Journal of Global Optimization **5**: 373–397, 1994.

Finding Maximum Likelihood Estimators for the Three-Parameter Weibull Distribution

ÉRIC GOURDIN
École Polytechnique de Montréal, Département de Mathématiques Appliquées, P.O. Box 6079, Station "Centre-ville", Montréal (Québec) H3C 3A7, Canada.

PIERRE HANSEN
GERAD and École des Hautes Études Commerciales, Département des Méthodes Quantitatives et Systèmes d'Information, 5255 avenue Decelles, Montréal (Québec) H3T 1V6, Canada.

and

BRIGITTE JAUMARD
GERAD and École Polytechnique de Montréal, Département de Mathématiques Appliquées, P.O. Box 6079, Station "Centre-ville", Montréal (Québec) H3C 3A7, Canada.

(Received: 14 September 1993; accepted: 14 March 1994)

Abstract. Much work has been devoted to the problem of finding maximum likelihood estimators for the three-parameter Weibull distribution. This problem has not been clearly recognized as a global optimization one and most methods from the literature occasionally fail to find a global optimum. We develop a global optimization algorithm which uses first order conditions and projection to reduce the problem to a univariate optimization one. Bounds on the resulting function and its first order derivative are obtained and used in a branch-and-bound scheme. Computational experience is reported. It is also shown that the solution method we propose can be extended to the case of right censored samples.

Key words: Global optimization, decomposition, maximum likelihood estimation, Weibull distribution.

1. Introduction

The Weibull distribution is extensively used in science and engineering, mainly to model behavior of humans, materials or systems increasingly subject to failure over time. Indeed, a survey of Kotz and Johnson [14] ranks the Weibull distribution as the third most used one in the statistical literature. Consequently, much work has been devoted to the problem of finding maximum likelihood estimators for the three-parameter Weibull distribution (when there are only two parameters, the problem is easy). Such maximum likelihood estimators have several desirable properties, recalled by Zanakis and Kyparisis [32]: consistency, asymptotic normality and asymptotic efficiency for large samples under fairly general assumptions.

Methods to find maximum likelihood estimators for the three-parameter Weibull distribution either consist in: (i) direct maximization of the likelihood or log-likelihood function or, (ii) determination of solutions of the first-order conditions for the log-likelihood function. Nonlinear programming and heuristic pattern search techniques applied by Zanakis [30, 31] pertain to class (i), the false position method used by Harter [10] and the modified quasi-linearization method of Wingo [28, 29] belong to class (ii). These and other methods are discussed in detail in the surveys of Zanakis and Kyparisis [32], and Panchang and Gupta [21]. The latter authors also discuss difficulties encountered by users of these various estimation procedures: there may be no interior stationary point, convergence may be to a saddle point or to a local maximum which is not a global one or even to a local minimum.

Such difficulties are typically those which are encountered when a method of optimization which only guarantees finding a local optimum is applied to a global optimization problem. Panchang and Gupta [21] recommend a palliative, already outlined by Lawless [15]: apply a grid search on the values of the location parameter (see next section) and solve each time the easy resulting two-parameter estimation problem. As no estimate of the slope of the implicit function considered is provided there is no guarantee to reach a point with an ε-optimal value, i.e., a point whose value differs from the optimal one by less than ε. Moreover as the behaviour of this function between evaluation points is unknown there is no guarantee that a solution close to the optimum one is obtained. Finally this procedure may be inordinately long, particularly if high precision is desired for the values of the parameters.

In this paper we propose a method for finding maximum likelihood parameters for the three-parameter Weibull distribution which guarantees that they are within any prescribed distance of the global optimum. It is based on using decomposition (or projection) to reduce the problem to a one-dimensional one and then exploiting bounds on the values of the log-likelihood function and its first order derivative in the remaining variable to curtail the search, within a branch-and-bound scheme.

The paper is organized as follows. The parameter estimation problem is stated in the next section. Some properties are given in Section 3. A decomposition scheme and an algorithm are described in Section 4. Extension to the case of right-censored samples and computational experience for test samples from the literature are presented in Section 5.

2. Problem Statement

We consider the three-parameter Weibull distribution, whose cumulative distribution function can be written:

$$F(x\,;\,u,v,w) = 1 - \exp\Bigl(\frac{-(x-u)^w}{v}\Bigr)$$

where u is the location parameter, v is the scale parameter and w is the shape parameter.

