

Rapport Scientifique No 350

par

**Luc Perreault
Bernard Bobée**

**Loi généralisée des valeurs extrêmes
Propriétés mathématiques et statistiques
Estimation des paramètres et des
quantiles X_T de période de retour T**

Mars 1992

**INRS-Eau
Université du Québec
C.P. 7500
Sainte-Foy, Québec
G1V 4C7**

TABLE DES MATIERES

Liste des tableaux et figures.....	iii
INTRODUCTION	1
CHAPITRE 1 : CARACTÉRISTIQUES GÉNÉRALES DES LOIS DES VALEURS EXTREMES.....	2
1.1 Distributions des valeurs extrêmes	2
1.2 Distribution GEV	3
CHAPITRE 2 : PROPRIÉTÉS STATISTIQUES DE LA LOI GÉNÉRALISÉE DES VALEURS EXTREMES	8
2.1 Loi GEV avec $k = 0$ (EV1, loi Gumbel).....	8
2.2 Loi GEV avec $k \neq 0$ (EV2 et EV3).....	9
CHAPITRE 3 : ESTIMATION DES PARAMETRES DE LA LOI GEV.....	14
3.1 Loi GEV avec $k = 0$ (EV1, loi Gumbel).....	15
3.1.1 Méthode du maximum de vraisemblance.....	15
3.1.2 Méthode des moments.....	17
3.1.3 Méthode des moments pondérés	18
3.2 Loi GEV avec $k \neq 0$ (EV2 et EV3).....	20
3.2.1 Méthode du maximum de vraisemblance.....	20
3.2.2 Méthode des moments.....	24
3.2.3 Méthode des moments pondérés	28
CHAPITRE 4 : ESTIMATION D'UN ÉVÉNEMENT DE PÉRIODE DE RETOUR $T X_T$ ET CALCUL DE LA VARIANCE.....	32
4.1 Loi GEV avec $k = 0$ (EV1, loi Gumbel).....	32
4.2 Loi GEV avec $k \neq 0$ (EV2 et EV3).....	34

CHAPITRE 5 : TEST DU RAPPORT DES VRAISEMBLANCES MAXIMALES POUR VÉRIFIER L'HYPOTHESE "$H_0 : k = 0$"	36
CONCLUSION	39
BIBLIOGRAPHIE	40
ANNEXE A : LA LOI GUMBEL : CAS LIMITE DE LA LOI GEV	43
ANNEXE B : DESCRIPTION DES MÉTHODES D'ESTIMATION	45
ANNEXE C : VARIANCES ET COVARIANCES DES MOMENTS DE L'ÉCHANTILLON	54

LISTE DES TABLEAUX ET FIGURES

Tableau 1. Variation de C_s en fonction de k	13
Figure 1. Forme de la f.d.p. de la loi GEV pour différentes valeurs de k	6
Figure 2. Forme de la f.d.p. de la loi EV1 pour différentes valeurs de α	6
Figure 3. Forme de la f.d.p. de la loi EV2 pour différentes valeurs de α	7
Figure 4. Forme de la f.d.p. de la loi EV3 pour différentes valeurs de α	7
Figure 5. Variation de C_s en fonction de k	12

INTRODUCTION

La construction et la planification des ouvrages hydrauliques sont reliées aux propriétés d'événements extrêmes comme par exemple, les niveaux ou les débits maximums de rivières et les précipitations maximales observées dans un intervalle de temps donné. Ces événements extrêmes peuvent être considérés comme des maximums observés parmi un ensemble assez large d'événements indépendants. Par exemple, le débit maximum annuel d'une rivière est le maximum d'un certain nombre de débits mesuré pendant l'année. Conséquemment, on peut raisonnablement supposer que ce type d'observations est distribué selon l'une des lois des valeurs extrêmes introduites par Fisher et Tippett (1928).

Ces auteurs ont distingué trois types de distributions : le Type 1 (EV1), le Type 2 (EV2) et le Type 3 (EV3). Jenkinson (1955) proposa ensuite la loi généralisée des valeurs extrêmes (GEV) dont la fonction de distribution regroupe ces trois types de lois. Suite aux recommandations en 1975 du "National Environment Research Council" (NERC), la loi GEV a été fréquemment utilisée dans l'analyse statistique des débits de crue pour les rivières de Grande-Bretagne. Ce document présente une synthèse des travaux relatifs aux aspects théoriques de cette loi. Les résultats présentés dans le présent rapport ont été utilisés pour introduire la loi GEV dans le logiciel AJUSTE.

Le premier chapitre présente la définition des lois EV1, EV2 et EV3 ainsi que la forme générale de la loi GEV. Les principales propriétés statistiques de cette distribution sont données au chapitre 2. Ensuite, le chapitre 3 est consacré à l'estimation des paramètres selon trois méthodes différentes: la méthode du maximum de vraisemblance (MXVR), la méthode des moments (MM), et la méthode des moments pondérés (MMP).

Le chapitre 4 traite de l'estimation d'un événement X_T de période de retour T donnée ainsi que de la variance asymptotique de l'estimateur obtenu pour les trois méthodes d'estimation considérées.

Enfin, le chapitre 5 est consacré à la construction d'un test statistique nous permettant d'identifier le type de loi des valeurs extrêmes susceptible de représenter le mieux un ensemble de données.

CHAPITRE 1

CARACTÉRISTIQUES GÉNÉRALES DES LOIS DES VALEURS EXTREMES

1.1 Distributions des valeurs extrêmes

Les distributions des valeurs extrêmes, introduites par Fisher et Tippett (1928), comprennent les trois lois caractérisées par les fonctions de distribution suivantes :

Type 1

$$G_1(x) = \exp \left\{ -\exp \left(-\frac{(x-m)}{\theta} \right) \right\}, \quad -\infty < x < \infty \quad (1.1)$$

Type 2

$$G_2(x) = 1 - \exp \left(-\frac{(x-m)^c}{\theta} \right), \quad x \geq m \quad (1.2)$$

Type 3

$$G_3(x) = \exp \left(-\frac{(m-x)^c}{\theta} \right), \quad x \leq m \quad (1.3)$$

où m , $\theta (> 0)$ et $c (> 0)$ sont les paramètres. Nous notons ces lois, tout au long du présent rapport, par EV1, EV2 et EV3 respectivement.

Le terme "valeurs extrêmes" est associé à ces distributions parce qu'elles peuvent être obtenues comme lois limites ($n \rightarrow \infty$) de la valeur maximale de n variables aléatoires indépendantes et identiquement distribuées. En remplaçant X par $-X$, les lois de probabilité des valeurs minimums sont obtenues. Elles sont donc aussi des distributions des valeurs extrêmes.

Même si les lois EV1, EV2 et EV3 sont appelées "lois des valeurs extrêmes", elles peuvent tout aussi bien être utilisées en pratique de la même manière que toute autre loi de probabilité. En effet, on peut ajuster l'une de ces distributions à des observations qui ne sont pas nécessairement des minimums ou des maximums.

1.2 Distribution GEV

Jenkinson (1955) a proposé ce qu'il a appelé la loi généralisée des valeurs extrêmes (GEV). Cette distribution a l'avantage de combiner les trois types de lois des valeurs extrêmes EV1, EV2 et EV3 développées par Fisher et Tippett. La forme générale de la fonction de distribution de cette loi peut s'écrire:

$$F(x) = \exp \left[- \left(1 - \frac{k}{\alpha} (x - u) \right)^{1/k} \right], \quad -\infty < x < +\infty \quad (1.4)$$

où u , α (> 0) et k sont les paramètres.

Chacun des types de lois des valeurs extrêmes est caractérisé par le signe du paramètre de forme k de la loi GEV. En effet,

- 1) pour $k \rightarrow 0$, $F(x)$ correspond à la loi EV1 (équation 1.1). L'équation (1.1) est en fait la limite, lorsque k tend vers zéro, de l'équation (1.4). Ce calcul est effectué à l'annexe A. La loi EV1 est mieux connue sous le nom de distribution Gumbel, et est de loin la plus utilisée des trois. Selon la reparamétrisation de Jenkinson, la fonction de distribution s'écrit

$$F(x) = \exp\left[-\exp\left(-\frac{(x-u)}{\alpha}\right)\right], -\infty < x < +\infty \quad (1.5)$$

- 2) pour $k < 0$, $F(x)$ correspond, à une reparamétrisation près, à la loi EV2 (équation 1.2). Dans ce cas, la variable X est bornée inférieurement par la valeur $u + \alpha/k$ avec $\alpha > 0$. Cette loi est reliée à la distribution EV1 par une transformation logarithmique. En effet, si X suit une loi EV2 alors $\ln(X)$ suit une loi Gumbel (ou EV1). C'est pourquoi la distribution EV2 est quelquefois appelée la loi log-Gumbel.
- 3) pour $k > 0$, $F(x)$ correspond, à une reparamétrisation près, à la loi EV3 (équation 1.3). Dans ce cas, la variable X est bornée supérieurement par la valeur $u + \alpha/k$ avec $\alpha > 0$. Cette distribution est reliée à la loi Weibull par un simple changement de signe: si X suit une loi EV3 alors $-X$ suit une loi de Weibull à 3 paramètres.

Ainsi, on distingue deux formes de fonction de distribution selon la valeur de k pour la loi GEV (équation 1.4 si $k \neq 0$ et équation 1.5 si $k = 0$). De la même façon, la fonction densité de probabilité prend deux formes obtenues en dérivant (1.4) et (1.5) :

$$f(x) = \frac{1}{\alpha} \exp\left[-\frac{(x-u)}{\alpha} - \exp\left(-\frac{(x-u)}{\alpha}\right)\right], \text{ si } k = 0 \quad (1.6)$$

$$f(x) = \frac{1}{\alpha} \left[1 - \frac{k}{\alpha}(x-u)\right]^{1/k-1} \exp\left[-\left(1 - \frac{k}{\alpha}(x-u)\right)^{1/k}\right], \text{ si } k \neq 0 \quad (1.7)$$

De plus, si nous notons respectivement par X_1 , X_2 et X_3 les variables aléatoires distribuées selon la loi GEV avec $k = 0$, $k < 0$ et $k > 0$, les variables réduites correspondantes s'écrivent :

$$Y_1 = \frac{(X_1 - u)}{\alpha} \quad (1.8)$$

$$Y_2 = 1 - \frac{(X_2 - u)}{\alpha} k \quad (1.9)$$

$$-Y_3 = 1 - \frac{(X_3 - u)}{\alpha} k \quad (1.10)$$

Leurs fonctions de distribution sont alors données par:

$$G(y_1) = \exp[-\exp(y_1)] \quad -\infty < y_1 < +\infty \quad (1.11)$$

$$G(y_2) = \exp[-y_2^{1/k}] \quad 0 < y_2 < +\infty \quad (1.12)$$

$$G(y_3) = \exp[-(-y_3)^{1/k}] \quad -\infty < y_3 < 0 \quad (1.13)$$

et leurs fonctions de densité de probabilité par:

$$g(y_1) = \exp[-y_1 - \exp(y_1)] \quad -\infty < y_1 < +\infty \quad (1.14)$$

$$g(y_2) = \frac{-y_2^{1/k-1}}{k} \exp[-y_2^{1/k}] \quad 0 < y_2 < +\infty \quad (1.15)$$

$$g(y_3) = \frac{-y_3^{1/k-1}}{k} \exp[-(-y_3)^{1/k}] \quad -\infty < y_3 < 0 \quad (1.16)$$

Les Figures 1 à 4 permettent de visualiser l'influence des paramètres sur la forme de la distribution GEV. La figure 1 montre l'allure de la f.d.p de la loi GEV pour des valeurs de k variant entre -0.6 et 0.6, u étant fixé à 5 et α à 2. Nous nous sommes restreint à cet ensemble de valeurs pour le paramètre k puisque les membres du NERC soulignent dans leur publication (NERC, 1975) qu'en pratique, pour les applications aux crues, k appartient à cet intervalle. La figure 2 montre l'influence du paramètre d'échelle α sur la f.d.p EV1 ($k = 0$), alors que les figures 3 et 4 démontrent cette influence dans les cas des distributions EV2 ($k < 0$) et EV3 ($k > 0$) respectivement.

