

Rapport Scientifique No 351

par

Luc Perreault
Bernard Bobée

Loi Weibull à deux paramètres
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	ii
INTRODUCTION	1
CHAPITRE 1 : PROPRIÉTÉS MATHÉMATIQUES ET STATISTIQUES DE LA LOI WEIBULL A DEUX PARAMETRES.....	3
1.1 Définition et propriétés mathématiques.....	3
1.2 Propriétés statistiques	5
CHAPITRE 2 : ESTIMATION DES PARAMETRES DE LA LOI WEIBULL.....	8
2.1 Méthode du maximum de vraisemblance.....	9
2.2 Méthode des moments.....	11
CHAPITRE 3 : ESTIMATION D'UN ÉVÉNEMENT DE PÉRIODE DE RETOUR $T X_T$ ET CALCUL DE LA VARIANCE.....	15
CHAPITRE 4 : LOI WEIBULL A TROIS PARAMETRES.....	17
CONCLUSION.....	21
BIBLIOGRAPHIE.....	22
ANNEXE A : DESCRIPTION DES MÉTHODES D'ESTIMATION.....	25

LISTE DES TABLEAUX ET FIGURES

Tableau 1. Revue de littérature non-exhaustive concernant l'estimation des paramètres de la loi Weibull à 3 paramètres	19
Figure 1. Forme de la f.d.p. de la loi Weibull pour différentes valeurs de c et $\alpha = 1$	4

INTRODUCTION

Diverses fonctions de densités de probabilité ont été proposées afin d'effectuer des analyses statistiques sur les débits de rivières dans le but d'estimer les extrêmes (crues ou étiages). Cette estimation est un outil important lors de la planification des aménagements hydrauliques.

Dans ce rapport, on présente la loi Weibull à deux paramètres. Cette loi de probabilité a été largement utilisée comme modèle probabiliste dans des études sur les temps de survie. En effet, il a été démontré, dans plusieurs domaines, que cette loi s'ajuste fort bien à ce type de données (Lawless, 1982). Toutefois, ce n'est pas cette propriété qui justifie l'intérêt d'employer cette loi en hydrologie, mais plutôt le lien qui l'associe aux lois des valeurs extrêmes. En effet, elle se prête bien à l'étude d'événements tels que les crues, les précipitations maximales ou minimales, les étiages, etc.. Mann (1968) discute de cette caractéristique de la loi Weibull.

Nous décrivons premièrement, dans le chapitre I du présent rapport, les propriétés mathématiques et statistiques de cette loi de probabilité. Nous donnons entre autres les principaux moments centrés et non-centrés de cette distribution. Le second chapitre traite des méthodes d'estimation des paramètres. Deux méthodes sont considérées : la méthode du maximum de vraisemblance et la méthode des moments. Pour chacune d'elles, nous donnons les expressions des estimateurs ainsi que les variances et covariances asymptotiques.

Au chapitre III, nous présentons l'estimation des quantiles de la loi Weibull et la variance asymptotique qui y est associée. L'estimation des quantiles est un outil privilégié pour les hydrologues. Ces valeurs, que l'on appelle en hydrologie les événements X_T de période de retour T (ou correspondant à une probabilité au dépassement $1/T$), permettent de planifier efficacement tout dimensionnement d'ouvrage hydraulique.

Enfin, nous présentons brièvement au chapitre IV la loi Weibull à trois paramètres. On y discute plus particulièrement du lien qui associe cette loi à la loi des valeurs extrêmes de Type 3 (EV3). Nous présentons aussi les principales références concernant l'estimation des paramètres de cette loi.

Les développements théoriques présentés dans ce rapport sont pour la plupart une synthèse des principaux résultats que l'on retrouve dans la littérature au sujet de cette loi de probabilité. Ce travail a été effectué dans le but d'ajouter la loi Weibull au logiciel AJUSTE dans le cadre d'un projet de partenariat entre Hydro-Québec et INRS-Eau.

CHAPITRE 1

PROPRIETES MATHÉMATIQUES ET STATISTIQUES DE LA LOI WEIBULL A DEUX PARAMETRES

1.1 Définition et propriétés mathématiques

Une variable aléatoire X est distribuée selon une loi Weibull à deux paramètres s'il existe des valeurs des paramètres $c (> 0)$ et $\alpha (> 0)$ telles que la variable

$$Y = \left(\frac{X}{\alpha}\right)^c \quad (1.1)$$

soit distribuée selon une loi exponentielle de fonction de densité de probabilité :

$$g(y) = e^{-y} \quad , y > 0 \quad (1.2)$$

La densité de probabilité de la variable X (loi Weibull) est donc :

$$f(x) = \frac{c}{\alpha} \left(\frac{x}{\alpha}\right)^{c-1} \exp\left[-\left(\frac{x}{\alpha}\right)^c\right] \quad , x > 0, \alpha > 0, c > 0 \quad (1.3)$$

et sa fonction de distribution :

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^c\right] \quad , x > 0 \quad (1.4)$$

La figure 1 montre la forme que peut prendre la fonction de densité de probabilité pour différentes valeurs du paramètre de forme c , lorsque le paramètre d'échelle α vaut 1.

