

**Record Number:** 11070  
**Author, Monographic:** Bureau, A./Perreault, L./Bobée, B.  
**Author Role:**  
**Title, Monographic:** Loi de Halphen type A : étude par simulation des estimateurs du maximum de vraisemblance  
**Translated Title:**  
**Reprint Status:**  
**Edition:**  
**Author, Subsidiary:**  
**Author Role:**  
**Place of Publication:** Québec  
**Publisher Name:** INRS-Eau  
**Date of Publication:** 1995  
**Original Publication Date:** Septembre 1995  
**Volume Identification:**  
**Extent of Work:** v, 81  
**Packaging Method:** pages incluant 3 annexes  
**Series Editor:**  
**Series Editor Role:**  
**Series Title:** INRS-Eau, rapport de recherche  
**Series Volume ID:** 459  
**Location/URL:**  
**ISBN:** 2-89146-397-8  
**Notes:** Rapport annuel 1995-1996  
**Abstract:** 15.00\$  
**Call Number:** R000459  
**Keywords:** rapport/ ok/ dl

**LOI DE HALPHEN TYPE A**

**ÉTUDE PAR SIMULATION DES  
ESTIMATEURS DU MAXIMUM  
DE VRAISEMBLANCE**

**LOI DE HALPHEN TYPE A**

**ÉTUDE PAR SIMULATION DES  
ESTIMATEURS DU MAXIMUM DE  
VRAISEMBLANCE**

**par**

**Alexandre Bureau**

**Luc Perreault**

**Bernard Bobée**

**Institut National de la Recherche Scientifique, INRS-Eau  
2800 rue Einstein, CP 7500, Sainte-Foy, Québec, G1V 4C7**

**Rapport de recherche No R-459**

**Septembre 1995**



# TABLE DES MATIÈRES

---

<b>LISTE DES TABLEAUX</b> .....	iii
<b>LISTE DES FIGURES</b> .....	v
<b>CHAPITRE 1 : Introduction</b> .....	<b>1</b>
1.1 Problématique .....	1
1.2 Procédure d'estimation.....	3
1.3 Objectifs .....	4
<b>CHAPITRE 2 : Caractéristiques de la simulation</b> .....	<b>5</b>
2.1 Choix de la méthode de génération .....	5
2.2 Choix des triplets de paramètres .....	9
2.3 Considérations sur les échantillons .....	15
<b>CHAPITRE 3 : Résultats</b> .....	<b>16</b>
3.1 Aperçu général des résultats .....	17
3.1.1. <i>Biais des estimateurs</i> .....	17
3.1.2. <i>Erreur quadratique moyenne des estimateurs</i> .....	19
3.1.3. <i>Distribution des estimateurs</i> .....	19
3.2 Erreur et biais relatif en fonction de $\alpha$ et de $v$ .....	22
3.2.1. <i>Biais et EQM en fonction de <math>\alpha</math></i> .....	22
3.2.2. <i>Biais et EQM en fonction de <math>v</math></i> .....	26
3.2.3. <i>Quelques cas extrêmes</i> .....	29
3.3 Erreur quadratique moyenne : variance et biais .....	31
3.4 Variances asymptotiques et expérimentales .....	34
3.5 Association entre les estimateurs des paramètres .....	36
3.6 Explications concernant le biais.....	37
<b>CHAPITRE 4 : Étude des problèmes d'estimation</b> .....	<b>39</b>
4.1 Considérations théoriques sur les absences de solution.....	39

4.2	Nombre d'échantillons sans solution en fonction de $\alpha$ , $\nu$ et $n$ .....	42
4.3	Effet des échantillons sans solution sur les résultats.....	45
4.4	Influence des solutions approximatives sur les résultats.....	48
<b>CHAPITRE 5</b>	<b>: Conclusions et recommandations .....</b>	<b>50</b>
5.1	Retour sur les résultats.....	50
5.2	Perspectives de recherche .....	51
5.3	Recommandations.....	53
<b>CHAPITRE 6</b>	<b>: Bibliographie.....</b>	<b>55</b>
<b>ANNEXE A</b>	<b>: Algorithme du traitement des échantillons simulés</b> .....	<b>57</b>
<b>ANNEXE B</b>	<b>: Code du programme <i>SIMULA</i>.....</b>	<b>61</b>
<b>ANNEXE C</b>	<b>: Manuel de l'utilisateur du programme <i>SIMULA</i> .</b>	<b>69</b>

## LISTE DES TABLEAUX

---

<b>Tableau 1.1:</b> Statistiques exhaustives de la loi de Halphen type A.....	2
<b>Tableau 2.1:</b> Résultats des tests d'adéquation pour la méthode UPM1.....	9
<b>Tableau 2.2:</b> Valeurs de $\alpha$ et de $\nu$ retenues pour la simulation.....	14
<b>Tableau 3.1:</b> Valeurs théoriques des quantiles lorsque $\alpha = 3, \nu = 2$ .....	17
<b>Tableau 3.2:</b> Part de variance dans l'erreur quadratique moyenne.....	33
<b>Tableau 3.3:</b> Moyennes et quantiles théoriques de quelques lois de Halphen type A.....	37
<b>Tableau 4.1:</b> Bornes de la dérivée de $\log L(\nu, \hat{\alpha}, \hat{m})$ aux points $-U$ et $U$ .....	40
<b>Tableau 4.2:</b> Nombre d'échantillons rejetés avant d'obtenir 500 répliqués.....	42
<b>Tableau 4.3:</b> Causes des rejets d'échantillons.....	43
<b>Tableau 4.4:</b> Effet du retrait des solutions approximatives sur le biais et la variance.....	49



## LISTE DES FIGURES

---

<b>Figure 2.1:</b> Distributions empiriques de taille 1000 générées par les méthodes LPM4 et INV2.....	7
<b>Figure 2.2:</b> Distributions empiriques de taille 1000 générées par la méthode UPM1.....	8
<b>Figure 2.3:</b> Estimation des paramètres $\alpha$ et $\nu$ de la loi de Halphen type A, ajustée aux observations de 103 stations hydrométriques du Québec et de l'Ontario .....	10
<b>Figure 2.4:</b> Coefficients de variation et d'asymétrie en fonction de $\alpha$ et de $\nu$ .....	12
<b>Figure 2.5:</b> Diagramme des coefficients de variation et d'asymétrie.....	13
<b>Figure 3.1:</b> Biais des estimateurs lorsque $\alpha = 3$ et $\nu = 2$ .....	18
<b>Figure 3.2:</b> Erreur quadratique moyenne des estimateurs lorsque $\alpha = 3$ et $\nu = 2$ .....	20
<b>Figure 3.3:</b> Distribution des estimateurs des paramètres ( $\alpha = 3$ et $\nu = 2$ ) .....	21
<b>Figure 3.4:</b> Distribution des estimateurs des quantiles ( $\alpha = 3$ et $\nu = 2$ ) .....	23
<b>Figure 3.5:</b> Biais relatif des estimateurs en fonction de $\alpha$ .....	24
<b>Figure 3.6:</b> Erreur relative des estimateurs en fonction de $\alpha$ .....	25
<b>Figure 3.7:</b> Biais relatif des estimateurs en fonction de $\nu$ .....	27
<b>Figure 3.8:</b> Erreur relative des estimateurs en fonction de $\nu$ .....	28
<b>Figure 3.9:</b> Biais relatif des estimateurs pour quelques cas extrêmes.....	30
<b>Figure 3.10:</b> Erreur relative lorsque $\alpha = 0,1$ et $\nu = 8$ ou $-8$ .....	32
<b>Figure 3.11:</b> Distribution de l'estimateur de $m$ lorsque $\alpha = 0,1$ et $\nu = 8$ ou $-8$ .....	32
<b>Figure 3.12:</b> Variances asymptotiques et expérimentales des estimateurs lorsque $\alpha=3$ et $\nu=2$ .....	35

**Figure 4.1:** Bornes des dérivées de  $\log L(\nu; \hat{\alpha}, \hat{m})$  aux points  $U$  et  $-U$ ..... 41

**Figure 4.2:** Distribution de l'estimateur de  $\nu$  lorsque  $\alpha = 0,1$  et  $\nu = 8$  ou  $-8$  ..... 46

**Figure 4.3:** Biais et EQM sur les paramètres lorsque  $\alpha = 0,1$  et  $\nu = 0$ ..... 47

**Figure 4.4:** Biais et EQM sur les paramètres lorsque  $\alpha = 10$  et  $\nu = 8$  ou  $-8$ ..... 47

# INTRODUCTION

---

## 1.1 Problématique

Un grand intérêt est porté en hydrologie aux débits extrêmes auxquels sont associés des probabilités au dépassement  $p$ . Dans le cas des débits de crue maximum annuelle, on veut évaluer le débit qui ne sera dépassé en moyenne qu'après un temps  $T = 1/p$  mesuré en années appelé une période de retour. Les événements extrêmes de période de retour  $T$  correspondent aux quantiles  $x_T$  de la loi de probabilité qui est utilisée comme modèle. L'ajustement d'une loi de probabilité à un échantillon de débits maximums annuels observés historiquement permet d'obtenir l'estimation  $\hat{x}_T$  de la valeur théorique du quantile inconnu  $x_T$ . On peut montrer que  $p = \Pr(X \geq x_T) = 1/T$  et que si  $F(x; \hat{\theta})$  désigne la fonction de répartition avec les paramètres estimés  $\hat{\theta}$ , alors le quantile estimé de période de retour  $T$  est donné par :

$$\hat{x}_T = F^{-1}\left(1 - 1/T; \hat{\theta}\right) \quad (1-1)$$

Cette procédure est appelée analyse de fréquence de crues (flood frequency analysis).

De manière générale, plusieurs lois de probabilité peuvent être utilisées pour modéliser des séries d'observations indépendantes mesurées sur les cours d'eau (débits, volumes, etc.) et déterminer des événements  $x_T$  de période de retour  $T$ . Cependant, peu d'entre elles ont été spécifiquement développées pour l'ajustement de séries de données hydrologiques. Étienne Halphen, hydrologue et statisticien français du milieu du siècle, a voulu remédier à cette lacune en développant de nouvelles lois de probabilités vérifiant certaines propriétés mathématiques et statistiques qu'il recherchait (Halphen, 1941; Morlat, 1956). Cependant, la complexité de leurs fonctions de densité de probabilité en a découragé l'utilisation à l'époque. On ne disposait pas de calculateurs assez performants pour effectuer rapidement les calculs numériques nécessaires à l'ajustement de ces lois et le sujet a été délaissé par les chercheurs pendant de nombreuses années. Depuis quelques années, grâce à des moyens informatiques plus avancés, de nouvelles recherches ont été entreprises.

On distingue trois types de lois de Halphen: A, B et  $B^{-1}$ . Le présent travail se limite à la loi de Halphen type A. La loi de type A correspond, à une reparamétrisation près, à la distribution appelée loi gaussienne inverse généralisée considérée, en particulier, par Barndorff-Nielsen *et al.* (1977) et étudiée en détail par Jørgensen (1982). La fonction de densité de probabilité d'une variable aléatoire  $X$  distribuée selon une loi de Halphen de Type A est donnée par:

$$f_A(x) = \frac{1}{2m^\nu K_\nu(2\alpha)} x^{\nu-1} \exp\left[-\alpha\left(\frac{x}{m} + \frac{m}{x}\right)\right], \quad x > 0 \quad (1-2)$$

où  $m (> 0)$  est un paramètre d'échelle,  $\alpha (> 0)$  et  $\nu \in \mathfrak{R}$  sont des paramètres de forme. La fonction  $K_\nu(2\alpha)$  est la fonction de Bessel modifiée du second type (Watson, 1966) et est définie par :

$$K_\nu(2\alpha) = \frac{1}{2m^\nu} \int_0^\infty x^{\nu-1} \exp\left[-\alpha\left(\frac{x}{m} + \frac{m}{x}\right)\right] dx \quad (1-3)$$

Le cas particulier  $\nu = 0$  correspond à la loi harmonique, première distribution considérée par Halphen. De plus, les lois gamma ( $G$ ) et gamma inverse ( $GI$ ) sont des cas limites de la loi de Halphen Type A pour des valeurs spécifiques des paramètres: si  $\alpha \rightarrow 0$ ,  $X \approx G$  pour  $\nu > 0$  et  $X \approx GI$  pour  $\nu < 0$ . Entre ces deux distributions limites, la loi de Type A admet une grande variété de formes toutes intéressantes pour représenter des données hydrologiques.

Chacune des lois de Halphen appartient à la classe des lois exponentielles qui ont d'intéressantes propriétés, particulièrement en ce qui concerne l'estimation des paramètres. En effet, ce type de distributions possède des statistiques exhaustives et complètes en nombre égal au nombre de paramètres. Soulignons que les lois de Halphen sont les seules lois à 3 paramètres utilisées en hydrologie appartenant à la classe des lois exponentielles. Pour la loi de Halphen type A, les statistiques exhaustives sont les moyennes arithmétique ( $A$ ), géométrique ( $G$ ) et harmonique ( $H$ ). Leur définition pour un échantillon  $x_1, x_2, \dots, x_n$  de taille  $n$ , est donnée au tableau 1.1.

**Tableau 1.1.** Statistiques exhaustives de la loi de Halphen type A

Statistiques exhaustives et complètes		
$H = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}\right)^{-1}$	$G = \exp\left(\frac{1}{n} \sum_{i=1}^n \log x_i\right)$	$A = \frac{1}{n} \sum_{i=1}^n x_i$

L'importance de ces statistiques exhaustives et complètes réside dans le fait que parmi les estimateurs sans biais, celui dont la variance est minimale est une fonction des statistiques exhaustives. De plus, d'après le théorème de Lehmann-Scheffé, cet estimateur est unique (Bickel et Doksum, 1977). Cependant, avec les lois de Halphen, il n'est pas possible d'exprimer explicitement cette fonction pour aucun des paramètres.

Une procédure a été développée à l'INRS-Eau pour estimer les paramètres et les quantiles des lois de Halphen par la méthode du maximum de vraisemblance (Perreault et Bobée, 1995). Les estimateurs obtenus sont fonctions des statistiques exhaustives, ils sont de plus convergents, ce qui implique que leur biais tend vers zéro lorsque la taille de l'échantillon tend vers l'infini. Ce sont les

raisons pour lesquelles cette méthode est la meilleure dont nous disposons. Cependant, pour des échantillons de petite taille, le biais peut être important, et la propriété de variance minimale non valable.

Il est donc intéressant d'étudier les propriétés (biais, variance) des estimateurs des paramètres et des quantiles. L'erreur quadratique moyenne, qui intègre à la fois le biais et la variance, est utile comme mesure globale de la qualité d'un estimateur. La comparaison de la variance expérimentale et de la variance asymptotique théorique qui constitue une borne inférieure à la variance d'un estimateur sans biais permet de juger si la méthode donne de bonnes estimations.

Par ailleurs, dans certaines circonstances, la procédure d'estimation basée sur la méthode du maximum de vraisemblance ne fournit pas de solution. Ce problème est assez sérieux pour mériter qu'on l'étudie en profondeur.

## 1.2 Procédure d'estimation

La compréhension des étapes de la méthode d'estimation est indispensable à l'interprétation de certains des résultats présentés dans ce rapport. Un exposé général en est fait ici et des aspects spécifiques en seront développés au besoin tout au long du rapport.

Il suffit en théorie de solutionner simultanément les 3 équations du maximum de vraisemblance pour estimer les valeurs inconnues des trois paramètres (Bickel et Doksum, 1977). En raison de problèmes numériques dans la résolution itérative de ce système d'équations non-linéaires, l'estimation est réalisée en pratique en deux étapes. D'abord, les paramètres  $\alpha$  et  $m$  sont estimés numériquement,  $\nu$  étant fixé. Ensuite, la fonction de vraisemblance logarithmique partielle  $\log L(\nu; \hat{\alpha}, \hat{m})$  est maximisée afin de déterminer l'estimation de  $\nu$ .

En fixant  $\nu$ , le système à résoudre est réduit à deux équations à deux inconnues et données par (1-4) et (1-5) :

$$D_A(\alpha, \nu) = \frac{K_{\nu+1}(2\alpha) K_{\nu-1}(2\alpha)}{K_{\nu}^2(2\alpha)} = \frac{A}{H} \quad (1-4)$$

$$m = A \frac{K_{\nu}(2\alpha)}{K_{\nu+1}(2\alpha)} \quad (1-5)$$

C'est sur la première équation que des restrictions s'appliquent à l'existence d'une solution. Ces limites sont expliqués en détail par Perreault et Bobée (1995). Il suffit de mentionner ici que

l'équation (1-4) n'admet de solution que si  $|\nu| \leq 1$  ou si  $AH^{-1} < |\nu|/(|\nu|-1)$ . En posant  $U = AH^{-1}/(AH^{-1} - 1)$  on a  $U > 1$  car  $H \leq A$  et on peut résumer les dernières conditions en disant qu'une solution à l'équation (1-4) existe si et seulement si  $|\nu| < U$ .

Dans l'équation (1-4), le paramètre  $\nu$  étant fixé et  $A/H$  étant une constante, la seule inconnue est le paramètre  $\alpha$ . Le paramètre  $m$ , qui n'intervient pas dans cette équation, est estimé à l'aide de l'équation 1-5 seulement à la fin de la procédure. En effet, l'inconnue  $\hat{m}$  peut être remplacée par son expression en  $\hat{\alpha}$  et  $\nu$  dans la fonction de vraisemblance logarithmique partiellement maximisée  $\log L(\nu, \hat{\alpha}, \hat{m})$  (équation 1-6); cette fonction ne dépend plus alors que de  $\hat{\alpha}$  et de  $\nu$ . Le maximum de cette fonction, à l'intérieur de l'intervalle  $[-U, U]$ , devient la valeur estimée de  $\nu$ . Finalement les valeurs estimées de  $\alpha$  et de  $\nu$  servent à estimer  $m$ .

$$\log L(\nu, \hat{\alpha}, \hat{m}) \propto \begin{cases} \nu \ln[\nu/A] - \ln \Gamma(\nu) + (\nu - 1) \ln G - \nu, & \text{si } \nu \geq U \text{ (Loi G)} \\ \ln \left[ \frac{G^{\nu-1}}{\hat{m}^\nu K_\nu(2\hat{\alpha})} \right] - \hat{\alpha} \left[ \frac{K_{\nu+1}(2\hat{\alpha})}{K_\nu(2\hat{\alpha})} + \frac{K_{\nu-1}(2\hat{\alpha})}{K_\nu(2\hat{\alpha})} \right], & \text{si } |\nu| < U \text{ (Loi A)} \\ -\nu \ln[-\nu H] - \ln \Gamma(-\nu) + (\nu - 1) \ln G + \nu, & \text{si } \nu \leq -U \text{ (Loi GI)} \end{cases} \quad (1-6)$$

### 1.3 Objectifs

La présente étude vise à évaluer la qualité de l'estimation des paramètres et des quantiles de la loi de Halphen type A. Plus précisément, les objectifs sont :

- d'étudier le biais et l'erreur quadratique moyenne des estimateurs du maximum de vraisemblance des paramètres de la loi de Halphen de type A en effectuant une simulation de Monte-Carlo;
- d'étudier le biais et l'erreur quadratique moyenne des quantiles des lois de Halphen type A dont les paramètres ont été estimés par la méthode du maximum de vraisemblance à partir des échantillons simulés;
- d'expliquer les problèmes rencontrés dans l'estimation des paramètres et des quantiles d'une loi de Halphen type A par la méthode du maximum de vraisemblance.

Au chapitre 2, les considérations techniques touchant la simulation et le choix des paramètres sont exposés. Les résultats de la simulation et les interprétations qui ont pu en être tirées sont présentés au chapitre 3. Le chapitre 4 est consacré aux problèmes associés à la méthode d'estimation, à leur ampleur, à leurs effets ainsi qu'à leur origine. Enfin, les conclusions et recommandations se retrouvent au 5<sup>e</sup> et dernier chapitre.

## 2 CARACTÉRISTIQUES DE LA SIMULATION

### 2.1 Choix de la méthode de génération

Afin de mener à bien la simulation, il était nécessaire de se doter d'un programme qui génère des observations provenant d'une loi de Halphen type A de paramètres ( $\alpha$ ,  $m$ ,  $\nu$ ) fixés. Ce problème a été étudié par Essadaoui *et al.* (1995) qui ont conçu plusieurs générateurs et les ont comparés. Les méthodes envisagées reposent soit sur l'inversion de la fonction de répartition, soit sur le procédé d'acceptation-rejet.

Le principal critère à considérer pour choisir un générateur est sa précision. Essadaoui *et al.* (1995) ont évalué ce critère en comparant la stabilité en fonction de la taille d'échantillon des statistiques exhaustives des échantillons (moyennes arithmétique, harmonique et géométrique) et leur convergence vers leurs valeurs théoriques.

Ce seul point de comparaison est insuffisant pour juger de la précision d'un générateur. Pour effectuer le choix de la meilleure méthode, des moyens supplémentaires ont été utilisés :

- Représentation des observations générées sur papier de probabilité;
- Application de tests d'adéquation aux données générées.

La représentation graphique d'échantillons ordonnés de taille 1000 permet d'examiner la distribution de façon globale et de détecter certains défauts des méthodes : sauts dans la distribution, queues déviantes, variations non aléatoires, etc.

Le logiciel *AJUSTE-II* (Perreault *et al.*, 1994) contient un module exécutant le test d'adéquation du khi-carré avec une loi de probabilité dont les paramètres sont estimés à partir de l'échantillon. Deux sources d'erreur se présentent alors : la méthode d'estimation et la génération de l'échantillon. Comme l'échantillon est produit par simulation d'une loi dont les paramètres sont fixés, il est possible de spécifier les paramètres de la loi dont on teste l'adéquation, et ainsi ne tenir compte que de l'erreur due à la génération. Aussi, le nombre de degrés de liberté est maximal, soit le nombre de classes moins un. Un programme permettant de réaliser un test d'adéquation du khi-carré avec des paramètres spécifiés a donc été conçu.

Cependant, même avec des paramètres fixés, la loi du khi-carré demeure une approximation de la loi de la statistique du test. De plus, la subdivision en classes du domaine de la loi est toujours arbitraire. Lorsque l'on connaît les paramètres de la loi, le test de Kolmogorov-

Smirnov est un test d'adéquation exact (D'Agostino and Stephens, 1986). Il offre l'avantage que la loi de sa statistique ne dépend pas de la loi ajustée et qu'on peut donc se référer à une table unique. Un module permettant le calcul de la statistique du test pour les lois de Halphen a été ajouté au programme d'ajustement (voir annexe B).

Les quatre méthodes identifiées par Essadaoui *et al.* (1995) comme les plus prometteuses ont été examinées et comparées. La méthode par inversion de la fonction de répartition appelée INV2 dans Essadaoui *et al.* (1995) s'adapte assez bien pour toutes les valeurs de paramètres mais la représentation graphique des observations simulées ordonnées montre un saut systématique dans la distribution à la probabilité cumulée 0,95 (voir figure 2.1a). Un défaut de ce genre saute aux yeux sur la distribution empirique des échantillons générés par la méthode d'acceptation-rejet avec fonctions majorante et minorante linéaires par morceaux LPM4. À un certain endroit, les observations s'élèvent au-dessus du prolongement de la courbe avant de revenir dans le tracé naturel, formant une bosse délimitée par deux cassures très marquées (voir figure 2.1b). Ces deux méthodes doivent donc être écartées d'emblée.

La méthode LPM9 s'apparente à la précédente par sa fonction majorante linéaire par morceaux mais s'en distingue par un découpage adapté à la forme de la loi. L'adéquation de ses observations simulées est satisfaisante pour une grande diversité de valeurs de paramètres, et elle ne présente aucun défaut majeur. Elle conviendrait très bien pour une simulation si la méthode suivante n'était encore meilleure.

La méthode de simulation retenue, appelée UPM1, est semblable à LPM9 mais sa fonction majorante est uniforme par morceaux. Avec une fonction majorante aussi simple, elle surpasse la précision de toutes les autres méthodes basées sur le même principe de découpage du domaine de définition.

L'ajustement graphique d'échantillons de 1000 observations ordonnées avec la loi estimée est particulièrement bon (voir figure 2.2). Même pour un  $\alpha$  très petit et  $\nu = 0$ , la courbe ajustée et les observations se suivent assez bien jusqu'aux quantiles extrêmes. Le test de Kolmogorov-Smirnov accepte avec de très forts seuils<sup>1</sup> l'hypothèse nulle «les observations proviennent de la loi de Halphen type A spécifiée». Quant au test du khi-carré, il accepte le plus souvent cette hypothèse, ne la rejetant que de peu dans de rares cas (voir tableau 2.1).

---

<sup>1</sup> Le terme «seuil» désigne ici la probabilité au dépassement de la valeur calculée de la statistique du test.