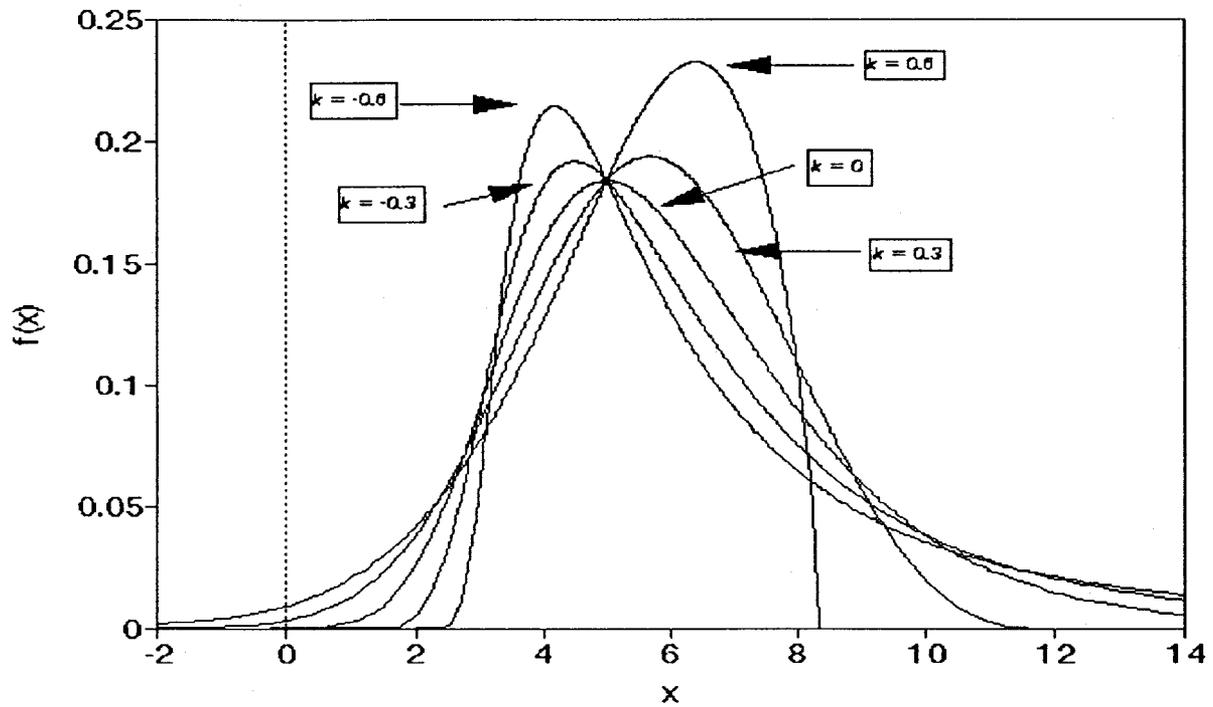


Figure 1. Forme de la f.d.p. de la loi GEV pour différentes valeurs de k .

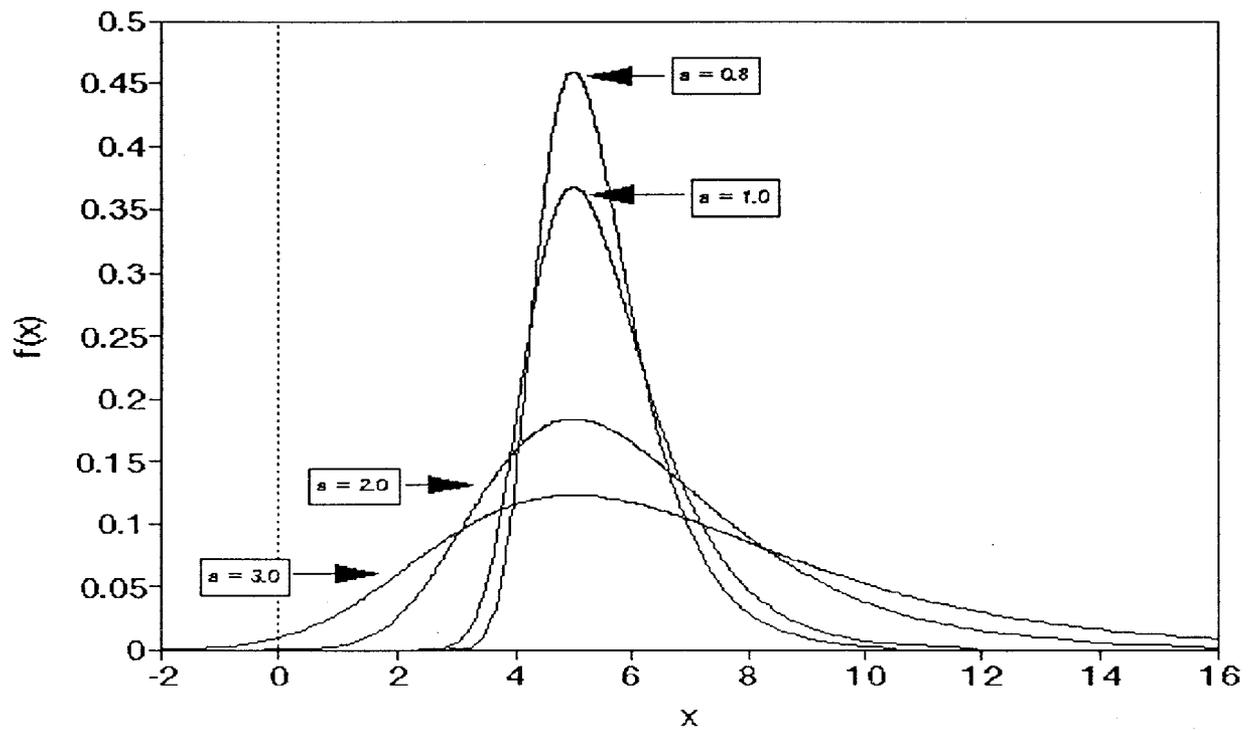


Figure 2. Forme de la f.d.p. de la loi EV1 pour différentes valeurs de α .

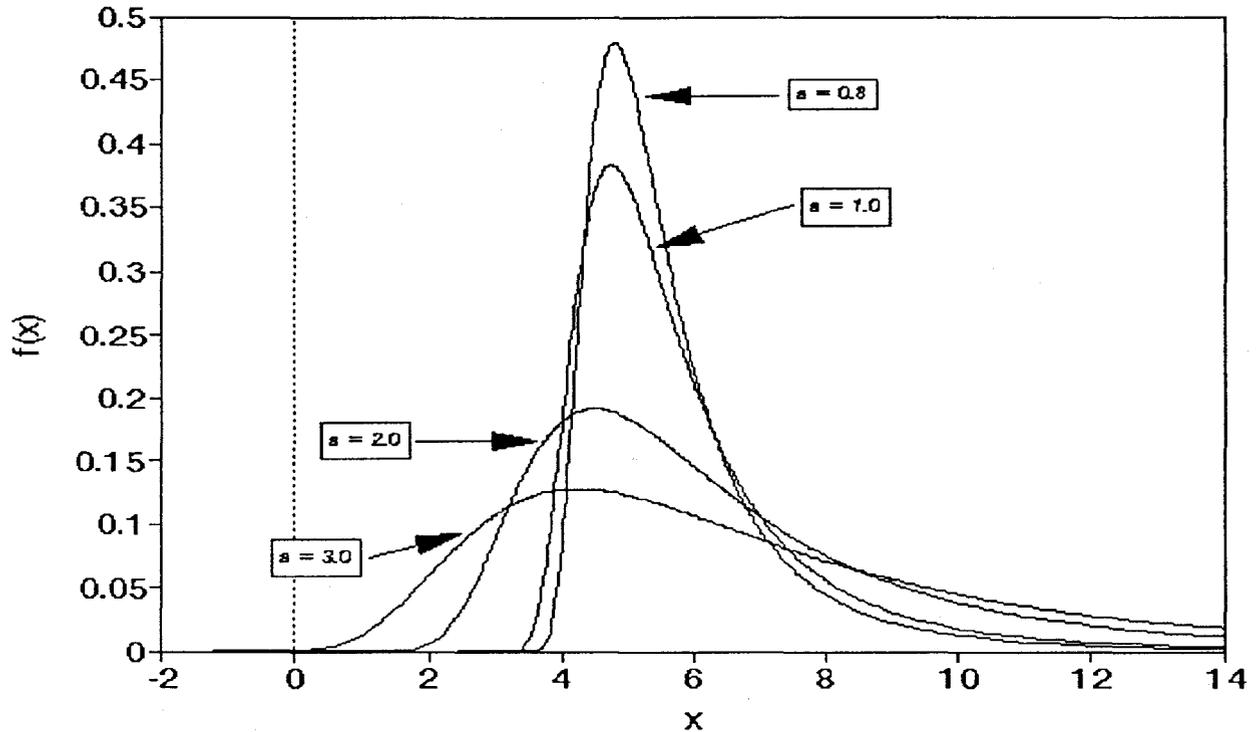


Figure 3. Forme de la f.d.p. de la loi EV2 pour différentes valeurs de α .

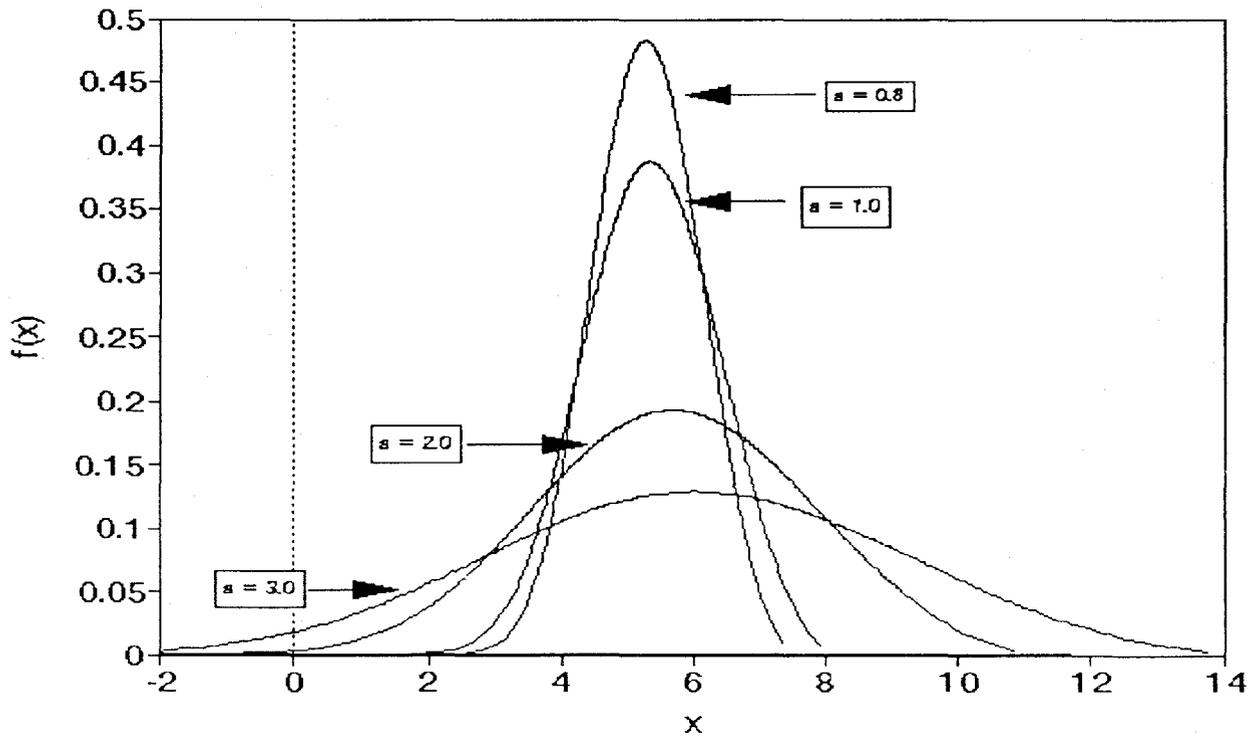


Figure 4. Forme de la f.d.p. de la loi EV3 pour différentes valeurs de α .

CHAPITRE 2

PROPRIETES STATISTIQUES DE LA LOI GENERALISEE DES VALEURS EXTREMES

Les relations donnant les principaux moments centrés et non centrés ainsi que les coefficients d'asymétrie et d'aplatissement de la loi GEV sont présentés dans ce chapitre.

2.1 Loi GEV avec $k = 0$ (EV1, loi Gumbel)

Soient X_1 une variable aléatoire distribuée selon une loi GEV avec $k = 0$ (éq. 1.6) et Y_1 la variable réduite correspondante (éq. 1.8) dont la fonction de distribution (1.11) ne dépend d'aucun paramètre. On peut montrer (Gumbel, 1960) que la moyenne $\mu_1(Y_1)$ et la variance $\mu_2(Y_1)$ de cette variable sont données respectivement par:

$$\mu_1(Y_1) = C \quad (2.1)$$

où C représente la constante d'Euler (= 0.5772..) et

$$\mu_2(Y_1) = \frac{\pi^2}{6} \quad (2.2)$$

La moyenne ainsi que la variance de X_1 sont alors déduites de la relation (1.8) et des équations (2.1) et (2.2). On obtient :

$$\mu_1(X_1) = \mu_1(u + \alpha Y_1) = u + \alpha \mu_1(Y_1) = u + C\alpha \quad (2.3)$$

$$\mu_2(X_1) = \mu_2(u + \alpha Y_1) = \alpha^2 \mu_2(Y_1) = \frac{\alpha^2 \pi^2}{6} \quad (2.4)$$

Enfin, les coefficients d'asymétrie (C_s) et d'aplatissement (C_k) sont donnés respectivement par

$$C_s(X_1) = \frac{\mu_3(X_1)}{\mu_2^{3/2}(X_1)} = 1.139 \quad (2.5)$$

et

$$C_k(X_1) = \frac{\mu_4(X_1)}{\mu_2^2(X_1)} = 2.400 \quad (2.6)$$

2.2 Loi GEV avec $k \neq 0$ (EV1 et EV2)

Les moments centrés d'ordre r par rapport à la moyenne $\mu_r(X_2)$ et $\mu_r(X_3)$ des variables aléatoires X_2 et X_3 distribuées selon les lois EV2 et EV3 sont facilement déduits des expressions donnant les moments non centrés d'ordre r des variables aléatoires réduites Y_2 et Y_3 (NERC, 1975) :

$$\mu'_r(Y_2) = E(Y_2^r) = \Gamma(1+rk) \quad (2.7)$$

et

$$\mu'_r(Y_3) = E(Y_3^r) = (-1)^r \Gamma(1+rk) \quad (2.8)$$

En effet, en utilisant les équations (1.9) et (1.10), on a :

$$X_2 = u + \frac{\alpha}{k} - \frac{\alpha}{k} Y_2 \quad (2.9)$$

et

$$X_3 = u + \frac{\alpha}{k} + \frac{\alpha}{k} Y_3 \quad (2.10)$$

Remarquons que les moments d'ordre r de la variable réduite Y_2 existent seulement si $k < -1/r$, et ceux de la variable réduite Y_3 seulement si $k > -1/r$. Les moments d'ordre r des lois EV2 (variable X_2) et EV3 (variable X_3) existent donc, respectivement, sous les mêmes conditions.