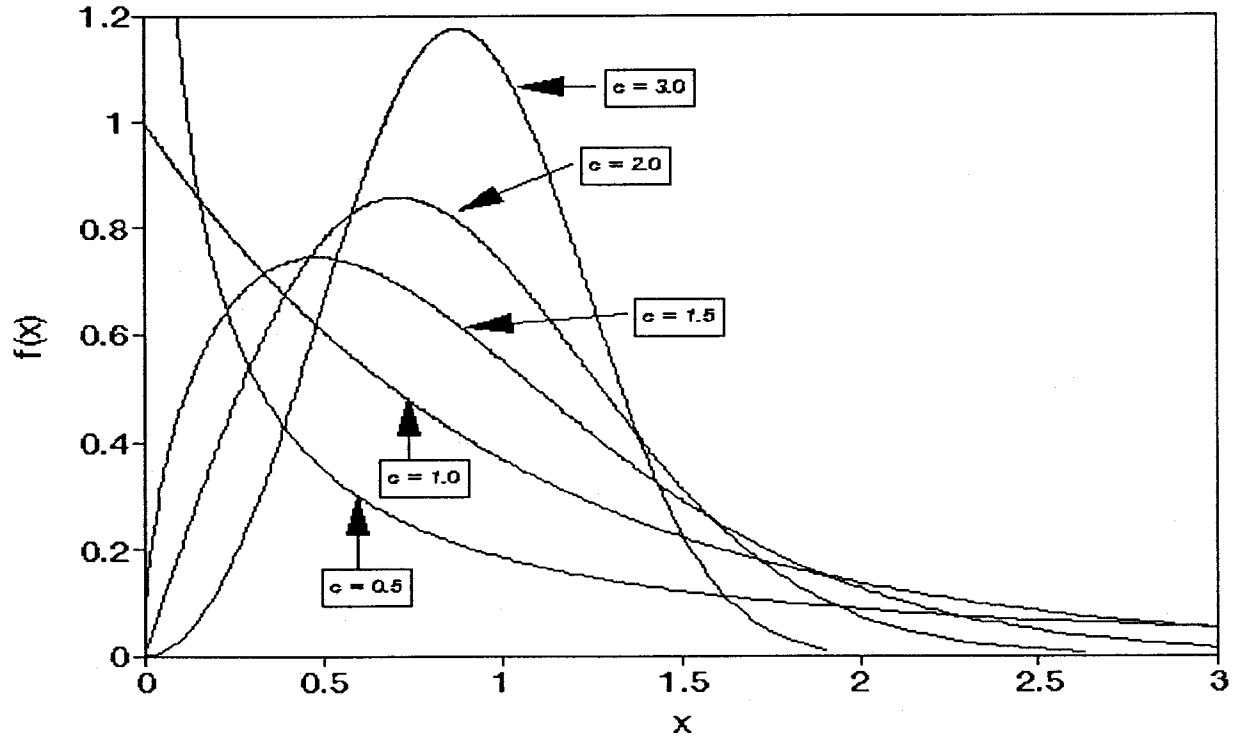


Figure 1. Forme de la f.d.p. de la loi Weibull pour différentes valeurs de c et $\alpha = 1$.

Pour $c > 1$, la densité de la loi Weibull tend vers zéro à mesure que x diminue. Dans ce cas, elle n'admet qu'un seul mode à :

$$x = \alpha \left(\frac{c - 1}{c} \right)^{1/c} \tag{1.5}$$

Cette valeur tend rapidement vers α lorsque c tend vers l'infini. Pour $0 < c < 1$, le mode est zéro et la densité est une fonction décroissante de x en tout point du domaine. Finalement, on peut montrer que la médiane de la loi Weibull est donnée par :

$$Med(x) = \alpha (\ln 2)^{1/c} \tag{1.6}$$

1.2 Propriétés statistiques

Les expressions mathématiques des principaux moments centrés et non-centrés ainsi que les coefficients d'asymétrie et d'aplatissement de la loi Weibull à deux paramètres sont présentés dans cette section.

Soit X une variable aléatoire distribuée selon une loi Weibull (éq. 1.3) de paramètres α et c . La variable aléatoire $Z = X/\alpha$ possède une fonction de densité de probabilité définie comme suit :

$$f(z) = c z^{c-1} \exp(-z)^c, \quad z > 0, c > 0 \quad (1.7)$$

La loi de Z dépend seulement du paramètre de forme c . Les coefficients d'asymétrie et d'aplatissement de la loi de Z sont donc les mêmes que ceux de la loi de Weibull (éq. 1.3) puisque, comme nous le verrons plus loin, ceux-ci sont indépendants du paramètre α . Les moments non-centrés correspondants à la loi Weibull à deux paramètres (éq. 1.3) sont facilement obtenus des moments de la loi de Z en utilisant la transformation $X = \alpha Z$.

