfrac{b_2}{n})p_n^* p_{n-2}^* - 2(a_1 + \tfrac{b_1}{n})(a_2 + \tfrac{b_2}{n})p_{n-1}^* p_{n-2}^* =$$

$$\tfrac{1}{m}\{p_n^* - (p_n^*)^2 + (a_1 + \tfrac{b_1}{n})[p_n^* - (a_2 + \tfrac{b_2}{n})p_{n-2}^*](1 - p_{n-1}^*)$$

$$+(a_2+\tfrac{b_2}{n})[p_n^* - (a_1+\tfrac{b_1}{n})p_{n-1}^*](1-p_{n-2}^*)+2(p_n^*)^2 - 2(a_1+\tfrac{b_1}{n})(a_2+\tfrac{b_2}{n})p_{n-1}^* p_{n-2}^*\} =$$

$$\tfrac{1}{m}\{p_n^* - (p_n^*)^2 + p_n^*[(a_1 + \tfrac{b_1}{n}) + (a_2 + \tfrac{b_2}{n})] - (p_n^*)^2$$

$$-(a_1 + \tfrac{b_1}{n})(a_2 + \tfrac{b_2}{n})[p_{n-1}^* + p_{n-2}^*] + 2(p_n^*)^2\} =$$

$$\tfrac{1}{m}\{p_n^*[1 + \textstyle\sum_{j=1}^2 c_{j,n}] - (p_{n-1}^* + p_{n-2}^*)\prod_{j=1}^2 c_{j,n}\}.$$

$$\bullet Cov(\varepsilon_1, \varepsilon_2) =$$

$$Cov[\hat{p}_1^* - (a_1 + \tfrac{b_1}{1})\hat{p}_0^*, \hat{p}_2^* - (a_1 + \tfrac{b_1}{2})\hat{p}_1^* - (a_2 + \tfrac{b_2}{2})\hat{p}_0^*] =$$

$$\tfrac{1}{m}\{-p_1^* p_2^* - (a_1+\tfrac{b_1}{2})p_1^*(1-p_1^*) + (a_2+\tfrac{b_2}{2})p_1^* p_0^* + (a_1+\tfrac{b_1}{1})p_0^* p_2^* - (a_1+\tfrac{b_1}{1})(a_1+\tfrac{b_1}{2})p_0^* p_1^*$$

$$+(a_1 + \tfrac{b_1}{1})(a_2 + \tfrac{b_2}{2})p_0^*(1 - p_0^*)\} =$$

$$\tfrac{1}{m}\{-p_2^* + (a_2 + \tfrac{b_2}{2})p_0^* + (a_1 + \tfrac{b_1}{1})p_2^* - (a_1 + \tfrac{b_1}{1})(a_1 + \tfrac{b_1}{2})p_1^*\} =$$

$$\tfrac{1}{m}\{c_{1,1}p_2^* - c_{1,2}p_1^*[1 + c_{1,1}]\}.$$

• Pour $n = 2, 3, 4, ..., w-1$, on a:

$Cov(\varepsilon_n, \varepsilon_{n+1}) =$

$Cov[\hat{p}_n^* - (a_1 + \frac{b_1}{n})\hat{p}_{n-1}^* - (a_2 + \frac{b_2}{n})\hat{p}_{n-2}^*, \hat{p}_{n+1}^* - (a_1 + \frac{b_1}{n+1})\hat{p}_n^* - (a_2 + \frac{b_2}{n+1})\hat{p}_{n-1}^*] =$

$\frac{1}{m}\{-p_n^* p_{n+1}^* - [p_{n+1}^* - (a_2 + \frac{b_2}{n+1})p_{n-1}^*](1 - p_n^*) + (a_2 + \frac{b_2}{n+1})p_n^* p_{n-1}^* +$

$(a_1 + \frac{b_1}{n})p_{n-1}^* p_{n+1}^* \; - (a_1 + \frac{b_1}{n})(a_1 + \frac{b_1}{n+1})p_{n-1}^* p_n^* +$

$(a_1 + \frac{b_1}{n})[p_{n+1}^* - (a_1 + \frac{b_1}{n+1})p_n^*](1 - p_{n-1}^*)$

$+(a_2 + \frac{b_2}{n})p_{n-2}^* p_{n+1}^* - (a_2 + \frac{b_2}{n})(a_1 + \frac{b_1}{n+1})p_{n-2}^* p_n^* - (a_2 + \frac{b_2}{n})(a_2 + \frac{b_2}{n+1})p_{n-2}^* p_{n-1}^*\} =$