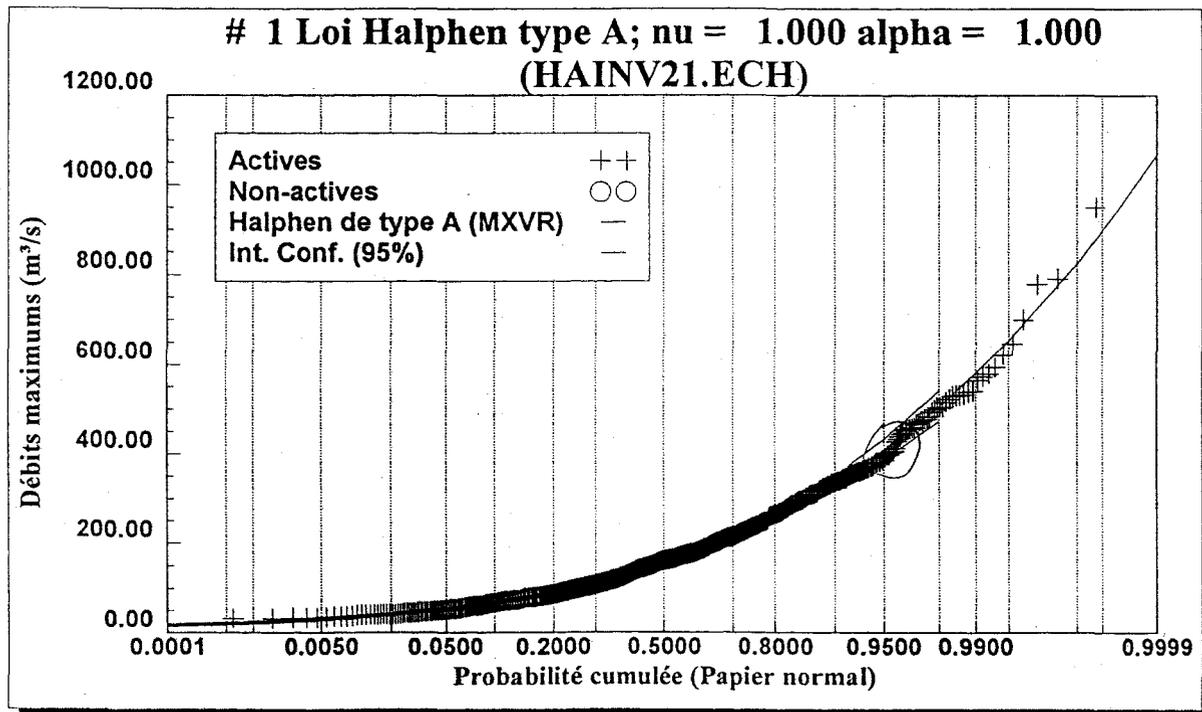
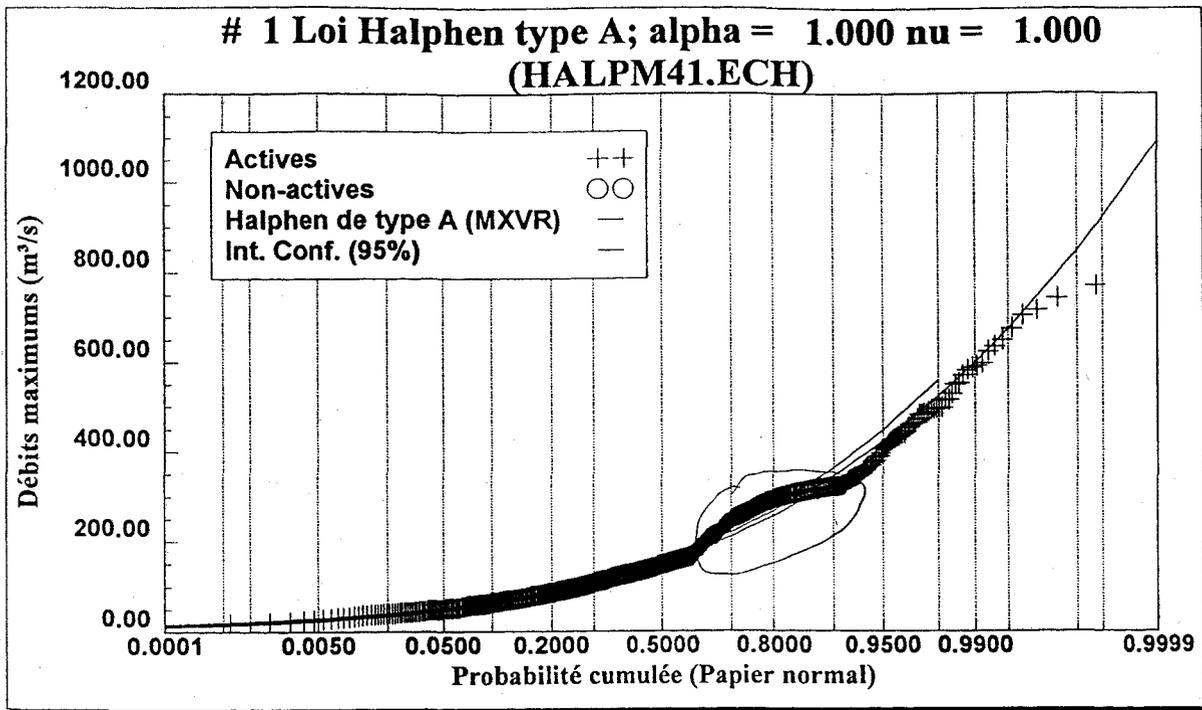


Figure 2.1. Distributions empiriques d'échantillons de taille 1000 générés par les méthodes LPM4 et INV2

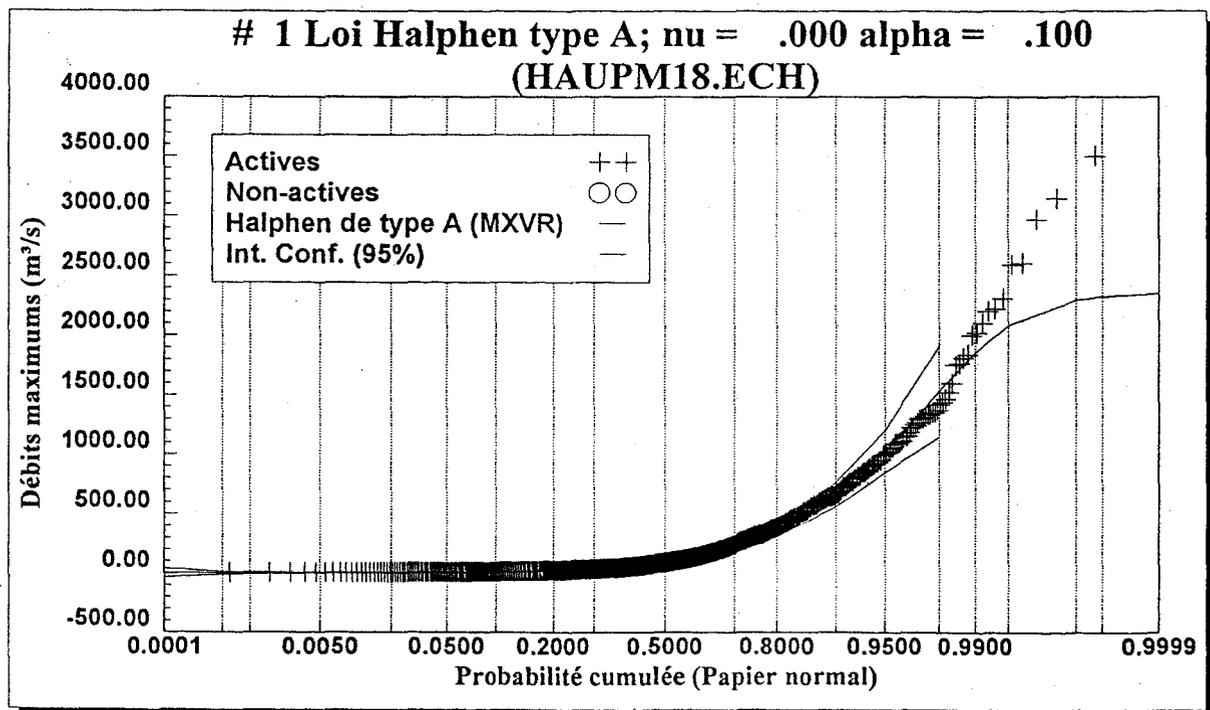
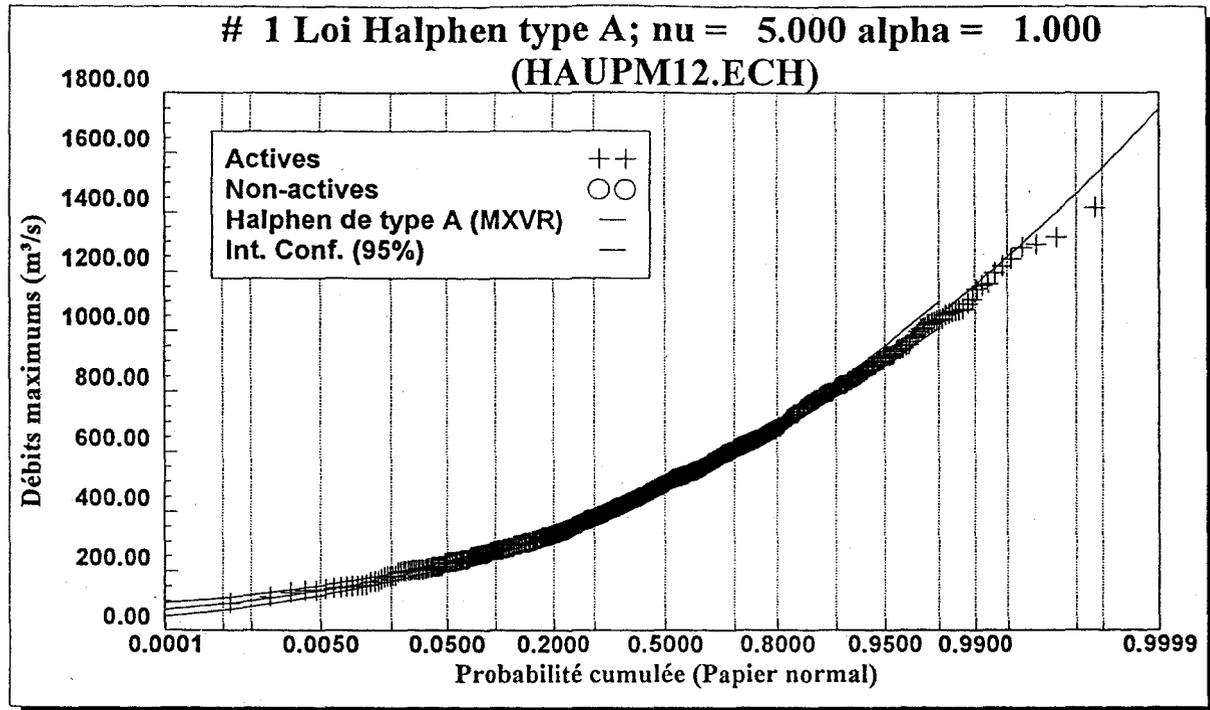


Figure 2.2. Distributions empiriques d'échantillons de taille 1000 générés par la méthode UPM1

Tableau 2.1. Résultats des tests d'adéquation pour la méthode UPM1

Valeurs des paramètres		Test du khi-carré		Test de Kolmogorov-Smirnov	
$\alpha$	$\nu$	Valeur de la statistique	Seuil observé	Valeur de la statistique	Seuil observé
0,1	0	52,202	0,0072	1,002	0,17
0,1	1	40,670	0,0925	0,748	0,63
1	0	36,082	0,2054	0,705	0,70
1	1	30,874	0,4216	0,697	0,71
5	0	47,180	0,0239	0,700	0,70
5	10	38,872	0,1286	0,651	0,78
10	1	30,440	0,4433	0,626	0,82

Avec toutes les méthodes de génération, le plus difficile semble d'obtenir des moyennes expérimentales qui coïncident parfaitement avec les moyennes théoriques. Pour la méthode UPM1, ces statistiques sont près de leurs valeurs théoriques, mais presque toujours un peu inférieures. Il est possible que cela influence légèrement l'estimation. Nous ne pouvons contourner cette situation, les autres méthodes étant pires à ce chapitre. Pour le cas limite  $\alpha = 0,1$  et  $\nu = 0$ , on peut noter une amélioration de la précision par rapport aux autres méthodes, ce qui se reflète dans les tests d'adéquation. Ces considérations montrent la très grande robustesse de la méthode UPM1 et l'excellente adéquation des observations qu'elle génère avec la loi dont elles doivent provenir.

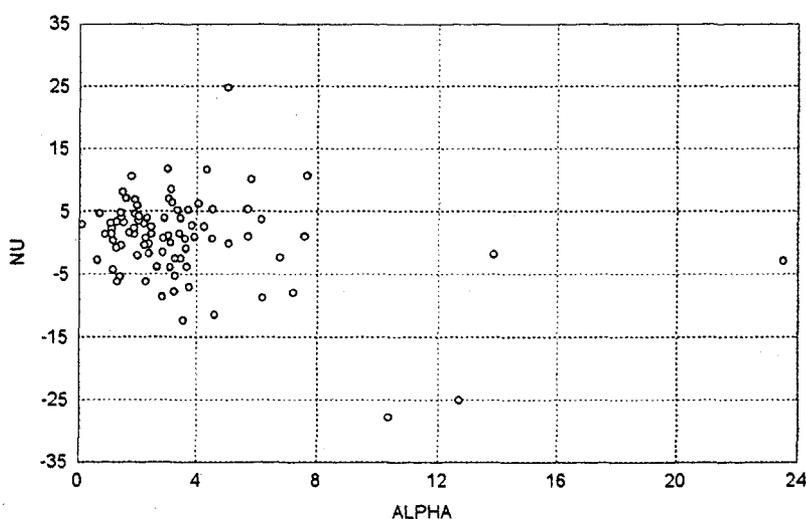
## 2.2 Choix des triplets de paramètres

Rappelons que la loi de Halphen type A possède un paramètre d'échelle ( $m$ ) et deux paramètres de forme ( $\alpha$  et  $\nu$ ). Le choix des triplets de paramètres à considérer dans la simulation est conditionné, en priorité, par des considérations hydrologiques. En effet, les paramètres doivent être sélectionnés de façon à ce que la forme des lois retenues s'apparente à celles observées en pratique. L'accent est donc mis sur le choix des couples  $(\alpha, \nu)$  après avoir réglé la question du paramètre d'échelle  $m$ .

La présence d'un paramètre d'échelle signifie que la loi aura la propriété suivante : Si  $X$  suit une loi  $HA(\alpha, m, \nu)$  alors  $kX$ , pour  $k$  positif, suit une loi  $HA(\alpha, km, \nu)$ . Cette propriété est utilisée en pratique dans les méthodes de génération, puisqu'il suffit de multiplier des observations d'une loi de Halphen de paramètre  $m = 1$  par une constante  $k$  pour obtenir des

observations d'une loi de Halphen de paramètre  $m = k$ . Remarquons aussi que, si on note par  $\underline{X}$  un vecteur de variables aléatoires indépendantes et identiquement distribuées, alors  $A(k\underline{X}) = kA(\underline{X})$ ,  $G(k\underline{X}) = kG(\underline{X})$  et  $H(k\underline{X}) = kH(\underline{X})$ . Un réarrangement des équations du maximum de vraisemblance (Perreault et Bobée, 1995) permet alors de constater que la multiplication des observations par une constante n'a aucune influence sur l'estimation des paramètres  $\alpha$  et  $\nu$  et que la valeur estimée du paramètre  $m$  est simplement multipliée par la constante. Ainsi on peut attribuer une valeur fixe à  $m$  pour toute la simulation sans perte de généralité. Nous avons choisi  $m = 100$  pour avoir assez de chiffres significatifs puisque le format de *AJUSTE-II* ne compte que deux chiffres après la virgule. Mais ensuite, tous les résultats sont calculés en ramenant la valeur de  $m$  à 1.

Pour sélectionner les couples  $(\alpha, \nu)$ , une première approche consiste à s'inspirer des estimations des paramètres de la loi de Halphen type A obtenues avec des séries d'observations hydrologiques. Un ajustement a été réalisé pour les débits maximums annuels de 103 stations hydrométriques du Québec et de l'Ontario (Perreault et Bobée, 1995). Ces estimations peuvent servir à donner un ordre de grandeur des valeurs de paramètres rencontrées dans la réalité. On constate qu'elles se retrouvent presque toutes entre -12 et 12 pour  $\nu$  et entre 0 et 8 pour  $\alpha$  (voir figure 2.3). Il ne faut toutefois pas considérer uniquement cette approche pour sélectionner des valeurs précises, puisque les estimateurs peuvent être biaisés.



**Figure 2.3.** Estimations des paramètres  $\alpha$  et  $\nu$  de la loi Halphen type A ajustée aux observations de 103 stations hydrométriques du Québec et de l'Ontario

Afin d'arrêter notre choix, une approche basée sur les coefficients de variation et d'asymétrie a été considérée. Il s'agit de choisir des valeurs de paramètres de telle sorte que les coefficients de variation ( $C_v$ ) et d'asymétrie ( $C_s$ ) théoriques des lois soient près des coefficients expérimentaux des échantillons de débits maximums annuels. De plus, d'autres couples de paramètres ont été choisis pour que les coefficients des lois atteignent les valeurs extrêmes que l'on est susceptible de rencontrer en hydrologie. La limite supérieure a été fixée à 2 pour le coefficient d'asymétrie et à 1 pour le coefficient de variation. La loi de Halphen type A étant une loi à asymétrie positive, la valeur limite inférieure est de 0 pour les deux coefficients.

Tout d'abord, l'étude a porté sur les valeurs théoriques de  $C_v$  et  $C_s$ . Cette étude donne une idée du comportement de ces coefficients théoriques en fonction de  $\alpha$  et de  $\nu$  et permet de déterminer le domaine des valeurs des paramètres qui donnent des coefficients dans la région admissible. Les courbes de niveau de  $C_v$  et de  $C_s$  ont été tracées dans le plan  $\alpha \times \nu$  (voir figure 2.4). Pour  $C_v$ , les courbes ressemblent à des demi-ellipses centrées à l'origine (voir figure 2.4a). Le  $C_v$  est décroissant avec  $\alpha$  et avec  $\nu$  positif et croissant avec  $\nu$  négatif. La courbe où  $C_v = 1$  passe près des points ( $\alpha = 0^+; \nu = 2$ ), ( $\alpha = 0,1; \nu = 0$ ) et ( $\alpha = 0^+; \nu = -2$ ).

Le comportement du  $C_s$  diffère grandement de celui du  $C_v$  (voir figure 2.4b). Même si le  $C_s$  est lui aussi décroissant avec  $\alpha$ , ses courbes de niveau ne sont pas symétriques par rapport à l'axe des  $\alpha$ . Le  $C_s$  atteint son maximum par rapport à  $\nu$  en un point qui est inférieur ou se rapproche de 0, sans être en dessous de -5. La décroissance du  $C_s$  est plus rapide à droite de ce point qu'à gauche. La courbe où  $C_s = 2$  passe près des points ( $\alpha = 0,1; \nu = 1$ ), ( $\alpha = 1,5; \nu = -2$ ) et ( $\alpha = 0; \nu = -8$ ).

Le diagramme mettant en relation les coefficients de variation et d'asymétrie est présenté à la figure 2.4. Les  $C_v$  et  $C_s$  des stations hydrométriques du Québec et de l'Ontario comptant plus de 40 années d'observations y sont situés et la zone de la loi de Halphen type A y est délimitée par les courbes de  $C_s$  et  $C_v$  des lois Gamma et Gamma inverse. On constate tout d'abord que la loi de Halphen type A ne couvre que les stations dont l'asymétrie des observations est la plus forte. De fait, lorsque le  $C_s$  diminue vers 0, les  $C_v$  admissibles pour la loi de Halphen type A sont de plus en plus petits et leur domaine se rétrécit. Ensuite, on remarque que pour obtenir des points près des limites des lois Gamma et Gamma inverse, il faut prendre de petites valeurs du paramètre  $\alpha$ , comme le prévoit la théorie ( $HA \rightarrow G$  ou  $IG$  si  $\alpha \rightarrow 0$ ).

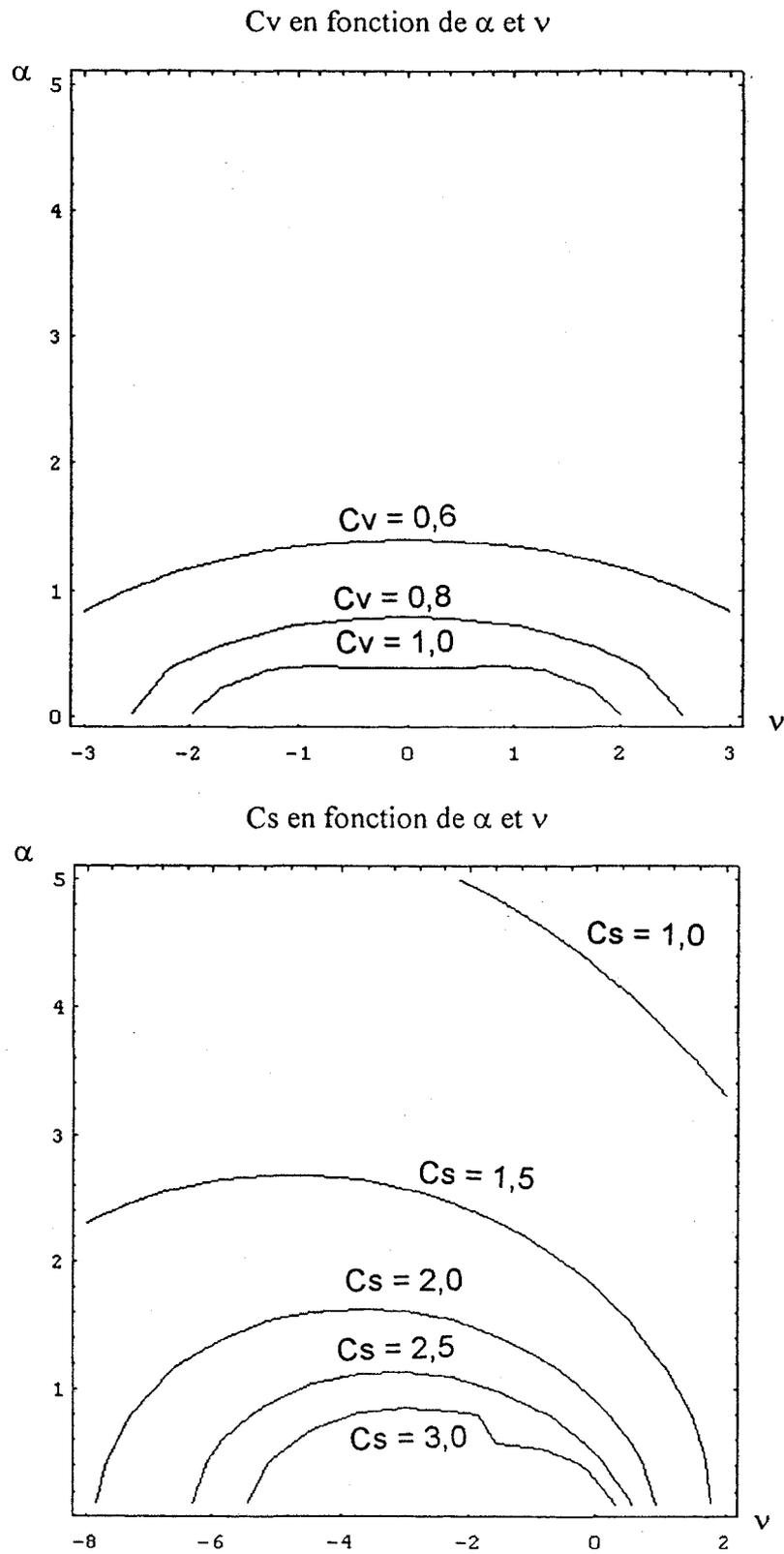


Figure 2.4 Coefficients de variation et d'asymétrie en fonction de  $\alpha$  et de  $v$

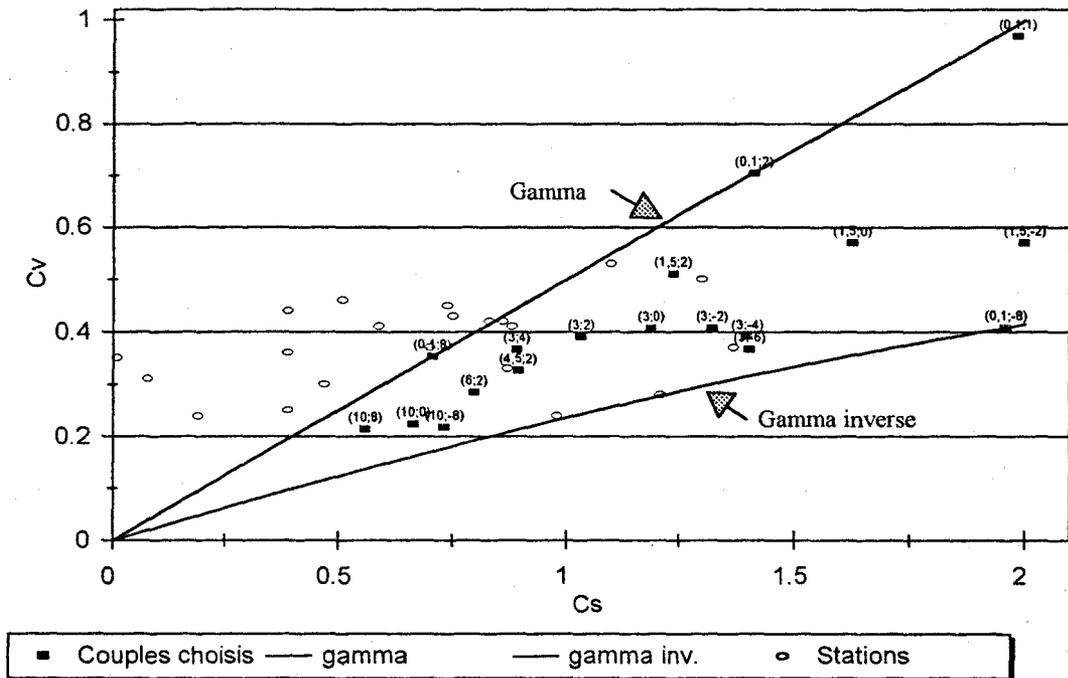


Figure 2.5. Diagramme des coefficients de variation et d'asymétrie

En plus de notre préoccupation de se rapprocher le plus possible de la réalité hydrologique, nous pouvons introduire d'autres critères d'intérêt pour l'analyse. En prenant plusieurs valeurs de  $\alpha$  équidistants en gardant  $\nu$  constant, il est possible d'étudier nos résultats en fonction de  $\alpha$ . On peut faire de même avec  $\nu$  pour un  $\alpha$  fixé, en s'arrangeant de plus pour prendre des valeurs de  $\nu$  opposées : au couple  $(\alpha, \nu)$  répond le couple  $(\alpha, -\nu)$ . Pour leur position centrale, il importe aussi de choisir quelques lois harmoniques ( $\nu = 0$ ).