Puisque Z^c est distribuée selon une loi exponentielle (éq. 1.2), le moment d'ordre r de Z est aussi le moment d'ordre r/c d'une variable de loi exponentielle (1.2). Ainsi,

$$\mu_r(Z) = E\{Z^r\} = E\{(Z^c)^{r/c}\} = \Gamma\left(1 + \frac{r}{c}\right) \quad (1.8)$$

car le moment d'ordre r d'une loi exponentielle (éq. 1.2) est $\Gamma(1 + r)$. On déduit alors aisément le moment non-centré d'ordre r de la variable X (loi Weibull) :

$$\begin{aligned} \mu_r(X) &= E\{X^r\} = E\{(\alpha Z)^r\} \\ &= \alpha^r \mu_r(Z) = \alpha^r \Gamma\left(1 + \frac{r}{c}\right) \end{aligned} \quad (1.9)$$

En particulier, la moyenne d'une variable aléatoire de loi Weibull à deux paramètres est donnée par

$$\mu_1(X) = \alpha \Gamma\left(1 + \frac{1}{c}\right) \quad (1.10)$$

On peut déduire le moment centré $\mu_r(X)$ d'ordre r à partir des moments non-centrés $\mu_r(X)$. Il suffit d'appliquer la relation (Kendall et Stuart, 1987) :

$$\mu_r(X) = \sum_{j=0}^r \binom{r}{j} \mu_{r-j}(X) (-\mu_1)^j(X) \quad (1.11)$$

Par conséquent, à partir des équations (1.9) et (1.11) on peut montrer, lorsque X suit une loi Weibull à deux paramètres, que sa variance est donnée par :

$$\mu_2(X) = \alpha^2 \left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right) \right] \quad (1.12)$$

et les moments centrés d'ordre 3 et 4 par :

$$\begin{aligned} \mu_3(X) = \alpha^3 & \left[\Gamma\left(1 + \frac{3}{c}\right) - 3\Gamma\left(1 + \frac{2}{c}\right)\Gamma\left(1 + \frac{1}{c}\right) \right. \\ & \left. + 2\Gamma^3\left(1 + \frac{1}{c}\right) \right] \end{aligned} \quad (1.13)$$

$$\begin{aligned} \mu_4(X) = \alpha^4 & \left[\Gamma\left(1 + \frac{4}{c}\right) - 4\Gamma\left(1 + \frac{3}{c}\right)\Gamma\left(1 + \frac{1}{c}\right) \right. \\ & \left. + 6\Gamma\left(1 + \frac{2}{c}\right)\Gamma^2\left(1 + \frac{1}{c}\right) - 3\Gamma^4\left(1 + \frac{1}{c}\right) \right] \end{aligned} \quad (1.14)$$

On remarque, en particulier, que la variance de cette loi est une fonction décroissante du paramètre c . En effet, à la limite, lorsque c devient grand ($c \rightarrow \infty$), la variance tend vers zéro ($\mu_2(X) \rightarrow 0$).

Enfin, des équations (1.12), (1.13) et (1.14), on déduit le coefficient d'asymétrie :

$$C_s = \frac{\Gamma\left(1 + \frac{3}{c}\right) - 3\Gamma\left(1 + \frac{2}{c}\right)\Gamma\left(1 + \frac{1}{c}\right) + 2\Gamma^3\left(1 + \frac{1}{c}\right)}{\left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right)\right]^{3/2}} \quad (1.15)$$

ainsi que le coefficient d'aplatissement :

$$C_K = \frac{\Gamma\left(1 + \frac{4}{c}\right) - 4\Gamma\left(1 + \frac{1}{c}\right)\Gamma\left(1 + \frac{3}{c}\right) + 6\Gamma^2\left(1 + \frac{1}{c}\right)\Gamma\left(1 + \frac{2}{c}\right) - 3\Gamma^4\left(1 + \frac{1}{c}\right)}{\left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right)\right]^2} \quad (1.16)$$

qui ne dépendent que du paramètre de forme c .

Remarquons que le coefficient d'asymétrie devient élevé pour de petites valeurs du paramètre c , tend vers zéro lorsque c tend vers 3.60 (Cohen, 1973), devient ensuite négatif, pour enfin tendre à nouveau vers zéro lorsque c approche l'infini.

CHAPITRE 2

ESTIMATION DES PARAMETRES DE LA LOI WEIBULL

Les méthodes d'estimation des paramètres les plus utilisées sont 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 Weibull d'autres méthodes ont été proposées, en particulier des méthodes d'estimation graphique et des méthodes basées sur les statistiques d'ordre. Citons entre autres les travaux de Dubey (1967) et Kappenman (1985). De plus, quelques auteurs ont proposé des modifications aux méthodes classiques du maximum de vraisemblance et des moments (Cohen et Whitten, 1982).

Nous considérons dans le présent rapport la méthode du maximum de vraisemblance pour ses propriétés asymptotiques bien connues (Lehmann, 1983) et la méthode des moments pour sa simplicité et parce que les estimations résultantes peuvent être utilisées comme valeurs initiales dans le processus itératif de la méthode du maximum de vraisemblance. Les principes généraux de ces deux méthodes, qui nous apparaissent adéquates pour la loi Weibull, sont décrits à l'annexe A.