$\frac{1}{m}\{-(a_1 + \frac{b_1}{n+1})p_n^* + p_{n+1}^* p_n^* + (a_1 + \frac{b_1}{n})p_{n+1}^* - (a_1 + \frac{b_1}{n})p_{n+1}^* p_{n-1}^* - (a_1 + \frac{b_1}{n})p_{n+1}^*$

$+(a_1 + \frac{b_1}{n})(a_2 + \frac{b_2}{n+1})p_{n-1}^* - (a_2 + \frac{b_2}{n})p_{n-2}^* p_{n+1}^*\} =$

$\frac{1}{m}\{c_{1,n}p_{n+1}^* - c_{1,n+1}p_n^*[1 + c_{1,n}]\}.$

$\bullet Cov(\varepsilon_1, \varepsilon_3) =$

$Cov[\hat{p}_1^* - (a_1 + \frac{b_1}{1})\hat{p}_0^*, \hat{p}_3^* - (a_1 + \frac{b_1}{3})\hat{p}_2^* - (a_2 + \frac{b_2}{3})\hat{p}_1^*] =$

$\frac{1}{m}\{-p_1^* p_3^* + (a_1 + \frac{b_1}{3})p_1^* p_2^* - [p_3^* - (a_1 + \frac{b_1}{3})p_2^*](1 - p_1^*)$

$+(a_1 + \frac{b_1}{1})p_0^* p_3^* - (a_1 + \frac{b_1}{1})(a_1 + \frac{b_1}{3})p_0^* p_2^* - (a_1 + \frac{b_1}{1})(a_2 + \frac{b_2}{3})p_0^* p_1^*\} = -\frac{c_{2,3}p_1^*}{m}.$

$\bullet$Pour $n = 2, 3, 4, ..., w-2$, on a:

$Cov(\varepsilon_n, \varepsilon_{n+2}) =$

$Cov[\hat{p}_n^* - (a_1 + \frac{b_1}{n})\hat{p}_{n-1}^* - (a_2 + \frac{b_2}{n})\hat{p}_{n-2}^*, \hat{p}_{n+2}^* - (a_1 + \frac{b_1}{n+2})\hat{p}_{n+1}^* - (a_2 + \frac{b_2}{n+2})\hat{p}_n^*] =$

$\frac{1}{m}\{-p_n^* p_{n+2}^* + (a_1 + \frac{b_1}{n+2})p_n^* p_{n+1}^* - [p_{n+2}^* - (a_1 + \frac{b_1}{n+2})p_{n+1}^*](1 - p_n^*)$

$+(a_1 + \frac{b_1}{n})p_{n-1}^* p_{n+2}^* - (a_1 + \frac{b_1}{n})(a_1 + \frac{b_1}{n+2})p_{n-1}^* p_{n+1}^* - (a_1 + \frac{b_1}{n})(a_2 + \frac{b_2}{n+2})p_{n-1}^* p_n^*$

$+(a_2 + \frac{b_2}{n})p_{n-2}^* p_{n+2}^* - (a_2 + \frac{b_2}{n})(a_1 + \frac{b_1}{n+2})p_{n-2}^* p_{n+1}^* - (a_2 + \frac{b_2}{n})(a_2 + \frac{b_2}{n+2})p_{n-2}^* p_n^*\} =$

$-\frac{c_{2,n+2} p_n^*}{m}$.

$\bullet$ Pour $k > 3$, on a:

$Cov(\varepsilon_1, \varepsilon_k) =$

$Cov[\hat{p}_1^* - (a_1 + \frac{b_1}{1})\hat{p}_0^*, \hat{p}_k^* - (a_1 + \frac{b_1}{k})\hat{p}_{k-1}^* - (a_2 + \frac{b_2}{k})\hat{p}_{k-2}^*] =$

$\frac{1}{m}\{-p_1^* p_k^* + p_1^*[(a_1 + \frac{b_1}{k})p_{k-1}^* + (a_2 + \frac{b_2}{k})p_{k-2}^*]$

$+(a_1 + \frac{b_1}{1})p_0^* p_k^* - (a_1 + \frac{b_1}{1})p_0^*[(a_1 + \frac{b_1}{k})p_{k-1}^* + (a_2 + \frac{b_2}{k})p_{k-2}^*]\} = 0.$

$\bullet$ Pour $j > 2$, on a:

$Cov(\varepsilon_n, \varepsilon_{n+j}) =$

$$Cov[\hat{p}_n^* - (a_1 + \tfrac{b_1}{n})\hat{p}_{n-1}^* - (a_2 + \tfrac{b_2}{n})p_{n-2}^*, \hat{p}_{n+j}^* - (a_1 + \tfrac{b_1}{n+j})\hat{p}_{n+j-1}^* -$$

$$(a_2 + \tfrac{b_2}{n+j})p_{n+j-2}^*] =$$

$$\tfrac{1}{m}\{-p_n^* p_{n+j}^* + p_n^*[p_{n+j}^* - (a_2 + \tfrac{b_2}{n+j})p_{n+j-2}^*] + (a_2 + \tfrac{b_2}{n+j})p_n^* p_{n+j-2}^* + p_{n+j}^* p_n^*$$

$$-(a_1 + \tfrac{b_1}{n})p_{n-1}^* p_{n+j}^* - (a_2 + \tfrac{b_2}{n})p_{n-2}^* p_{n+j}^*\} = 0.$$

APPENDICE B

PROGRAMMES S-PLUS

• Le programme suivant génère des échantillons de la convolution présentée au chapitre 4:

$rconvolution \leftarrow function(m)$ (On donne comme paramètre la taille de l'échantillon)

$\{z \leftarrow cumsum(convolution);$

$v \leftarrow matrix(nrow = n, ncol = 1);$

$for(i$ in 1:n)$\{a \leftarrow runif(1);$

$for(j$ in 1: 9)$(if$ a $<z[j])$

$\{$v[i]$\leftarrow j - 1$;break$\}\}\}$;as.vector(v)$\}$.

• La fonction suivante permet de calculer la matrice de variance-covariance des termes d'erreurs:

$S1 \leftarrow function(p1, p2, a1, b1, b2)$

$\{c(p1*(1+(a1+b1)), (a1+b1)*p2-(a1+b1/2)*p1*(1+(a1+b1)), -(b2/3)*p1,$

$0, 0, 0, 0, 0)\}$

$S2 \leftarrow function(p0, p1, p2, p3, a1, b1, b2)$

$\{c((a1+b1)*p2-(a1+b1/2)*p1*(1+(a1+b1)), p2*(1+(a1+b1/2)+(b2/2))-$

$(p1 - p0) * ((a1 + b1/2) * (b2/2)), (a1 + b1/2) * p3 - (a1 + b1/3) * p2*$

$(1 + (a1 + b1/2)), (b2/4) * p2, 0, 0, 0, 0)\}$

$S3 \leftarrow function(p1, p2, p3, p4, a1, b1, b2)$

$\{c(-(b2/3) * p1, (a1 + b1/2) * p3 - (a1 + b1/3) * p2 * (1 + (a1 + b1/2)), p3*$

$(1 + (a1 + b1/3) + (b2/3)) - (p2 - p1) * ((a1 + b1/3)*$

$b2/3)), (a1+b1/3)*p4-(a1+b1/4)*p3*(1+(a1+b1/3)), -(b2/5)*p3, 0, 0, 0)\}$

$S4 \leftarrow function(p2, p3, p4, p5, a1, b1, b2)$

$\{c(0, -(b2/4) * p2, (a1 + b1/3) * p4 - (a1 + b1/4) * p3 * (1 + (a1 + b1/3)), p4*$

$(1 + (a1 + b1/4) + (b2/4)) - (p3 - p2) * ((a1 + b1/4)*$

$(b2/4)), (a1+b1/4)*p5-(a1+b1/5)*p4*(1+(a1+b1/4)), -(b2/6)*p4, 0, 0)\}$

$S5 \leftarrow function(p3, p4, p5, p6, a1, b1, b2)$

$\{c(0, 0, -(b2/5) * p3, (a1 + b1/4) * p5 - (a1 + b1/5) * p4 * (1 + (a1 + b1/4)), p5*$

$(1 + (a1 + b1/5) + (b2/5)) - (p4 - p3) * ((a1 + b1/5)*$

$(b2/5)), (a1 + b1/5) * p6 - (a1 + b1/6) * p5 * (1 + (a1 + b1/5)), -(b2/7) * p5, 0)\}$

$S6 \leftarrow function(p4, p5, p6, p7, a1, b1, b2)$

{c(0,0,0,-(b2/6)*p4, (a1+b1/5)*p6-(a1+b1/6)*p5*(1+(a1+b1/5)), p6*

$(1 + (a1 + b1/6) + (b2/6)) - (p5 - p4) * ((a1 + b1/6)*$

$(b2/6)), (a1 + b1/6) * p7 - (a1 + b1/7) * p6 * (1 + (a1 + b1/6)), -(b2/8) * p6)\}$