Tous ces éléments permettent maintenant de sélectionner les couples  $\alpha$  et  $\nu$  des lois de Halphen type A qui seront utilisés pour la simulation. Autour d'un couple central  $\alpha = 3$ ,  $\nu = 2$  de  $C_v = 0,4$  et  $C_s = 1$  (voir figure 2.5), on construit une série de  $\alpha$   $\{0,1; 1,5; 3; 4,5; 6\}$  en gardant  $\nu = 2$ , et une série de  $\nu$   $\{-6; -4; -2; 0; 2; 4\}$  en gardant  $\alpha = 3$ . Tous ces couples donnent des lois dont les  $C_v$  et  $C_s$  sont assez centrés sur la région de valeurs admissibles.

Les couples de paramètres de lois aux coefficients extrêmes sont ensuite choisis. Pour un  $C_s$  près de 2, le point  $(0,1;1)$  correspond à un  $C_v$  près de 1 ( $C_v$  de la loi Gamma), le point  $(0,1; -8)$  correspond à un  $C_v$  près de  $\sqrt{2} - 1$  ( $C_v$  de la loi Gamma inverse), et le point  $(1,5;-2)$  est entre les deux. Le couple  $(0,1;8)$  a été retenu car il est l'opposé de  $(0,1;-8)$  et que cette loi est près de la loi Gamma.

Pour obtenir des lois correspondant à de faibles  $C_v$  et  $C_s$ , il faut augmenter les valeurs de  $\alpha$  et de  $\nu$ . Cependant, pour des raisons expliquées au chapitre 4, la proportion d'échantillons simulés pour lesquels les conditions nécessaires à l'obtention d'une solution ne sont pas remplies (voir section 1.2) s'accroît démesurément. Les plus grandes valeurs qu'il a été possible de prendre sont  $\alpha = 10$  et  $\nu = 8$ . Le couple opposé (10, -8) a également été retenu.

Dans l'intention de générer plus de lois harmoniques, les couples (10;0) et (1,5;0) ont été ajoutés à la liste. Enfin, même si la valeur de son  $C_s$  est irréaliste (près de 4), la loi de paramètres (0,1;0) a été considérée car elle a la propriété que des solutions des équations du maximum de vraisemblance existent pour tous ses échantillons.

Le tableau 2.2 contient tous les couples retenus avec les coefficients de variation et d'asymétrie de lois associées.

**Tableau 2.2.** Valeurs de  $\alpha$  et de  $\nu$  retenues pour la simulation

$\alpha$	$\nu$	$C_s$	$C_v$
0,1	-8	1,9579	0,4081
0,1	0	3,9699	1,6747
0,1	1	1,9836	0,9688
0,1	2	1,4140	0,7038
0,1	8	0,7071	0,3535
1,5	-2	2,0005	0,5702
1,5	0	1,6262	0,5702
1,5	2	1,2371	0,5103
3	-6	1,4034	0,3663
3	-4	1,3985	0,3911
3	-2	1,3223	0,4067
3	0	1,1875	0,4067
3	2	1,0327	0,3911
3	4	0,8920	0,3663
4,5	2	0,8950	0,3263
6	2	0,7985	0,2850
10	-8	0,7326	0,2178
10	0	0,6649	0,2235
10	8	0,5609	0,2143

## 2.3 Considérations sur les échantillons

Diverses questions se posaient en ce qui a trait aux échantillons à utiliser. Elles concernent leur taille, leur nombre et leur validation.

La principale préoccupation dans le choix des tailles d'échantillons était encore une fois la concordance avec la réalité. Il est très rare que les séries de données de crues annuelles dépassent un siècle. Elles sont le plus souvent bien plus courtes. Nous avons donc fixé la plus petite taille à 20 et la plus grande à 100. Afin de pouvoir étudier le comportement des estimateurs en fonction de la taille de l'échantillon, 5 tailles multiples de 20 ont été retenues : 20, 40, 60, 80 et 100.

Les critères principaux pour fixer le nombre de réplicats étaient la similarité avec des études par simulation que l'on retrouve dans la littérature (par exemple Cohen and Whitten, 1980; Kappenman, 1984) et le temps de calcul impliqué. Un nombre de 500 réplicats est apparu approprié pour rendre notre étude crédible. Par ailleurs, nous avons dû composer avec le fait que la méthode utilisée ne produit pas d'estimations pour tous les échantillons (voir section 1.2). L'origine de ces absences de solution sera détaillée au chapitre 4. On peut déjà dire que leur proportion n'est pas négligeable et qu'elle varie selon les valeurs de  $\alpha$  et de  $\nu$ .

Pour ces raisons, l'approche qui consiste à tirer nos résultats des estimations réussies parmi un nombre fixe d'échantillons a été écartée. Nous avons préféré poursuivre la procédure d'estimation de façon à atteindre le nombre de solutions effectives fixé au départ (500 réplicats), en puisant dans une réserve d'échantillons supplémentaires. Les deux façons de faire se retrouvent dans la littérature. (Cohen and Whitten, 1980; Hoshi *et al.*, 1984; Kappenman, 1984)

Préalablement à l'estimation, il est souhaitable de vérifier si les échantillons générés proviennent bel et bien d'une loi de Halphen type A. Les mêmes outils sont disponibles que lors du choix de la méthode de génération. Seuls les tests d'adéquation sont cette fois-ci employés. Il s'agit de voir si le seuil de 5% du test est respecté, c'est-à-dire si sur un grand nombre d'échantillons de même loi, environ 5% seront rejetés. Cela est respecté pour toutes les lois de Halphen type A simulées, renforçant ainsi le choix de la méthode uniforme par morceaux UPM1. Les échantillons rejetés n'ont pas été exclus de la simulation, puisque leur rejet découle de l'erreur de type I du test.

### 3 RÉSULTATS

---

Ce chapitre est consacré à la présentation des résultats qui ont été obtenus par simulation. Pour chaque couple de paramètres  $(\alpha, \nu)$ , 500 échantillons de la loi de Halphen type A ont été générés et des statistiques de base ont été calculées. Les biais absolu (B) et relatif ( $B_{rel}$ ) pour une quantité  $\theta$  quelconque sont définis respectivement par les équations 3-1 et 3-2.

$$B = \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_i - \theta) \quad (3-1)$$

$$B_{rel} = B/\theta \quad (3-2)$$

où  $\hat{\theta}_i$  désigne l'estimation de  $\theta$ .

L'erreur quadratique moyenne (EQM) d'une quantité  $\theta$  est donnée par l'équation 3-3. Il est à noter que l'erreur quadratique moyenne est égale à la variance plus le biais au carré. On préfère exprimer cette mesure globale en prenant sa racine carrée pour se ramener aux mêmes unités que la quantité estimée, et en la divisant par cette quantité pour que toutes les erreurs soient à la même échelle. Nous appelons erreur relative la nouvelle mesure obtenue (équation 3-4).

$$EQM = \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_i - \theta)^2 \quad (3-3)$$

$$\text{Erreur relative} = \sqrt{EQM/\theta} \quad (3-4)$$

La section 3.1 présente ces statistiques pour une loi de valeurs de paramètres choisies ( $\alpha=3$ ,  $\nu=2$ ). Les résultats sont ensuite étudiés en fonction de  $\alpha$  et de  $\nu$  à la section 3.2, où quelques cas extrêmes sont aussi considérés. Dans la section suivante, l'erreur quadratique moyenne est décomposée en variance et biais. La variance expérimentale ainsi obtenue est comparée avec la variance de la loi asymptotique dans la section qui suit. L'association entre les estimateurs est étudiée à l'aide de la mesure de corrélation à la section 3.5. Tous les éléments sont alors réunis pour tenter de donner quelques interprétations des résultats dans la dernière section du chapitre.

### 3.1 Aperçu général des résultats

Une fois les échantillons générés et le traitement effectué, des résultats sont disponibles pour 19 lois de Halphen type A correspondant aux 19 jeux de paramètres retenus (Tableau 2.2). Parmi ces lois, plusieurs donnent des résultats semblables. Ainsi, quelques extraits de ces résultats suffisent pour se faire une idée sur l'ensemble. Nous avons donc décidé de présenter dans ce qui suit les résultats concernant la loi de paramètres  $\alpha = 3$  et  $\nu = 2$  qui sont représentatifs de ceux observés pour la plupart des lois. De plus, cette distribution admet des valeurs de quantiles et de coefficients  $C_s$  et  $C_v$  assez centrales ( $C_s$  de 1,  $C_v$  de 0,4; voir figure 2.4). Les quantiles sont donnés au tableau 3.1.

**Tableau 3.1.** Valeurs théoriques des quantiles lorsque  $\alpha = 3$ ,  $\nu = 2$

x2	136.670
x5	188.847
x20	253.290
x100	320.457
x1000	410.438

Les valeurs du biais ou de l'erreur quadratique moyenne sont reportées sur des graphiques en fonction de la taille d'échantillon ( $n$ ) en abscisse (figures 3.1 et 3.2). Les 5 points ainsi obtenus pour un paramètre ou un quantile sont reliés afin de donner une idée de l'évolution de la mesure en fonction de  $n$ .

#### 3.1.1 Biais des estimateurs (Figure 3.1)

Le biais de l'estimateur de  $m$  est positif, celui des estimateurs de  $\alpha$  et de  $\nu$  est négatif (figure 3.1a). En fonction de la taille de l'échantillon, on observe, en valeur absolue, la décroissance du biais sur  $m$  et sur  $\nu$ . Cette décroissance est plus régulière pour  $\nu$  que pour  $m$ . Le biais de l'estimateur de  $\alpha$  reste quant à lui stable.

Le biais relatif sur  $m$  et  $\nu$  (figure 3.1b) est, en valeur absolue, de l'ordre de 100% pour une taille d'échantillon de 20. Ce biais considérable est tout à fait typique de ce que nous avons rencontré au cours de la simulation des lois de Halphen type A. L'estimateur de  $\alpha$  par contre n'est biaisé que d'environ 10%.

En ce qui a trait aux quantiles, nous trouvons un biais relatif de moins de 5% (figure 3.1d), beaucoup moindre que pour les paramètres. Cette observation est valable pour toutes les lois de Halphen type A simulées et nous tenterons de l'expliquer à la section 3.6. On note par

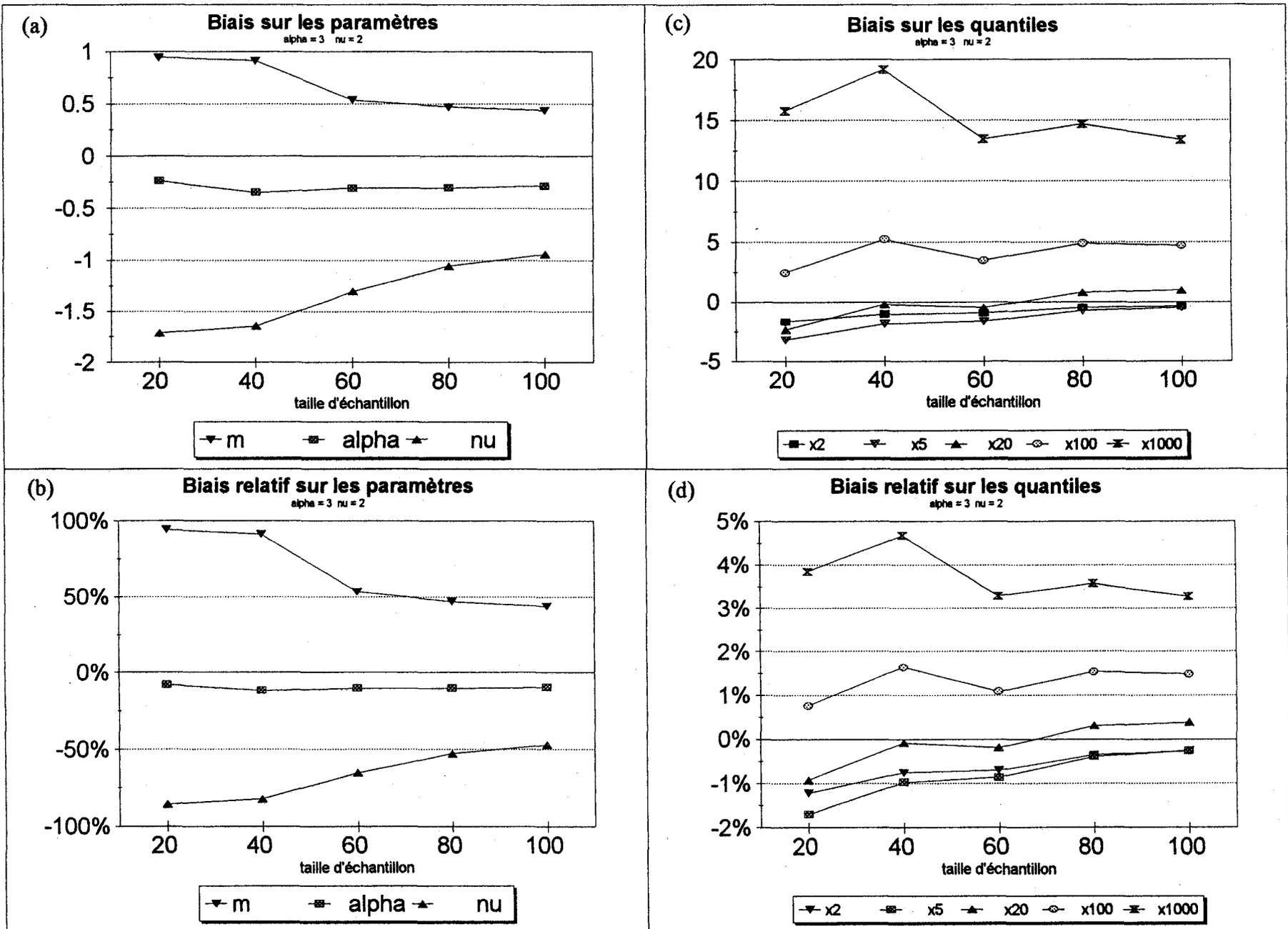


Figure 3.1. Biais des estimateurs lorsque alpha = 3 et nu = 2

ailleurs (figures 3.1c et d) que le biais sur les grands quantiles ( $x_{100}$ ,  $x_{1000}$ ) est positif alors que celui sur les petits quantiles ( $x_2$ ,  $x_5$ ) est négatif. Entre eux, le biais sur le quantile  $x_{20}$  passe du négatif au positif avec l'augmentation de  $n$ . L'estimateur du quantile  $x_{1000}$  est le plus biaisé, mais le biais relatif n'est pas nécessairement de plus en plus petit à mesure que le quantile diminue. Enfin, il n'y a pas de tendance décroissante régulière en fonction de la taille de l'échantillon.

### 3.1.2 Erreur quadratique moyenne des estimateurs (Figure 3.2)

L'EQM sur  $\alpha$  et  $\nu$  décroît régulièrement avec  $n$  (figure 3.2a). Cette décroissance est en gros proportionnelle à  $1/n$  (ou à  $1/\sqrt{n}$  pour l'erreur relative). Le comportement de l'erreur de  $m$  est beaucoup plus irrégulier. L'erreur relative des estimateurs de  $m$  et de  $\nu$  (figure 3.2b) est d'environ 3 pour une taille d'échantillon de 20 et d'environ 2 pour une taille 100. L'erreur sur  $\alpha$  passe d'environ 0,5 pour  $n = 20$  à environ 0,25 pour  $n = 100$ . Ces caractéristiques sont communes à la plupart des lois de Halphen type A considérées dans cette étude.

Comme c'est le cas avec le biais, l'erreur relative des estimateurs des quantiles (figure 3.2c) est nettement inférieure à celle des estimateurs des paramètres. Elle se situe à moins de 25% pour  $n = 20$  et à moins de 15% pour  $n = 100$ . L'erreur décroît régulièrement à mesure que la taille d'échantillon augmente et à mesure que la période de retour  $T$  diminue.

### 3.1.3 Distribution des estimateurs

Pour une loi de paramètres  $\alpha$  et  $\nu$  spécifiés, et pour une taille fixée, on peut réaliser un histogramme des estimations d'un paramètre ou d'un quantile (figure 3.3 et 3.4). Cela permet d'examiner la distribution de l'estimateur qui, autant dans le cas des paramètres que des quantiles, devrait tendre vers une loi normale (propriété des estimateurs du maximum de vraisemblance). Les tailles d'échantillons 20 et 100 ont été choisies.

La distribution des estimateurs des paramètres  $\alpha$  et  $\nu$  a les allures d'une loi normale, mais cette loi normale est décentrée par rapport à la valeur du paramètre (figure 3.3). Le biais n'est donc pas dû à quelques valeurs extrêmes mais bien à un décalage de toute la distribution. L'estimation de  $m$  a aussi une distribution décentrée, et elle souffre en plus d'une asymétrie positive à cause de la présence de quelques estimations extrêmes.

Dans tous les cas, la distribution est mieux centrée, moins dispersée et plus symétrique avec les échantillons de taille 100 qu'avec ceux de taille 20.

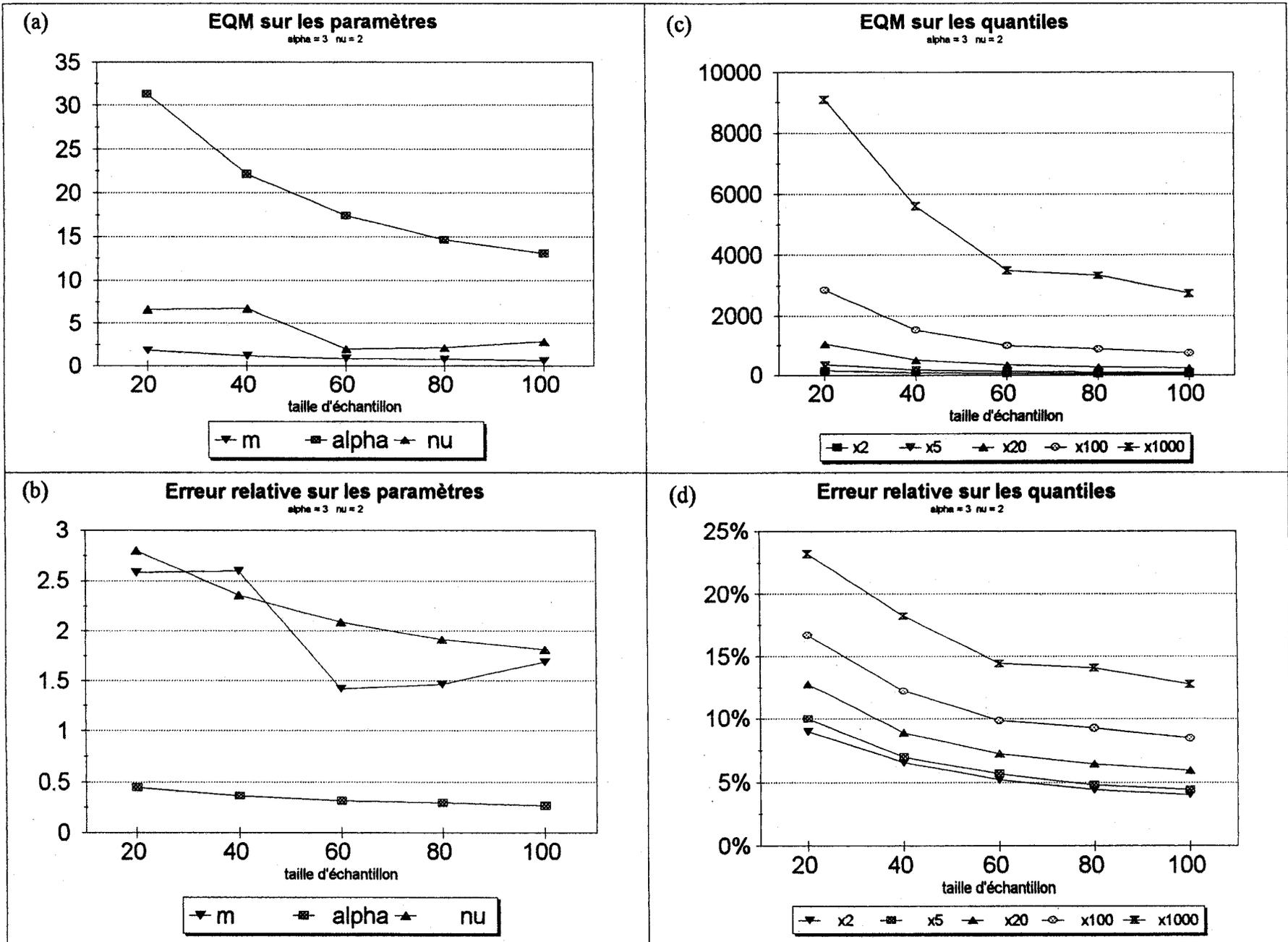


Figure 3.2. Erreur quadratique moyenne des estimateurs lorsque alpha = 3 et nu = 2

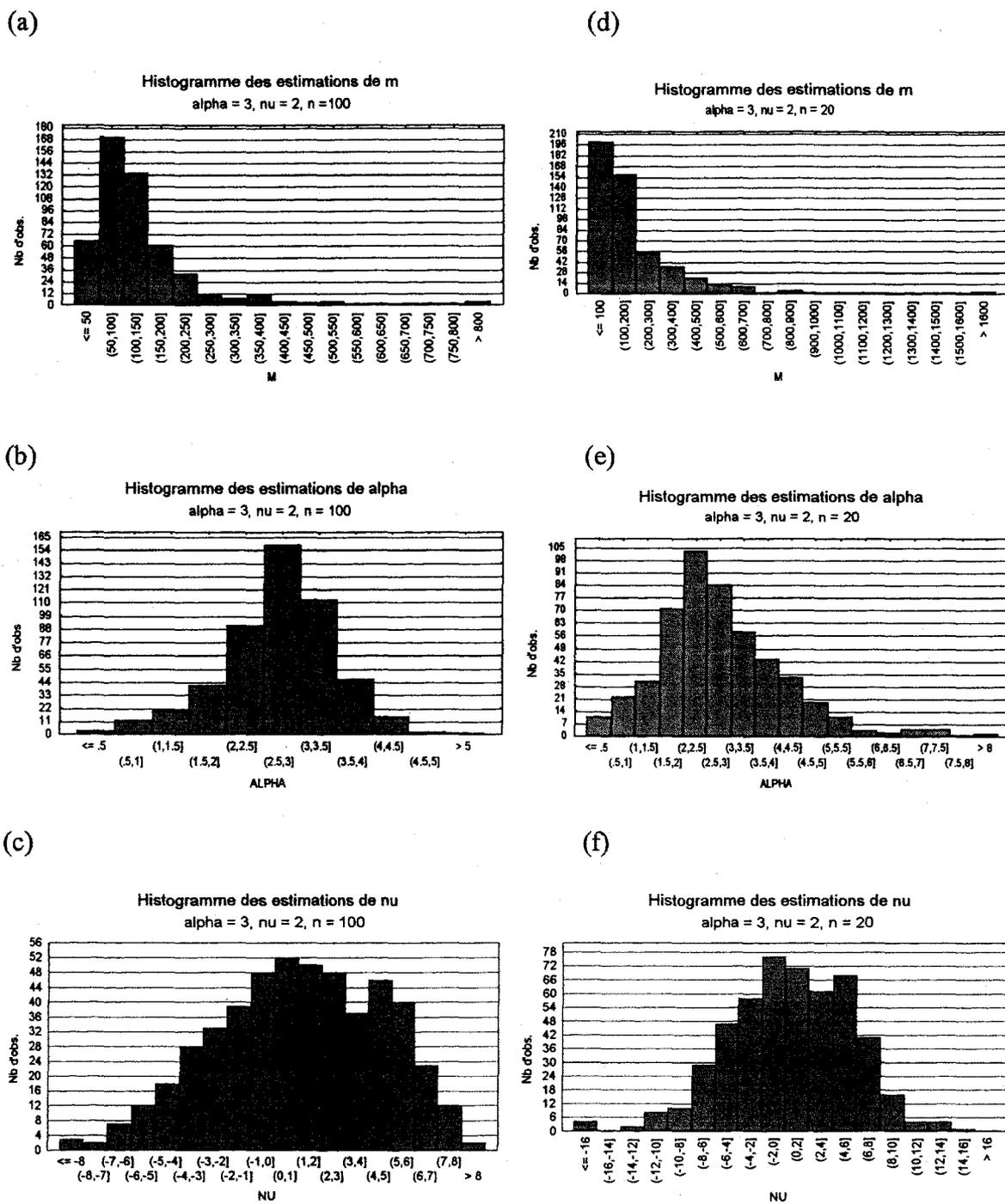


Figure 3.3 Distribution des estimateurs des paramètres ( $\alpha = 3$ ,  $\nu = 2$  et  $m = 100$ )

Le fait que les estimateurs des quantiles soient moins biaisés que ceux des paramètres se reflète dans leurs distributions (figure 3.4 et tableau 3.1). Notons que la distribution de l'estimateur de la médiane ( $x_2$ ) est bien symétrique et prend la forme de cloche caractéristique de la loi normale. Toutefois, plus le quantile est élevé, plus l'asymétrie positive de la distribution est grande. Dans ces cas, la convergence est plus lente vers la loi normale.