On déduit, en posant $r = 1$, la moyenne de ces distributions:

$$\begin{aligned}\mu_1(X_2) = E(X_2) &= u + \frac{\alpha}{k} - \frac{\alpha}{k}E(Y_2) & (2.11) \\ &= u + \frac{\alpha}{k} - \frac{\alpha}{k}\Gamma(1+k) \\ &= u + \frac{\alpha}{k}[1 - \Gamma(1+k)]\end{aligned}$$

et

$$\begin{aligned}\mu_1(X_3) = E(X_3) &= u + \frac{\alpha}{k} + \frac{\alpha}{k}E(Y_3) & (2.12) \\ &= u + \frac{\alpha}{k} + \frac{\alpha}{k}(-1)\Gamma(1+k) \\ &= u + \frac{\alpha}{k}[1 - \Gamma(1+k)]\end{aligned}$$

On utilise ces résultats pour donner la définition générale des moments centrés par rapport à la moyenne d'ordre r pour X_2 et X_3 en fonction des variables réduites Y_2 et Y_3 .

En utilisant les relations (2.9) et (2.11), on a :

$$\begin{aligned}\mu_r(X_2) &= E\{[X_2 - E(X_2)]^r\} & (2.13) \\ &= (-1)^r \left(\frac{\alpha}{k}\right)^r E\{[Y_2 - \Gamma(1+k)]^r\}\end{aligned}$$

et en utilisant les expressions (2.10) et (2.12), on a:

$$\begin{aligned}\mu_r(X_3) &= E\{[X_3 - E(X_3)]^r\} \\ &= \left(\frac{\alpha}{k}\right)^r E\{[Y_3 + \Gamma(1+k)]^r\}\end{aligned}\tag{2.14}$$

Ainsi, à partir des équations (2.13) et (2.14), on peut montrer en utilisant les relations (2.7) et (2.8) que lorsque X suit une loi GEV avec $k \neq 0$ (c'est-à-dire une loi EV2 ou EV3), sa variance ($r = 2$) est donnée, quelque soit le signe du paramètre k , par:

$$\mu_2(X) = \left(\frac{\alpha}{k}\right)^2 [\Gamma(1+2k) - \Gamma^2(1+k)]\tag{2.15}$$

et les moments d'ordre 3, 4, 5 et 6 par:

$$\begin{aligned}\mu_3(X) &= -\left(\frac{\alpha}{k}\right)^3 [\Gamma(1+3k) - 3\Gamma(1+2k)\Gamma(1+k) \\ &\quad + 2\Gamma^3(1+k)]\end{aligned}\tag{2.16}$$

$$\begin{aligned}\mu_4(X) &= \left(\frac{\alpha}{k}\right)^4 [\Gamma(1+4k) - 4\Gamma(1+k)\Gamma(1+3k) \\ &\quad + 6\Gamma^2(1+k)\Gamma(1+2k) - 3\Gamma^4(1+k)]\end{aligned}\tag{2.17}$$

$$\begin{aligned}\mu_5(X) &= -\left(\frac{\alpha}{k}\right)^5 [\Gamma(1+5k) - 5\Gamma(1+k)\Gamma(1+4k) + 10\Gamma^2(1+k)\Gamma(1+3k) \\ &\quad - 10\Gamma^3(1+k)\Gamma(1+2k) + 4\Gamma^5(1+k)]\end{aligned}\tag{2.18}$$

$$\begin{aligned}\mu_6(X) &= \left(\frac{\alpha}{k}\right)^6 [\Gamma(1+6k) - 6\Gamma(1+k)\Gamma(1+5k) + 15\Gamma^2(1+k)\Gamma(1+4k) \\ &\quad - 20\Gamma^3(1+k)\Gamma(1+3k) + 15\Gamma^4(1+k)\Gamma(1+2k) \\ &\quad - 5\Gamma^6(1+k)]\end{aligned}\tag{2.19}$$

Enfin, de ces résultats, on peut déduire le coefficient d'asymétrie, $C_s(X) = \mu_3(X)/\mu_2^{3/2}(X)$:

$$C_s(X) = \frac{|k| [\Gamma(1+3k) - 3\Gamma(1+2k)\Gamma(1+k) + 2\Gamma^3(1+k)]}{k [\Gamma(1+2k) - \Gamma^2(1+k)]^{3/2}} \quad (2.20)$$

ainsi que le coefficient d'aplatissement, $C_k(X) = \mu_4(X)/\mu_2^2(X)$:

$$C_k(X) = \frac{[\Gamma(1+4k) - 4\Gamma(1+k)\Gamma(1+3k) + 6\Gamma^2(1+k)\Gamma(1+2k) - 3\Gamma^4(1+k)]}{[\Gamma(1+2k) - \Gamma^2(1+k)]^2} \quad (2.21)$$

qui ne dépendent que du paramètre de forme k .

Les coefficients d'asymétrie, pour une série de valeurs du paramètre k , sont présentés dans le Tableau 1. Il est important de noter ici que pour des valeurs de $k < 0$ (EV2), l'asymétrie de la population est supérieure à 1.139 (asymétrie de la distribution EV1) et que pour des valeurs de $k > 0$ (EV3), elle est inférieure à 1.139. De plus, la relation entre C_s et k est illustrée graphiquement à la Figure 5.

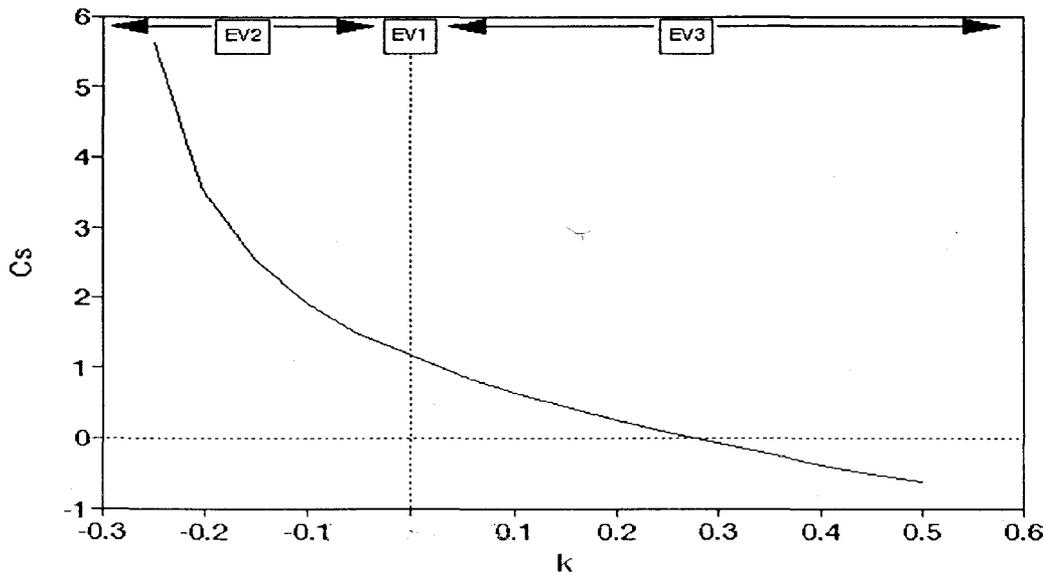


Figure 5. Variation C_s de en fonction de k .

Tableau 1. Variation de C_s en fonction de k .

k	C_s
-0.25	5.605138
-0.20	3.535071
-0.15	2.530205
-0.10	1.910333
-0.05	1.473894
0.00	1.139000
0.05	0.867965
0.10	0.637637
0.15	0.435743
0.20	0.367551
0.25	0.087237
0.30	-0.068742
0.35	-0.216653
0.40	-0.375519
0.45	-0.496340
0.50	-0.631111

CHAPITRE 3

ESTIMATION DES PARAMETRES DE LA LOI GEV

Les méthodes d'estimation des paramètres les plus utilisées sont en pratique la méthode du maximum de vraisemblance (MXVR) et la méthode des moments (MM). La première est employée généralement à cause de ses propriétés asymptotiques fort intéressantes, la seconde pour sa simplicité.

En ce qui concerne la loi GEV deux autres méthodes sont souvent utilisées : la méthode des sextiles, développée par Jenkinson (1969), et la méthode des moments pondérés introduite par Greenwood et al. (1979). Cette dernière méthode est fortement recommandée par Hosking et al. (1985) pour l'ajustement de la loi GEV. Ces auteurs soulèvent entre autres le fait que la méthode des sextiles implique un choix arbitraire (pourquoi des sextiles plutôt que des quartiles ou des octiles), et qu'elle possède des propriétés statistiques mal connues. Ces mêmes auteurs mentionnent aussi que la justification de la méthode du maximum de vraisemblance est basée sur la théorie des grands échantillons (méthode asymptotiquement optimale) et que son efficacité n'a pas été démontrée pour les échantillons de petite taille.

Nous considérons dans le présent rapport trois méthodes d'estimation des paramètres pour la loi GEV : la méthode du maximum de vraisemblance pour ses propriétés asymptotiques, la méthode des moments (MM) pour sa simplicité et parce que les estimations résultantes sont utilisées comme valeurs initiales dans le processus itératif de la méthode du maximum de vraisemblance (voir section 3.1), et la méthode des moments pondérés (MMP) que recommandent Hosking et al. (1985). Les principes généraux des trois méthodes sont décrits à l'annexe B.

Nous présentons dans ce chapitre les développements et les résultats menant aux différents estimateurs des paramètres de la loi GEV pour les trois méthodes considérées. Les variances et covariances asymptotiques des estimateurs obtenus sont également calculées.

3.1 Loi GEV avec $k = 0$ (EV1, loi Gumbel)

3.1.1 Méthode du maximum de vraisemblance

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes de fonction de densité de probabilité (1.6). La vraisemblance logarithmique de l'échantillon x_1, x_2, \dots, x_n , une réalisation de X_1, X_2, \dots, X_n , est donnée par :

$$\ln L(\alpha, u) = -n \ln(\alpha) - \sum_{i=1}^n \frac{x_i - u}{\alpha} - \sum_{i=1}^n \exp\left[-\frac{x_i - u}{\alpha}\right] \quad (3.1)$$

En prenant la dérivée de $\ln L(\alpha, u)$ par rapport à α et u , on obtient le système d'équations qui permet de maximiser cette fonction :

$$n - \sum_{i=1}^n \frac{x_i - u}{\alpha} + \sum_{i=1}^n \frac{x_i - u}{\alpha} \exp\left[-\frac{x_i - u}{\alpha}\right] = 0 \quad (3.2)$$

$$-n + \sum_{i=1}^n \exp\left[-\frac{x_i - u}{\alpha}\right] = 0 \quad (3.3)$$

On peut toutefois simplifier ce système en écrivant l'équation (3.3) sous la forme :

$$u = -\alpha \ln\left\{\frac{1}{n} \sum_{i=1}^n \exp\left[-\frac{x_i}{\alpha}\right]\right\} \quad (3.4)$$

et en remplaçant u dans (3.2) par (3.4) on obtient :

$$\alpha = \bar{x} - \left\{\sum_{i=1}^n x_i \exp\left[-\frac{x_i}{\alpha}\right]\right\} \left\{\sum_{i=1}^n \exp\left[-\frac{x_i}{\alpha}\right]\right\}^{-1} \quad (3.5)$$

Les équations (3.4) et (3.5) sont équivalentes au système (3.2) et (3.3). On peut ainsi, pour un échantillon donné x_1, x_2, \dots, x_n , déterminer $\hat{\alpha}$ par (3.5) et ensuite connaissant $\hat{\alpha}$ en déduire \hat{u} par (3.4).

Pour résoudre (3.5), on doit employer une méthode itérative comme celle de Newton-Raphson. Toutefois, Phien (1987) souligne que la méthode numérique proposée par Clarke (1973) converge plus rapidement. Cette méthode est expliquée en détail dans Phien (1987) et est utilisée dans le logiciel AJUSTE.