$S7 \leftarrow function(p5, p6, p7, p8, a1, b1, b2)$

{c(0,0,0,0,-(b2/7)*p5, (a1+b1/6)*p7-(a1+b1/7)*p6*(1+(a1+b1/6)), p7*

$(1 + (a1 + b1/7) + (b2/7)) - (p6 - p5) * ((a1 + b1/7)*$

$(b2/7)), (a1 + b1/7) * p8 - (a1 + b1/8) * p7 * (1 + (a1 + b1/7)))\}$

$Sigma \leftarrow function(p0, p1, p2, p3, p4, p5, p6, p7, p8, a1, b1, b2)$

{rbind(S1(p1,p2,a1,b1,b2),S2(p0,p1,p2,p3,a1,b1,b2), S3(p1,p2,p3,p4,a1,b1,b2),

S4(p2,p3,p4,p5,a1,b1,b2),S5(p3,p4,p5,p6,a1,b1,b2),S6(p4,p5,p6,p7,a1,b1,b2),

S7(p5,p6,p7,p8,a1,b1,b2),S8(p5,p6,p7,p8,a1,b1,b2))}

- Fonction principale:

$RE \leftarrow function(n)$

$\{EE \leftarrow rconvolution(n);$

$r0 \leftarrow sum(EE == 0)/n;$

$r1 \leftarrow sum(EE == 1)/n;$

$r2 \leftarrow sum(EE == 2)/n;$

$r3 \leftarrow sum(EE == 3)/n;$

$r4 \leftarrow sum(EE == 4)/n;$

$r5 \leftarrow sum(EE == 5)/n;$

$r6 \leftarrow sum(EE == 6)/n;$

$r7 \leftarrow sum(EE == 7)/n;$

$r8 \leftarrow sum(EE == 8)/n;$

$Y \leftarrow c(r1, r2, r3, r4, r5, r6, r7, r8);$

$x1 \leftarrow c(r0, r1, r2, r3, r4, r5, r6, r7);$

$x2 \leftarrow c(r0, r1/2, r2/3, r3/4, r4/5, r5/6, r6/7, r7/8);$

$x3 \leftarrow c(0, r1/2, r2/3, r3/4, r4/5, r5/6, r6/7, r7/8);$

$X \leftarrow cbind(x1, x2, x3);$

$teta \leftarrow solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y;$

$p1 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[1,];$

$p2 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[2,];$

$p3 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[3,];$

$p4 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[4,];$

$p5 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[5,];$

$p6 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[6,];$

$p7 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[7,];$

$p8 \leftarrow (X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y)[8,];$

$p0 \leftarrow 1 - sum(X\% * \%solve(t(X)\% * \%X)\% * \%t(X)\% * \%Y);$

$a1 \leftarrow teta[1,];$

$b1 \leftarrow teta[2,];$

$b2 \leftarrow teta[3,];$

$S \leftarrow Sigma(p0, p1, p2, p3, p4, p5, p6, p7, p8, a1, b1, b2);$

$u \leftarrow solve(t(X)\% * \%solve(S)\% * \%X)\% * \%t(X)\% * \%solve(S)\% * \%Y;$

$m \leftarrow 0;$

$while((abs(teta[1,] - u[1,]) > 1e - 06)|(abs(teta[2,] - u[2,]) > 1e - 06)|$

$(abs(teta[1,] - u[1,]) > 1e - 06))$

$\{m \leftarrow m + 1; teta \leftarrow u;$

$a1 \leftarrow teta[1,];$

$b1 \leftarrow teta[2,];$

$b2 \leftarrow teta[3,];$

$p1 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[1,];$

$p2 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[2,];$

$p3 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[3,];$

$p4 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[4,];$

$p5 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[5,];$

$p6 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[6,];$

$p7 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[7,];$

$p8 \leftarrow (X\%*\%solve(t(X)\%*\%solve(S)\%*\%X)\%*\%t(X)\%*\%solve(S)\%*\%$

$Y)[8,];$

$S \leftarrow Sigma(p0, p1, p2, p3, p4, p5, p6, p7, p8, a1, b1, b2);$

$u \leftarrow solve(t(X)\% * \%solve(S)\% * \%X)\% * \%t(X)\% * \%solve(S)\% * \%Y;$

}; print(u); print(S);print(m); solve(t(X)%*%solve(S/n)%*%X)}