## 3.2 Erreur et biais relatifs en fonction de $\alpha$ et de $v$

La ressemblance des résultats pour plusieurs valeurs de paramètres n'exclut pas des variations d'une valeur à l'autre. Des couples  $(\alpha, v)$  ont été sélectionnés pour étudier les résultats en fonction de  $\alpha$  et en fonction de  $v$  (voir section 2.2). Les résultats pertinents concernent l'erreur et le biais relatifs puisque les valeurs du paramètre étudié et des quantiles varient. Il faut considérer les tailles d'échantillons une à la fois. Puisque les résultats sont très semblables d'une taille à l'autre, on ne s'attarde qu'aux tailles extrêmes 20 et 100. Pour chacune de ces tailles, les résultats pour les valeurs de  $\alpha$  de 0,1 à 6 avec  $v$  fixé à 2 sont portés sur un graphique et les points d'un même paramètre ou quantile sont reliés entre eux (figure 3.5 et 3.6). Il est fait de même avec la série des valeurs de  $v$  qui s'étend de -6 à 4 pour un  $\alpha$  fixé à 3 (figure 3.7 et 3.8).

### 3.2.1 Biais et EQM en fonction de $\alpha$ (Figure 3.5 et 3.6)

Les figures 3.5a et 3.5b montrent le biais sur les paramètres en fonction de  $\alpha$ . Le biais de l'estimateur de  $v$  est négatif et son importance en valeur absolue s'accroît avec  $\alpha$ . Le biais des estimateurs de  $\alpha$  et de  $m$  est très élevé à  $\alpha = 0,1$  et il décroît radicalement à partir de  $\alpha = 1,5$ . Pour  $\alpha > 1,5$ , le biais sur  $\alpha$  devient négatif et s'intensifie graduellement. Le biais sur  $m$  reste positif et augmente progressivement lorsque  $\alpha$  augmente. Il est possible que le biais sur ces deux paramètres atteigne un minimum en une valeur de  $\alpha$  entre 0,1 et 1,5.

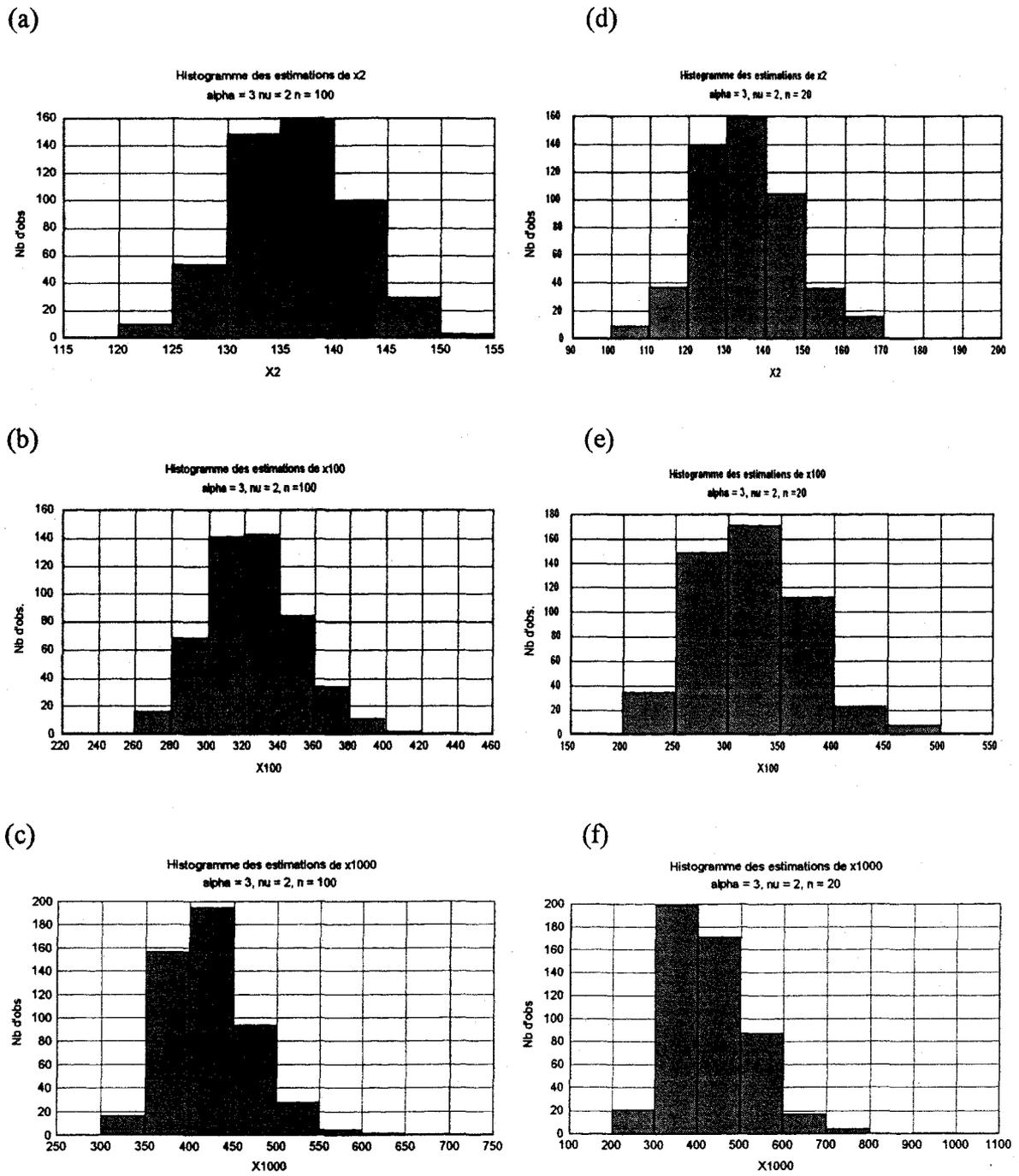


Figure 3.4 Distribution des estimateurs des quantiles ( $\alpha = 3, \nu = 2$  et  $m = 100$ )

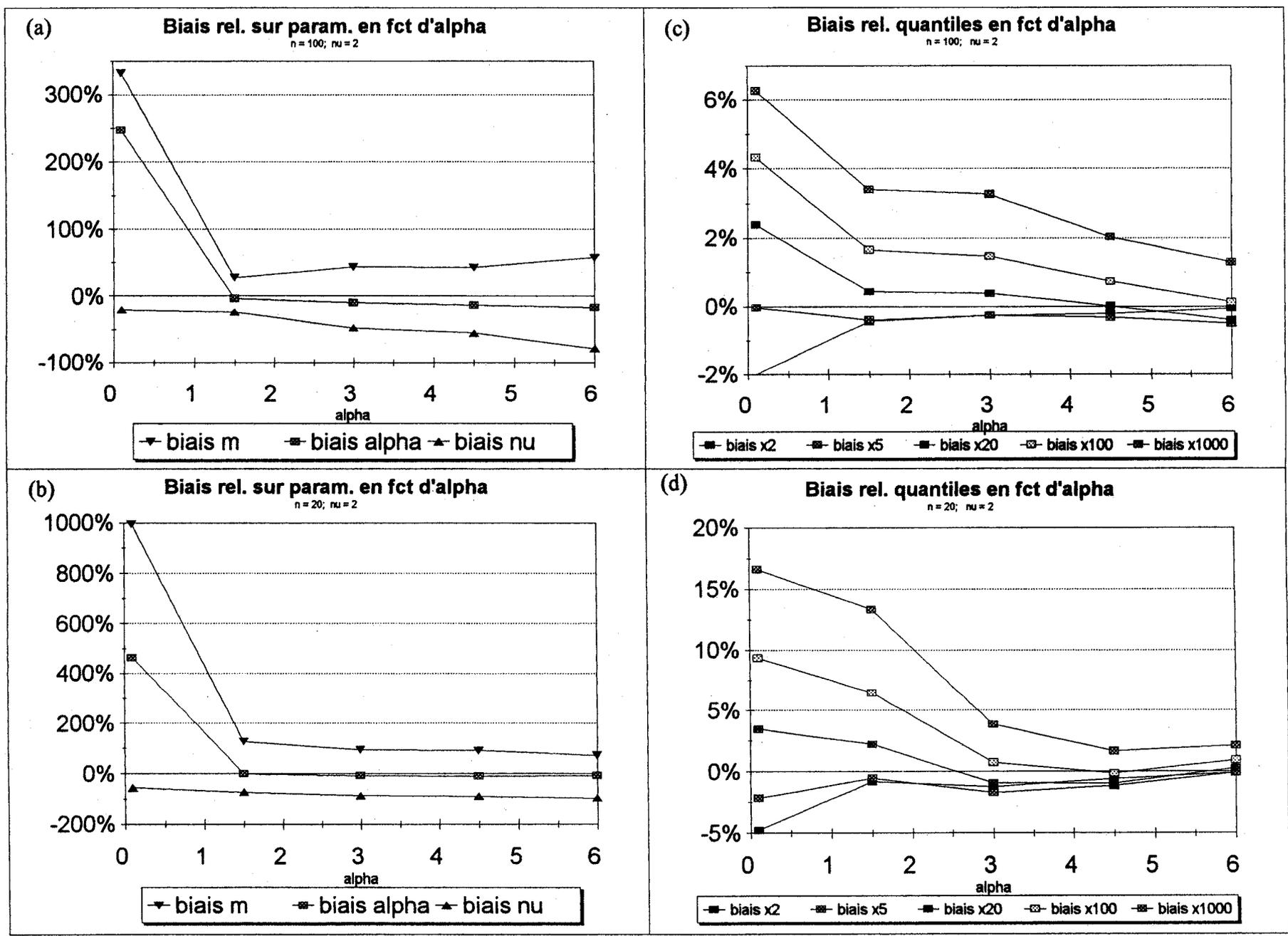


Figure 3.5 Biais relatif des estimateurs en fonction de alpha

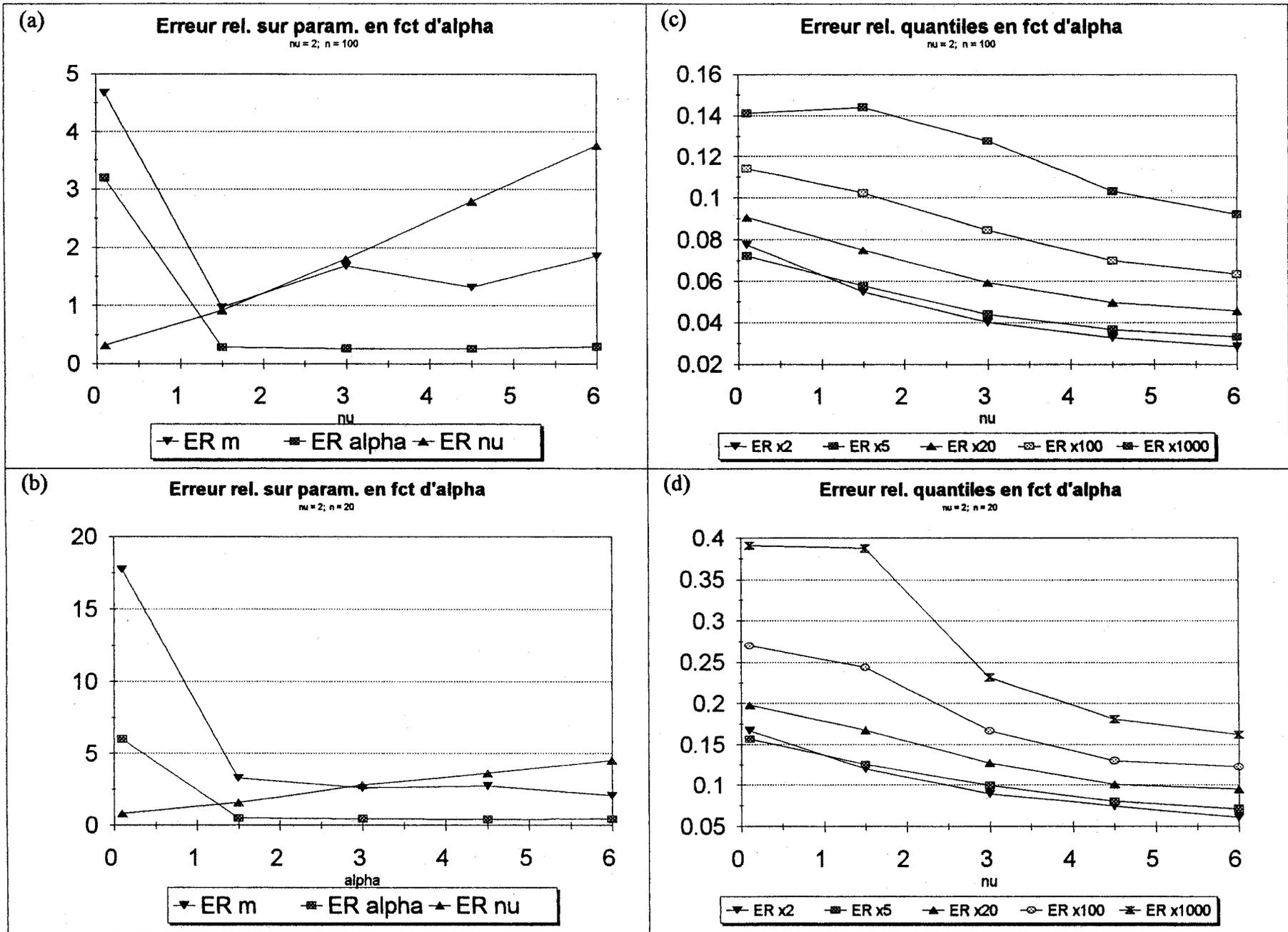


Figure 3.6 Erreur relative des estimateurs en fonction de alpha

Les estimateurs des quantiles voient leur biais décroître de façon générale en fonction de  $\alpha$  (figures 3.5c et 3.5d). Cependant, on peut observer certaines irrégularités. Par exemple, le biais de l'estimateur de  $x_5$  est pratiquement nul à  $\alpha = 0,1$  avec  $n = 100$ .

Aux figures 3.6a et 3.6b, on observe une croissance de l'erreur relative sur  $v$  qui semble proportionnelle à  $\alpha$ . Les erreurs des estimateurs de  $\alpha$  et de  $m$  ont pour leur part un comportement semblable au biais. Très élevées en  $\alpha = 0,1$ , elles décroissent brutalement entre 0,1 et 1,5 et par la suite restent généralement stables. Pour leur part, les erreurs relatives de tous les estimateurs des quantiles décroissent lorsque  $\alpha$  augmente de 0,1 à 6 (figures 3.6c et 3.6d).

### 3.2.2 Biais et EQM en fonction de $v$ (Figures 3.7 et 3.8)

Les remarques concernant le biais sur les paramètres en fonction de  $v$  sont faites d'après les figures 3.7a et 3.7b. La valeur  $v = -6$  a été ajoutée en cours de la simulation après avoir remarqué que le biais des estimateurs de  $m$  et de  $\alpha$  était presque nul à  $v = -4$ , laissant supposer un changement de signe autour de ce point. Effectivement, le biais sur  $m$  qui est positif et croissant à droite de  $v = -4$  devient négatif à  $v = -6$ . Le biais sur  $\alpha$ , qui est négatif dans l'intervalle  $[-4, 4]$ , devient positif en  $v = -6$  et tout laisse croire que la même chose se produit au-delà de  $v = 4$ , la courbe du biais en fonction de  $v$  semblant symétrique autour de  $v = 0$ . Ces constatations sont discutées plus en détail à la section 3.6. Pour le biais relatif sur  $v$ , nous n'avons pas de résultat en  $v = 0$ , mais les autres points indiquent un biais assez stable.

Une particularité est aussi rattachée au biais des estimateurs des quantiles (figures 3.7c et 3.7d). Pour les grands quantiles ( $x_{20}$ ,  $x_{100}$ ,  $x_{1000}$ ) le biais est du même signe que  $v$  et pour les petits quantiles ( $x_2$ ,  $x_5$ ) le biais est de signe opposé à  $v$ . Il se produit donc une inversion autour de 0, point où le biais est le plus faible.

Il semble y avoir un comportement asymptotique de l'erreur relative sur  $v$  qui tend vers l'infini lorsque  $v$  tend vers 0 (figures 3.8a et 3.8b). Elle diminue tandis que l'erreur sur  $\alpha$  croît légèrement lorsqu'on s'éloigne de part et d'autre de  $v = 0$ . Quant à l'estimateur de  $m$ , son erreur relative est en dents de scie pour  $n = 100$  alors qu'elle est croissante pour  $n = 20$ .

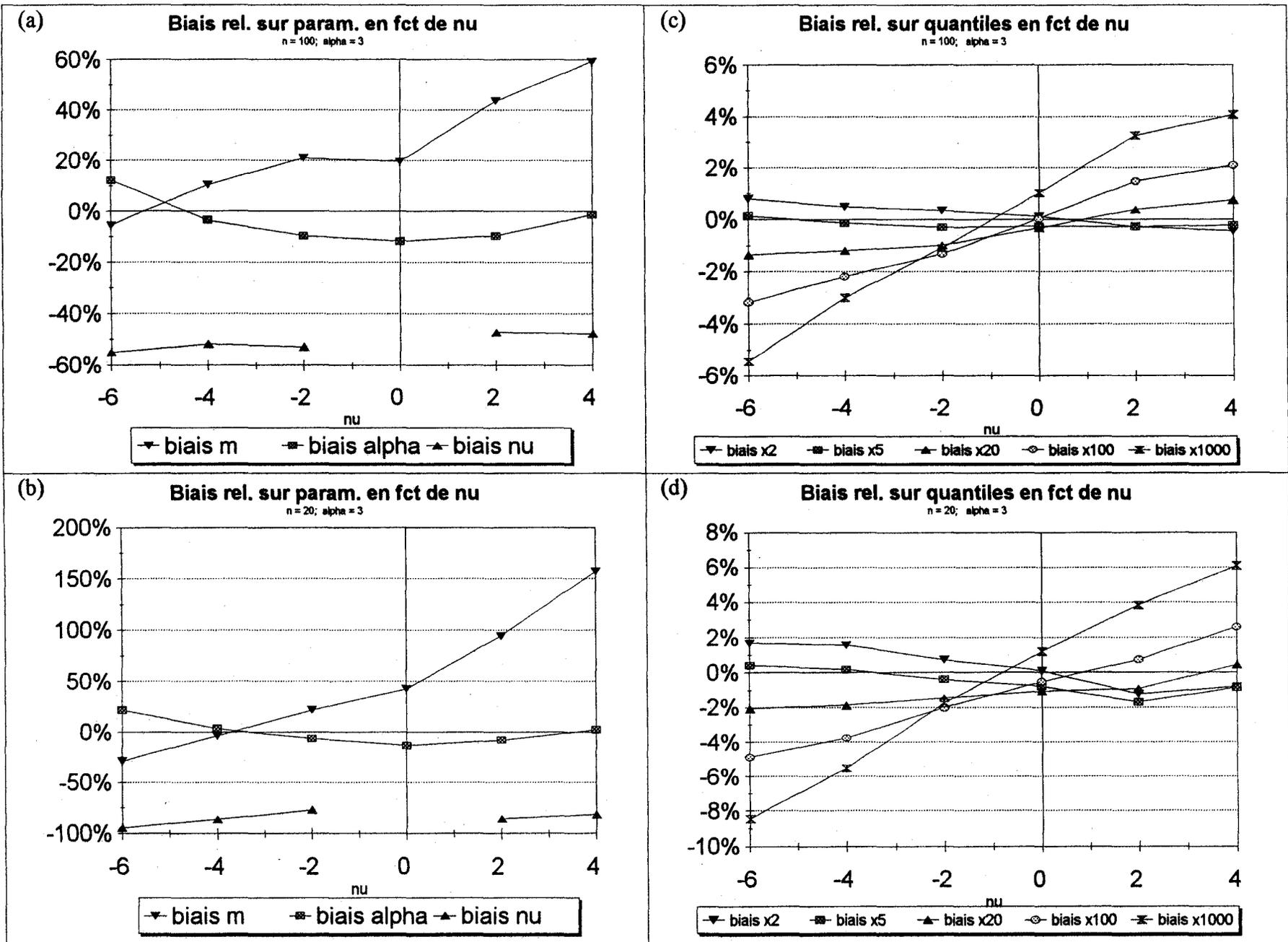


Figure 3.7 Biases relatifs des estimateurs en fonction de nu

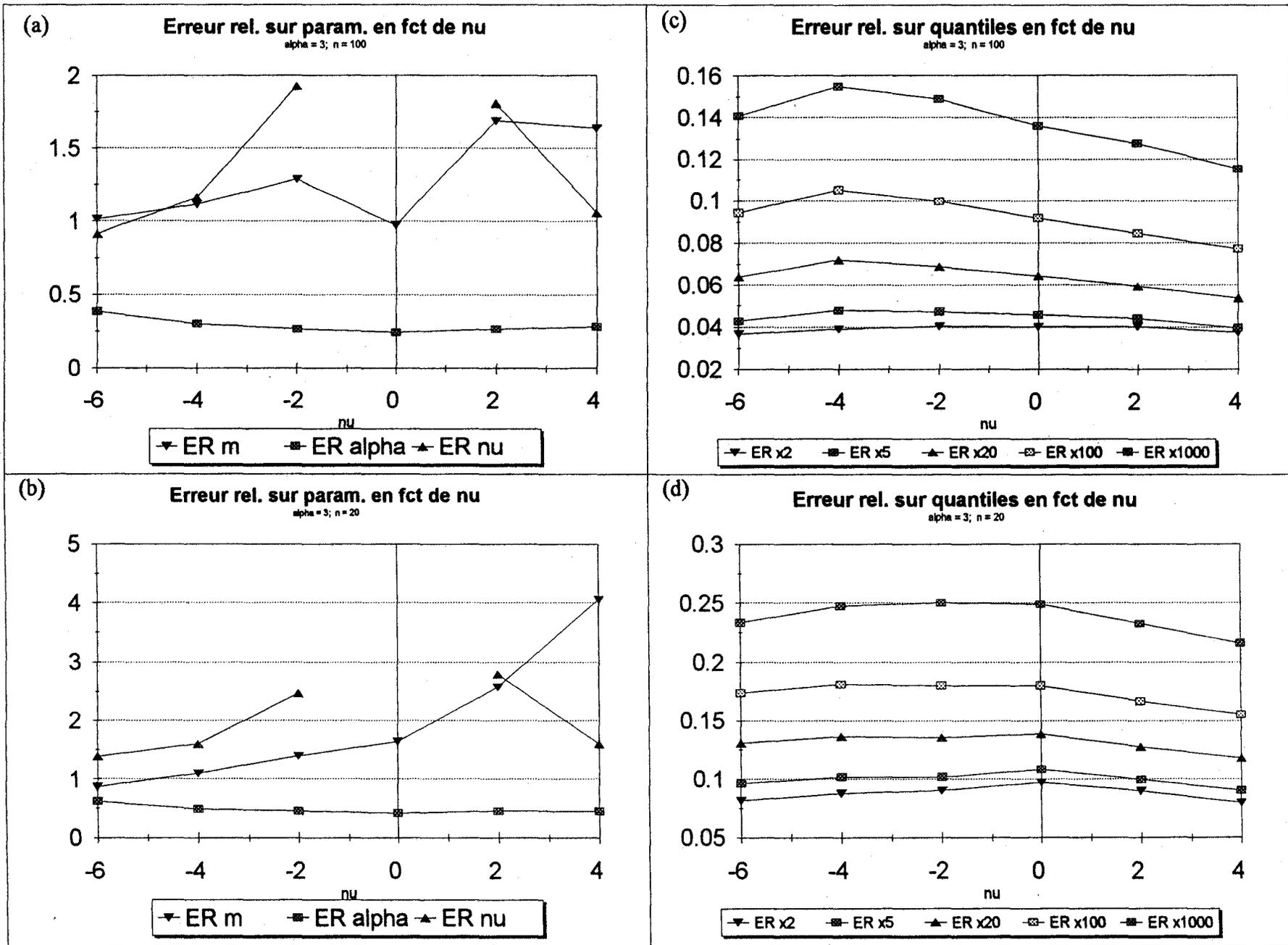


Figure 3.8

Erreur relative des estimateurs en fonction de  $\nu$

L'erreur relative des estimateurs des quantiles ne subit pas de profondes variations en fonction de  $v$  (figures 3.8c et 3.8d). Elle semble décroître lorsque  $v$  s'éloigne de 0 positivement. Du côté négatif, un sommet incertain est atteint autour de -4, mais la variation de part et d'autre est très faible. Ce maximum coïncide avec le point où le biais sur  $m$  et sur  $\alpha$  est minimal.