Les estimateurs du maximum de vraisemblance sont asymptotiquement non-biaisés. Par contre, pour une taille d'échantillon finie, ils sont généralement biaisés. En particulier, les estimateurs \hat{u} et $\hat{\alpha}$ obtenus dans le cas de la loi EV1 possèdent un biais (Lowery et Nash, 1970). Une bonne correction réduisant ce biais a été proposée par Fiorentio et Gabriele (1984), et les estimateurs corrigés résultants sont donnés par :

$$\alpha^* = \frac{n}{n-0.8} \hat{\alpha} \quad (3.6)$$

et

$$u^* = \alpha^* \ln \left\{ n \left(\sum_{i=1}^n \exp \left[-\frac{x_i}{\alpha^*} \right] \right)^{-1} \right\} - \frac{0.7}{n} \alpha^* \quad (3.7)$$

Cette correction, qui permet de réduire significativement le biais, a été introduite dans le logiciel AJUSTE.

Les variances et la covariance asymptotiques des estimateurs du maximum de vraisemblance sont obtenues en inversant la matrice d'information de Fisher définie à l'annexe B. Nous obtenons pour la loi Gumbel (voir Phien, 1987) des expressions explicites des variances et de la covariance de \hat{u} et $\hat{\alpha}$. Elles sont données par:

$$Var(\hat{\alpha}) = 0.608 \frac{\alpha^2}{n} \quad (3.8)$$

$$Var(\hat{u}) = 1.109 \frac{\alpha^2}{n} \quad (3.9)$$

$$Cov(\hat{u}, \hat{\alpha}) = 0.257 \frac{\alpha^2}{n} \quad (3.10)$$

De ces trois expressions, on déduit aisément les variances et covariances asymptotiques des estimateurs corrigés (Phien, 1987) :

$$Var(\alpha^*) = \left[\frac{n}{n-0.8} \right]^2 Var(\hat{\alpha}) \quad (3.11)$$

$$Var(u^*) = \left(\frac{\alpha^2}{n} \right) \left[1.109 - \frac{0.360}{n-0.8} + \frac{0.298}{n^2} \right] \quad (3.12)$$

et

$$Cov(u^*, \alpha^*) = \left(\frac{\alpha^2}{n} \right) \left[\frac{0.257n}{n-0.8} - \frac{0.426}{n} \right] \quad (3.13)$$

Remarquons que comme on peut s'y attendre, pour de grandes valeurs de n , les variances et la covariance de u^* et α^* sont équivalentes à celles de \hat{u} et $\hat{\alpha}$.

3.1.2 Méthode des moments

La méthode des moments appliquée à la loi Gumbel consiste à évaluer les deux premiers moments théoriques (moyenne et écart-type) de cette loi aux moments estimés à partir de l'échantillon (moyenne \bar{x} et écart-type s de l'échantillon). On obtient ainsi, d'après les équations (2.3) et (2.4), le système suivant :

$$\bar{x} = u + C\alpha \quad (3.14)$$

$$s = \frac{\pi}{\sqrt{6}}\alpha \quad (3.15)$$

d'où l'on déduit aisément les estimateurs des moments :

$$\hat{\alpha} = 0.77969s \quad (3.16)$$

$$\hat{u} = \bar{x} - 0.45005s \quad (3.17)$$

Les variances et la covariance asymptotiques des estimateurs des moments sont obtenus à partir des équations (3.16) et (3.17), et des résultats de Lowery et Nash (1970, p.272) concernant les variances et la covariance asymptotiques de \bar{X} et S . Après quelques calculs nous obtenons :

$$\begin{aligned} Var(\hat{u}) &= Var(\bar{X}) + 6\left(\frac{C}{\pi}\right)^2 Var(S) - 2C\left(\frac{\sqrt{6}}{\pi}\right)Cov(\bar{X}, S) \\ &= \left(\frac{\alpha^2}{n}\right)1.16779 \end{aligned} \quad (3.18)$$

$$Var(\hat{\alpha}) = \left(\frac{6}{\pi^2}\right)Var(S) = \left(\frac{\alpha^2}{n}\right)1.10005 \quad (3.19)$$

$$Cov(\hat{u}, \hat{\alpha}) = \left(\frac{\alpha^2}{n}\right)0.095848 \quad (3.20)$$

3.1.3 Méthode des moments pondérés

La méthode des moments pondérés (annexe B) a été développée pour les lois pouvant être définies de façon explicite par leur fonction de distribution inverse. On remarque aisément que la loi Gumbel peut s'exprimer explicitement sous forme de fonction inverse de la façon suivante :

$$X(F) = u - \alpha \ln(-\ln F) \quad (3.21)$$

Pour cette loi, seulement deux moments pondérés sont nécessaires afin d'estimer les deux paramètres:

$$\beta_r = E[XF^r], \quad r = 0,1 \quad (3.22)$$

On peut montrer (Phien, 1987) que ceux-ci sont donnés par :

$$\beta_r = \frac{1}{1+r} \{u + \alpha[\ln(1+r) + C]\} \quad (3.23)$$

Ainsi,

$$\beta_0 = u + C\alpha \quad (3.24)$$

et

$$\beta_1 = \frac{1}{2}[u + \alpha(\ln 2 + C)] \quad (3.25)$$

En remplaçant dans ces équations, β_0 et β_1 par leurs estimateurs respectifs b_0 et b_1 (voir Annexe B), on obtient, après quelques calculs :

$$\hat{\alpha} = \frac{2b_1 - b_0}{\ln 2} \quad (3.26)$$

$$\hat{u} = b_0 - C\hat{\alpha} \quad (3.27)$$

Il est à noter que ces estimateurs obtenus par la méthode des moments pondérés sont équivalents à ceux proposés par Downton (1966). Les estimateurs développés par cet auteur sont construits à partir de combinaisons linéaires de statistiques d'ordre. Ils ont comme principal avantage la déduction fort simple de leur variance asymptotique. Vue cette équivalence, les variances et la covariance asymptotiques calculées par cet auteur pour ces estimateurs sont applicables ici pour la méthode des moments pondérés. Les variances et la covariance de \hat{u} et $\hat{\alpha}$ obtenues par Downton (1966) sont :

$$Var(\hat{\alpha}) = \frac{\alpha^2}{n(n-1)} [0.804627n - 0.185527] \quad (3.28)$$

$$Var(\hat{u}) = \frac{\alpha^2}{n(n-1)} [1.112825n - 0.906557] \quad (3.29)$$

$$\text{Cov}(\hat{u}, \hat{\alpha}) = -\frac{\alpha^2}{n(n-1)}[0.228707n - 0.586058] \quad (3.30)$$

3.2 Loi GEV avec $k \neq 0$ (EV2 et EV3)

3.2.1 Méthode du maximum de vraisemblance

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes dont la fonction de densité de probabilité est donnée par (1.7), et une réalisation x_1, x_2, \dots, x_n de ces variables. Si l'on considère la variable auxiliaire :

$$y(x) = -\frac{1}{k} \ln \left[1 - \frac{k}{\alpha} (x - u) \right] \quad (3.31)$$

ce qui est équivalent à :

$$1 - \frac{k}{\alpha} (x - u) = e^{-ky(x)} \quad (3.32)$$

la relation (1.7) devient :

$$f(x) = \frac{1}{\alpha} (e^{-ky(x)})^{1/k-1} \exp[-(e^{-ky(x)})^{1/k}] \quad (3.33)$$

ce qui s'écrit encore :

$$f(x) = \frac{1}{\alpha} e^{(k-1)y(x)} \exp(-e^{-y(x)}) \quad (3.34)$$

Le logarithme de la fonction de vraisemblance est alors :

$$\ln L(x; u, \alpha, k) = -n \ln(\alpha) - (1-k) \sum_{i=1}^n y(x_i) - \sum_{i=1}^n e^{-y(x_i)} \quad (3.35)$$

Les estimateurs du maximum de vraisemblance sont obtenus en maximisant cette fonction, c'est-à-dire en annulant simultanément les trois dérivées partielles de $\ln L$ par rapport aux 3 paramètres. Selon la notation de Jenkinson (1969), on peut montrer que le système d'équation à résoudre est :

$$\frac{\partial \ln L(x; u, \alpha, k)}{\partial u} = \frac{Q}{\alpha} \quad (3.36)$$

$$\frac{\partial \ln L(x; u, \alpha, k)}{\partial \alpha} = \frac{1}{\alpha} \frac{P+Q}{k} \quad (3.37)$$

$$\frac{\partial \ln L(x; u, \alpha, k)}{\partial k} = \frac{1}{k} \left(R - \frac{P+Q}{k} \right) \quad (3.38)$$

où

$$P = n - \sum_{i=1}^n e^{-y(x_i)} \quad (3.39)$$

$$Q = \sum_{i=1}^n e^{-y(x_i)+ky(x_i)} - (1-k) \sum_{i=1}^n e^{ky(x_i)} \quad (3.40)$$

$$R = n - \sum_{i=1}^n y(x_i) + \sum_{i=1}^n y(x_i) e^{-y(x_i)} \quad (3.41)$$

Ce système d'équations ne peut être résolu explicitement. C'est pourquoi nous devons utiliser une méthode itérative pour trouver la solution.

Supposons que l'estimateur du maximum de vraisemblance du vecteur $\theta = (u, \alpha, k)^T$ soit $\hat{\theta} = (\hat{u}, \hat{\alpha}, \hat{k})^T$, et que $\theta^{(j)}$ est la solution du système obtenue à la j^{ème} itération. Les éléments du vecteur

$$-\Delta \ln L(\theta^{(j)}) = \left(-\frac{\partial \ln L(u, \alpha, k)}{\partial u}, -\frac{\partial \ln L(u, \alpha, k)}{\partial \alpha}, -\frac{\partial \ln L(u, \alpha, k)}{\partial k} \right)^T$$

à la jⁱème itération peuvent être développés en séries de Taylor autour de $\hat{\alpha}$ (voir NERC, 1975). On obtient, en ne retenant que les trois premiers termes (dérivées secondes), la formule itérative qui s'exprime sous forme matricielle de la façon suivante :

$$\theta^{(j+1)} = \theta^{(j)} + V(\hat{\theta})^{-1} \Delta \ln L(\theta^{(j)}) \quad (3.42)$$

où $\Delta \ln L(\theta^{(j)})$ est le vecteur des dérivées premières évaluées en $\theta^{(j)}$ et $V(\hat{\theta})$ est la matrice suivante:

$$V(\hat{\theta}) = \begin{pmatrix} -\frac{\partial^2 \ln L}{\partial u^2} & -\frac{\partial^2 \ln L}{\partial u \partial \alpha} & -\frac{\partial^2 \ln L}{\partial u \partial k} \\ -\frac{\partial^2 \ln L}{\partial \alpha \partial u} & -\frac{\partial^2 \ln L}{\partial \alpha^2} & -\frac{\partial^2 \ln L}{\partial \alpha \partial k} \\ -\frac{\partial^2 \ln L}{\partial k \partial u} & -\frac{\partial^2 \ln L}{\partial k \partial \alpha} & -\frac{\partial^2 \ln L}{\partial k^2} \end{pmatrix} \quad (3.43)$$

dont les éléments sont les dérivées secondes de la fonction de vraisemblance logarithmique affectées du signe moins et évaluées en $\hat{\theta}$. Puisque l'estimateur du maximum de vraisemblance $\hat{\theta}$ n'est pas connu, une approximation de cette matrice est nécessaire pour appliquer la procédure itérative.

Jenkinson (1969) propose d'approximer la matrice $V(\hat{\theta})$ par la matrice d'information de Fisher $I_f(\theta^{(j)})$ dont les éléments, les espérances mathématiques des dérivées secondes de la fonction de vraisemblance logarithmique affectées du signe moins, sont évaluées en $\theta^{(j)}$. Toutefois, Perscott et Walden (1983) mentionnent que la convergence vers $\hat{\theta}$ est plus rapide en utilisant $V(\theta^{(j)})$ plutôt que $I_f(\theta^{(j)})$. Les éléments de V sont calculés à partir des équations (3.36) à (3.41) et ils sont donnés explicitement dans Perscott et Walden (1983). Hosking (1985) donne un programme FORTRAN qui permet l'application de la procédure itérative présentée plus haut. Le logiciel AJUSTE utilise ce programme presque intégralement.

Les variances et covariances asymptotiques des estimateurs du maximum de vraisemblance sont obtenues en inversant la matrice d'information de Fisher I_f (voir annexe B). En utilisant la notation de Jenkinson (1969) la matrice I_f^{-1} s'écrit:

$$I_f^{-1} = \begin{pmatrix} \text{Var}(\hat{u}) & \text{Cov}(\hat{u}, \hat{\alpha}) & \text{Cov}(\hat{u}, \hat{k}) \\ \text{Cov}(\hat{u}, \hat{\alpha}) & \text{Var}(\hat{\alpha}) & \text{Cov}(\hat{\alpha}, \hat{k}) \\ \text{Cov}(\hat{u}, \hat{k}) & \text{Cov}(\hat{\alpha}, \hat{k}) & \text{Var}(\hat{k}) \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \alpha^2 b & \alpha^2 h & \alpha f \\ \alpha^2 h & \alpha^2 a & \alpha g \\ \alpha f & \alpha g & c \end{pmatrix} \quad (3.44)$$

où les valeurs de a, b, c, f, g et h ne dépendent que de k et ont été tabulées (Jenkinson, 1969) pour un certain nombre de valeurs de ce paramètre. Une interpolation est donc nécessaire en général. Pour éviter cela, il est préférable d'évaluer exactement les termes de la matrice I_f , laquelle est finalement inversée pour obtenir les variances et covariances des estimateurs.