Nous présentons dans ce chapitre les développements et les résultats menants aux différents estimateurs des paramètres de la loi Weibull pour les deux méthodes considérées. Les calculs des variances et covariances asymptotiques des estimateurs obtenus sont aussi donnés.

2.1 Méthode du maximum de vraisemblance

Soient n variables aléatoires X_1, X_2, \dots, X_n indépendantes tirées de la loi Weibull dont la fonction de densité de probabilité est donnée par (1.3). La vraisemblance logarithmique d'une réalisation x_1, x_2, \dots, x_n est donnée par :

$$\ln L(\alpha, c) = n \ln(c) - n \ln(\alpha) - \sum_{i=1}^n \left(\frac{x_i}{\alpha} \right)^c + (c-1) \sum_{i=1}^n \ln \left(\frac{x_i}{\alpha} \right) \quad (2.1)$$

Lorsque $c < 1$, la loi Weibull est une fonction décroissante de x en tout point du domaine (voir Figure 1) et les propriétés asymptotiques usuelles de la méthode du maximum de vraisemblance ne s'appliquent pas. Dans ce cas, les estimateurs du maximum de vraisemblance existent, mais leurs variances et covariances asymptotiques n'ont plus de sens puisque le déterminant de la matrice d'information de Fisher (annexe A) devient négatif, fournissant ainsi des variances négatives (Ravenis, 1964). De plus, des problèmes numériques peuvent survenir. En pratique, il n'est pas recommandé d'utiliser la méthode du maximum de vraisemblance lorsqu'il y a de fortes raisons de croire que le paramètre de forme c est inférieur à 1. Pour examiner une telle hypothèse, on peut par exemple tracer un histogramme des données.

Lorsque $c = 1$, la densité (1.3) correspond à celle d'une loi exponentielle à un paramètre et l'estimateur du maximum de vraisemblance de α est optimal (estimateur non-biaisé de variance minimum). Dans ce cas, $\hat{\alpha} = \bar{X}$, la moyenne arithmétique des observations.

Lorsque $c > 1$, il suffit pour déterminer les estimateurs du maximum de vraisemblance de résoudre le système d'équations obtenu en annulant les dérivées partielles, par rapport aux paramètres, de la fonction de vraisemblance logarithmique (voir annexe A). Après quelques calculs, nous obtenons le système suivant :

$$\alpha = \left(\frac{1}{n} \sum_{i=1}^n x_i^c \right)^{1/c} \quad (2.2)$$

$$c = \left[\left(\sum_{i=1}^n x_i^c \ln x_i \right) \left(\sum_{i=1}^n x_i^c \right)^{-1} - \frac{1}{n} \sum_{i=1}^n \ln x_i \right]^{-1} \quad (2.3)$$

Puisque l'équation (2.3) ne dépend que du paramètre c , il suffit de la résoudre pour trouver \hat{c} et ensuite de substituer c dans l'équation (2.2) par cette valeur afin d'obtenir $\hat{\alpha}$. Toutefois, la résolution de l'équation (2.3) nécessite l'emploi d'une méthode numérique. Nous utilisons dans le logiciel AJUSTE une méthode classique, celle de Newton-Raphson, dont la procédure est décrite dans Johnson et Riess (1982).

La convergence de ce type de méthode itérative est fortement influencée par les valeurs de départ choisies (initialisation de la procédure). Zanakis (1979b) propose d'utiliser comme valeur de départ c_o l'estimateur suivant :

$$c_o = \frac{\ln[\ln(1 - p_k)/\ln(1 - p_i)]}{\ln[X_{(k)}/X_{(i)}]} \quad (2.4)$$

où :

$p_i = 0.16731$ et $p_k = 0.97366$, deux probabilités empiriques qui minimisent la variance asymptotique de c_o (Dubey, 1967);

$X_{(i)}$ et $X_{(k)}$, les statistiques d'ordre correspondantes (les quantiles)

Cette proposition fait suite aux résultats obtenus par le même auteur (Zanakis, 1979a). Une comparaison de plusieurs estimateurs simples des paramètres de la loi Weibull a montré que c_o est le plus performant.

D'autres auteurs ont proposé diverses valeurs de départ pour le paramètre c , citons en particulier Cohen (1965) et, Cohen et Whitten (1982). Ces auteurs s'appuient sur les relations existant entre les valeurs du paramètre et différentes statistiques comme le coefficient de variation et le coefficient d'asymétrie. Toutefois, les essais que nous avons effectués semblent montrer que c_o permet une convergence plus rapide de la méthode de Newton-Raphson.