Given a sample of N ($\geq$ 1) observations $x_1, x_2, ..., x_N$, a common way of estimating the parameters u, v and w is to maximize the log-likelihood function:

$$\begin{cases} N \ln\left(\frac{w}{v}\right) - \frac{1}{v} \sum_{j=1}^{N} (x_j - u)^w + (w-1) \sum_{j=1}^{N} \ln(x_j - u) & \text{if } w \neq 1 \\ -N \ln v - \frac{1}{v} \sum_{j=1}^{N} (x_j - u) & \text{if } w = 1 \end{cases} \tag{1}$$

subject to the following constraints:

$$0 \leq u \leq x_j \quad j = 1, 2, ..., N$$
$$v > 0$$
$$w > 0 .$$

Note that L is defined by a separate expression at $w = 1$ as the term $(w - 1) \sum_{j=1}^{N} \ln(x_j - u)$ is undefined, even by continuity, when w goes to 1 and u goes to $\min_{j=1,...,N} x_j$. For $w \neq 1$, the function L takes its values on $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, by continuity when u goes to $\min_{j=1,2,...,N} x_j$. However, if we assume $0 \leq u < \min_{j=1,2,...,N} x_j$, the log-likelihood function is well defined (again taking its values on $\bar{\mathbb{R}}$) by the expression:

$$L(u, v, w) = N \ln\left(\frac{w}{v}\right) - \frac{1}{v} \sum_{j=1}^{N} (x_j - u)^w + (w-1) \sum_{j=1}^{N} \ln(x_j - u) \tag{2}$$

for $v > 0$ and $w \geq 0$.

Solution of this problem is easy if $w < 1$. Indeed the first partial derivative of L with respect to u,

$$\frac{\partial L}{\partial u}(u, v, w) = \frac{w}{v} \sum_{j=1}^{N} (x_j - u)^{w-1} - (w-1) \sum_{j=1}^{N} \frac{1}{x_j - u}$$

is strictly positive for $0 \leq u < \min_{j=1,2,...,N} x_j$ and $v > 0$. Hence the function L increases monotonously with respect to u and goes to $+\infty$ when u goes to $\min_{j=1,2,...,N} x_j$. Such an optimal solution in which $L(u, v, w)$ takes an infinite value and u only depends on the smallest observation of the sample does not appear to be realistic. In such a case, using other estimators such as modified moment ones may be preferable (see, e.g., Cohen, Whitten and Ding [5]). In this paper we only consider maximum log-likelihood estimators and therefore assume from now on that $w \geq 1$. (An optimal value of $w^* = 1$ will thus be an indication that alternative estimators should be considered.)

Finding parameters is usually achieved by solving the following log-likelihood equations:

$$\frac{\partial L}{\partial u}(u, v, w) = 0 , \qquad \frac{\partial L}{\partial v}(u, v, w) = 0 , \qquad \frac{\partial L}{\partial w}(u, v, w) = 0 .$$

However, as discussed above, solving these equations may only provide a local maximum, or even a solution for which the objective function is locally minimum with respect to variable u.

3. Properties

We now present properties which will be used in the decomposition scheme embedded in the exact algorithm. One of the most significant results of this section is the existence of a value $u_0 \in [0, \min_{j=1,2,\ldots,N} x_j]$ on the left of which there is no local optimum. The two first properties are known (McCool [18]; Rockette, Antle and Klimko [27]) as local optimality conditions. When necessary we complete the proofs to show they are in fact global optimality conditions.

Maximization of the log-likelihood function L (as defined by (1)) can be expressed as:

$$\begin{aligned} \max_{u,v,w} \;& L(u,v,w) \\ \text{subject to: } & u \in U \\ & v \in V \\ & w \in W \end{aligned} \tag{3}$$

where $U = [0, \min_{j=1,2,\ldots,N} x_j]$, $V =]0, +\infty[$ and $W = [1, +\infty[$.

If we first project the (u, v, w)-space on the (u, w)-space (Geoffrion [7]), this problem becomes:

$$\begin{aligned} \max_{u,w} \;& L^*(u,w) \\ \text{subject to: } & u \in U \\ & w \in W \end{aligned}$$

where:

$$L^*(u,w) = \max_{v \in V} L(u,v,w)\,. \tag{4}$$

It is easy to obtain an analytical expression for the exact solution of the inner optimization problem (4).