### 3.2.3 Quelques cas extrêmes

À partir de ces constatations, il est possible d'anticiper les résultats pour des lois dont les valeurs de paramètres sont extrêmes. Ainsi, on peut déduire que le biais sur les quantiles est minimisé lorsque  $\alpha$  est très grand et que  $v = 0$ . Effectivement, c'est avec le couple (10,0) que les biais les plus faibles sont observés sur les quantiles : moins de 0,8% (figure 3.9a). Les variations désordonnées en fonction de  $n$  ne sont pas inquiétantes avec un biais si faible.

Les problèmes aussi peuvent être prévus. Le biais sur les estimateurs des paramètres  $\alpha$  et  $m$  est susceptible de croître considérablement lorsque  $\alpha$  est petit et que  $|v|$  est grand. Les résultats de la loi de paramètre  $\alpha = 0,1$  et  $v = 8$  et de son opposé de paramètres  $\alpha = 0,1$  et  $v = -8$  se distinguent en effet par des estimations de  $\alpha$  et de  $m$  qui n'ont rien à voir avec les vraies valeurs de ces paramètres.

Avec le couple (0,1;8) (figure 3.9b), le biais sur  $m$  atteint 9000% pour une taille d'échantillons de 20, ce qui veut dire que le paramètre  $m$  est estimé en moyenne à plus de 90 fois sa vraie valeur. Ce biais diminue à un peu moins de 5000% pour une taille de 100. Le couple (0,1;-8) (figure 3.9c) est aux antipodes pour l'estimation de  $m$ . Cette fois le biais se situe à -97% pour toutes les tailles d'échantillons. Le paramètre  $m$  n'est donc estimé qu'à 3% de sa valeur. Le biais de l'estimateur de  $\alpha$  pour ces deux lois dépasse 3000% avec  $n = 20$  et ne descend guère en deçà quand la taille augmente. Sur  $v$ , le biais prend des valeurs plus habituelles (mais néanmoins considérables) entre -80% et -60%.

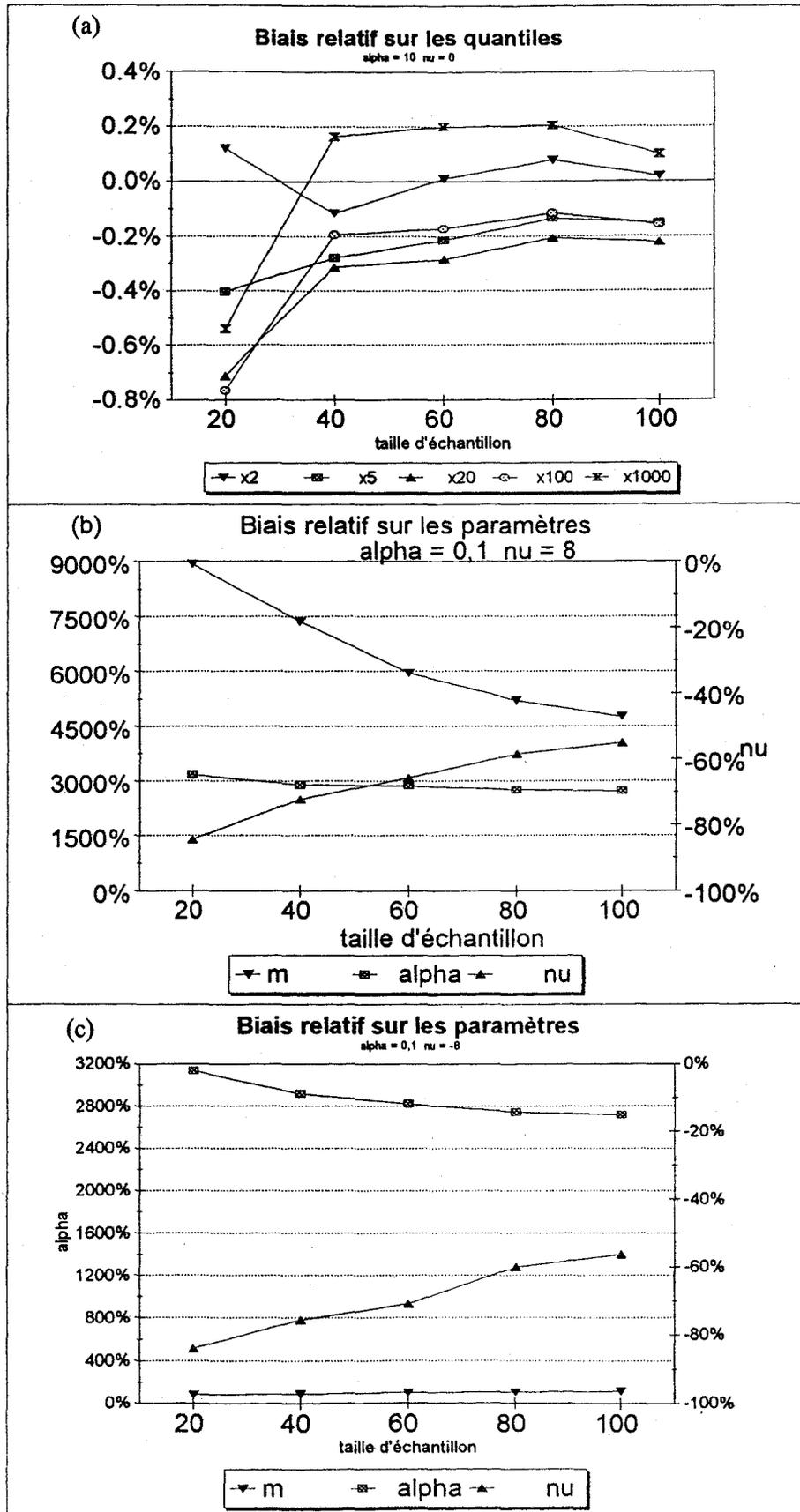


Figure 3.9 Biais relatif des estimateurs pour quelques cas extrêmes

Un tel biais sur  $\alpha$  et sur  $m$  fait grossir l'erreur d'estimation de ces paramètres. Voici ce qu'on constate (figure 3.10):

- Avec  $\alpha = 0,1$  et  $v = 8$ , l'erreur relative sur  $m$  avoisine 140 pour  $n = 20$  et 60 pour  $n = 100$ .
- Avec  $\alpha = 0,1$  et  $v = -8$ , l'erreur relative sur  $m$  est d'environ 1 pour toutes les tailles, soit à peu près égale au biais en valeur absolue.
- Pour les deux lois, l'erreur relative de l'estimateur de  $\alpha$  atteint 35 pour  $n = 20$  et décroît lentement quand  $n$  augmente.

Ces erreurs et ces biais disproportionnés ne sont pas l'effet de seulement quelques estimations singulières. Les histogrammes des estimations de  $m$  pour les deux lois considérées (figure 3.11) montrent bien que toute la distribution de l'estimateur est située loin de la valeur à estimer, c'est-à-dire 100.

Avec de tels biais et de telles erreurs sur les paramètres, il serait prévisible que les estimateurs des quantiles s'écartent des vraies valeurs. Ce n'est pourtant pas ce qui arrive. Le biais se maintient à moins de 20% et l'erreur relative à moins de 30% pour une taille de 20.

### 3.3 Erreur quadratique moyenne : variance et biais

L'erreur quadratique moyenne peut s'exprimer selon l'équation (3-5) :

$$EQM = Var + B^2 \quad (3-5)$$

Cette décomposition explique pourquoi l'erreur quadratique moyenne est considérée comme une mesure globale de la qualité d'un estimateur. Elle permet aussi d'étudier l'importance relative du biais et de la variance.

La mesure retenue pour étudier ce point est la proportion de variance, exprimée en pourcentage de l'erreur quadratique moyenne des estimateurs. Ces proportions ont été calculées pour toutes les lois simulées et pour les 5 tailles d'échantillon. Pour une même loi, les parts du biais et de la variance sont pratiquement constantes pour toutes les tailles d'échantillons. Aussi le tableau 3.2 contient la moyenne pour toutes les tailles, dont la variance n'excède pas quelques dixièmes de un pour cent.

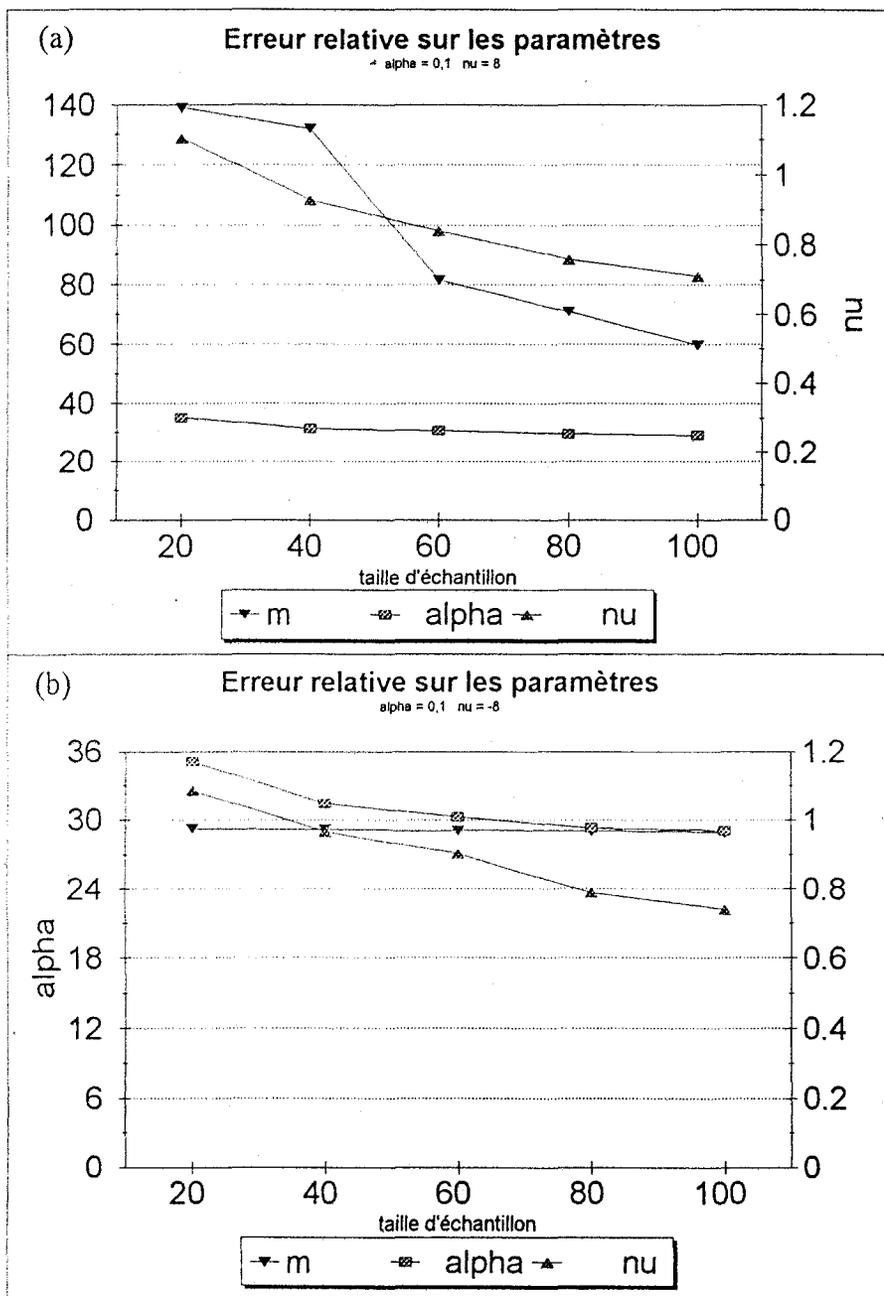


Figure 3.10 Erreur relative lorsque  $\alpha = 0,1$  et  $\nu = 8$  ou  $-8$

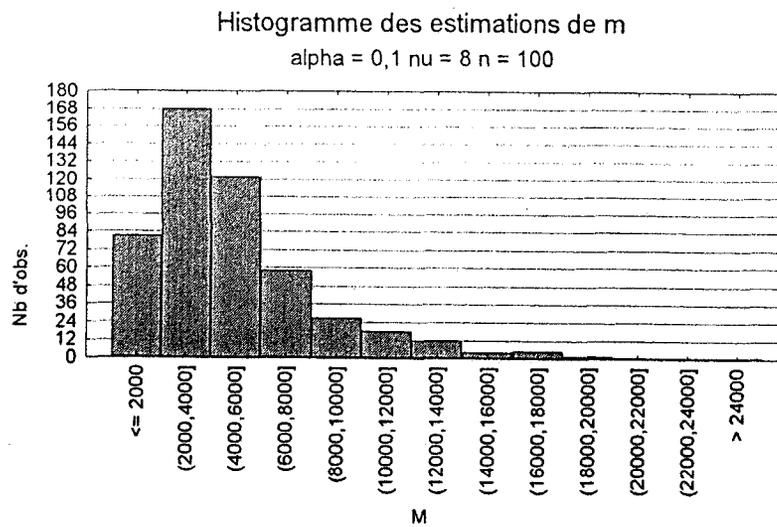
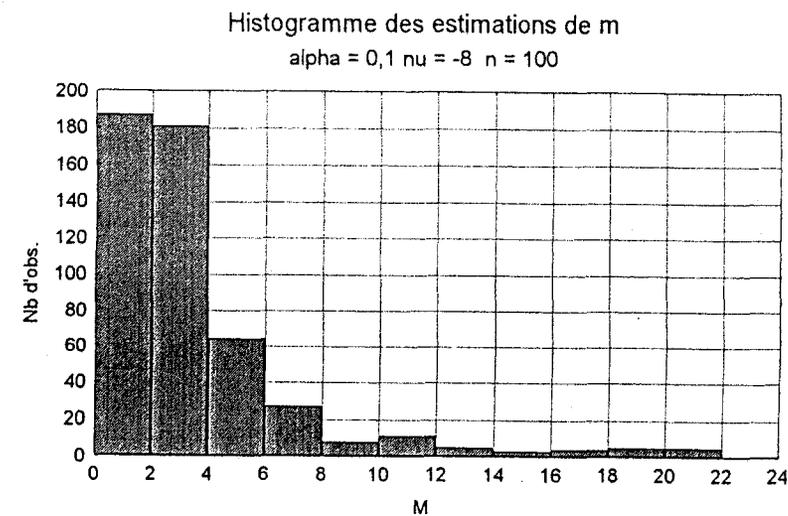


Figure 3.11 Distribution de l'estimateur de  $m$  lorsque  $\alpha = 0,1$  et  $\nu = 8$  ou  $-8$

Tableau 3.2. Part de variance dans l'erreur quadratique moyenne

Valeur des paramètres		Part de la variance dans l'erreur quadratique moyenne							
$\alpha$	$\nu$	$\hat{m}$	$\hat{\alpha}$	$\hat{\nu}$	$x_2$	$x_5$	$x_{20}$	$x_{100}$	$x_{1000}$
0.1	-8	0.1%	14.4%	40.4%	92.3%	99.6%	89.7%	68.9%	49.0%
0.1	0	96.3%	96.3%	100.0%	98.6%	99.5%	99.8%	99.7%	99.2%
0.1	1	82.0%	81.4%	88.4%	98.2%	99.3%	99.6%	97.8%	96.0%
0.1	2	57.8%	39.6%	59.5%	92.8%	99.3%	95.3%	87.4%	81.7%
0.1	8	51.4%	14.4%	39.9%	93.2%	98.8%	94.2%	82.3%	72.6%
1.5	-2	98.5%	99.4%	84.8%	97.4%	99.9%	97.1%	96.7%	98.5%
1.5	0	95.4%	89.2%	99.7%	99.9%	99.0%	99.2%	99.9%	98.8%
1.5	2	89.8%	99.0%	86.8%	99.3%	99.6%	99.1%	95.2%	91.1%
3	-6	96.3%	88.7%	57.8%	94.4%	99.8%	95.7%	88.3%	83.1%
3	-4	99.4%	99.2%	75.8%	97.7%	100.0%	97.5%	95.3%	95.2%
3	-2	97.6%	88.8%	91.1%	99.4%	99.7%	98.4%	98.5%	99.5%
3	0	93.8%	80.1%	99.7%	99.9%	99.8%	99.8%	99.7%	98.6%
3	2	88.6%	90.1%	90.8%	98.8%	98.4%	99.7%	98.2%	94.5%
3	4	85.3%	99.6%	76.0%	98.4%	99.3%	98.8%	93.8%	87.9%
4.5	2	89.1%	83.7%	94.4%	99.5%	98.9%	99.8%	99.2%	96.8%
6	2	88.4%	78.7%	94.7%	99.9%	99.8%	99.7%	98.3%	95.7%
10	-8	99.2%	88.6%	80.6%	99.2%	100.0%	99.1%	97.9%	97.0%
10	0	92.1%	83.7%	99.9%	99.9%	99.6%	99.5%	99.8%	99.9%
10	8	84.2%	89.8%	80.3%	99.4%	99.6%	99.5%	97.4%	93.9%

On tire principalement du tableau 3.1 les observations suivantes :

- Règle générale, la proportion de variance est beaucoup plus grande que la proportion de biais;
- La proportion de variance est plus grande pour les estimateurs des quantiles que pour ceux des paramètres;
- La proportion de variance est plus grande avec les petits quantiles qu'avec les grands; Elle atteint presque 100% avec la médiane ( $x_2$ );
- Lorsque  $\alpha$  augmente ou que  $\nu$  s'approche de 0 :
  - a) la part de variance pour l'estimateur de  $\nu$  augmente
  - b) la part de variance pour l'estimateur de  $\alpha$  diminue
- Lorsque les estimateurs des paramètres sont les plus biaisés, soit lorsque  $\alpha$  est petit ( $\alpha = 0,1$ ) et  $\nu$  grand ( $\nu = \pm 8$ ), la part de variance devient inférieure à la part de biais. Parfois même, la part de variance devient négligeable.

En dehors des cas extrêmes, la variance demeure beaucoup plus importante que le biais. Cette observation tend à montrer que le biais n'est pas si considérable qu'il n'apparaît à première vue.

### 3.4 Variances asymptotiques et expérimentales

L'objectif de minimiser la variance est poursuivi lors de l'estimation d'une quantité. Il faut se rappeler que l'estimation des paramètres de la loi de Halphen type A est basée sur les statistiques exhaustives et qu'une fonction de ces statistiques est l'estimateur sans biais de variance minimale. D'autre part, une borne minimale à la variance d'un estimateur est donnée, lorsque celui-ci est sans biais, par l'inverse du produit de l'information de Fisher par le nombre d'observations (Bickel and Doksum, 1977). Pour les estimateurs des quantiles, une telle borne minimale est une fonction des variances et covariances des estimateurs des paramètres et de la fonction de répartition de probabilité (Bobée et Ashkar, 1991). Rappelons qu'avec un estimateur du maximum de vraisemblance, qui est asymptotiquement non-biaisé, cette borne minimale est la variance de la loi asymptotique de l'estimateur, soit la loi normale. Cette variance est donc atteinte lorsque la taille d'échantillon tend vers l'infini.

Il est maintenant connu que les estimateurs du maximum de vraisemblance des paramètres et des quantiles sont biaisés. La comparaison de leur variance observée avec la variance asymptotique des estimateurs sans biais n'en est pas moins instructive.

Les courbes de la variance expérimentale et de la variance asymptotique ont été tracées sur un même graphique pour les estimateurs des trois paramètres et de trois quantiles. Quelques graphiques représentatifs sont reproduits à la figure 3.12. La variance formant la majeure partie de l'erreur quadratique moyenne, on remarque que les courbes de la variance expérimentales rappellent celles de l'EQM (voir figure 3.2a). La variance asymptotique est proportionnelle à  $1/n$  avec la constante de proportionnalité égale à l'inverse de l'information de Fisher.

En ce qui concerne le paramètre  $m$ , la variance expérimentale est supérieure à la variance asymptotique, sauf lorsque le biais sur  $m$  est négatif. La variance expérimentale de l'estimateur de  $\alpha$  est inférieure à sa variance asymptotique, excepté pour les lois où  $v = 0$ , et il en est de même pour l'estimateur de  $v$  à l'exception des lois où  $\alpha$  est petit.

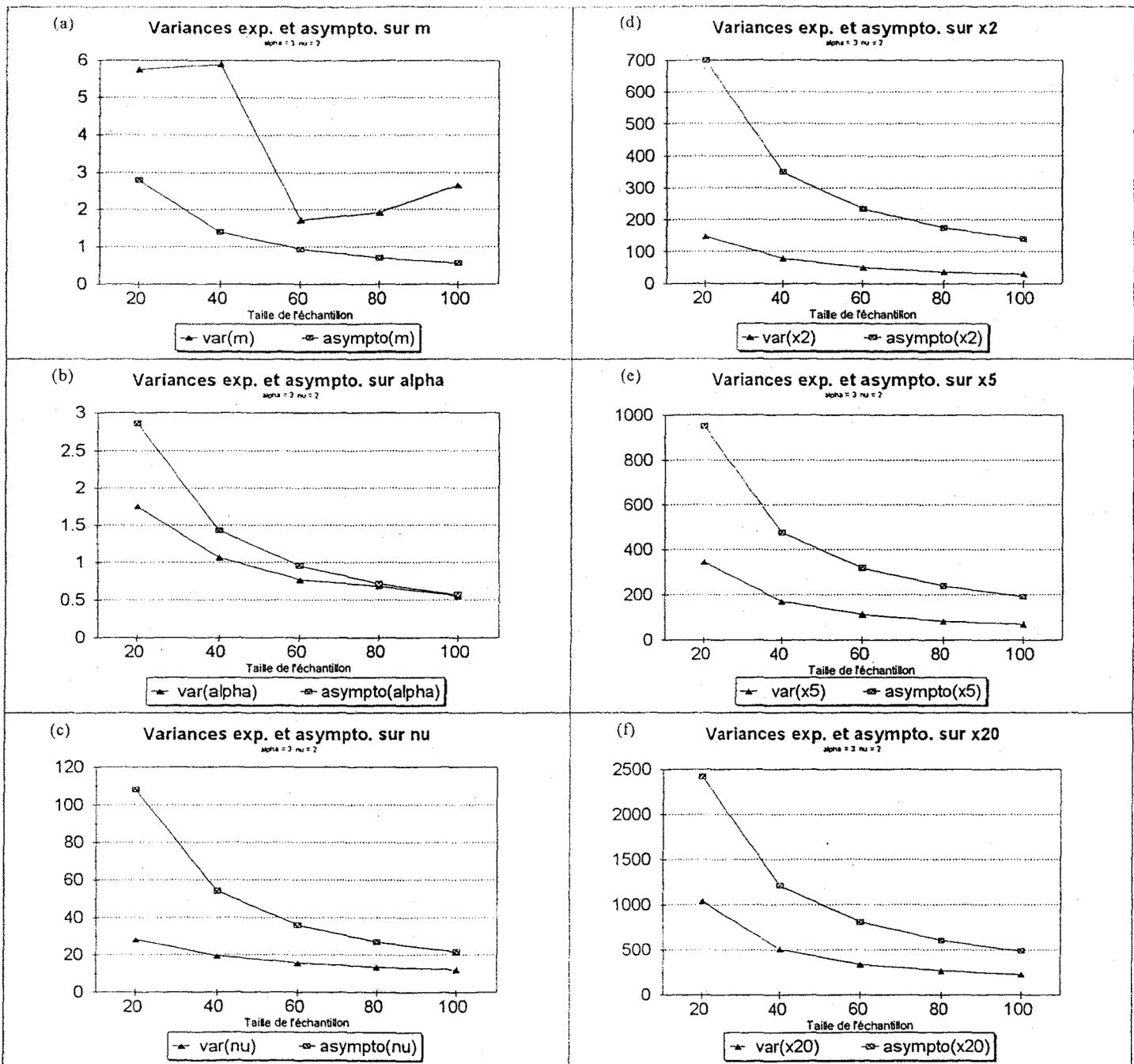


Figure 3.12. Variance asymptotiques et expérimentales des estimateurs lorsque  $\alpha=3$  et  $\nu=2$

Les variances asymptotiques théoriques ne sont calculées que pour les quantiles  $x_2$ ,  $x_5$  et  $x_{20}$ . Pour ces trois quantiles, quelles que soient les valeurs des paramètres de la loi, la variance expérimentale est inférieure à la variance asymptotique

Ces observations portent à penser que pour les quantiles et pour certains intervalles de valeurs des paramètres, les estimateurs du maximum de vraisemblance ont une variance inférieure à la borne minimale pour la variance des estimateurs sans biais, et que comme le biais est petit comparé à la variance, ils seraient "meilleurs" que des estimateurs sans biais. Cette interprétation est mise en doute si on considère que le calcul numérique des variances asymptotiques par l'inversion de la matrice d'information de Fisher est imprécis, voire parfois erroné. Ainsi, lorsque  $\alpha = 0,1$  et  $\nu = \pm 8$ , les variances asymptotiques des paramètres sont négatives. Les covariances prennent quelquefois des valeurs démesurées. Ainsi, le calcul des variances asymptotiques sur les quantiles à partir de ces éléments imprécis peut être complètement faussé. C'est d'ailleurs à cause de leurs valeurs démesurées que les estimations des variances asymptotiques pour les grands quantiles ont été retirées de *AJUSTE-II*. L'amélioration des méthodes numériques employées, et en particulier l'inversion de matrice, est donc un préalable à toute étude ultérieure de la question.