Les variances et la covariance asymptotiques des estimateurs du maximum de vraisemblance, qui peuvent être calculées lorsque $c > 1$, sont obtenues en inversant la matrice d'information de Fisher définie à l'annexe A. La matrice d'information de Fisher, dans ce cas, est donnée par :

$$I_f(\hat{\alpha}, \hat{c}) = \begin{pmatrix} \frac{nc^2}{\alpha^2} & -\frac{n}{\alpha}\Psi(2) \\ -\frac{n}{\alpha}\Psi(2) & \frac{n}{c^2}\{\Psi'(1) + \Psi^2(2)\} \end{pmatrix} \quad (2.5)$$

où $\Psi(\cdot)$ et $\Psi'(\cdot)$ sont respectivement les fonctions digamma et trigamma. Les valeurs de ces fonctions sont tabulées pour différents arguments dans Abramowitz et Stegun (1972). En inversant cette matrice et en substituant les valeurs des fonctions $\Psi(\cdot)$ et $\Psi'(\cdot)$, on obtient les variances et la covariance asymptotiques des estimateurs :

$$Var(\hat{\alpha}) = \frac{\alpha^2}{nc^2} 1.108665 \quad (2.6)$$

$$Var(\hat{c}) = \frac{c^2}{n} 0.607927 \quad (2.7)$$

$$Cov(\hat{\alpha}, \hat{c}) = \frac{\alpha}{n} 0.257022 \quad (2.8)$$

2.2 Méthode des moments

La méthode des moments est généralement utilisée parce qu'elle est très simple à appliquer. Elle permet aussi de fournir des estimations pour les paramètres lorsque la méthode du maximum de vraisemblance ne converge pas. De plus, la méthode des moments peut être utilisée pour initialiser les paramètres dans le cadre du processus itératif du maximum de vraisemblance.

La méthode des moments appliquée à la loi Weibull à deux paramètres consiste à égaliser les deux premiers moments (moyenne et variance) de l'échantillon aux moments théoriques correspondants :

$$\bar{x} = \mu_1(X) \quad (2.9)$$

$$s^2 = \mu_2(X) \quad (2.10)$$

où :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.11)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.12)$$

A partir des équations (2.9), (2.10) et de (1.10) et (1.12), on déduit le système d'équations à résoudre :

$$\bar{x} = \alpha \Gamma\left(1 + \frac{1}{c}\right) \quad (2.13)$$

$$s = \alpha \left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right) \right]^{1/2} \quad (2.14)$$

On peut toutefois simplifier ce système en éliminant le paramètre α . Il suffit de prendre le rapport des équations (2.13) et (2.14). On obtient ainsi une équation qui ne dépend que du paramètre de forme c :

$$\frac{s}{\bar{x}} \Gamma\left(1 + \frac{1}{c}\right) = \left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right) \right]^{1/2} \quad (2.15)$$

Cette équation ne peut être résolue de manière explicite et on doit utiliser une procédure itérative. On utilise encore ici la méthode de Newton-Raphson pour obtenir \hat{c} , et on détermine ensuite $\hat{\alpha}$ à l'aide de l'équation (2.13). L'initialisation de la procédure est la même que celle utilisée pour la méthode du maximum de vraisemblance (c_0).

Les variances et la covariance asymptotiques des estimateurs obtenus par la méthode des moments sont calculées en appliquant la méthode décrite à l'annexe A, c'est-à-dire en utilisant l'expression matricielle suivante :

$$V_p = V^{-1}V_m \quad (2.16)$$

Pour la loi Weibull à deux paramètres, les éléments de V_p sont les variances et la covariance asymptotiques recherchées de $\hat{\alpha}$ et \hat{c} , et ceux de V_m , les variances et la covariance asymptotiques des moments de l'échantillon \bar{X} et S^2 . Les éléments de la matrice V sont fonction des dérivées premières des deux moments de la population par rapport aux paramètres. Selon la notation de l'annexe A, ces dérivées sont les termes A_{ij} suivants :

$$A_{rj} = \frac{\partial M_r}{\partial \theta_j} \quad i, j = 1, 2 \quad (2.17)$$

où $\theta_1 = \alpha$, $\theta_2 = c$, et M_1 et M_2 sont respectivement la moyenne et la variance données aux équations (1.10) et (1.12). Nous obtenons ici :

$$A_{11} = \Gamma\left(1 + \frac{1}{c}\right) \quad (2.18)$$

$$A_{12} = -\frac{\alpha}{c^2} \Psi\left(1 + \frac{1}{c}\right) \Gamma\left(1 + \frac{1}{c}\right) \quad (2.19)$$

$$A_{21} = 2\alpha \left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right) \right] \quad (2.20)$$

$$A_{22} = -\frac{2\alpha^2}{c^2} \left[\Psi\left(1 + \frac{2}{c}\right) \Gamma\left(1 + \frac{2}{c}\right) - \Psi\left(1 + \frac{1}{c}\right) \Gamma^2\left(1 + \frac{1}{c}\right) \right] \quad (2.21)$$

Maintenant, avant de donner les expressions explicites des éléments du vecteur V_m , rappelons les définitions suivantes:

$$C_s = \frac{\mu_3}{\mu_2^{3/2}} \quad (2.22)$$

$$C_k = \frac{\mu_4}{\mu_2^2} \quad (2.23)$$

On peut montrer (voir Perreault et Bobée, 1992, Annexe C) que les éléments du vecteur V_m sont donnés dans l'ordre par:

$$\text{Var}(\bar{X}) = \frac{\mu_2}{n} \quad (2.24)$$

$$\text{Var}(S^2) = \frac{\mu_2^2(C_k - 1)}{n} \quad (2.25)$$

$$\text{Cov}(\bar{X}, S^2) = \frac{\mu_3}{n} = \frac{\mu_2^{3/2} \cdot C_s}{n} \quad (2.26)$$

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

CHAPITRE 3

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

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

$$Prob\{X < X_p\} = p \quad (3.1)$$

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

$$Prob\{X > X_T\} = \frac{1}{T} \quad (3.2)$$

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 Weibull à deux paramètres ainsi que la détermination des variances asymptotiques qui y sont associées.