PROPOSITION 1. *(Rockette, Antle and Klimko [27].) For given $u \in U$ and $w \in W$, let $\hat{v} = \sum_{j=1}^{N} (x_j - u)^w / N$. Then*:

$$\forall u \in U \quad \forall w \in W \quad \max_{v \in V} L(u,v,w) = L(u,\hat{v},w)\,.$$

Proof. We consider the maximization of function L with respect to v, the other variables being held fixed. The first order condition can be stated as:

$$\frac{\partial L}{\partial v}(u,v,w) = -\frac{N}{v} + \frac{1}{v^2}\sum_{j=1}^{N}(x_j - u)^w = 0\,.$$

Function $\partial L/\partial v$ is equal to zero at $\hat{v} = \sum_{j=1}^{N} (x_j - u)^w/N$, positive for $v < \hat{v}$ and negative for $v > \hat{v}$. Moreover, $\hat{v} > 0$ unless all x_j coincide with u, in which case $\hat{v} = 0$ (by continuity on $\bar{\mathbb{R}}$). So, assuming at least two sample points are distinct,

$\hat{v} \in V =]0, +\infty[$. Hence $\hat{v}$ is the global maximum of the function L with respect to v for u and w fixed. □

Thus problem (3) can be restated as a problem in the variables u and w only:

$$\begin{aligned} &\max\ L^*(u,w)\ \left(=\ L(u,\hat{v},w)\right) \\ &\text{subject to :}\qquad u \in U \\ &\qquad\qquad\qquad\quad w \in W\,. \end{aligned} \tag{5}$$

Projecting now the (u, w)-space on the u-space, problem (5) reduces to the univariate problem:

$$\max_{u \in U} L^{**}(u)\,.$$

where:

$$L^{**}(u) = \max_{w \in W} L^*(u,w)\,. \tag{6}$$

Again it is easy to compute the global solution of the inner problem (6).

THEOREM 1. *Let* $\hat{w} \in W$ *satisfy* $L^*(u,\hat{w}) = \max_{w \in W} L^*(u,w)$.
If $\partial L^*/\partial w\,(0,1) < 0$ *then* $\hat{w} = 1$.
If $\partial L^*/\partial w\,(0,1) \geq 0$, *then there exists a unique value* $u_0 \in [0, \min_{j=1,2,\ldots,N} x_j[$ *satisfying* $\partial L^*/\partial w\,(u,1) = 0$ *such that:* $\forall u > u_0\ \ \hat{w} = 1$; $\forall u \leq u_0\ \ \hat{w}$ *is equal to the unique root of* $\partial L^*/\partial w\,(u,w) = 0$ *on* $[1, +\infty[$.

Proof. We first consider the case when $u = \min_{j=1,2,\ldots,N} x_j$ and then the maximization of function L^* with respect to w, the variable u being fixed at a value in $[0, \min_{j=1,2,\ldots,N} x_j[$ (which allows the use of the analytical expression (2) for the log-likelihood function L).

If $u = \min_{j=1,2,\ldots,N} x_j$ then, $L(u,v,w)$ is equal to $-\infty$ if $w \neq 1$, and $L(u,v,w) \in \mathbb{R}$ if $w = 1$. Since we maximize $L(u,v,w)$, it follows that $\hat{w} = 1$ when $u = \min_{j=1,2,\ldots,N} x_j$.

If $u \in [0, \min_{j=1,2,\ldots,N} x_j[$, computing $\hat{w}$ leads to examine the root(s), if they exist, of:

$$\frac{\partial L^*}{\partial w}(u,w) = \frac{N}{\left(w \sum\limits_{j=1}^{N} e^{wz_j}\right)^2} \left(\sum_{j=1}^{N} e^{wz_j}(1 - wz_j)\right) = 0 \tag{7}$$

where

$$z_j = \ln(x_j - u) - \frac{1}{N} \sum_{j=1}^{N} \ln(x_j - u)\,.$$

As $w \sum_{j=1}^{N} e^{wz_j}$ is never equal to zero, a solution of equation (7), if it exists, is a solution of:

$$\sum_{j=1}^{N} e^{wz_j}(1 - wz_j) = 0 . \tag{8}$$

We now show that this last equation has at most one solution. The derivative of the left-hand side of (8) with respect to w is equal to:

$$-w \sum_{j=1}^{N} z_j^2 \, e^{wz_j} \tag{9}$$

which is non-positive for $w \geq 1$. As the left-hand side of (8) is a continuous function, going to $-\infty$ when w goes to $+\infty$, equation (9) has exactly one root if and only if there exists a value z_j such that:

$$\sum_{j=1}^{N} e^{z_j}(1 - z_j) \geq 0 \tag{10}$$

or equivalently, if there exists u such that:

$$\frac{\partial L^*}{\partial w}(u, 1) \geq 0 . \tag{11}$$

The left-hand side of (10) can be rewritten:

$$\sum_{j=1}^{N} (x_j - u)\, e^{-\frac{1}{N}\sum_{j=1}^{N} \ln(x_j - u)} \left(1 - \ln(x_j - u) + \frac{1}{N} \sum_{j=1}^{N} \ln(x_j - u)\right)$$

or equivalently,

$$\frac{1}{e^{\frac{1}{N}\sum_{j=1}^{N} \ln(x_j-u)}} \, \varphi(u) ,$$

where:

$$\varphi(u) = \sum_{j=1}^{N} (x_j - u) \left(1 - \ln(x_j - u) + \frac{1}{N} \sum_{j=1}^{N} \ln(x_j - u)\right) .$$