Prescott et Walden (1980) donnent les expressions théoriques des éléments de I_f en termes de fonctions gamma (Γ) et digamma (Ψ). Nous reproduisons ici les équations fournies par ces auteurs:

$$E \left[-\frac{\partial^2 \ln L}{\partial u^2} \right] = \frac{n}{\alpha^2} p \quad (3.45)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial \alpha^2} \right] = \frac{n}{\alpha^2 k^2} [1 - 2\Gamma(2-k) + p] \quad (3.46)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial k^2} \right] = \frac{n}{k^2} \left[\frac{\pi^2}{6} - \left(1 - C - \frac{1}{k}\right)^2 + \frac{2q}{k} + \frac{p}{k^2} \right] \quad (3.47)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial u \partial \alpha} \right] = \frac{n}{\alpha^2 k} [p - \Gamma(2-k)] \quad (3.48)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial u \partial k} \right] = -\frac{n}{\alpha k} \left(q + \frac{p}{k} \right) \quad (3.49)$$

$$E \left[-\frac{\partial^2 \ln L}{\partial \alpha \partial k} \right] = \frac{n}{\alpha^2 k^2} \left\{ 1 - C - \frac{[1 - \Gamma(2-k)]}{k} - q - \frac{p}{k} \right\} \quad (3.50)$$

où C est la constante d'Euler (0.5772),

$$p = (1-k)^2 \Gamma(1-2k) \quad (3.51)$$

$$q = \Gamma(2-k) \left[\psi(1-k) - \frac{(1-k)}{k} \right] \quad (3.52)$$

3.2.2 Méthode des moments

La méthode des moments, pourtant d'emploi courant en ce qui concerne plusieurs lois, est peu utilisée pour l'ajustement de la loi GEV lorsque k est différent de zéro. Toutefois, elle peut fournir des estimations pour les paramètres de cette loi lorsque la méthode du maximum de vraisemblance ne converge pas. De plus, tel que mentionné plus haut, on utilise la méthode des moments pour fournir des valeurs de départ pour les paramètres dans le cadre du processus itératif du maximum de vraisemblance.

La méthode des moments appliquée à la loi GEV consiste à égaliser la moyenne, l'écart-type et le coefficient d'asymétrie de l'échantillon aux valeurs correspondantes de la population :

$$\bar{x} = \mu_1(X) \quad (3.53)$$

$$s = \sqrt{\mu_2(X)} \quad (3.54)$$

$$\hat{C}_s = \frac{\mu_3(X)}{[\mu_2(X)]^{3/2}} \quad (3.55)$$

où

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.56)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.57)$$

et

$$\hat{C}_s = \frac{n^2}{(n-1)(n-2)} \left(\frac{m^3}{S^3} \right) \quad (3.58)$$

avec

$$m^3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (3.59)$$

A partir des équations (3.53), (3.54), (3.55) et de (2.11), (2.15) et (2.20), on déduit le système d'équations suivant :

$$\bar{x} = u + \frac{\alpha}{k} \{1 - \Gamma(1+k)\} \quad (3.60)$$

$$s = \frac{\alpha}{|k|} \{\Gamma(1+2k) - \Gamma^2(1+k)\}^{1/2} \quad (3.61)$$

$$\hat{C}_s = -\frac{|k| [\Gamma(1+3k) - 3\Gamma(1+2k)\Gamma(1+k) + 2\Gamma^3(1+k)]}{k [\Gamma(1+2k) - \Gamma^2(1+k)]^{3/2}} \quad (3.62)$$

Évidemment, on ne peut résoudre ce système de manière explicite. On doit utiliser une procédure itérative. Remarquons toutefois que l'équation (3.62) ne dépend que du paramètre k . Il suffit alors de résoudre cette équation par la méthode de Newton-Raphson pour obtenir l'estimateur des moments \hat{k} , et ensuite de déterminer $\hat{\alpha}$ et \hat{u} à l'aide des équations suivantes déduites de (3.60) et (3.61) :

$$\hat{\alpha} = |\hat{k}| \{\Gamma(1+2\hat{k}) - \Gamma^2(1+\hat{k})\}^{-1/2} s \quad (3.63)$$

$$\hat{u} = \bar{x} - \frac{\hat{\alpha}}{\hat{k}} \{1 - \Gamma(1+\hat{k})\} \quad (3.64)$$

Les variances et covariances asymptotiques des estimateurs issus de la méthode des moments sont obtenus en appliquant la méthode décrite à l'annexe B, c'est-à-dire en utilisant l'expression matricielle suivante :

$$V_p = V^{-1} V_m \quad (3.65)$$

Pour la loi GEV, les éléments de V_p sont les variances et covariances asymptotiques recherchées de \hat{u} , $\hat{\alpha}$ et \hat{k} , et ceux de V_m , les variances et covariances asymptotiques des moments de l'échantillon \bar{X} , S et \hat{C}_s . La matrice V , pour sa part, est de dimension 6×6 et ses éléments sont fonction des dérivées premières, par rapport aux paramètres, des trois moments de la population. Selon la notation de l'annexe B, ces dérivées sont les termes A_{rj} suivants :

$$A_{rj} = \frac{\partial M_r}{\partial \theta_j} \quad (3.66)$$

où $\theta_1 = u$, $\theta_2 = \alpha$, $\theta_3 = k$, et M_1 , M_2 et M_3 sont respectivement la moyenne, l'écart-type et le coefficient d'asymétrie donnés aux équations (2.11), (2.15) et (2.20). Nous obtenons ici :

$$A_{11} = 1 \quad (3.67)$$

$$A_{12} = \frac{1}{k} \{1 - \Gamma(1+k)\} \quad (3.68)$$

$$A_{13} = \frac{\alpha}{k^2} \{ \Gamma(1+k) - k\Gamma(1+k)\psi(1+k) - 1 \} \quad (3.69)$$

$$A_{22} = \frac{1}{|k|} \{ \Gamma(1+2k) - \Gamma^2(1+k) \}^{1/2} \quad (3.70)$$

$$A_{23} = \frac{|k|}{k} \frac{\alpha}{k^2} \left\{ \frac{k[\Gamma(1+2k)\psi(1+2k) - \Gamma^2(1+k)\psi(1+2k)] - [\Gamma(1+2k) - \Gamma^2(1+k)]}{[\Gamma(1+2k) - \Gamma^2(1+k)]^{1/2}} \right\} \quad (3.71)$$

$$A_{33} = -\frac{|k|}{k} \frac{C}{A^{3/2}} \left[\frac{1}{C} \frac{\partial C}{\partial k} - \frac{3}{2A} \frac{\partial A}{\partial k} \right] \quad (3.72)$$

où

$$C = \Gamma(1+3k) - 3\Gamma(1+2k)\Gamma(1+k) + 2\Gamma^3(1+k) \quad (3.73)$$

$$A = \Gamma(1+2k) - \Gamma^2(1+k) \quad (3.74)$$

$$\frac{\partial C}{\partial k} = 3\{\Gamma(1+3k)\Psi(1+3k) - \Gamma(1+2k)\Gamma(1+k)[\Psi(1+k) + 2\Psi(1+2k)] + 2\Gamma^3(1+k)\Psi(1+k)\} \quad (3.75)$$

$$\frac{\partial A}{\partial k} = 2[\Gamma(1+2k)\Psi(1+2k) - \Gamma^2(1+k)\Psi(1+k)] \quad (3.76)$$

et $A_{21} = A_{31} = A_{32} = 0$.

Maintenant, avant de donner les expressions explicites des éléments du vecteur V_m , définissons les rapports de moments suivants:

$$C_3 = \frac{\mu_3}{\mu_2^{3/2}} = C_s \quad (3.77)$$

$$C_4 = \frac{\mu_4}{\mu_2^2} = C_k \quad (3.78)$$

$$C_5 = \frac{\mu_5}{\mu_2^{5/2}} \quad (3.79)$$

$$C_6 = \frac{\mu_6}{\mu_2^3} \quad (3.80)$$

On montre aisément (voir Annexe C) que les éléments du vecteur V_m sont donnés dans l'ordre par:

$$Var(\bar{X}) = \frac{\mu_2}{n} \quad (3.81)$$

$$Var(S) = \frac{\mu_2(C_4 - 1)}{4n} \quad (3.82)$$

$$Var(C_s) = \frac{1}{n} \left[9 + C_6 - 3C_3C_5 + C_4 \left(\frac{9}{4}C_3^2 - 6 \right) + \frac{35}{4}C_3^2 \right] \quad (3.83)$$

$$Cov(\bar{X}, S) = \frac{\mu_2 C_3}{2n} \quad (3.84)$$

$$Cov(\bar{X}, C_s) = \frac{\sqrt{\mu_2}}{n} \left[C_4 - 3 \left(1 + \frac{C_3^2}{2} \right) \right] \quad (3.85)$$

$$Cov(S, C_s) = \frac{\sqrt{\mu_2}}{n} \left[\frac{C_5}{2} - \frac{C_3}{4} (3C_4 + 5) \right] \quad (3.86)$$

Enfin, pour obtenir le vecteur V_p des variances et des covariances des estimateurs déduits de la méthode des moments, on construit la matrice V à partir des termes A_j (voir annexe B) et on post-multiplie son inverse par le vecteur V_m (vecteur des variances et covariances des moments estimés).

3.2.3 Méthode des moments pondérés

Lorsque k est différent de zéro, on peut exprimer sous forme inverse la fonction de distribution de la loi GEV donnée par la relation (1.4), de la façon suivante :

$$X(F) = u + \frac{\alpha}{k} [1 - (-\ln F)^k] \quad (3.87)$$

Les moments pondérés β_r de cette loi s'expriment sous la forme générale suivante (Hosking et al, 1985):

$$\beta_r = \frac{1}{1+r} \left\{ u + \frac{\alpha}{k} \left[1 - \frac{\Gamma(1+k)}{(1+r)^k} \right] \right\} \quad (3.88)$$

où r représente l'ordre du moment. On peut donc exprimer à partir de cette équation les moments pondérés d'ordre 0, 1 et 2 nécessaires à la détermination des estimateurs \hat{u} , $\hat{\alpha}$ et \hat{k} des paramètres u , α et k . Ensuite, après quelques manipulations algébriques on déduit les relations suivantes:

$$\beta_0 = u + \frac{\alpha}{k} [1 - \Gamma(1+k)] \quad (3.89)$$

$$2\beta_1 - \beta_0 = \frac{\alpha}{k} [\Gamma(1+k)(1-2^{-k})] \quad (3.90)$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1 - 3^{-k}}{1 - 2^{-k}} \quad (3.91)$$

La méthode d'estimation des paramètres utilisant ces moments pondérés consiste à trouver les valeurs de u , α et k qui sont solutions du système formé des équations (3.89) à (3.91) lorsque β_0 , β_1 et β_2 sont remplacées par leurs estimateurs non biaisés (b_0 , b_1 et b_2) obtenus à partir de l'échantillon (voir annexe B).

Ainsi, on peut déterminer l'estimateur des moments pondérés \hat{k} de k à partir de l'équation:

$$\frac{3b_2 - b_0}{2b_1 - b_0} = \frac{1 - 3^{-k}}{1 - 2^{-k}} \quad (3.92)$$

qui ne dépend que de ce paramètre, pour ensuite déduire les estimateurs \hat{u} et $\hat{\alpha}$ de u et α à l'aide des relations :

$$\hat{\alpha} = \hat{k} \left[\frac{2b_1 - b_0}{\Gamma(1+\hat{k})(1-2^{-\hat{k}})} \right] \quad (3.93)$$

$$\hat{u} = b_0 + \frac{\hat{\alpha}}{\hat{k}} [\Gamma(1+\hat{k}) - 1] \quad (3.94)$$

La solution de l'équation (3.92) requiert l'utilisation de méthodes itératives. Toutefois, Hosking et al. (1985) mentionnent que, puisque le terme de droite de l'équation (3.92) est presque linéaire pour des valeurs de k comprises entre -0.5 et 0.5 (ce qui est généralement le cas en pratique), il est possible d'utiliser des approximations polynômiales. Ces auteurs proposent d'utiliser la relation :

$$\hat{k} = 7.8590c + 2.9554c^2 \quad (3.95)$$

où

$$c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\log 2}{\log 3} \quad (3.96)$$

Les variances et covariances des paramètres estimés par la méthode des moments pondérés sont aussi obtenus à partir de l'expression matricielle présentée à l'équation (3.65) et décrite à l'annexe B (Hosking et al. 1985). En utilisant la même notation, les termes A_{rj} sont donnés ici par:

$$A_{rj} = \frac{\partial M_r}{\partial \theta_j}, \quad r \text{ et } j = 1, 2, 3 \quad (3.97)$$

où $\theta_1 = u$, $\theta_2 = \alpha$, $\theta_3 = k$, et M_1 , M_2 et M_3 sont respectivement β_0 , β_1 et β_2 obtenus de l'équation (3.88). On déduit alors que :

$$A_{r1} = \frac{1}{1+r} \quad (3.98)$$

$$A_{r2} = \frac{1}{k(1+r)} \left[1 - \frac{\Gamma(1+k)}{(1+r)^k} \right] \quad (3.99)$$

$$A_{r3} = \frac{\alpha}{k^2(1+r)} \left\{ \frac{\Gamma(1+k)}{(1+r)^k} + k\Gamma(1+k) \left[\frac{\ln(1+r) - \Psi(1+k)}{(1+r)^k} \right] - 1 \right\} \quad (3.100)$$