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 dont la fonction de densité de probabilité donnée par (1.3). Alors, des équations (1.4) et (3.2), on déduit aisément que :

$$X_T = \alpha \left[-\ln\left(\frac{1}{T}\right) \right]^{\frac{1}{c}} \quad (3.3)$$

Pour estimer cette quantité, il suffit de remplacer dans l'équation (3.3), les paramètres α et c par leurs estimateurs respectifs $\hat{\alpha}$ et \hat{c} présentés au chapitre 2. Nous obtenons alors le débit de période de retour T estimé :

$$\hat{X}_T = \hat{\alpha} \left[-\ln\left(\frac{1}{T}\right) \right]^{\frac{1}{\hat{c}}} \quad (3.4)$$

Cet estimateur sera celui du maximum de vraisemblance ou des moments selon la méthode utilisée pour déterminer $\hat{\alpha}$ et \hat{c} (Sections 2.1 et 2.2).

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

$$Var(\hat{X}_T) = \left(\frac{\partial X_T}{\partial \alpha} \right)^2 Var(\hat{\alpha}) + \left(\frac{\partial X_T}{\partial c} \right)^2 Var(\hat{c}) + 2 \left(\frac{\partial X_T}{\partial \alpha} \right) \left(\frac{\partial X_T}{\partial c} \right) Cov(\hat{\alpha}, \hat{c}) \quad (3.5)$$

Or, en posant

$$Y_T = -\ln\left(\frac{1}{T}\right) \quad (3.6)$$

et en évaluant les dérivées partielles, l'équation (3.5) se réduit à :

$$Var(\hat{X}_T) = Y_T^{2c} \left[Var(\hat{\alpha}) + \frac{\alpha^2}{c^4} (\ln Y_T)^2 Var(\hat{c}) - 2 \frac{\alpha}{c^2} \ln Y_T Cov(\hat{\alpha}, \hat{c}) \right] \quad (3.7)$$

En remplaçant $Var(\hat{\alpha})$, $Var(\hat{c})$ et $Cov(\hat{\alpha}, \hat{c})$ par les expressions obtenues au Chapitre 2, on peut déduire la variance asymptotique de \hat{X}_T pour les deux méthodes d'estimation considérées dans ce travail.

CHAPITRE 4

LOI DE WEIBULL A TROIS PARAMETRES

Nous avons mentionné au début du présent rapport que la principale raison qui a justifié l'emploi de la loi Weibull en hydrologie est le lien qui l'associe aux lois des valeurs extrêmes. En effet, plusieurs problèmes de construction et de planification d'ouvrages hydrauliques sont reliés aux propriétés d'événements extrêmes comme par exemple, les niveaux ou les débits maximums et minimums des rivières observés dans un intervalle de temps donné. Le terme "valeurs extrêmes" est associé à ces distributions parce qu'elle peuvent être obtenues comme loi limite ($p \rightarrow \infty$) de la valeur extrême (maximum ou minimum) de p variables aléatoires indépendantes et identiquement distribuées. Ainsi, les événements hydrologiques extrêmes ont souvent été considérés comme des variables aléatoires distribuées selon l'une des trois lois des valeurs extrêmes introduites par Fisher et Tippett (1928). En particulier, on a souvent utilisé, afin d'ajuster les débits maximums annuels tirés de p mesures (par exemple $p = 365$ débits journaliers), la loi des valeurs extrêmes de Type 3 (EV3) dont la fonction de distribution est donnée par :

$$G(x) = \exp\left(-\frac{(m-x)^\xi}{\theta}\right), x \leq m \quad (4.1)$$

où m , θ (> 0) et ξ (> 0) sont les paramètres. Cette distribution est la loi limite des maximums.

Toutefois, les hypothèses de base d'un tel modèle ne sont pas théoriquement respectées (p grand, débits journaliers indépendants). Ainsi, on ne peut conclure que les lois des valeurs extrêmes (EV3, par exemple) sont plus adéquates théoriquement pour ajuster les débits maximums annuels. On peut tout de même les utiliser comme n'importe quelle autre loi puisqu'aucune distribution n'est plus justifiée théoriquement qu'une autre pour ajuster ce type de données.