It follows that condition (10) is equivalent to:

there exists u such that $\varphi(u) \geq 0$.

Let us examine the roots of φ on $[0, \min_{j=1,2,\ldots,N} x_j[$. After simplification, the first order derivative of $\varphi(u)$ can be written:

$$\varphi'(u) = -\frac{1}{N}\sum_{j=1}^{N}(x_j - u)\sum_{j=1}^{N}\frac{1}{x_j - u},$$

which is non-positive for all $u \in [0, \min_{j=1,2,\ldots,N} x_j[$.

Since φ is a continuous non-increasing function on $[0, \min_{j=1,2,\ldots,N} x_j[$ and $\varphi(u)$ goes to $-\infty$ as u goes to $\min_{j=1,2,\ldots,N} x_j$,[1] there exists a unique value u_0 satisfying $\varphi(u) = 0$ on $[0, \min_{j=1,2,\ldots,N} x_j[$ if and only if $\varphi(0) \geq 0$, or equivalently $\partial L^*/\partial w\,(0,1) \geq 0$ using inequality (11).

Therefore:

(a) if $\dfrac{\partial L^*}{\partial w}(0,1) < 0$, inequality (10) is never satisfied.
It follows that: $\forall u \in U\ \ \forall w \in W\ \ \partial L^*/\partial w\,(u,w) < 0$ and hence, $\hat{w} = 1$.

(b) if $\dfrac{\partial L^*}{\partial w}(0,1) \geq 0$, there exists $u_0 \in [0, \min_{j=1,2,\ldots,N} x_j[$ such that

(i) $\forall u \in [0, u_0[\qquad \dfrac{\partial L^*}{\partial w}(u,1) \geq 0$

(ii) $\forall u \in [u_0, \min_{j=1,2,\ldots,N} x_j] \qquad \dfrac{\partial L^*}{\partial w}(u,1) < 0\,.$

This implies that, in case (a): $\forall w \in W\ \ \partial L^*/\partial w\,(u,w) < 0$, and hence, $\hat{w} = 1$. In case (b), there exists $\hat{w} \in W$ such that:

$$\forall u \in [0, u_0[\qquad \begin{cases} \forall w \in [1, \hat{w}[& \dfrac{\partial L^*}{\partial w}(u,w) > 0 \\ \dfrac{\partial L^*}{\partial w}(u,\hat{w}) = 0 & \\ \forall w \in]\hat{w}, +\infty[& \dfrac{\partial L^*}{\partial w}(u,w) < 0\,. \end{cases}$$

Hence $\hat{w}$ is a maximum of $L^*(u,w)$ for fixed u. □

---

[1] $\varphi(u)$ can be rewritten as $\ln(x_k - u)\,\varphi_1(u) + \varphi_2(u)$ where:

$$x_k = \min_{j=1,2,\ldots,N} x_j\,,$$

$$\varphi_1(u) = \frac{N-1}{N}(x_k - u) + \frac{1}{N}\sum_{\substack{j=1\\ j\neq k}}^{N}(x_j - u)\,,$$

and

$$\varphi_2(u) = \sum_{j=1}^{N}(x_j - u)\left(1 + \frac{1}{N}\sum_{\substack{j=1\\ j\neq k}}^{N}\ln(x_j - u)\right) + \sum_{\substack{j=1\\ j\neq k}}^{N}(x_j - u)\ln(x_j - u)\,.$$

Assuming not all observation points are equal, we deduce $\varphi_1(x_k) > 0$, hence $\lim_{u\to x_k} \ln(x_k - u)\,\varphi_1(u) = -\infty$. Since $0 \leq \lim_{u\to x_k} \varphi_2(u) < +\infty$, we obtain $\lim_{u\to x_k} \varphi(u) = -\infty$.