### 3.5 Association entre les estimateurs des paramètres

La covariance est une mesure de l'association entre deux variables aléatoires comme le sont les estimateurs. Cette mesure étant sujette à l'ordre de grandeur des quantités estimées, on lui préfère sa mesure standardisée, la corrélation.

Les observations suivantes sont faites à propos de l'association entre les estimateurs des paramètres. La corrélation expérimentale entre les estimateurs de  $\alpha$  et de  $\nu$  est de signe opposé au signe du paramètre  $\nu$  de la loi. La grandeur de cette corrélation est très variable. Les estimateurs de  $\nu$  et de  $m$  sont fortement corrélés négativement. Entre les estimateurs de  $\alpha$  et de  $m$ , la corrélation expérimentale est négative et plus ou moins forte, sauf lorsque  $\alpha$  est petit et que  $\nu > 0$ . Dans ce cas, elle est fortement positive.

Le point le plus important à retenir de ceci est le fort lien entre les estimateurs des paramètres. Diverses interprétations du signe des corrélations pourraient être tirées éventuellement. À titre d'exemple, la forte corrélation négative entre les estimateurs de  $m$  et de  $\nu$  signifie que la sous-estimation de  $\nu$  s'accompagne de la surestimation de  $m$ .

Ces corrélations jouent probablement un rôle important et pourraient expliquer le fait que même si les estimateurs des paramètres sont fortement biaisés, ceux des quantiles le sont moins. Ce point sera développé plus longuement à la section suivante.

### 3.6 Explications concernant le biais

L'examen de la méthode d'estimation permet de voir que l'estimation des paramètres  $m$  et  $v$  découle de l'estimation de  $\alpha$  et donc de la solution de l'équation  $D_A(\alpha, v) = A/H$ . De tels liens entre les estimateurs des paramètres montrent que l'erreur commise en estimant  $\alpha$  a des conséquences sur l'estimation de  $v$  et de  $m$ . Les corrélations discutées à la section 3.5 illustrent également les liens entre ces estimateurs.

Une loi à trois paramètres est très souple et peut prendre une grande variété de formes. Réciproquement, deux formes très voisines de la loi peuvent avoir des valeurs de paramètres très différentes. La présence de deux paramètres de forme dont les effets peuvent se compenser rend cette situation possible. Dans le cas de la loi de Halphen type A, la compensation implique aussi le paramètre d'échelle  $m$  étant donné que les valeurs des paramètres de forme ont une grande influence sur l'échelle de la distribution. Le tableau 3.2 permet de le constater. Ce tableau présente les moyennes et certains quantiles théoriques pour diverses valeurs du couple de paramètres  $(\alpha, v)$ , avec  $m$  fixé à 100.

**Tableau 3.3.** Moyennes et quantiles théoriques de quelques lois de Halphen type A

$\alpha$	$v$	A	G	H	$X_2$	$X_{20}$	$X_{1000}$
0,1	-8	1,428	1,332	1,250	1,304	2,511	5,070
0,1	0	272,49	100,00	36,70	100,00	1132,88	4047,97
0,1	8	8001,4	7507,1	7001,7	7670,7	13149,5	19627,6
3	-4	59,37	55,43	51,89	54,81	103,29	182,81
3	0	108,03	100,00	92,56	100,00	190,52	326,30
3	4	192,70	180,42	168,45	182,44	323,62	503,17
10	-8	69,85	68,27	66,73	68,08	97,47	134,19
10	0	102,47	100,00	97,59	100,00	143,85	196,24
10	8	149,85	146,49	143,16	146,88	207,24	275,62

Il ressort de ce tableau que :

- Les moyennes théoriques et les quantiles sont croissants en  $v$ .
- La dispersion diminue quand  $\alpha$  augmente.

L'explication du comportement de biais de l'estimateur de  $m$  peut maintenant être donnée. L'estimation de  $m$  est le produit de deux éléments : la moyenne arithmétique et le rapport de deux fonctions de Bessel (équation 1-5). Ce rapport est l'inverse de la fonction  $R(\alpha, \nu)$  étudiée par Jørgensen (1982). On constate que la moyenne arithmétique varie beaucoup entre des lois de Halphen type A de même paramètre  $m$ . Si les paramètres  $\alpha$  et  $\nu$  sont bien estimés, le rapport des fonctions de Bessel est tel qu'il ramène l'estimation près de la vraie valeur du paramètre  $m$ . Mais on observe que la valeur de l'estimateur de  $\nu$  est systématiquement plus près de 0 que la valeur théorique de  $\nu$  et que  $\alpha$  est surestimé lorsqu'il est petit (voir par exemple les figures 3.5a et 3.9b). Dans ces conditions le rapport des fonctions de Bessel reste près de 1 et l'estimation de  $m$  est à peu près égale à la moyenne arithmétique des observations. Ceci illustre que l'estimation de  $m$  compense pour la mauvaise estimation de  $\alpha$  ou de  $\nu$ .

Le passage de valeurs de  $m$  sous-estimées à des valeurs surestimées devrait s'effectuer lorsque  $m$  est égal à la moyenne théorique de la loi, ce qui se produit à  $\nu = -1/2$ . Dans la pratique, ce passage s'effectue à une valeur de  $\nu$  inférieure, ce qui serait possiblement dû encore une fois au biais de l'estimateur de  $\nu$ .

Il a été répété à plusieurs reprises que le biais et l'erreur quadratique moyenne des quantiles demeurent minime malgré les biais énormes sur les paramètres. À ce point-ci, l'explication de ce phénomène est possible. Des estimations de paramètres très éloignées de leur vraie valeur donnent néanmoins une loi dont la densité de probabilité est voisine de la vraie loi des observations. Les quantiles de la loi ajustée sont donc près de ceux de la vraie loi, d'où un biais et une erreur quadratique moindre.

## **4 ÉTUDE DES PROBLÈMES D'ESTIMATION**

Les estimateurs du maximum de vraisemblance des paramètres de la loi de Halphen type A existent et sont uniques, tel que démontré par Jørgensen (1982), mais ils ne sont pas nécessairement les solutions exactes des équations du maximum de vraisemblance. La méthode d'estimation qui consiste à résoudre ces équations ne donne donc pas toujours une solution. D'autre part, les limites des méthodes numériques utilisées entraînent que certaines solutions ne sont qu'approximatives. Ces problèmes sont suffisamment sérieux pour en justifier un examen attentif.

### **4.1 Considérations théoriques sur les absences de solution**

Pour comprendre l'origine des problèmes d'estimation, il est nécessaire de revenir à la procédure d'estimation, et plus particulièrement, d'examiner le comportement de la statistique  $U$  (voir section 1.2).

La statistique  $U = AH^{-1}/(AH^{-1} - 1)$  est une mesure de la concentration des observations. En effet, plus les observations sont concentrées, plus les moyennes arithmétiques et harmoniques sont rapprochées, ce qui fait augmenter la valeur de  $U$ .

Parmi les paramètres de la loi de Halphen type A,  $\alpha$  est un paramètre de concentration. Il est donc naturel que la valeur de  $\alpha$  ait un effet sur la localisation de  $U$ . Les valeurs de  $U$  rencontrées sont plus près de 1 (valeur minimale de  $U$ ) pour les lois dont le paramètre  $\alpha$  est petit, et plus élevées pour les lois dont le paramètre  $\alpha$  est grand. L'intervalle  $[-U, U]$  est donc réduit et on s'attend à ce que le nombre d'échantillons pour lesquels le maximum de la fonction de vraisemblance partiellement maximisée  $\log L(\nu, \hat{\alpha}, \hat{m})$  est en dehors de l'intervalle soit plus important. Lorsque cela se produit, il a été démontré que les estimateurs du maximum de vraisemblance sont ceux d'une loi Gamma ou Gamma inverse selon le signe de  $\nu$  (Jørgensen 1982). Ce résultat est en accord avec la démonstration que la loi de Halphen type A tend vers une de ces lois lorsque  $\alpha$  tend vers 0.

Dans la pratique cependant, on s'aperçoit que la fréquence des estimations en dehors de l'intervalle  $[-U, U]$  décroît quand  $\alpha$  approche 0 et s'accroît quand  $\alpha$  augmente. Cette situation, contraire aux attentes, a été étudiée.

Pour savoir si la fonction  $\log L(\nu, \hat{\alpha}, \hat{m})$  atteint son maximum à l'intérieur ou à l'extérieur de  $[-U, U]$  les dérivées par rapport à  $\nu$  de la fonction aux points  $-U$  et  $U$  sont calculées.

Des bornes inférieure et supérieure à cette dérivée sont aisément trouvées (Jørgensen 1982). Ces bornes sont représentées à la figure 4.1 et sont données au tableau 4.1.

**Tableau 4.1.** Bornes de la dérivée par rapport à  $\nu$  de  $\log L(\nu, \hat{\alpha}, \hat{m})$  aux points  $-U$  et  $U$

Point	$-U$	$U$
Borne Supérieure	$n[\psi(U) - \ln(U-1)]$	$n[\ln(U) - \psi(U)]$
Borne Inférieure	$n[\psi(U) - \ln(U)]$	$n[\ln(U-1) - \psi(U)]$

Le symbole  $\psi$  désigne la fonction digamma définie par l'équation 4-1.

$$\Psi(u) = \frac{d \ln \Gamma(u)}{du} \quad (4-1)$$

Quelques constatations peuvent être faites:

- Les dérivées peuvent prendre des valeurs positives ou négatives pour toute valeur de  $U$ .
- C'est lorsque  $U$  tend vers 0 que le domaine des valeurs des dérivées est le plus large.
- Les dérivées aux points  $U$  et  $-U$  tendent vers 0 lorsque  $U$  tend vers l'infini, ce qui signifie que la fonction  $\log L(\nu, \hat{\alpha}, \hat{m})$  s'aplatit.

L'absence de solution est causée en dernière analyse par une dérivée de signe positif au point  $U$  ou négatif au point  $-U$ , indiquant que le maximum de  $\log L(\nu, \hat{\alpha}, \hat{m})$  est en dehors de l'intervalle  $[-U, U]$ . L'augmentation de la fréquence de ce problème lorsque  $U$  est grand peut provenir d'une instabilité réelle du maximum de la fonction  $\log L(\nu, \hat{\alpha}, \hat{m})$ . Cependant, comme la fonction est très aplatie et la valeur de la dérivée très près de 0, une évaluation imprécise de la fonction digamma peut faire changer le signe de la dérivée et causer des rejets injustifiés.

La valeur du paramètre  $\nu$  de la loi joue aussi un grand rôle dans le succès ou l'échec de la méthode d'estimation. Son influence est prévisible : plus la valeur de  $\nu$  s'éloigne de 0, plus le maximum de la fonction  $\log L(\nu, \hat{\alpha}, \hat{m})$  est susceptible d'être en dehors de l'intervalle  $[-U, U]$ .

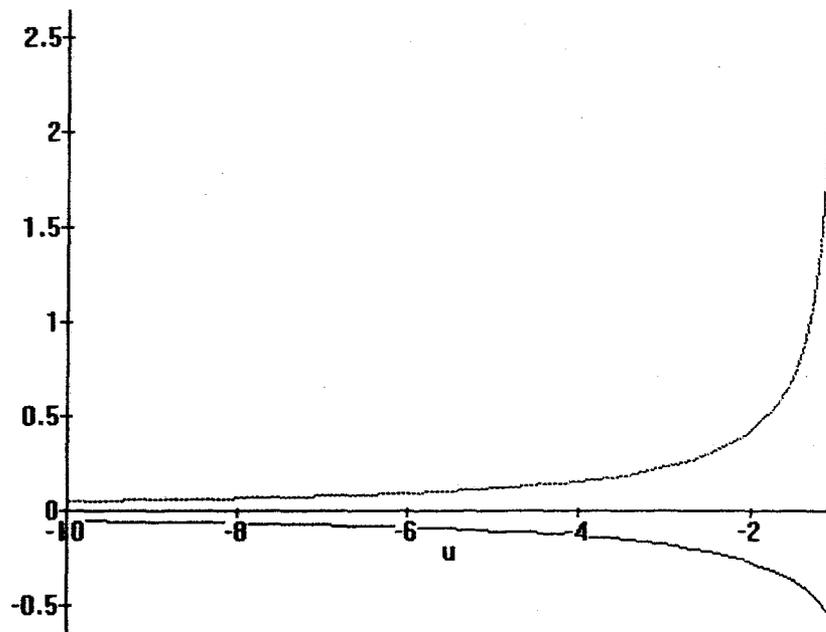
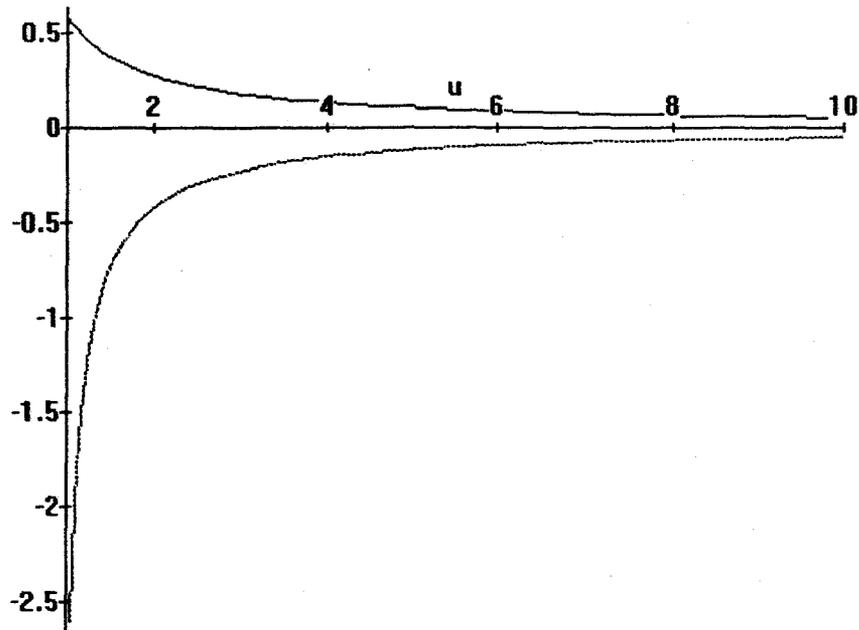


Figure 4.1 Bornes des dérivées de  $\log L(u, \hat{\alpha}, \hat{m})$  aux points  $U$  et  $-U$

## 4.2 Nombre d'échantillons sans solution en fonction de $\alpha$ , $\nu$ et $n$ .

La fréquence des échantillons pour lesquels la procédure d'estimation est incapable de fournir une solution est extrêmement variable en fonction des valeurs de  $\alpha$  et de  $\nu$  et de la taille des échantillons. Le nombre d'échantillons écartés avant d'obtenir 500 estimations réussies est donné au tableau 4.2 pour toutes les lois simulées et toutes les tailles d'échantillons considérées.

**Tableau 4.2.** Nombre d'échantillons rejetés avant d'obtenir 500 réplicats

Valeurs des paramètres		Taille des échantillons					Total
$\alpha$	$\nu$	20	40	60	80	100	
0.1	-8	486	415	362	380	418	2061
0.1	0	0	0	0	0	0	0
0.1	1	27	17	11	9	7	71
0.1	2	148	179	188	186	179	880
0.1	8	510	410	376	378	391	2065
1.5	-2	214	99	77	65	47	502
1.5	0	134	39	25	13	7	218
1.5	2	211	122	76	55	39	503
3	-6	448	288	255	212	206	1409
3	-4	373	240	162	153	113	1041
3	-2	359	188	108	83	62	800
3	0	341	170	97	60	46	714
3	2	314	173	107	96	83	773
3	4	373	227	177	154	133	1064
4.5	2	434	254	172	130	107	1097
6	2	592	340	267	202	156	1557
10	-8	913	585	461	350	321	2630
10	0	800	538	421	287	270	2316
10	8	849	533	422	390	324	2518

Comme il a été expliqué à la section 4.1, le nombre d'échantillons sans solution augmente lorsque  $\alpha$  s'accroît et lorsque  $\nu$  s'éloigne de 0. Logiquement, parmi les couples de

paramètres des lois simulées, le nombre de rejets est minimal pour le couple (0,1;0). En fait, dans ce cas la méthode d'estimation a toujours donné une solution.

Dès que l'on prend  $\alpha$  ou  $|\nu|$  plus élevé, le nombre d'échantillons sans solution se met à croître. Les sommets sont atteints avec les couples (10,8) et (10,-8) où près de 1400 échantillons sont nécessaires pour obtenir 500 répliqués de taille 20 (voir tableau 4.2). C'est la raison essentielle pour laquelle des valeurs plus élevées des paramètres  $\alpha$  et  $\nu$  n'ont pas été incluses dans l'étude.

L'effet d'une augmentation de la taille d'échantillon est en général de faire diminuer le nombre de rejets, et cette diminution est alors assez marquée. Cependant, avec un  $\alpha$  petit ( $\alpha = 0,1$ ) et  $\nu$  différent de 0, le nombre d'échantillons sans solution fluctue d'une taille à l'autre sans tendance à la baisse, à l'exception d'un saut entre les tailles 20 et 40.

Une absence de solution est causée par une des deux situations suivantes: soit le maximum de la vraisemblance est atteint en un point inférieur à la borne  $-U$ , soit ce maximum est atteint en un point supérieur à la borne  $U$ . Il est possible de distinguer pour chaque échantillon écarté la cause de son rejet. Il est prévisible que, pour une même valeur de  $\alpha$ , lorsque  $\nu > 0$ , le nombre de rejets causés par une estimation de  $\nu$  supérieure à  $U$  soit beaucoup plus grand que le nombre de rejets dus à une estimation inférieure à  $-U$ , et que l'inverse se produise lorsque  $\nu < 0$ . On peut même avancer qu'avec  $\alpha = 0,1$ , comme aucun rejet n'est enregistré à  $\nu = 0$ , la totalité des rejets devraient provenir du dépassement de la borne du même signe que  $\nu$ . Pourtant, il arrive qu'avec les couples (0,1;8), des échantillons soient rejetés parce que la borne de signe contraire est dépassée (voir tableau 4.3). Ceci indique que le maximum de la fonction de vraisemblance logarithmique partielle est très instable, à tout le moins lorsque  $|\nu|$  est grand.

**Tableau 4.3.** Causes des rejets d'échantillons

Valeurs des paramètres		Situation du maximum de $\log L(\nu, \hat{\alpha}, \hat{m})$		Total
$\alpha$	$\nu$	$< -U$	$> U$	
0,1	-8	1900 (92%)	161 (8%)	2061
0,1	0	0	0	0
0,1	8	166 (8%)	1899 (92%)	2065
3	-4	784 (75%)	257 (25%)	1041
3	0	376 (53%)	338 (47%)	714
3	4	201 (19%)	863 (81%)	1064

Le fait d'identifier simplement la tendance de la fréquence des rejets en fonction des valeurs de paramètres et de la taille de l'échantillon reste assez vague. On peut se demander lequel des paramètres  $\alpha$  ou  $v$  a le plus d'effet sur le nombre d'échantillons sans solution, ou s'il y a interaction entre ces facteurs. Dans le cours de la simulation, il était utile de prédire le nombre de rejets qui allaient survenir avec un nouveau couple de paramètres, en se basant sur les résultats antérieurs. Cela permettait de prévoir suffisamment d'échantillons de remplacement. Pour obtenir plus d'information sur la fréquence des rejets, des modèles de régression linéaire ont été utilisés. De tels modèles sont bien sûr imparfaits, surtout avec le petit nombre de valeurs simulées de chaque paramètre. De plus, le nombre de rejets obéit à une loi binomiale négative, et la qualité de l'approximation de cette distribution par la loi normale est inconnue. Il est aussi douteux que les variances des erreurs soient égales. En raison de cela, on ne peut obtenir d'intervalles de confiance fiables.

Les différentes tailles d'échantillons ont été traitées séparément. Il s'est avéré que le modèle approprié diffère d'une taille à l'autre. Il est aussi très influencé par l'ajout de nouvelles données comme cela a pu être constaté au fur et à mesure de l'avancement de la simulation. Les prévisions faites dans le courant de ce processus ont néanmoins été assez précises.

Les formes les plus courantes du modèle sont du type :

$$NR = b_0 + b_{11}\alpha + b_{22}v^2 + e$$

ou

$$NR = b_0 + b_{11}\alpha + b_{12}\alpha^2 + b_{22}v^2 + e$$

où NR désigne le nombre de rejets. Les coefficients  $b_{ij}$  sont très variables. Le terme  $e$  est le terme d'erreur.

Malgré toutes les restrictions applicables, quelques conclusions intéressantes peuvent être tirées de l'exercice:

- Le nombre de rejets dans tous les modèles ajustés est proportionnel au carré de  $v$ . Il y a donc symétrie autour de  $v = 0$ . La croissance pourrait être d'un ordre encore plus élevé que le carré.
- Il n'y a pas d'interaction observable entre les facteurs que sont les paramètres  $\alpha$  et  $v$ .
- Le nombre de rejets pour une petite taille d'échantillon est plus influencé par la valeur de  $\alpha$  alors que pour une grande taille, il est plus influencé par la valeur de  $v$ .

### 4.3 Effet des échantillons sans solution sur les résultats

Il est raisonnable de penser que le rejet de plusieurs échantillons pour lesquels la méthode d'estimation ne donne pas de solution et leur remplacement influencent les résultats de la simulation.

L'estimation de  $v$  est la première affectée. Le domaine de valeurs possibles de l'estimateur de  $v$  est limité par la valeur de la statistique  $U$ . Cette statistique est variable d'un échantillon à l'autre, mais elle n'en demeure pas moins une contrainte à l'estimation. Il suffit de regarder l'histogramme des estimations de  $v$  avec une petite valeur de  $\alpha$  (figure 4.2). On y voit que la distribution y apparaît tronquée de ses valeurs les plus élevées. La restriction imposée par la valeur de  $U$  biaise l'estimation de  $v$  et réduit sa variance. À la section 3.6, il a été vu comment la mauvaise estimation d'un paramètre se répercute sur les autres.

Un point de vue plus général peut être adopté pour étudier la question. On peut comparer les résultats entre des lois qui ont le plus grand nombre de rejets ( $\alpha = 10$  et  $v = \pm 8$ ; figure 4.4) et la loi de paramètres  $\alpha = 0,1$  et  $v = 0$  pour laquelle aucun rejet n'est enregistré (figure 4.3). L'échelle du biais et de l'erreur quadratique moyenne des estimateurs dépend surtout de la valeur des paramètres et ne doit pas être prise en compte. Ce sont les formes des courbes en fonction de la taille de l'échantillon qui sont révélatrices. Il est visible que la décroissance est beaucoup moins rapide et plus irrégulière pour les lois enregistrant plus de rejets que pour la loi sans problème d'estimation. Par contre, pour le biais et l'erreur quadratique moyenne des quantiles (qui ne sont pas présentées ici) les courbes sont conformes à ce qui est observé typiquement dans cette simulation.