Les expressions explicites pour les éléments du vecteur V_m (variances et covariances asymptotiques des moments pondérés estimés) sont obtenues en appliquant les résultats de Chernoff *et al.* (1967) concernant la loi asymptotique des combinaisons linéaires de statistiques d'ordre. Ces expressions dépendent de la fonction G définie de la manière suivante :

$$G(x) = {}_2F_1(k, 2k; 1+k; -x) \quad \text{avec } G(0) = 1 \quad (3.101)$$

où ${}_2F_1$ est la fonction hypergéométrique (Abramowitz et Stegun, 1970, p.556). En effet, on peut montrer que les variances et covariances asymptotiques des moments pondérés estimés sont données par (Hosking et al., 1985) :

$$Var(b_r) = \frac{1}{n} v_{rr} \quad (3.102)$$

$$Cov(b_r, b_{r,r+1}) = \frac{1}{n} v_{r,r+1} \quad (3.103)$$

et

$$Cov(b_r, b_{r,r+s}) = \frac{1}{n} v_{r,r+s} \quad (3.104)$$

où

$$v_{rr} = \frac{\alpha^2}{k^2(1+r)^{2k}} \left[\Gamma(1+2k)G\left(\frac{r}{1+r}\right) - \Gamma^2(1+k) \right] \quad (3.105)$$

$$v_{r,r+1} = \frac{\alpha^2}{2k^2} \left\{ \frac{\Gamma(1+2k)G\left(\frac{r}{2+r}\right)}{(2+r)^{2k}} + \frac{[(1+r)^{-k} - 2(2+r)^{-k}]\Gamma^2(1+k)}{(1+r)^k} \right\} \quad (3.106)$$

et

$$v_{r,r+s} = \frac{\alpha^2}{2k^2} \left\{ \frac{\Gamma(1+2k)G\left(\frac{r}{r+s+1}\right)}{(r+s+1)^{2k}} - \frac{\Gamma(1+2k)G\left(\frac{r+1}{r+s}\right)}{(r+s)^k} \right. \\ \left. + \frac{2[(r+s)^{-k} - (r+s+1)^{-k}]\Gamma^2(1+k)}{(1+r)^k} \right\} \quad (3.107)$$

De ces relations, on déduit aisément les variances et covariances asymptotiques des moments estimés qui nous intéressent en remplaçant r et s adéquatement. On obtient alors le vecteur V_m . Enfin, pour obtenir les variances et covariances des estimateurs, il suffit de construire la matrice V à partir des termes A_{rj} (voir annexe B) et de post-multiplier son inverse par le vecteur V_m .

CHAPITRE 4

ESTIMATION D'UN ÉVÉNEMENT DE PÉRIODE DE RETOUR T X_T ET CALCUL DE LA VARIANCE

Le quantile X_p d'ordre p est la valeur dont la probabilité au non-dépassement est p . Plus précisément, X_p est tel que :

$$Prob\{X < X_p\} = p \quad (4.1)$$

En hydrologie l'estimation des quantiles est utilisée fréquemment, en particulier pour analyser les séries de débits maximums annuels. Dans ce contexte, les hydrologues appellent cette statistique le débit de période de retour T , et le note X_T . Ils le définissent comme étant la valeur dont la probabilité d'être dépassée est $1/T$ où T est une intervalle de temps donné. Ainsi, X_T est tel que

$$Prob\{X > X_T\} = \frac{1}{T} \quad (4.2)$$

La connaissance de cette valeur est importante entre autres pour le dimensionnement de nouveaux ouvrages hydrauliques et aussi pour la gestion d'ouvrages déjà existants.

Nous présentons, dans ce chapitre, l'estimation des quantiles (débits de période de retour T) de la loi GEV ainsi que la détermination des variances asymptotiques qui y sont associées.

4.1 Loi GEV avec $k = 0$ (loi Gumbel ou EV1)

Soit un échantillon aléatoire de taille n x_1, x_2, \dots, x_n (par exemple des débits maximums annuels mesurés sur n années) tiré d'une loi Gumbel dont la fonction de densité de probabilité est donnée par (1.6). Alors, des équations (1.5) et (4.2), on déduit aisément que :

$$X_T = u - \alpha \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \quad (4.3)$$

Pour estimer cette quantité, il suffit de remplacer dans l'équation (4.3), les paramètres u et α par leurs estimateurs respectifs \hat{u} et $\hat{\alpha}$ présentés au chapitre 3. Nous obtenons alors l'estimation \hat{X}_T du débit de période de retour T :

$$\hat{X}_T = \hat{u} - \hat{\alpha} \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \quad (4.4)$$

Cet estimateur sera celui du maximum de vraisemblance, des moments ou des moments pondérés selon la méthode utilisée pour déterminer \hat{u} et $\hat{\alpha}$ (section 3.1).

En appliquant le théorème central limite (Lehmann, 1983) à la statistique \hat{X}_T , on déduit la variance asymptotique de cet estimateur:

$$Var(\hat{X}_T) = \left(\frac{\partial X_T}{\partial u} \right)^2 Var(\hat{u}) + \left(\frac{\partial X_T}{\partial \alpha} \right)^2 Var(\hat{\alpha}) + 2 \left(\frac{\partial X_T}{\partial u} \right) \left(\frac{\partial X_T}{\partial \alpha} \right) Cov(\hat{u}, \hat{\alpha}) \quad (4.5)$$

Or, en posant

$$Y_T = -\ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \quad (4.6)$$

et en évaluant les dérivées partielles, l'équation (4.5) se réduit à :

$$Var(\hat{X}_T) = Var(\hat{u}) + Y_T^2 Var(\hat{\alpha}) + 2Y_T Cov(\hat{u}, \hat{\alpha}) \quad (4.7)$$

En remplaçant $Var(\hat{u})$, $Var(\hat{\alpha})$ et $Cov(\hat{u}, \hat{\alpha})$ par les expressions obtenues à la section 3.1, on obtient la variance asymptotique de \hat{X}_T pour les trois méthodes d'estimation considérées dans ce travail. Plus précisément, on obtient respectivement pour la méthode du maximum de vraisemblance (MV1), la méthode du maximum de vraisemblance avec correction du biais (MV2), la méthode des moments (MO) et la méthode des moments pondérés (MP).

$$Var_{MV1}(\hat{X}_T) = \left(\frac{\alpha^2}{n} \right) [1.109 + 0.514Y_T + 0.608Y_T^2] \quad (4.8)$$

$$Var_{MV2}(\hat{X}_T) = \left(\frac{\alpha^2}{n} \right) \left\{ \left[1.109 - \frac{0.360}{n-0.8} + \frac{0.928}{n^2} \right] + \left[0.608 \left(\frac{n}{n-0.8} \right)^2 \right] Y_T^2 \right. \\ \left. + 2 \left[\left(\frac{0.257n}{n-0.8} \right) - \frac{0.426}{n} \right] Y_T \right\} \quad (4.9)$$

$$Var_{MO}(\hat{X}_T) = \left(\frac{\alpha^2}{n} \right) [1.168 + 0.192Y_T + 1.100Y_T^2] \quad (4.10)$$

et

$$Var_{MP}(\hat{X}_T) = \left(\frac{\alpha^2}{n(n-1)} \right) [(1.1128n - 0.9066) - (0.4574n - 1.1722)Y_T \\ + (0.8046n - 0.1855)Y_T^2] \quad (4.11)$$

4.2 Loi GEV avec $k \neq 0$ (loi EV2 et EV3)

Étant donné un échantillon aléatoire de taille n x_1, x_2, \dots, x_n tiré d'une loi GEV avec $k \neq 0$ dont la fonction de densité de probabilité est donnée par la relation (1.7), on déduit, des équations (1.4) et (4.2), que :

$$X_T = u + \frac{\alpha}{k} \left\{ 1 - \left[-\ln \left(1 - \frac{1}{T} \right) \right]^k \right\} \quad (4.12)$$

De la même façon qu'à la section précédente, pour obtenir un estimateur du débit de période de retour T , il suffit de remplacer dans l'équation (4.12) les paramètres u , α et k par leurs estimateurs respectifs issus de la méthode du maximum de vraisemblance, de la méthode des moments ou de la méthode des moments pondérés (section 3.2) :

$$\hat{X}_T = \hat{u} + \frac{\hat{\alpha}}{\hat{k}} \left\{ 1 - \left[-\ln \left(1 - \frac{1}{T} \right) \right]^{\hat{k}} \right\} \quad (4.13)$$

D'après le théorème central limite (Lehmann, 1983), la variance asymptotique de \hat{X}_T est donnée par :

$$\begin{aligned} Var(\hat{X}_T) &= \left(\frac{\partial X_T}{\partial u}\right)^2 Var(\hat{u}) + \left(\frac{\partial X_T}{\partial \alpha}\right)^2 Var(\hat{\alpha}) + \left(\frac{\partial X_T}{\partial k}\right)^2 Var(\hat{k}) \\ &+ 2\left(\frac{\partial X_T}{\partial u}\right)\left(\frac{\partial X_T}{\partial \alpha}\right)Cov(\hat{u}, \hat{\alpha}) + 2\left(\frac{\partial X_T}{\partial u}\right)\left(\frac{\partial X_T}{\partial k}\right)Cov(\hat{u}, \hat{k}) + 2\left(\frac{\partial X_T}{\partial \alpha}\right)\left(\frac{\partial X_T}{\partial k}\right)Cov(\hat{\alpha}, \hat{k}) \end{aligned} \quad (4.14)$$

Les dérivées partielles sont calculées à partir de l'équation (4.12) en posant:

$$W_T = \frac{1}{k} \left\{ 1 - \left[-\ln\left(1 - \frac{1}{T}\right) \right]^k \right\} = \frac{X_T - u}{\alpha} \quad (4.15)$$

En substituant les dérivées obtenues dans l'équation (4.14), on obtient:

$$\begin{aligned} Var(\hat{X}_T) &= Var(\hat{u}) + Var(\hat{\alpha})W_T^2 + \left(\alpha \frac{\partial W_T}{\partial k}\right)^2 Var(\hat{k}) + 2W_T Cov(\hat{u}, \hat{\alpha}) \\ &+ 2\alpha \left(\frac{\partial W_T}{\partial k}\right) Cov(\hat{u}, \hat{k}) + 2W_T \alpha \left(\frac{\partial W_T}{\partial k}\right) Cov(\hat{\alpha}, \hat{k}) \end{aligned} \quad (4.16)$$

où $\partial W_T / \partial k$ est donnée par :

$$\frac{\partial W_T}{\partial k} = \frac{1}{k} \left[\ln\left(1 - \frac{1}{T}\right) \right]^k [\ln(W_T) + W_T] \quad (4.17)$$

Enfin, pour déterminer la variance asymptotique du débit maximum annuel de période de retour T , on doit remplacer $Var(\hat{u})$, $Var(\hat{\alpha})$, $Var(\hat{k})$, $Cov(\hat{u}, \hat{\alpha})$, $Cov(\hat{u}, \hat{k})$ et $Cov(\hat{\alpha}, \hat{k})$ par les valeurs déduites des méthodes décrites à la section 3.2.

CHAPITRE 5

TEST DU RAPPORT DES VRAISEMBLANCES MAXIMALES POUR VÉRIFIER L'HYPOTHESE " $H_0 : k = 0$ "

On a déjà noté que la loi GEV regroupe en une seule expression les trois types de lois des valeurs extrêmes (EV1, EV2 et EV3) développées par Fisher et Tippett (1928). En effet, si le paramètre k est nul la loi GEV est une loi Gumbel (EV1), s'il est négatif une loi EV2, et s'il est positif une loi EV3. Cette propriété est particulièrement intéressante en pratique puisqu'on est souvent amené à utiliser une loi des valeurs extrêmes sans en connaître a priori le type. Or, les méthodes présentées au chapitre 3, permettent d'estimer le paramètre k , donc de nous informer sur le type de loi des valeurs extrêmes le plus approprié aux observations.

En particulier, lorsque \hat{k} est voisin de zéro, il est préférable de reprendre l'analyse des données en ajustant directement une loi de Gumbel. En effet, puisque cette loi ne possède que deux paramètres, l'estimation de ceux-ci ainsi que celle des quantiles X_T sera plus précise. Il est donc intéressant de développer une procédure nous permettant de tester l'hypothèse " $H_0 : k = 0$ " contre " $H_1 : k \neq 0$ " lors de l'ajustement de la loi GEV.

Le test proposé dans ce chapitre est un test du rapport des vraisemblances maximales. Cette procédure de construction de tests d'hypothèses est reconnue pour ses propriétés asymptotiques fort intéressantes (Lehmann, 1986). En particulier, pour de grandes tailles d'échantillon, les tests issus de cette procédure sont uniformément les plus puissants, c'est-à-dire qu'ils rejettent plus souvent l'hypothèse H_0 que tout autre test lorsque l'hypothèse H_1 est vraie. De plus, les résultats d'une étude comparative sur 13 tests menée par Hosking (1984) pour examiner l'hypothèse H_0 montre que le test du rapport des vraisemblances maximales donne les meilleurs résultats, et ce pour des tailles d'échantillon comprises entre 25 et 500 observations.