Lorsque la variable aléatoire X est distribuée selon une loi EV3, on peut montrer (Johnson et Kotz, 1970) que la variable $-X$ suit une loi Weibull à trois paramètres de fonction de distribution :

$$F(x) = 1 - \exp\left[-\left(\frac{x - m}{\alpha}\right)^c\right], \quad x > m \quad (4.2)$$

Cette distribution est donc la loi limite des minimums, et lorsque le paramètre d'origine m est nul, nous retrouvons la loi Weibull à deux paramètres.

Les différentes propriétés mathématiques et statistiques de la loi Weibull à trois paramètres peuvent être déduites aisément de celles de la loi des valeurs extrêmes de Type 3 (Perreault et Bobée, 1992) en effectuant un simple changement de variable.

L'estimation des paramètres de cette loi est généralement effectuée en utilisant la méthode du maximum de vraisemblance puisque, lorsque $c > 2$, les conditions de régularité pour les estimateurs sont vérifiées (Harter, 1971). Ceci implique, dans ce cas, que les estimateurs du maximum de vraisemblance possèdent les propriétés usuelles de convergence, d'efficacité et de normalité asymptotique (Lehmann, 1983). Lorsque $c \leq 1$, la plus petite observation de l'échantillon devient un estimateur hyper-efficace (voir Lehmann, 1983, pour la définition) pour le paramètre d'origine m (Dubey, 1966), mais aucun estimateur du maximum de vraisemblance existe pour les deux autres paramètres. Pour $1 < c < 2$, les estimateurs du maximum de vraisemblance existent mais les variances et covariances asymptotiques n'ont plus de sens puisque le déterminant de la matrice d'information de Fisher (annexe A) devient négatif, fournissant ainsi des variances négatives (Ravenis, 1964). Nous avons observé le même type de problème avec la loi Weibull à deux paramètres lorsque $c < 1$ (Section 2.1). Enfin, si $c = 2$, le déterminant de la matrice d'information est nul et les variances asymptotiques des estimateurs sont indéterminées.

Plusieurs auteurs se sont penchés sur l'estimation des paramètres de la loi Weibull à trois paramètres. Le tableau qui suit donne les principales références concernant ce problème.

Tableau 1. Revue de littérature non-exhaustive concernant l'estimation des paramètres de la loi Weibull à 3 paramètres.

Auteurs	Procédure proposée
Ravenis (1964)	Méthodes d'estimation graphiques des paramètres
Harter et Moore (1965)	Proposent un algorithme permettant de solutionner le système non-linéaire
Harter et Moore (1967)	Donnent les variances asymptotiques des estimateurs du maximum de vraisemblance
Dubey (1967)	Propose un estimateur pour le paramètre d'origine qui est indépendant des deux autres
Wingo (1973)	Propose un algorithme permettant de solutionner le système non-linéaire
Zanakis (1977)	Compare différents algorithmes de résolution du système
Zanakis (1979a)	Propose et compare des estimateurs simples basés sur les statistiques d'ordre
Zanakis (1979b)	Propose un algorithme pour résoudre le système non-linéaire
Cohen et Whitten (1982)	Modifient les méthodes du maximum de vraisemblance et des moments
Mann (1984)	Revue des ouvrages traitant des problèmes pratiques rencontrés lors de l'estimation des paramètres
Smith (1985)	Discute des conditions permettant d'obtenir les propriétés usuelles des estimateurs du maximum de vraisemblance
Kappenman (1985)	Propose des estimateurs simples qui ne nécessitent pas la résolution d'un système d'équations non-linéaires

Ce cours chapitre n'est qu'un survol des principales caractéristiques de la loi Weibull à trois paramètres. Nous vous invitons à consulter les différents ouvrages cités pour plus de détails.

CONCLUSION

Ce rapport effectue une synthèse des propriétés mathématiques et statistiques de la loi Weibull à deux paramètres présentées dans différentes études. Deux méthodes classiques d'estimation des paramètres ont été présentées : la méthode du maximum de vraisemblance et la méthode des moments. 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, nous avons introduit, au dernier chapitre, la loi Weibull à trois paramètres. Ce chapitre est en fait une revue de la littérature non-exhaustive concernant les différentes propriétés de cette loi.

Tous les développements touchant la loi Weibull à deux paramètres ont été effectués dans le but d'introduire cette loi dans le logiciel AJUSTE. 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 Weibull, ce travail devrait permettre de faciliter l'utilisation en pratique de cette loi de probabilité qui possède d'intéressantes propriétés.

BIBLIOGRAPHIE

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

Cohen, A.C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and censored samples. *Technometrics*, **5**, 579-588.

Cohen, A.C. (1973). The reflected Weibull distribution. *Technometrics*, **15**, 867-873.

Cohen, A.C. et Whitten B. (1982). Modified maximum likelihood and modified moment estimators in the three-parameter Weibull distribution. *Commun. Statist.-Theor. Meth.*, **11**, 2631-2656.

Dubey, S.D. (1966). Hyper-efficient estimator of the location parameter of the Weibull laws. *Naval Research Logistics Quarterly*, **13**, 253-263.

Dubey, S.D. (1967). Some percentiles estimators for Weibull parameters. *Technometrics*, **9**, 119-129.

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.