Rappelons que lorsqu'il y a des rejets, les petits échantillons sont rejetés plus fréquemment que les grands (voir section 4.2). On peut avancer que le biais et l'erreur quadratique moyenne sur les paramètres sont moins élevés pour les tailles d'échantillons plus petites que ce qu'ils seraient si aussi peu d'échantillons de petite taille étaient rejetés que d'échantillons de grande taille. Le remplacement d'un plus grand nombre d'échantillons entraînerait donc une réduction du biais et de l'erreur quadratique moyenne. Cette explication contredit partiellement celle donnée en début de section, et sa validité est mise en doute si on tient compte que les valeurs des paramètres influent beaucoup sur les résultats. D'ailleurs, on peut difficilement distinguer l'effet de la valeur des paramètres de celui des échantillons sans solution. Une manière d'éviter la difficulté consisterait à dire que puisque le nombre de rejets

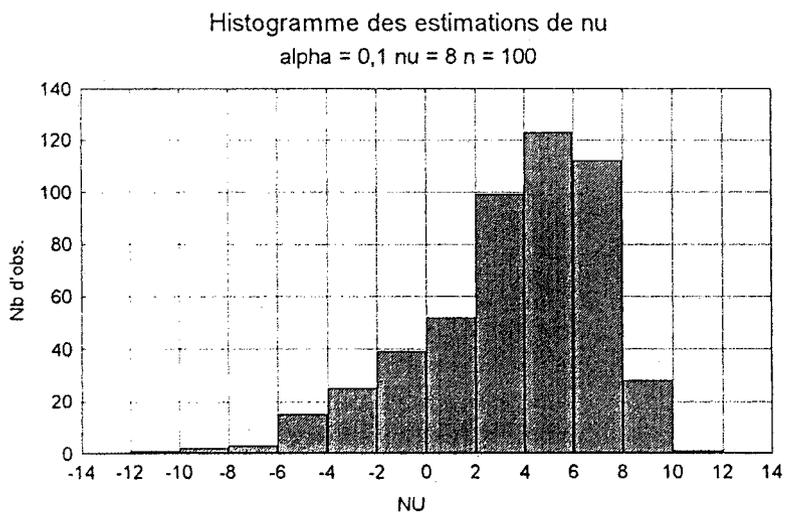
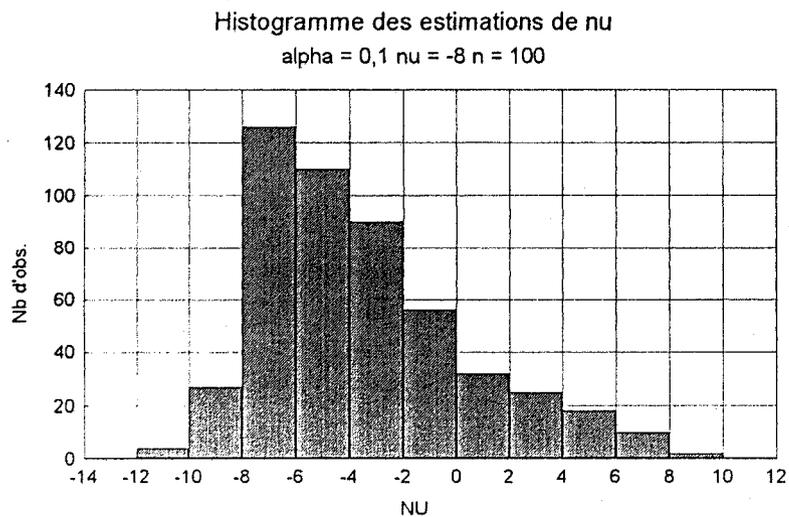


Figure 4.2 Distribution de l'estimateur de nu lorsque alpha = 0,1 et nu = 8 ou -8

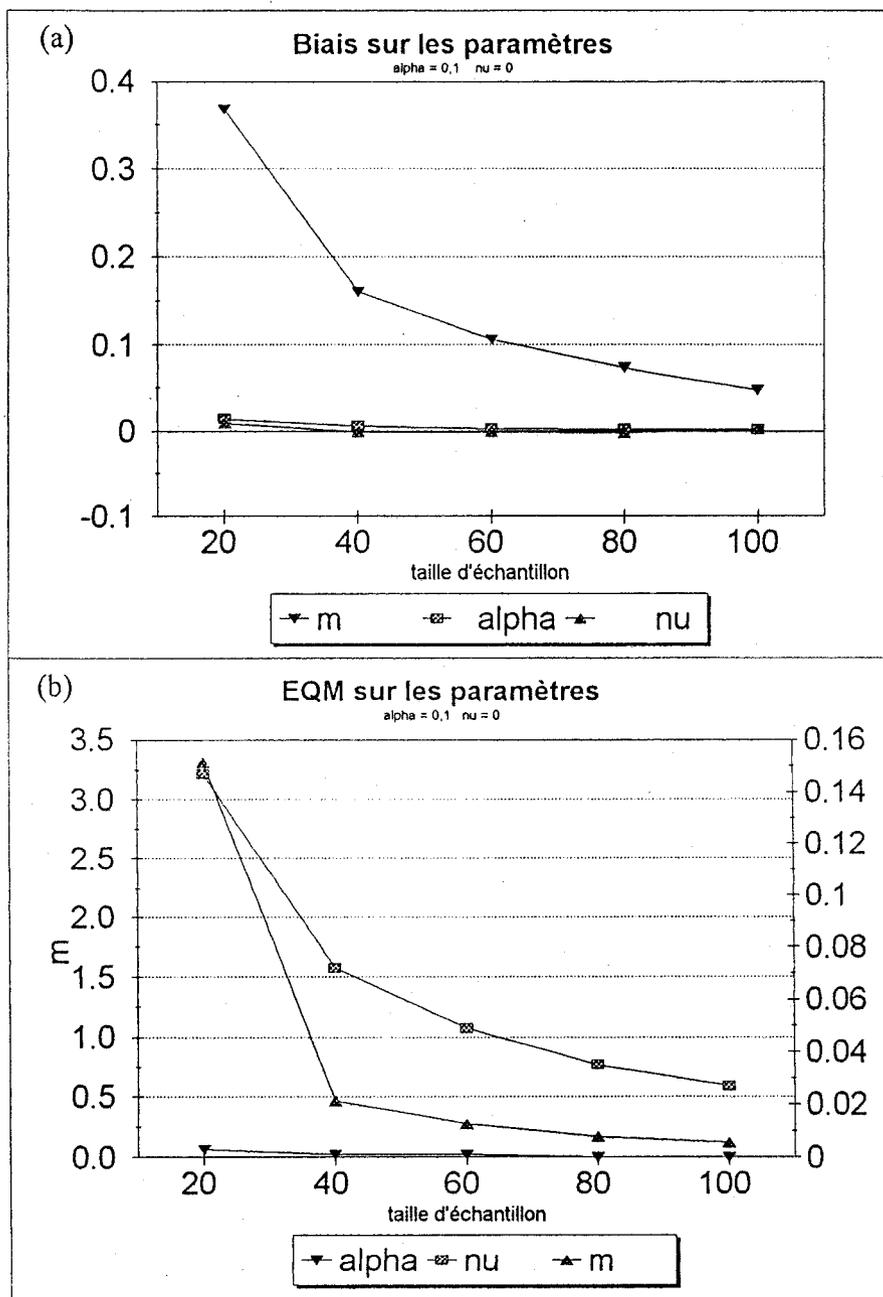


Figure 4.3 Biais et EQM sur les paramètres lorsque alpha = 0,1 et nu =

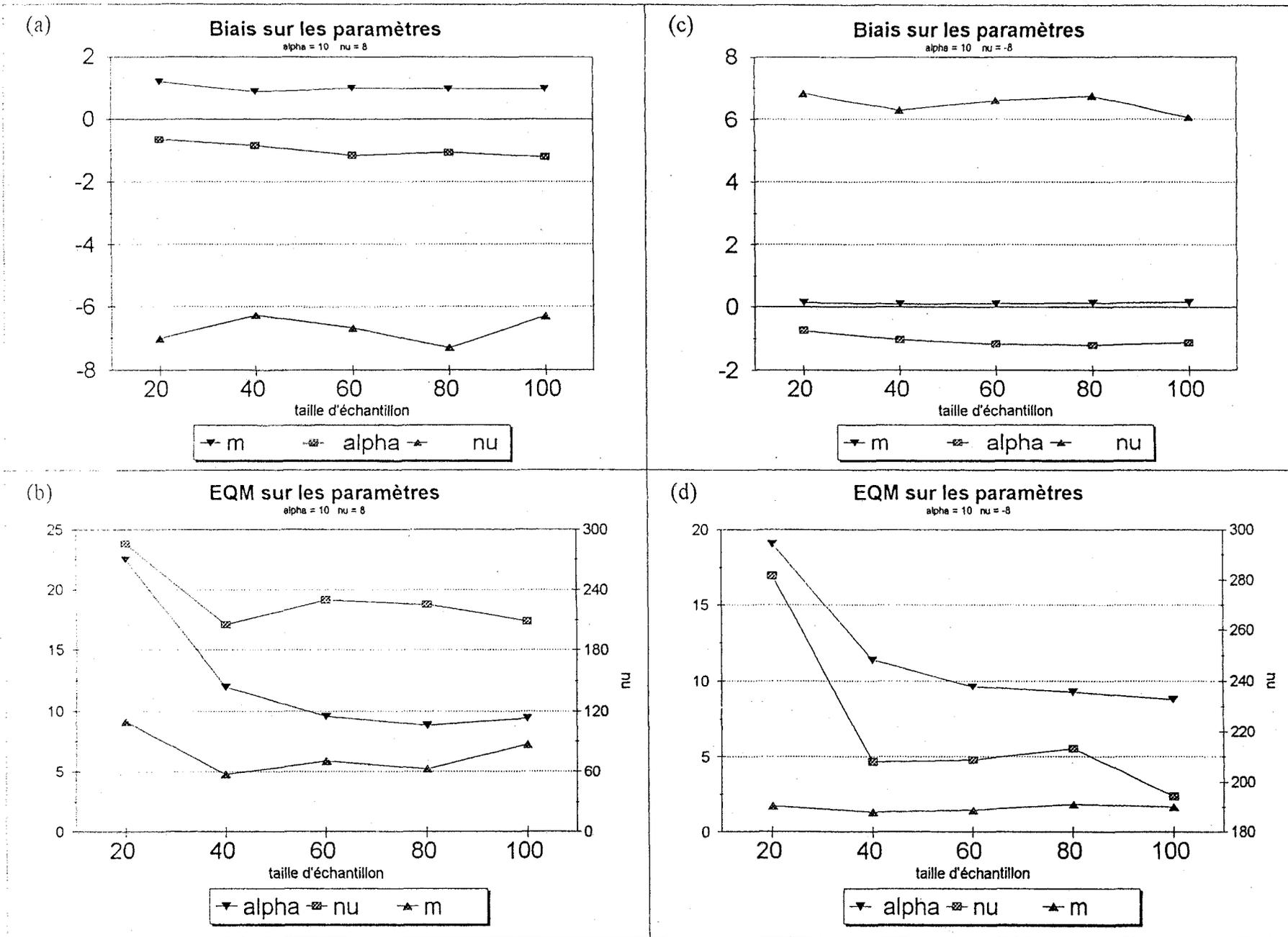


Figure 4.4 Biais et EQM sur les paramètres lorsque  $\alpha = 10$  et  $\nu = 8$  ou  $-8$

dépend lui-même de la valeur des paramètres, ce dernier facteur est en bout de ligne le seul à influencer les résultats.

#### **4.4 Influence de solutions approximatives sur les résultats**

Les solutions approximatives sont un autre problème qui surgit lors de l'estimation. Il ne s'agit pas cette fois de l'absence de solution pour une raison théorique. Une solution est bel et bien obtenue et le problème est numérique.

Les équations du maximum de vraisemblance sont construites en égalant les moyennes théoriques aux moyennes échantillonnales. Normalement, la résolution du système d'équations de la loi de Halphen type A devrait faire en sorte qu'en substituant aux paramètres leurs estimations dans l'expression des moyennes théoriques, celles-ci soient égales aux moyennes échantillonnales (Les statistiques exhaustives). Or, comme la résolution du système se fait numériquement, l'égalité n'est jamais parfaite, et il arrive que l'écart soit important. Quand le seuil de tolérance fixé à un dix-millième de la moyenne échantillonnale est dépassé, la solution est qualifiée d'approximative.

L'influence des valeurs des paramètres  $\alpha$  et  $\nu$  sur la fréquence de ces solutions approximatives n'a pas été étudiée en détail, mais il semble que la fréquence augmente avec  $\alpha$  et n'est pas influencée par la valeur de  $\nu$ . Les solutions approximatives représentent entre 0 et 20% des estimations selon les valeurs des paramètres. Elles sont plus fréquentes avec les petits échantillons qu'avec les grands.

Les solutions approximatives sont les estimations extrêmes en ce sens qu'elles sont dans les queues inférieures et supérieures des distributions des estimations. Leur retrait de l'ensemble des estimations a pour effet de réduire la variance de manière plus ou moins importante (de moins de 1% à plus de 50%) dépendant du paramètre et de sa valeur. Le biais est très peu affecté quant à lui. Un exemple des différences entre le biais et la variance des estimations avec et sans les solutions approximatives est donné au tableau 4.3.

**Tableau 4.3.** Effet du retrait des solutions approximatives sur le biais et la varianceLoi de paramètres  $\alpha=3$  et  $\nu = 2$ 

Taille	Rapport du biais sans les solutions approximatives sur le biais avec ces solutions			Rapport de la variance sans les solutions approximatives sur la variance avec ces solutions		
	m	$\alpha$	$\nu$	m	$\alpha$	$\nu$
20	78.93%	74.52%	96.28%	42.53%	84.62%	76.90%
40	83.66%	82.08%	103.38%	49.74%	90.89%	91.24%
60	102.64%	94.23%	106.26%	99.85%	96.32%	96.59%
80	101.76%	95.77%	103.83%	100.01%	96.52%	98.62%
100	91.77%	95.52%	100.99%	68.14%	95.31%	97.80%

Lorsque la distribution est très décalée par rapport à la vraie valeur du paramètre, il se trouve que les solutions approximatives dans une des queues de la distribution sont les plus près de cette valeur (voir par exemple la figure 3.11). Il n'est donc pas nécessairement souhaitable de les rejeter *a priori*.

## 5 CONCLUSIONS ET RECOMMANDATIONS

---

### 5.1 Retour sur les résultats

Un des premiers objectifs de cette étude était de déterminer l'importance du biais dans l'estimation des paramètres et des quantiles de la loi de Halphen type A pour de petites tailles d'échantillons. Il a été observé que pour les tailles inférieures ou égales à 100, les estimateurs du maximum de vraisemblance des paramètres sont biaisés, et que ceux des quantiles le sont aussi mais dans une moindre mesure (cf. chapitre 3). La propriété de variance minimale d'un estimateur sans biais fonction des statistiques exhaustives ne peut donc pas s'appliquer, bien que la loi de Halphen type A appartienne à la classe des lois exponentielles (cf. section 1.1).

De manière générale, le biais sur les paramètres  $v$  et  $m$  est plus important que le biais sur  $\alpha$ , et il diminue lorsque la taille de l'échantillon augmente. Le biais sur  $\alpha$  est plutôt stable d'une taille d'échantillon à l'autre (cf. section 3.1).

Dans tous les cas considérés, le paramètre  $v$  est en moyenne estimé plus près de 0 que sa vraie valeur, ce qui donne un biais relatif négatif. Les bornes imposées au domaine des estimations admissibles de  $v$  expliquent cette situation (cf. section 4.3).

Les estimations du paramètre  $m$  passent de valeurs beaucoup trop petites à des valeurs beaucoup trop grandes. Elles suivent de près la moyenne arithmétique des observations, ce qui ne devrait pas être le cas si les paramètres  $\alpha$  et  $v$  étaient correctement estimés. Cela fait partie d'un mécanisme de compensation impliquant autant les paramètres de forme que le paramètre d'échelle  $m$ , et qui permet que la forme de la loi ajustée reste près de celle de la vraie loi, malgré des estimations biaisées des paramètres (cf. section 3.6). Il en résulte des estimateurs des quantiles beaucoup moins biaisés que ceux des paramètres. Ce biais oscille d'une taille d'échantillon à l'autre sans décroître de façon marquée. On remarque que lorsque  $v$  est supérieur à 0, le biais des grands quantiles est positif et celui des petits quantiles négatif. Leurs signes s'inversent quand  $v$  est inférieur à 0 (cf. section 3.2).

En plus du biais, il faut se préoccuper de la variance lorsqu'on estime une quantité. Les deux sont combinés dans une seule mesure : l'erreur quadratique moyenne. L'observation que le biais sur les quantiles ou les paramètres a peu de poids comparé à la variance dans l'erreur quadratique moyenne, sauf en cas de biais extrême, vient relativiser l'importance du biais (cf. section 3.3).

L'erreur quadratique moyenne des estimateurs des paramètres dépasse de beaucoup celle des estimateurs des quantiles. Avec la grande majorité des lois de Halphen type A simulées, l'erreur quadratique moyenne décroît avec l'augmentation de la taille de l'échantillon pour tous les paramètres et tous les quantiles. Elle est en général plus grande sur les paramètres  $m$  et  $v$  que sur  $\alpha$ . Enfin, plus un quantile est grand, plus son erreur quadratique moyenne, toutes proportions gardées, est grande (cf. section 3.1).

Il était déjà connu avant de commencer cette étude que les équations du maximum de vraisemblance ne peuvent pas toujours être résolues, et qu'alors la loi Gamma ou Gamma inverse doit être substituée à la loi de Halphen type A. La simulation a permis d'apprendre que de telles absences de solution sont très fréquentes. Cette fréquence est croissante avec  $\alpha$ , ce qui est paradoxal puisque c'est lorsque  $\alpha$  tend vers 0 que la loi de Halphen type A tend vers la loi Gamma ou Gamma inverse (cf. section 4.1). La fréquence des absences de solution est aussi croissante avec  $|v|$ , ce qui était attendu à cause de l'origine même des rejets, soit le dépassement de la borne à l'estimation de  $v$ .

Lorsqu'une solution existe, les moyennes théoriques doivent être égalées aux moyennes échantillonales en ayant recours à des méthodes numériques. Certaines solutions sont qualifiées d'approximatives parce que la différence entre les moyennes théoriques calculées avec les estimations des paramètres et les moyennes expérimentales dépasse un seuil de tolérance. Les estimations en question sont aux extrémités inférieures et supérieures des distributions des estimateurs (cf. section 4.4).

## 5.2 Perspectives de recherche

Les quelques paragraphes précédents constituent un survol des principaux résultats. À l'intérieur de la durée limitée du stage, et parce qu'il s'agit d'une première étude par simulation de l'estimation des paramètres de la loi de Halphen type A, de nombreuses données n'ont pu être étudiées et de nombreux résultats n'ont pu être analysés. Tout au long du rapport, plusieurs questions ont été soulevées et plusieurs pistes laissées inexplorées. Une liste de sujets pouvant être approfondis lors d'études ultérieures peut ainsi être dressée.

- Le choix des valeurs à donner aux paramètres  $\alpha$  et  $v$  lors de la simulation a été motivé par une étude des coefficients de variation et d'asymétrie de la loi de Halphen type A (cf. section 2.2). Les relations entre les résultats et ces coefficients ont été examinées pour constater que les résultats extrêmes ne sont pas nécessairement associés à des valeurs de

Cs et Cv extrêmes. Les valeurs des paramètres jouent beaucoup plus sur la qualité des estimations. Une étude systématique des résultats en fonction des Cs et Cv pourrait nous en apprendre davantage.

- Les coefficients de variation et d'asymétrie échantillonnaires sont calculés pour chaque réplicat. Après l'ajustement d'une loi de Halphen type A, les coefficients théoriques de la loi ajustée sont calculés en fonction des estimations des paramètres. La comparaison des valeurs échantillonnaires et théoriques serait particulièrement utile pour juger si le choix d'une loi peut se justifier par les valeurs des Cs et Cv des échantillons.
- Il a été mis en évidence que les estimateurs du maximum de vraisemblance des paramètres de la loi de Halphen type A sont biaisés. Une méthode permettant de réduire le biais comme la méthode du Jackknife pourrait être appliquée. L'effet de cette méthode sur la variance et le temps de calcul nécessaire devront être pris en compte pour juger des résultats obtenus.
- Il a été souligné que le signe du biais des quantiles s'inverse lorsque le signe de  $v$  change. La raison de ce comportement reste à expliquer.
- Le programme de traitement des échantillons simulés calcule la variance asymptotique pour chaque estimation d'un paramètre ou d'un quantile. La présentation à la section 3.4 de la variance asymptotique pour les valeurs théoriques des paramètres a révélé l'imprécision de ces calculs. Lorsque leur fiabilité sera assurée, un important matériel deviendra disponible. En plus de la comparaison des variances asymptotiques et expérimentales, la précision des intervalles de confiance asymptotiques estimés devrait faire l'objet d'une étude.
- Les corrélations expérimentales entre les estimateurs des paramètres sont souvent fortes, tel qu'il a été observé à la section 3.5. Une analyse plus approfondie de ces statistiques est encore à faire. Il serait intéressant de les comparer aux valeurs asymptotiques théoriques. Cela permettrait entre autres de constater si l'estimation modifie la corrélation entre les estimateurs des paramètres. Le signe des corrélations est aussi un point à étudier.
- L'impossibilité de résoudre les équations du maximum de vraisemblance implique qu'une loi Gamma ou Gamma inverse peut être plus appropriée que la loi de Halphen de type A. On peut envisager d'ajuster une de ces lois, selon le cas, aux échantillons sans solution. Les quantiles de ces lois pourraient même être comparés à ceux des lois de Halphen type

A, en gardant cependant à l'esprit que l'erreur d'estimation est moindre avec deux paramètres qu'avec trois.

- La méthode du maximum de vraisemblance a été privilégiée parce qu'elle donne un estimateur de variance minimale parmi les estimateurs sans biais. Cependant, du fait que les estimateurs obtenus pour la loi de Halphen type A sont biaisés, les propriétés d'optimalité des estimateurs qui sont fonction des statistiques exhaustives ne sont plus garanties. On peut donc envisager d'employer une méthode d'estimation basée sur les moments afin de comparer ses performances à celles de la méthode du maximum de vraisemblance. Éventuellement, une telle méthode pourrait constituer une alternative pour l'estimation des paramètres de la loi de Halphen type A.

### 5.3 Recommandations

Les résultats qui ont pu être tirés de la simulation sont très diversifiés et l'analyse a touché un grand nombre d'aspects de ces résultats. Cela ne se prête pas à des recommandations globales, mais plutôt à des suggestions précises.

- La maximisation de la fonction de vraisemblance logarithmique  $\log L(\nu, \alpha, m)$  avec les trois paramètres est une opération très lourde requérant un temps de calcul relativement long. Il se peut aussi que le maximum de cette fonction se situe en dehors de l'intervalle des valeurs admissibles pour  $\nu$ . Il est à noter que cela se produit surtout quand la fonction est aplatie et que son maximum est instable et peu proéminent. En revanche, la maximisation partielle de la fonction  $\log L(\nu, \alpha, m)$  pour  $\alpha$  et  $m$  est simple et rapide. Lorsque la valeur de  $\nu$  est connue grossièrement, il est recommandé de fixer  $\nu$  à cette valeur et d'estimer seulement  $\alpha$  et  $m$ . Il est aussi envisageable de répéter l'opération avec plusieurs valeurs de  $\nu$  voisines et de comparer l'adéquation obtenue, par exemple par des méthodes graphiques.
- L'estimation du paramètre  $\alpha$  est toujours la première à être effectuée, et qu'elle consiste à résoudre une équation non-linéaire. Aussi, l'amélioration de la précision de l'estimation des paramètres doit commencer par l'amélioration de la précision de l'estimation de  $\alpha$ .
- Le calcul des variances et covariances asymptotiques manque de précision. Le problème réside principalement dans l'inversion de la matrice d'information de Fisher qui doit être revue.

- L'obtention ou non d'une estimation des paramètres dépend de la dérivée de la fonction de vraisemblance logarithmique partiellement maximisée aux bornes de l'intervalle des valeurs admissibles pour  $v$ . Étant données les constatations faites sur les fréquences de rejets, il importe de s'assurer que la dérivée est correctement calculée pour que des erreurs numériques ne soient pas la cause des absences de solution.

## 6 BIBLIOGRAPHIE

---

- Barndorff-Nielsen, O., Blaesild, P., et Halgreen, C. (1977). First hitting time models for the generalised inverse gaussian distribution. *Stochastic Processes and their Applications*, 7: 49-54.
- Bickel, P.J. et Doksum, K.A. (1977). *Mathematical Statistics*. Holden-Day Inc., Oakland, CA, 492 p.
- Bobée B. et F. Ashkar (1991). *The Gamma and Derived Distributions Applied in Hydrology*. Water Resources Publications, Littleton, CO, Yevjevitch (Ed.) : 202 p.
- Cohen, A. C. and B. J. Whitten (1980). Estimation in the Tree-Parameter Lognormal Distribution. *Journal of the American Statistical Association*, 75: 399-404.
- D'Agostino, R.B. and M.A. Stephens (1986). *Goodness-of-fit Techniques*. Marcel-Dekker, New York, NY, 560 p.
- Essadaoui, M., L. Perreault et B. Bobée (1995). *Générateur d'échantillons pour la loi de Halphen de type A*. Rapport de Recherche, INRS-Eau (en préparation)
- Halphen, E. (1941). Sur un nouveau type de courbe de fréquence. *Compte rendus de l'académie des Sciences, Tome 213*, Pp.633-635.
- Hoshi, K., J. R. Stedinger and S. J. Burges (1984) Estimation of Log-normal Quantiles: Monte-Carlo Results and First-order Approximations. *Journal of Hydrology*, 71: 1-30.
- Jørgensen B. (1982). *Statistical properties of the generalized inverse gaussian distribution*. Lecture Notes in Statistics, No 9.
- Kappenman, R. F. (1984). Estimation for the three-parameter Weibull, lognormal, and gamma distributions. *Computational Statistics & Data Analysis*, 3: 11-23.
- Morlat G. (1956). Lois de probabilité de Halphen. *Revue de la Statistique Appliquée*, Pp. 21-43.

Perreault, L., Bobée, B. et Legendre, P. (1994). *Rapport Général du Logiciel AJUSTE-II : Théorie et Application*. Rapport de Recherche R-421, INRS-Eau.

Perreault L. et Bobée, B. (1995). *Les lois de Halphen: Théorie et Application en Hydrologie*. Rapport de Recherche, INRS-Eau (en préparation).

Watson, G.N. (1966). *A treatise on the theory of Bessel functions*. Cambridge University Press.

## ANNEXE A

### Algorithme du traitement des échantillons simulés

---

#### Algorithme général

Début

(1<sup>re</sup> partie)

Génération de  $kN$  échantillons de  $n_{\max}$  obs. d'une loi de Halphen type A de param.  $\alpha, m$  et  $\nu$  fixés

Écriture des échantillons dans un fichier

(2<sup>e</sup> partie)

Calculer et écrire dans un fichier les variances/covariances asymptotiques théoriques des estimateurs des param.