La statistique T du test déduit de cette procédure compare les vraisemblances maximales sous les hypothèses H_0 (les observations proviennent d'une loi Gumbel) et H_1 (les observations proviennent d'une loi GEV avec $k \neq 0$). Plus précisément,

$$T = \frac{L_{H_0}(\hat{u}, \hat{\alpha})}{L_{H_1}(\hat{u}, \hat{\alpha}, \hat{k})} \quad (5.1)$$

où $L_{H_0}(\hat{u}, \hat{\alpha})$ et $L_{H_1}(\hat{u}, \hat{\alpha}, \hat{k})$ sont respectivement les fonctions de vraisemblance de la loi Gumbel et de la loi GEV calculées pour les valeurs des paramètres estimées par la méthode du maximum de vraisemblance. On remarque que de grandes valeurs de T soutiennent l'hypothèse H_0 (plus "vraisemblable" que H_1) tandis que de petites valeurs de cette statistique incitent à la rejeter.

Pour déterminer la valeur critique du test nous permettant de rejeter l'hypothèse nulle à un niveau de signification donné, il faut déterminer la loi de probabilité de la statistique T sous l'hypothèse nulle (H_0). Cette loi étant fort complexe, on emploie en général une approximation. On peut montrer, pour le test du rapport des vraisemblances maximales décrit plus haut, que la statistique $t = -2 \ln(T)$ est asymptotiquement distribuée selon une loi chi-deux à un degré de liberté. La nouvelle statistique t , qui est équivalente à T , peut se réécrire après quelques manipulations algébriques de la façon suivante:

$$t = 2\{\ln L_{H_1}(\hat{u}, \hat{\alpha}, \hat{k}) - \ln L_{H_0}(\hat{u}, \hat{\alpha})\} \quad (5.2)$$

où $\ln L_{H_0}(\hat{u}, \hat{\alpha})$ et $\ln L_{H_1}(\hat{u}, \hat{\alpha}, \hat{k})$ sont respectivement les logarithmes de la fonction de vraisemblance de la loi Gumbel (équation 3.1) et de la loi GEV (équation 3.35) évaluées en \hat{u} et $\hat{\alpha}$, les estimations obtenues par la méthode du maximum de vraisemblance.

Cette statistique définit donc un test dont la valeur critique, pour un niveau de signification donné α , est le $(1 - \alpha)^{\text{ième}}$ quantile de la loi de chi-deux à un degré de liberté. Plus précisément, la règle de décision est la suivante :

On rejette l'hypothèse H_0 lorsque la statistique calculée à l'aide des observations est plus grande que le quantile : si $\hat{k} < 0$, on peut en conclure que les observations proviennent d'une loi EV2 et si $\hat{k} > 0$, qu'elles proviennent de la loi EV3.

On accepte l'hypothèse H_0 (k n'est pas significativement différent de zéro) lorsque la statistique est inférieure au quantile : les observations sont alors distribuées selon une loi Gumbel.

Le logiciel AJUSTE donne la possibilité à l'utilisateur d'effectuer ce test lors de l'ajustement de la loi GEV. Selon les résultats obtenus, on peut ensuite décider d'ajuster les observations à l'aide d'une loi Gumbel (section 3.1) ou d'utiliser directement les estimations déduites de la loi GEV (section 3.2). Notons que le logiciel utilise une statistique t modifiée proposée par Lawley (1956). La loi de cette statistique, que nous notons t^* , définie de la manière suivante :

$$t^* = \left(1 - \frac{2.8}{n}\right)t \quad (5.3)$$

est approchée plus efficacement par la loi du chi-deux que la loi de la statistique t . L'approximation est donc meilleure, en particulier pour des échantillons de petite taille. Evidemment, cette modification ne change en rien la règle de décision présentée plus haut.

CONCLUSION

Ce rapport effectue une synthèse des propriétés mathématiques et statistiques de la loi GEV présentées dans différentes études. Trois méthodes d'estimation des paramètres ont été présentées : la méthode du maximum de vraisemblance, la méthode des moments et la méthode des moments pondérés. Pour chacune des méthodes d'estimation l'étude théorique des variances et des covariances asymptotiques des estimateurs a été effectuée. De plus, nous avons déterminé les estimateurs des quantiles X_T et leur variance asymptotique. Enfin, un test permettant de sélectionner le type de loi des valeurs extrêmes a été développé.

Tous ces développements ont été effectués dans le but d'introduire dans le logiciel AJUSTE la loi GEV et la loi Gumbel. Chacune des méthodes d'estimation mentionnées plus haut ont été intégrées au programme AJUSTE (logiciel permettant de faire l'ajustement automatique d'une distribution théorique choisie à une série de données observées). Ce logiciel fournit aussi les quantiles estimés pour 21 probabilités au dépassement ainsi que les intervalles de confiance asymptotiques associés à ces événements pour divers niveaux de confiance.

En tentant de faire le point sur la loi GEV, ce travail devrait permettre de faciliter l'utilisation de cette loi de probabilité qui possède des propriétés intéressantes et qui est susceptible d'une application plus large au Canada.

BIBLIOGRAPHIE

Abramowitz, M. et Stegun, I.A. (1970). *Handbook of Mathematical Functions*. Dover Publications, Inc., New York : 1046 p.

Chernoff, H., Gastwirth, J.L. et Johns, M.V. (1967). Asymptotic distribution of functions of order statistics with applications to estimation. *Ann. Math. Statist.*, **38**, 52-72.

Clarke, R.T. (1973). Mathematical models in hydrology. FAO, Rome, *Irrig. Drain. Pap.*, **19**, 133-136.

Downton, F. (1966). Linear estimates of parameters in extreme value distribution. *Technometrics*, **8**, 3-17.

Fiorentino, M. et Gabriele, S. (1984). A correction for the bias of maximum likelihood estimators of Gumbel parameters. *J. Hydrol.*, **73**, 39-49.

Fisher, R.A. et Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190.

Greenwood, J.A., Landwerh, J.M., Matalas, N.C. and J.R. Wallis (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Wat. Resour. Res.*, **15**, 1049-1054.

Gumbel, E.J. (1960). *Statistics of extremes*. Columbia University Press, New York, N.Y.

Hosking, J.R.M., Wallis, J.R. et E.F. Wood (1985). Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technometrics*, **27**, 251-261.

Hosking, J.R.M. (1984). Testing whether the shape parameter is zero in the generalized extreme-value distribution. *Biometrika*, **71**, 367-374.

Hosking, J.R.M. (1985). Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics*, **34**, 301-310.

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart. J. Roy. Meteor. Soc.*, **81**, 158-171.

Jenkinson, A.F. (1969). *Statistics of Extremes*, Technical Note 98, World Meteorological Office, Geneva.

Kendall, M.G. et Stuart, A. (1987). *Advanced Theory of Statistics*. Oxford University Press, New York : 604 p.

Landwehr, J.M., Matalas, N.C. and J.R. Wallis (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Wat. Resour. Res.*, **15**, 1055-1064.

Lawley, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295-303.

Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York : 497 p.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses*. Wiley, New York : 600 p.

Lowery, M.D., Nash, J.E., (1970). A comparison of methods of fitting the double exponential distribution. *J. Hydrol.*, **10**, 259-275.

Natural Environment Research Council (1975). *Flood Studies Report (Volume I)*, London : 550 p.

Phien, H.N. (1987). A review of methods of parameter estimation for the extreme value type-1 distribution. *J. Hydrol.*, **90**, 251-268.

Prescott, P., et Walden, A.T. (1980). Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, **67**, 723-724.

Prescott, P., et Walden, A.T. (1983). Maximum-likelihood estimation of the three-parameter generalized extreme-value distribution from censored samples. *J. Statist. Comput. Simul.*, **16**, 241-250.

ANNEXE A

**LA LOI GUMBEL :
CAS LIMITE DE LA LOI GEV**

La fonction de distribution $F(x)$ de la loi GEV pour $k \neq 0$ est donnée à l'équation (1.4). On montre ici que la fonction de distribution $G(x)$ de la loi Gumbel (GEV avec $k = 0$, équation 1.5) est obtenue en prenant la limite, lorsque k tend vers zéro, de $F(x)$:

$$\begin{aligned} \lim_{k \rightarrow 0} F(x) &= \lim_{k \rightarrow 0} \exp\left[-\left(1 - \frac{k}{\alpha}(x-u)\right)^{1/k}\right] & (A.1) \\ &= \lim_{k \rightarrow 0} \exp\left\{-\exp\left[\frac{1}{k} \ln\left(1 - \frac{k}{\alpha}(x-u)\right)\right]\right\} \\ &= \exp\left\{-\exp\left[\lim_{k \rightarrow 0} \frac{1}{k} \ln\left(1 - \frac{k}{\alpha}(x-u)\right)\right]\right\} \end{aligned}$$

Or,

$$\lim_{k \rightarrow 0} \frac{1}{k} \ln\left(1 - \frac{k}{\alpha}(x-u)\right) = \frac{0}{0} \quad (A.2)$$

La limite de cette fonction est donc indéterminée. Par contre, en appliquant la règle de l'Hospital, on obtient :

$$\begin{aligned} \lim_{k \rightarrow 0} \frac{1}{k} \ln\left(1 - \frac{k}{\alpha}(x-u)\right) &= \lim_{k \rightarrow 0} \frac{\frac{d}{dk} \ln\left(1 - \frac{k}{\alpha}(x-u)\right)}{\frac{d}{dk} k} & (A.3) \\ &= \lim_{k \rightarrow 0} -\frac{x-u}{\alpha - k(x-u)} \\ &= -\left(\frac{x-u}{\alpha}\right) \end{aligned}$$

On déduit ensuite de (A.1) et (A.3) le résultat :

$$\lim_{k \rightarrow 0} F(x) = \exp\left\{-\exp\left[-\left(\frac{x-u}{\alpha}\right)\right]\right\} = G(x) \quad (A.4)$$

ANNEXE B

DESCRIPTION DES METHODES D'ESTIMATION

Dans cette section nous décrivons brièvement les trois méthodes d'estimation considérées dans le présent rapport, et présentons le calcul de la matrice des variances et covariances asymptotiques qui est nécessaire à la détermination de $Var(\hat{X}_T)$, la variance asymptotique de l'événement de période de retour T . Pour plus de détail, nous vous invitons à consulter les ouvrages cités.

B.1 Méthode du maximum de vraisemblance

Soit un échantillon aléatoire de taille n x_1, x_2, \dots, x_n tiré d'une loi de fonction de densité de probabilité $f(x; \theta_1, \theta_2, \dots, \theta_p)$ à p paramètres. Les estimateurs du maximum de vraisemblance sont obtenus en maximisant par rapport à $\theta_1, \theta_2, \dots, \theta_p$ la fonction de vraisemblance L telle que :

$$L(\theta_1, \theta_2, \dots, \theta_p) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_p) \quad (\text{B.1})$$

En pratique, pour des raisons de simplicité, on peut être amené à maximiser plutôt le logarithme de cette fonction :

$$\ln L(\theta_1, \theta_2, \dots, \theta_p) = \sum_{i=1}^n \ln f(x_i; \theta_1, \theta_2, \dots, \theta_p) \quad (\text{B.2})$$

Maximiser (B.1) ou (B.2) par rapport aux paramètres $\theta_i, i = 1, \dots, p$, est équivalent puisque l'on a :

$$\frac{\partial \ln L}{\partial \theta_i} = \frac{1}{L} \frac{\partial L}{\partial \theta_i}$$

Pour déterminer les estimateurs du maximum de vraisemblance, il suffit alors de résoudre le système d'équations :

$$\begin{aligned}\frac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_p)}{\partial \theta_1} &= 0 \\ \frac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_p)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_p)}{\partial \theta_p} &= 0\end{aligned}$$

Souvent, ce système n'admet pas de solution explicite, et on doit le résoudre numériquement. En général, on utilise une méthode itérative du type Newton-Raphson.

Les propriétés asymptotiques des estimateurs obtenus par la méthode du maximum de vraisemblance sont bien connues. En particulier, ces estimateurs sont convergents, asymptotiquement non-biaisés et asymptotiquement efficaces (voir Lehmann, 1983). De plus, si la fonction de vraisemblance admet un seul maximum, les variables aléatoires $\sqrt{n}(\hat{\theta}_1 - \theta_1)$, $\sqrt{n}(\hat{\theta}_2 - \theta_2)$, ..., $\sqrt{n}(\hat{\theta}_k - \theta_k)$ sont asymptotiquement distribuées selon une loi normale multivariée de moyenne $(0, 0, \dots, 0)$ et de matrice variance-covariances Σ dont les éléments correspondent à ceux de l'inverse de la matrice d'information de Fisher I_f :

$$(I_f)_{ij} = -E\left(\frac{\partial^2 \ln L(\theta_1, \theta_2, \dots, \theta_p)}{\partial \theta_i \partial \theta_j}\right), \quad (i, j) \in \{1, 2, \dots, p\}^2 \quad (\text{B.3})$$

B.2 Méthode des moments

Soit un échantillon aléatoire de taille n x_1, x_2, \dots, x_n tiré d'une loi $F(x; \theta_1, \theta_2, \dots, \theta_p)$ à p paramètres. L'espérance mathématique $E(X^r)$ est définie de la manière suivante :

$$E(X^r) = \int_{-\infty}^{+\infty} x^r f(x; \theta_1, \theta_2, \dots, \theta_p) dx, \quad r = 1, 2, 3, \dots \quad (\text{B.4})$$

où $f(\cdot)$ est la fonction de densité de probabilité de la loi $F(\cdot)$. $E(X^r)$ est le moment théorique non centré d'ordre r . Le moment non centré d'ordre r de l'échantillon, qui est un estimateur de $E(X^r)$, est défini de la façon suivante :

$$m_r = \frac{1}{n} \sum_{i=1}^n x_i^r \quad (\text{B.5})$$

Les estimateurs obtenus par la méthode des moments sont les valeurs $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ sont la solution du système de p équations :

$$\begin{aligned} E(X) &= m_1 \\ E(X^2) &= m_2 \\ &\vdots \\ E(X^p) &= m_p \end{aligned}$$

où $E(X^i)$, $i = 1, \dots, p$, sont des fonctions des paramètres. Ici aussi, la résolution de ce système d'équations nécessite en général l'emploi de méthodes numériques. Notons que de façon équivalente, nous pouvons obtenir les estimateurs de la méthodes des moments en utilisant les moments centrés plutôt que les moments non-centrés.