Harter, H.L. et Moore, A.H. (1965). Maximum likelihood estimation of the parameters of the gamma and Weibull populations from censored samples. *Technometrics*, **7**, 639-643.

Harter, H.L. et Moore, A.H. (1967). Asymptotic variances and covariances of maximum likelihood estimators from censored samples, of the parameters of the gamma and Weibull. *Annals of Math. Statist.*, **38**, 557-571.

Harter, H.L. (1971). *Some optimization problems in parameter estimation, in Optimizing Methods in Statistics*. J.S. Rustagi (Ed.), Academic Press, New York.

Johnson, N.L. et Kotz, S. (1970). *Continuous univariate distributions-1*. Wiley, New York.

Johnson, W.L. et Riess, R.D. (1982). *Numerical Analysis*. Addison Wesley Publishing Company.

Kappenman, R.F. (1985). Estimation for the three-parameter Weibull, lognormal, and gamma distributions. *Computational Statistics & Data Analysis*, 3, 11-23.

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

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

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

Mann, N.R. (1968). Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions. *Technometrics*, 10, 231-256.

Mann, N.R. (1984). Statistical estimation of the parameters of the Weibull and Frechet distributions. Dans *Statistical Extremes and Applications*. J. Tiago de Oliveira, pp.231-256, Dordrecht Reidel.

Perreault, L et Bobée, B. (1992). Loi généralisée des valeurs extrêmes : Propriétés mathématiques et statistiques, estimation des paramètres et des quantiles X_T de période de retour T . *Rapport Scientifique No 350*, INRS-Eau.

Ravenis, J.V.J. (1964). Estimating Weibull distribution parameters. *Electrotechnology*, 46-54.

Smith, L.S. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72, 67-90.

Wingo, D.R. (1973). Solution of the three-parameter Weibull equations by constrained modified quasilinearization. *IEEE Transactions on Reliability*, 22, 96-102.

Zanakis, S.H. (1977). Computational experience with some non-linear optimization algorithms in deriving MLE for the three-parameter Weibull distribution. *Algorithmic Methods in Probability*, TIMS Studies in Management Sciences, **7**, North Holland Publisher.

Zanakis, S.H. (1979a). A Monte Carlo study of some simple estimators of the three-parameter Weibull distribution. *Journal of Statistical Computation and Simulation*, **9**, 101-116.

Zanakis, S.H. (1979b). Extended pattern search with transformation for the three-parameter Weibull distribution. *Management Science*, **25**, 1149-1161.

ANNEXE A

DESCRIPTION DES METHODES D'ESTIMATION

Dans cette section nous décrivons brièvement les deux méthodes d'estimation considérées dans le présent rapport. Pour plus de détail, nous vous invitons à consulter les ouvrages cités.

A.1 Méthode du maximum de vraisemblance

Soit un échantillon aléatoire de taille n x_1, x_2, \dots, x_n tiré d'une loi $F(x; \theta_1, \theta_2, \dots, \theta_k)$ à k paramètres. Les estimateurs du maximum de vraisemblance sont obtenus en maximisant par rapport à $\theta_1, \theta_2, \dots, \theta_k$ la fonction suivante :

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

où $f(\cdot)$ est la fonction de densité de probabilité de la loi $F(\cdot)$.

En pratique, pour des raisons de simplicité, on maximise plutôt le logarithme de cette fonction :

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

Remarquons que la fonction de vraisemblance $L(\cdot)$ et son logarithme $\ln L(\cdot)$ atteignent leur maximum aux mêmes valeurs de $\theta_1, \theta_2, \dots, \theta_k$ puisque $\frac{\partial \ln L}{\partial \theta_i} = \frac{1}{L} \frac{\partial L}{\partial \theta_i}$. Maximiser (A.1) ou (A.2) est donc équivalent.

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

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

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

Les propriétés asymptotiques des estimateurs issus de 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 Fischer I_f :

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

A.2 Méthode des moments

Soit un échantillon aléatoire de taille n x_1, x_2, \dots, x_n tiré d'une loi $F(x; \theta_1, \theta_2, \dots, \theta_k)$ à k paramètres. L'espérance mathématique $E(X^r)$ définie comme suit:

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

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

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

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

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

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 :

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

où M_r et M_q sont les moments théoriques (équation A.4) de la loi pour r et $q \in \{1, 2, \dots, k\}$.

En supposant que les deux premiers moments sont suffisants pour estimer les paramètres de la loi (c'est le cas pour la loi Weibull à deux paramètres), l'équation peut s'écrire de la façon suivante :

$$\begin{bmatrix} \text{Var}(m_1) \\ \text{Var}(m_2) \\ \text{Cov}(m_1, m_2) \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix} \cdot \begin{bmatrix} \text{Var}(\hat{\theta}_1) \\ \text{Var}(\hat{\theta}_2) \\ \text{Cov}(\hat{\theta}_1, \hat{\theta}_2) \end{bmatrix}$$

Sous forme matricielle, on a :

$$V_m = V \cdot V_p$$

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

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

où

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

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 la covariance asymptotiques des estimateurs des moments (le vecteur V_p) par la relation :

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