Calculer et écrire dans un fichier les quantiles théoriques sélectionnés

Calculer et écrire dans un fichier les variances théoriques des quantiles sélectionnés

Tant que  $\min(ht[i]) \leq N$

  Tant que  $i < \text{NbTailles}$

    Lire un échantillon de taille  $n[i]$

    Si  $ht[i] \leq N$

      Estimer les paramètres et écrire les estimations dans un fichier

      Estimer les var./covar. asympto. des estimateurs des param. et les écrire dans un fichier

      Estimer les quantiles sélectionnés et écrire les estimations dans un fichier

      Estimer les variances/covariances asympto. des quantiles et les écrire dans un fichier

      Mettre à jour les moyennes et erreurs quadratiques moyennes des estimations des param.

      Mettre à jour les moyennes et erreurs quadratiques moyennes des estimations des quantiles

    Fin si

    Incrémenter  $ht[i]$

  Incrémenter  $i$

  Fin Tant que  $i$

Fin Tant que  $h$

Écrire les moyennes et erreurs quadratiques moy. des estimations des paramètres dans un fichier

Écrire les moyennes et erreurs quadratiques moy. des estimations des quantiles dans un fichier

Fin

**Algorithme détaillé de la 2<sup>e</sup> partie****Début**

Lire les noms de fichiers à utiliser ou à créer

Lire les valeurs des paramètres  $\theta[0]$ ,  $\theta[1]$  et  $\theta[2]$ 

Ouvrir en écriture les fichiers estimation, var\_estimee, quantile, var\_est\_xt et fich\_err

Ouvrir en lecture le fichier donnee

Écriture des valeurs des paramètres dans estimation

Calculer les variances/covariances asymptotiques des paramètres

Écrire les variances/covariances des paramètres dans le fichier var\_theo

Calculer les quantiles  $xt[10]$ ,  $xt[8]$ ,  $xt[6]$ ,  $xt[4]$  et  $xt[2]$ 

Écrire les valeurs des quantiles sélectionnés dans quantile

Calculer les variances/covariances asymptotiques des quantiles  $xt[10]$ ,  $xt[8]$ ,  $xt[6]$ ,  $xt[4]$  et  $xt[2]$ 

Écrire les variances/covariances des quantiles sélectionnés dans var\_est\_xt

Pour  $i$  de 0 à  $NB\_XT\_SELECT - 1$      $xt\_select[i] = xt[10-2*i]$ **Fin**Initialiser  $moy[][]$ ,  $varexp[][][]$ ,  $moy\_xt[][]$  et  $varexp\_xt[][][]$  à 0 et  $ht[]$  et  $h$  à 1Tant que  $\min(ht[i]) \leq N$  et qu'on peut lire une ligne de donnee    Initialiser  $j$  à 0

Initialiser l'échantillon

    écrire le numéro d'échantillon  $h$  dans fich\_err

Lire et afficher l'entête de l'échantillon

    Pour  $i$  de 0 à  $NB\_TAILLES - 1$         Tant que  $j < n[i]$  et fichier pas fini        Lire une observation de l'échantillon  $h$  dans donnee

Si fin d'échantillon alors

 $i = NB\_TAILLES$ 

interrompre tant que

Fin si

    Incrémenter  $j$ Fin Tant que  $j$

```

Si ht[i] <= N
  Calculer les estimations m, alpha et nu des paramètres
  Écrire m, alpha et nu dans le fichier estimation
  Calculer les variances/covariances asymptotiques de m, alpha et nu
  Écrire les variances/covariances de m, alpha et nu dans le fichier var_estimee
  Calculer les quantiles xt[10], xt[8], xt[6], xt[4] et xt[2]
  Écrire les valeurs des quantiles sélectionnés dans quantile
  Calculer les variances/covariances asympto. des quantiles xt[10], xt[8], xt[6], xt[4] et xt[2]
  Écrire les variances/covariances des quantiles sélectionnés dans var_est_xt
  theta_est[0] = m
  theta_est[1] = alpha
  theta_est[2] = nu
  Pour k de 0 à NB_PARAMS - 1
    moy[k][i] = moy[k][i] + (theta_est[k] - moy[k][i]) / ht[i]
    Pour l de 0 à k
      varexp[l][k][i] = varexp[l][k][i] + (theta_est[k]-theta[k])*(theta_est[l]-theta[l]) / N
    Fin l
  Fin k
  Pour k de 0 à NB_XT_SELECT - 1
    moy_xt[k][i] = moy_xt[k][i] + (xt[10-2*k] - moy_xt[k][i]) / ht[i]
    varexp_xt[l][k][i] = varexp_xt[l][k][i] + (xt[10-2*k] - xt_select[k])*(xt[10-2*k] -
xt_select[k]) / N
  Fin k
  incrémenter ht[i]
Fin si
Afficher fin de l'échantillon
Fin i
incrémenter h
Fin tant que min(ht[i])
Fermer donnee, estimation, var_estimee, quantile, var_est_xt et fich_err

```

```
Ouvrir en écriture esperance, var_exp, esper_xt et var_exp_xt
Pour i de 0 à NB_TAILLES - 1
  Pour k de 0 à NB_PARAMS - 1
    Écrire moy[k][i] dans le fichier esperance
    Pour l de 0 à NB_PARAMS - 1
      Écrire varexp[l][k][i] dans le fichier var_exp
    Fin l
  Fin k
  Pour k de 0 à NB_XT_SELECT - 1
    Écrire moy_xt[k][i] dans esper_xt
    Écrire varexp_xt[k][i] dans var_exp_xt
  Fin k
Fin i
Fermer esperance, var_exp, esper_xt et var_exp_xt
Fin
```

## ANNEXE B

### Code du programme *SIMULA*

---

#### Module Simula.cpp

```
/* Simulation d'une lois de Halphen type A pour l'étude de ses estimateurs
du maximum de vraisemblance */
```

```
#include <stdio.h>
#include <string.h>
#include <time.h>
#include "halphena.h"
#include "echant.h"
#include "simula.h"

int main (void)
{
char chemin_entree[80], chemin_result[80], Ligne[80];
char *fin_nom;
BOOLEEN fini;
time_t debut;
FILE *donnee, *estimation, *esperance, *var_estimee, *var_exp,
    *quantile, *esper_xt, *var_est_xt, *var_exp_xt;
unsigned int h, i, j, k, l, nb_qui_manque, N, ht[NB_TAILLES];
FLOAT theta[NB_PARAMS], theta_est[NB_PARAMS], xt_select[NB_XT_SELECT],
    moy[NB_PARAMS][NB_TAILLES],
varexp[NB_PARAMS][NB_PARAMS][NB_TAILLES],
    moy_xt[NB_XT_SELECT][NB_TAILLES], varexp_xt[NB_XT_SELECT][NB_TAILLES];

// Présentation du programme
puts ("\tProgramme de traitement d'échantillons d'une loi de Halphen type A\n");
// Lecture des noms et répertoires des fichiers à utiliser ou à créer
puts ("Entrez le nom du fichier de données précédé de son répertoire: ");
scanf ("%s", chemin_entree);
puts ("\nSpécifiez le nom général des fichiers de sortie précédé du répertoire");
puts ("(sans les extensions qui seront ajoutées par le programme) : ");
scanf ("%s", chemin_result);
```

```

// Lecture du nombre de répliquats
printf ("\nEntrez le nombre de répliquats : ");
scanf ("%u", &N);
// Lecture des valeurs des paramètres de la loi simulée
printf ("\nEntrez la valeur du paramètre m : ");
scanf ("%lf", &theta[0]);
printf ("\nEntrez la valeur du paramètre alpha : ");
scanf ("%lf",&theta[1]);
printf ("\nEntrez la valeur du paramètre nu : ");
scanf ("%lf",&theta[2]);
debut = time(NULL);

/* Ouverture des fichiers */
donnee = fopen (chemin_entree, "rt");
verifie (donnee != NULL, "ouverture fichier de données");
rewind (donnee);
// Validation du nom des fichiers de résultats
fin_nom = strchr (chemin_result, '.');
if (fin_nom == NULL)
    { strcat (chemin_result, ".par");
      fin_nom = strchr (chemin_result, '.');
    }
else strcpy (fin_nom, ".par");
estimation = fopen (chemin_result, "wt");
verifie (estimation != NULL, "ouverture fichier des estimations");
strcpy (fin_nom, ".var");
var_estimee = fopen (chemin_result, "wt");
verifie (var_estimee != NULL, "ouverture fichier des variances estimées");
strcpy (fin_nom, ".xt");
quantile = fopen (chemin_result, "wt");
verifie (quantile != NULL, "ouverture fichier des quantiles");
strcpy (fin_nom, ".vxt");
var_est_xt = fopen (chemin_result, "wt");
verifie (var_est_xt != NULL, "ouverture fichier des variances estimées des quantiles");
strcpy (fin_nom, ".err");
fich_err = fopen (chemin_result, "wt");
verifie (fich_err != NULL, "ouverture fichier des messages d'erreur");

// Écriture des entêtes de fichiers
fputs ("Ech Taille  m  alpha  nu\n", estimation);
fputs ("Ech Taille var(m) cov(m,alpha) var(alpha) cov(m,nu) cov(alpha,nu) var(nu)\n",
      var_estimee);
fputs ("Ech Taille  x2  x5  x20  x100  x1000\n", quantile);
fputs ("Ech Taille var(x2) var(x5) var(x20) var(x100) var(x1000)\n",
      var_est_xt);

```

```
/* Assignation des éléments du vecteur des valeurs des paramètres
   aux estimés des paramètres */
ha.m = theta[0];
ha.alpha = theta[1];
ha.nu = theta[2];

// Écriture des valeurs des paramètres en première ligne de estimation
fprintf (estimation, "theorique%8.3f%8.3f%8.3f\n", ha.m, ha.alpha, ha.nu);

/* Calcul des variances/covariances asymptotiques théoriques
   et écriture en première ligne de var_estimee */
ech.n = 100;
ha.calcule_bessel (ha.alpha, ha.nu);
fprintf (var_estimee, "theorique");
traite_var_asympto (var_estimee);

// Calcul des quantiles théoriques et écriture en première ligne de quantile
fprintf (quantile, "theorique");
traite_xt_select (quantile);

/* Calcul des variances des quantiles théoriques et écriture
   en première ligne de var_xt_estimee */
fprintf (var_est_xt, "theorique");
traite_var_xt (var_est_xt);

/* Assignation des quantiles théoriques aux éléments du vecteur de quantiles
   sélectionnés */
for (i = 0; i < NB_XT_SELECT; i++)
  xt_select[i] = ha.xt[10 - 2*i];

/*Initialisation des matrices de mesures expérimentales */
for (i = 0; i < NB_TAILLES; i++)
  {
  for (j = 0; j < NB_PARAMS; j++)
    {
    moy[j][i] = 0.0;
    for (k = 0; k < NB_PARAMS; k++)
      varexp[k][j][i] = 0.0;
    }
  for (j = 0; j < NB_XT_SELECT; j++)
    moy_xt[j][i] = varexp_xt[j][i] = 0.0;
  ht[i] = 1;
  }
}
```

```

// Affichage du temp de calcul
printf ("\nTemps de calcul initial : %5.0f secondes\n", difftime (time (NULL), debut));

/* Exécution des calculs de la simulation */
h = 1;
    /* Lecture de la ligne de titre de l'échant. */
while ( poursuivre(ht, N) && LIT_FICHER (Ligne, donnee) != NULL)
{
    // Initialisations
    j = nb_qui_manque = 0;
    fini = NON;
    ech.initialise();

    // Délimitation des échantillons dans le fichier d'erreur
    fprintf (fich_err, "\nEchantillon no. %u\n", h);

    printf ("\n%s",Ligne); // Affichage du titre
    LIT_FICHER(Ligne, donnee ); // Lecture de l'ordre, on ne fait rien avec.

    for (i = 0; i < NB_TAILLES; i++)
    {
        while (j < n[i]&& !feof(donnee))
        { // Lecture d'une observation
            LIT_FICHER(Ligne, donnee );
            if ( Ligne[0] == '.' && Ligne[1] == '.' )
                {i = NB_TAILLES;
                 fini = OUI; break; }
            // Découpage de la ligne lue
            nb_qui_manque += ech.decoupe_ligne(Ligne);
            j++;
        }
        if (ht[i] <= N)
        {
            prob_ajuste = 0;
            ha.estimate_parametres();
            if (!prob_ajuste)
            {
                // Écriture des paramètres estimés
                fprintf (estimation, "%3u %4u %8.3f%8.3f%8.3fn",
                    h, ech.n, ha.m, ha.alpha, ha.nu);
                // Calcul des variances asymptotiques des estimateurs des paramètres
                fprintf (var_estimee, "%3u %4u ", h, ech.n);
                traite_var_asympto (var_estimee);
            }
        }
    }
}

```

```

// Calcul des quantiles sélectionnés à partir des estimés des paramètres
fprintf (quantile, "%3u %4u ", h, ech.n);
traite_xt_select (quantile);
// Calcul des variances asymptotiques des estimateurs des quantiles
fprintf (var_est_xt, "%3u %4u ", h, ech.n);
traite_var_xt (var_est_xt);
// Assignation des estimés aux éléments du vecteur de paramètres estimés
theta_est[0] = ha.m;
theta_est[1] = ha.alpha;
theta_est[2] = ha.nu;
/* Mise à jour des moyennes et variances expérimentales des estimateurs
des paramètres avec les estimés tirés de l'échantillon */
for (k = 0; k < NB_PARAMS; k++)
{
moy[k][i] += (theta_est[k] - moy[k][i])/ht[i];
for (l = 0; l <= k; l++)
varexp[l][k][i] += (theta_est[l] - theta[l])*(theta_est[k] - theta[k])/N;
}
/* Mise à jour des moyennes et variances expérimentales des estimateurs
des quantiles avec les estimés tirés de l'échantillon */
for (k = 0; k < NB_XT_SELECT; k++)
{
moy_xt[k][i] += (ha.xt[10 - 2*k] - moy_xt[k][i])/ht[i];
varexp_xt[k][i] += (ha.xt[10 - 2*k] - xt_select[k])
* (ha.xt[10 - 2*k] - xt_select[k])/N;
}
ht[i] ++;
} // fin si
} // fin si
} // fin pour
// Message d'avertissement si des observations manquent
if (nb_qui_manque > 0) printf ("%u observations n'ont pu être incluses dans \
l'échantillon.\n",nb_qui_manque);
// Message de fin de l'échantillon
if (!fini) LIT_FICHER(Ligne, donnee);
verifie ( (Ligne[0] == '-' && Ligne[1] == '-'), "fin d'échantillon" );
printf ("%s",Ligne);
printf ("Temps de calcul cumulatif : %5.0f secondes\n",
difftime(time(NULL),debut));
h++;
}

```

```

/* fermeture des fichiers ouverts et ouverture de nouveaux fichiers */
fclose (donnee);
fclose (estimation);
fclose (var_estimee);
fclose (quantile);
fclose (var_est_xt);
fclose (fich_err);

strcpy (fin_nom, ".moy");
esperance = fopen (chemin_result, "wt");
strcpy (fin_nom, ".exp");
var_exp = fopen (chemin_result, "wt");
strcpy (fin_nom, ".mxt");
esper_xt = fopen (chemin_result, "wt");
strcpy (fin_nom, ".xxt");
var_exp_xt = fopen (chemin_result, "wt");

// Écriture des moyennes, variances et covariances expérimentales
if (esperance != NULL && var_exp != NULL && esper_xt != NULL && var_exp_xt != NULL)
{ // Écriture des entêtes de fichiers
  fputs ("Taille m alpha nu\n", esperance);
  fputs ("Taille var(m) cov(m,alpha) var(alpha) cov(m,nu) cov(alpha,nu) var(nu)\n",
    var_exp);
  fputs ("Taille x2 x5 x20 x100 x1000\n", esper_xt);
  fputs ("Taille var(x2) var(x5) var(x20) var(x100) var(x1000)\n", var_exp_xt);
  for (i = 0; i < NB_TAILLES; i++)
  {
    fprintf (esperance, "%4u ", n[i]);
    fprintf (var_exp, "%4u ", n[i]);
    fprintf (esper_xt, "%4u ", n[i]);
    fprintf (var_exp_xt, "%4u ", n[i]);
    for (k = 0; k < NB_PARAMS; k++)
    {
      fprintf (esperance, "%8.3f ", moy[k][i]);
      for (l = 0; l <= k; l++)
        fprintf (var_exp, "%10.3f ", varexp[l][k][i]);
    }
  }
  for (k = 0; k < NB_XT_SELECT; k++)
  {
    fprintf (esper_xt, "%8.3f ", moy_xt[k][i]);
    fprintf (var_exp_xt, "%10.3f ", varexp_xt[k][i]);
  }
}

```

```
fprintf (esperance, "\n");
fprintf (var_exp, "\n");
fprintf (esper_xt, "\n");
fprintf (var_exp_xt, "\n");
}
}
// Fermeture des fichiers ouverts
fclose (var_exp);
fclose (esperance);
fclose (var_exp_xt);
fclose (esper_xt);
return 1;
}

// Fonctions

void verifie (int condition, char *message)
{
/* Fonction vérifiant si une condition est remplie
Effet : interruption du programme si la condition n'est pas remplie */
if (condition) return;
else { fprintf(stderr, "ERREUR %s\n", message);
exit(1); }
}

void traite_var_asympto(FILE *fichier)
{
/* Fonction d'estimation et d'écriture des variances asymptotiques
Effet : écriture dans fichier */
unsigned int k,l;
ha.estimate_variance_param();
for (k = 1; k <= NB_PARAMS; k++)
for (l = 1; l <= k; l++)
fprintf (fichier, "%11.3f", ha.VC(k,l));
fprintf (fichier, "\n");
}

void traite_xt_select(FILE *fichier)
{
unsigned int i;
FLOAT dernp, // DERNIERE PROBABILITE AU NON DEPASSEMENT CALCULEE
dernq; // DERNIER QUANTILE CALCULE
```

```

// INITIALISATION DE LA DERNIERE PROB. ET DU DERNIER QUANTILE CALCULE
demp = 0;
dernq = 0;
// ESTIMATION DE LA DENSITE CUMULEE INVERSE POUR LES DIFFERENTES PROB.
AU DEP.
for (i = 10; i >= 2; i -= 2)
{
    ha.xt[i] = ha.cumulee_inverse(1-prob_dep[i],demp,dernq);
    fprintf (fichier, "%8.3f ", ha.xt[i]);
    if (ha.xt[i] == DONNEE_MANQUANTE)
        { demp = 0;
          dernq = 0; }
    else { demp = 1-prob_dep[i];
          dernq = ha.xt[i]; }
}
fprintf (fichier, "\n");
}

void traite_var_xt (FILE *fichier)
{
    unsigned int i;
    ha.estimate_variance_xt ();
    for (i = 10; i >= 2; i -= 2)
        fprintf (fichier, "%10.3f ", ha.var_xt[i]);
    fprintf (fichier, "\n");
}

BOOLEEN poursuivre (unsigned int *ht, unsigned int N)
{
    unsigned int i;
    for (i = 0; i < NB_TAILLES; i++)
        if (ht[i] <= N) return VRAI; // s'il reste des répliquats à traiter
    return FAUX; // sinon, la simulation est terminée
}

```

# ANNEXE C

## Manuel de l'utilisateur du programme *SIMULA*

### C-1 Utilité

Le programme *SIMULA* a pour fonction de traiter des échantillons simulés d'une loi de Halphen type A. Il effectue :

- L'estimation des paramètres;
- L'estimation des variances et covariances asymptotiques de ces paramètres;
- L'estimation de quantiles sélectionnés;
- L'estimation des variances asymptotiques de ces quantiles;
- Le calcul de diverses statistiques expérimentales.

### C-2 Programmation

*SIMULA* a été programmé en langage C++. Il est constitué d'un programme principal contenu dans le fichier *simula.cpp* et des modules du logiciel *AJUSTE-II* relatifs à la loi de Halphen type A. Le fichier exécutable *simula.exe* a été compilé avec Borland C 4.0 pour Windows. Il doit être lancé à l'intérieur de Windows. Il pourrait également être compilé pour DOS sans modification puisque aucune fonctionnalité graphique n'est utilisée.

### C-3 Modules utilisés

Pour compiler le programme, les modules de *AJUSTE-II* suivants sont nécessaires dans leur version originale :

- **bessel**
- **fam\_half**
- **halphena**
- **liste**
- **lois**
- **mathemat**
- **matrice**

En plus, il faut compiler des modules modifiés :

- Le module **echant** doit inclure la fonction *decoupe\_ligne* en plus des fonctions d'origine.
- Le module **erreur** doit comprendre la fonction *ecrit\_erreur*, qui admet deux arguments entiers lorsque XVT n'est pas défini.
- Le module **trad\_err**, conçu pour le programme *SIMULA*. Il est inclus dans **erreur.cpp**.

## C-4 Paramètres du programme

### C-4.1 Tailles des échantillons

Le programme s'exécute pour plusieurs tailles d'échantillons différentes. Il faut lui soumettre des échantillons de la taille maximale. Le programme lit le nombre d'observations correspondant à la plus petite taille, effectue les calculs puis ajoute les observations suivantes à l'échantillon lu pour atteindre la deuxième taille d'échantillon, et ainsi de suite jusqu'à la taille maximale.

Si un échantillon trop petit est soumis, les calculs s'effectueront seulement pour les tailles d'échantillons inférieures ou égale au nombre d'observations entrées. Si un échantillon trop grand est soumis, une erreur survient et le programme est interrompu.

Le nombre de tailles d'échantillons a été fixé à 5, soit 20, 40, 60, 80 et 100. Ces constantes se trouvent dans **simula.h**. Pour que des modifications de ces constantes dans le code prennent effet, il faut recompiler le programme.

### C-4.2 Nombre de réplicats

Le nombre de réplicats est entré par l'utilisateur au début de l'exécution du programme. Le programme s'exécute jusqu'à ce qu'il ait réussi exactement le nombre spécifié d'estimations des paramètres pour toutes les tailles d'échantillons ou que la fin du fichier soit atteinte. De cette manière, si on fournit plus d'échantillons que le nombre de réplicats spécifié, le programme puise dans les échantillons supplémentaires pour compenser les absences de solution.

### C-4.3 Valeurs des paramètres

Les valeurs des paramètres entrées au début de l'exécution du programme servent à calculer les valeurs théoriques des résultats. De fausses valeurs n'affectent pas les résultats de la simulation.

### C-4.4 Quantiles estimés

Les quantiles estimés sont les quantiles 0,5; 0,8; 0,95; 0,99 et 0,999 correspondant à des périodes de retour de 2, 5, 20, 100 et 1000 ans respectivement. Ils ont été sélectionnés parmi les quantiles estimés par *AJUSTE-II* contenus dans le vecteur de constantes *prob\_dep*. Pour modifier la sélection, il faut changer la fonction *traite\_xt\_select* dans *simula.cpp*.

### C-4.5 Fichier d'entrée

Le nom du fichier d'entrée contenant les échantillons simulés doit être donné par l'utilisateur avec le répertoire où il se trouve au début de l'exécution.

Ce fichier doit être un fichier texte (ASCII) et les échantillons doivent respecter le format *AJUSTE-II*. Toutefois, rien ne doit être inscrit après la ligne de fin d'échantillon ( $n + 3$ ). Les échantillons doivent se suivre sans espace entre eux. (La ligne de titre de l'échantillon suivant doit venir tout de suite après la ligne de fin d'échantillon).

### C-4.6 Fichiers de sortie

L'utilisateur doit spécifier un répertoire qui recevra les fichiers de sortie et un nom générique pour ceux-ci. Une extension y sera ajoutée pour chaque fichier produit. Si une extension est ajoutée par l'utilisateur, elle sera effacée et remplacée.

Les fichiers produits sont :

**fichier.par** : estimations des paramètres;

**fichier.var** : estimations des variances et covariances asymptotiques des estimateurs des paramètres;

**fichier.xt** : estimations des quantiles sélectionnés;

**fichier.vxt** : estimations des variances et covariances asymptotiques des estimateurs des quantiles sélectionnés;

Ces fichiers sont tous structurés de la même manière. La première ligne est une ligne d'entête, la deuxième affiche les résultats obtenus avec les valeurs théoriques des paramètres pour la taille d'échantillons maximale. Chacune des lignes suivantes contient les estimations pour une taille d'un échantillon. La première colonne de la ligne donne le numéro d'échantillon, la seconde le nombre d'observations considérées.

**fichier.moy** : moyennes expérimentales des estimations des paramètres;

**fichier.exp** : erreurs quadratiques moyennes des estimations des paramètres;

**fichier.mxt** : moyennes expérimentales des estimations des quantiles;

**fichier.xxt** : erreur quadratique moyenne des estimations des quantiles;

Les fichiers de cette deuxième série ont aussi une structure commune. La première ligne est encore une ligne d'entête. Les résultats sont donnés par taille d'échantillons sur les autres lignes. Chaque ligne commence en identifiant la taille d'échantillon.

**fichier.err** : messages d'erreur et d'avertissement.

Dans ce fichier sont reportés les messages expliquant les erreurs qui surviennent lors de l'estimation des paramètres et les messages avertissant que la solution est approximative. Ce sont les mêmes messages qui apparaissent à l'écran lors de l'exécution. Une ligne indiquant le numéro de l'échantillon précède les messages concernant cet échantillon, et ce pour tous les échantillons.