Les variances et covariances asymptotiques des estimateurs déduits de la méthode des moments sont reliées à celles des moments de l'échantillon. Plus précisément, on a :

$$Cov(m_r, m_q) = \sum_{j=1}^p \left(\frac{\partial M_r}{\partial \theta_j} \right) \left(\frac{\partial M_q}{\partial \theta_j} \right) Var(\hat{\theta}_j) + \sum_{j=1}^p \sum_{\substack{i=1 \\ i \neq j}}^p \left(\frac{\partial M_r}{\partial \theta_j} \right) \left(\frac{\partial M_q}{\partial \theta_i} \right) Cov(\hat{\theta}_j, \hat{\theta}_i)$$

où $M_r = E(X^r)$ et $M_q = E(X^q)$ sont les moments théoriques donnés par l'équation (B.4) de la loi pour r et $q \in \{1, 2, \dots, p\}$.

Dans le cas des lois à trois paramètres (comme GEV), les trois premiers moments non centrés sont suffisants pour estimer les paramètres. L'application de la dernière équation conduit au système suivant :

$$\begin{bmatrix} Var(m_1) \\ Var(m_2) \\ Var(m_3) \\ Cov(m_1, m_2) \\ Cov(m_1, m_3) \\ Cov(m_2, m_3) \end{bmatrix} = \begin{bmatrix} V_{11} & \dots & V_{16} \\ \cdot & & \cdot \\ V_{61} & \dots & V_{66} \end{bmatrix} \cdot \begin{bmatrix} Var(\hat{\theta}_1) \\ Var(\hat{\theta}_2) \\ Var(\hat{\theta}_3) \\ Cov(\hat{\theta}_1, \hat{\theta}_2) \\ Cov(\hat{\theta}_1, \hat{\theta}_3) \\ Cov(\hat{\theta}_2, \hat{\theta}_3) \end{bmatrix}$$

ce qui peut s'écrire sous forme matricielle :

$$V_m = V \cdot V_p$$

V_m étant le vecteur des variances et covariances asymptotiques des moments de l'échantillon, V_p celui des variances et covariances asymptotiques des estimateurs des paramètres et V la matrice définie comme suit :

$$V = \begin{bmatrix} A_{11}^2 & A_{12}^2 & A_{13}^2 & 2A_{11}A_{12} & 2A_{11}A_{13} & 2A_{12}A_{13} \\ A_{21}^2 & A_{22}^2 & A_{23}^2 & 2A_{21}A_{22} & 2A_{21}A_{23} & 2A_{22}A_{23} \\ A_{31}^2 & A_{32}^2 & A_{33}^2 & 2A_{31}A_{32} & 2A_{31}A_{33} & 2A_{32}A_{33} \\ A_{11}A_{21} & A_{12}A_{22} & A_{13}A_{23} & (A_{11}A_{22} + A_{12}A_{21}) & (A_{11}A_{23} + A_{13}A_{21}) & (A_{12}A_{23} + A_{13}A_{22}) \\ A_{11}A_{31} & A_{12}A_{32} & A_{13}A_{33} & (A_{11}A_{32} + A_{12}A_{31}) & (A_{11}A_{33} + A_{13}A_{31}) & (A_{12}A_{33} + A_{13}A_{32}) \\ A_{21}A_{31} & A_{22}A_{32} & A_{23}A_{33} & (A_{21}A_{32} + A_{22}A_{31}) & (A_{21}A_{33} + A_{23}A_{31}) & (A_{22}A_{33} + A_{23}A_{32}) \end{bmatrix}$$

où

$$A_{rj} = \frac{\partial M_r}{\partial \theta_j} \quad (r, j = 1, 2, 3)$$

Les termes A_{ij} de la matrice V peuvent être déterminés à partir des relations entre les moments M_r et les paramètres θ_j obtenues par l'équation (B.4).

Les éléments du vecteur V_m peuvent être obtenus pour tout moment de l'échantillon à partir des expressions générales définies dans Kendall et Stuart (1987). Une fois V_m et V déterminés, on déduit aisément les variances et covariances asymptotiques des estimateurs des moments (le vecteur V_p) par la relation :

$$V_p = V^{-1} \cdot V_m$$

B.3 Méthode des moments pondérés

La méthode des moments pondérés a été introduite par Greenwood et al (1979). Elle a été développée principalement pour les lois qui peuvent être exprimées explicitement sous forme inverse:

$$X = X(F) \tag{B.6}$$

Soit un échantillon aléatoire de taille n x_1, x_2, \dots, x_n tiré d'une telle loi $F(x; \theta_1, \theta_2, \dots, \theta_p)$ à p paramètres (par exemple la loi GEV). Les moments pondérés sont définis de la manière suivante:

$$M_{l,r,k} = E\{X^l F^r (1-F)^k\} = \int_0^1 [x(F)]^l F^r (1-F)^k dF \tag{B.7}$$

où l, r et k sont des nombres entiers et F la fonction de distribution. Toutefois, en pratique, on considère généralement les moments pondérés suivants correspondant au cas particulier où $l = 1$ et $k = 0$:

$$\beta_r = M_{1,r,0} = E\{XF^r\}, r = 0, 1, 2, \dots \tag{B.8}$$

Dans cette expression on remarque que la variable X est toujours utilisée à la puissance un, ce qui, pour plusieurs lois, simplifie les expressions analytiques des moments pondérés.

Etant donné l'échantillon ordonné de taille n , $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, une réalisation des statistiques d'ordre $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, la statistique b_r telle que

$$b_r = \frac{1}{n} \sum_{j=1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{(j)} \quad r = 0, 1, 2, \dots \quad (\text{B.9})$$

est un estimateur non-biaisé de β_r (Landwehr *et al.*, 1979). De façon analogue à la méthode des moments classique, les estimateurs des moments pondérés $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ constituent la solution du système des p équations suivantes :

$$\begin{aligned} \beta_1 &= b_1 \\ \beta_2 &= b_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ \beta_p &= b_p \end{aligned}$$

La résolution de ce système peut, ici aussi, nécessiter en pratique l'utilisation de méthodes numériques.

Le calcul des variances et covariances asymptotiques des estimateurs déduits de la méthode des moments pondérés peut être effectué de la même façon que dans la méthode classique des moments. En effet, on a :

$$Cov(b_r, b_q) = \sum_{j=1}^k \left(\frac{\partial \beta_r}{\partial \theta_j} \right) \left(\frac{\partial \beta_q}{\partial \theta_j} \right) Var(\hat{\theta}_j) + \sum_{j=1}^k \sum_{\substack{i=1 \\ i \neq j}}^k \left(\frac{\partial \beta_r}{\partial \theta_j} \right) \left(\frac{\partial \beta_q}{\partial \theta_i} \right) Cov(\hat{\theta}_j, \hat{\theta}_i)$$

En supposant que les trois premiers moments pondérés ($r = 0, 1$ et 2) sont suffisants pour estimer les paramètres de la loi (c'est le cas pour la loi GEV), la procédure matricielle décrite à la section B.2 peut être appliquée. Ici, les éléments de la matrice V sont fonction des termes A_{rj} définis par :

$$A_{rj} = \frac{\partial \beta_r}{\partial \theta_j} \quad (r = 0, 1, 2 \quad ; \quad j = 1, 2, 3)$$

et le vecteur V_m est constitué des variances et covariances asymptotiques des trois premiers moments pondérés de l'échantillon (b_0, b_1 et b_2).

Pour déterminer les variances et covariances asymptotiques des moments pondérés de l'échantillon, il suffit d'appliquer le résultat de Chernoff *et al.* (1967). Ces auteurs ont montré, sous certaines conditions de régularité, qu'une combinaison linéaire de statistiques d'ordre L_n définie comme suit :

$$L_n = \frac{1}{n} \sum_{j=1}^n c_{nj} X_{(j)} \tag{B.10}$$

est asymptotiquement distribuée selon une loi normale de moyenne :

$$E(L_n)$$

et de variance :

$$\sigma^2(J, F) = 2 \int \int_{x < y} J\{F(x)\} J\{F(y)\} F(x) \{1 - F(y)\} dx dy$$

où la fonction $J(u)$ ($u \in [0, 1]$) , appelée fonction de poids, est la façon usuelle d'écrire les coefficients c_{nj} dans la littérature :

$$c_{nj} = J\left\{\frac{j}{n+1}\right\}$$

La statistique b_r peut être écrite sous forme (B.10) où les poids c_{nj} sont :

$$c_{nj} = \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} \quad (\text{B.11})$$

Lorsque $n \rightarrow \infty$, $j \rightarrow \infty$ et $j/n \rightarrow \theta$ ($0 < \theta < 1$), on montre aisément que

$$c_{nj} \approx \left\{ \frac{j}{n+1} \right\}^r \Rightarrow J(u) = u^r \quad (\text{B.12})$$

et qu'ainsi b_r satisfait les conditions du théorème I de Chernoff *et al.* (1967). On déduit donc, en appliquant le résultat, que :

$$n^{-1} \text{Var}(b_r) = 2n^{-1} \int \int_{x < y} \{F(x)\}^r \{F(y)\}^r F(x) \{1-F(y)\} dx dy$$

Une généralisation de ce théorème peut être appliquée à toute combinaison linéaire de statistiques L_n . Par conséquent, on peut déterminer la covariance asymptotique des moments pondérés de l'échantillon. Plus précisément, on a :

$$\text{Cov}(b_r, b_s) = \frac{1}{2}(g_{rs} + g_{sr})$$

où

$$g_{rs} = 2 \int \int_{x < y} \{F(x)\}^r \{F(y)\}^s F(x) \{1-F(y)\} dx dy$$

ANNEXE C

**VARIANCES ET COVARIANCES DES
MOMENTS DE L'ECHANTILLON**

Cette annexe donne les principaux résultats permettant de déduire les variances et covariances des moments de l'échantillon (éq. 3.81 à 3.86).

Soient μ_r et μ_r les moments de la population non centrés et centrés par rapport à la moyenne. Considérons de plus un échantillon aléatoire x_1, x_2, \dots, x_n . Les estimateurs respectifs de μ_r et μ_r :

$$m_r' = \frac{1}{n} \sum_{i=1}^n x_i^r \quad (\text{C.1})$$

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - m_1)^r \quad (\text{C.2})$$

On peut démontrer (Kendall et Stuart, 1987, Chap. 10), en supposant l'existence des moments appropriés, les relations générales suivantes :

$$\text{Cov}(m_q', m_r') = \frac{\mu_{q+r} - \mu_q' \mu_r'}{n} \quad (\text{C.3})$$

$$\text{Cov}(m_r, m_q) = \frac{\mu_{r+q} - \mu_r \mu_q + r q \mu_2 \mu_{r-1} \mu_{q-1} - r \mu_{r-1} \mu_{q+1} - q \mu_{r+1} \mu_{q-1}}{n} \quad (\text{C.4})$$

d'où l'on déduit, en posant $q = r$:

$$\text{Var}(m_r') = \frac{\mu_{2r} - \mu_r'^2}{n} \quad (\text{C.5})$$

$$\text{Var}(m_r) = \frac{\mu_{2r} - \mu_r^2 + r^2 \mu_2 \mu_{r-1}^2 - 2r \mu_{r-1} \mu_{r+1}}{n}$$

Si l'on ajoute à ces relations le résultat suivant :

$$\text{Var}\left(\frac{m_r}{m_q}\right) = \left[\frac{E(m_r)}{E(m_q)}\right]^2 \left[\frac{\text{Var}(m_r)}{E^2(m_r)} + \frac{\text{Var}(m_q)}{E^2(m_q)} - \frac{2\text{Cov}(m_r, m_q)}{E(m_r)E(m_q)} \right] \quad (\text{C.6})$$

que l'on peut aussi appliquer aux moments non centrés (m_r), où $E(.)$ désigne l'espérance mathématique, nous avons tout pour démontrer les relations données en (3.81) à (3.86).

Nous ne donnons pas les détails des calculs ici, nous vous référons plutôt à Kendall et Stuart (1987) où ces résultats sont présentés.