It follows, from Theorem 1, that:

PROPOSITION 2. *Let* $\hat{w} \in W$ *satisfying* $L^*(u, \hat{w}) = \max_{w \in W} L^*(u, w)$. *Define* u_0 *as the root of* $\partial L^*/\partial w\,(u, 1) = 0$ *if* $\partial L^*/\partial w\,(0, 1) > 0$, *and* $u_0 = 0$ *otherwise. Then,*

$$\begin{cases} \forall u \in [0, u_0] & L^{**}(u) = L^*(u, \hat{w}) \\ & \quad \text{where } \hat{w} \text{ is the root of } \dfrac{\partial L^*}{\partial w}(u, w) = 0 \\ \forall u \in]u_0, \min\limits_{j=1,2,\ldots,N} x_j] & L^{**}(u) = L^*(u, 1)\,, \ i.e., \ \hat{w} = 1\,. \end{cases}$$

We now show that the maximization of $L^{**}(u)$ is straightforward when $u \in]u_0, \min_{j=1,2,\ldots,N} x_j]$:

PROPOSITION 3.

$$\max_{u \in]u_0, \min\limits_{j=1,2,\ldots,N} x_j]} L^{**}(u) = L^{**}(\min_{j=1,2,\ldots,N} x_j)\,.$$

Proof. When $u_0 < u < \min_{j=1,2,\ldots,N} x_j$, the expression of $L^{**}(u)$ simplifies to $L^*(u, 1) = N \ln N - N - N \ln\left(\sum_{j=1}^{N} (x_j - u)\right)$, and its first derivative is equal to $N^2/\left(\sum_{j=1}^{N} (x_j - u)\right)$. As L^{**} is non-decreasing for $u_0 < u < \min_{j=1,2,\ldots,N} x_j$, and is continuous to the left of $\min_{j=1,2,\ldots,N} x_j$, the result follows. □

We now provide some additional properties which will be useful to design an algorithm which outputs ε-optimal parameter values.

PROPOSITION 4. *(Rockette, Antle and Klimko [27].) The implicit function* $\hat{w}(u)$ *defined as the solution of* $\partial L^*/\partial w\,(u, w) = 0$ *is non-increasing.*

Proof. See [27]. □

PROPOSITION 5. *The implicit function* $\hat{w}(u)$ *is of class* C^1 *on an open set* Ω *containing* $[0, u_0]$.

Proof. We proved in Theorem 1 that:

$$\forall \tilde{u} \in [0, u_0] \quad \exists! \tilde{w} \in W \text{ such that } \frac{\partial L^*}{\partial w}(\tilde{u}, \tilde{w}) = 0\,,$$

or equivalently, such that:

$$f(\tilde{u}, \tilde{w}) = 0\,, \text{ where } f(u, w) = \sum_{j=1}^{N} e^{w z_j}(1 - w z_j)$$

$$\text{and } z_j = \ln(x_j - u) - \frac{1}{N} \sum_{j=1}^{N} \ln(x_j - u)\,.$$

The function f is clearly of class C^1 on $U \times W$ and for every $(u, w) \in U \times W$, $\partial f/\partial w(u, w) = -w \sum_{j=1}^{N} z_j^2 e^{wz_j} \neq 0$.

Hence, the basic theorem on implicit functions (see Arnaudies and Fraysse [2], p. 221) applied for every $\tilde{u} \in [0, u_0]$, states that there exists a neighborhood $V_{\tilde{u}}$ of $\tilde{u}$ and a function $w_{\tilde{u}}$ of class C^1 on $V_{\tilde{u}}$, such that $\tilde{w} = w_{\tilde{u}}(\tilde{u})$ and $\forall u \in V_{\tilde{u}}$ $f\Big(u, w_{\tilde{u}}(u)\Big) = 0$.

The implicit functions theorem also states that if there are two implicit functions $w_{\tilde{u}_1}$ and $w_{\tilde{u}_2}$ taking the same value at some point in $V_{\tilde{u}_1} \cap V_{\tilde{u}_2}$, then the two functions are the same (unicity property, Arnaudies and Fraysse [2], p. 221). Therefore, if we set $\Omega = \bigcup_{\tilde{u} \in [0, u_0]} V_{\tilde{u}}$, the function $\hat{w}$ defined by $\hat{w}(\tilde{u}) = \tilde{w}$ is of class C^1 on Ω which contains $[0, u_0]$. □

PROPOSITION 6. *The function* $\hat{v}(u, w) = \sum_{j=1}^{N} (x_j - u)^w / N$ *is non-increasing with respect to* u *and convex with respect to* w *on* $U \times W$.

Proof. When $0 < u < \min_{j=1,2,\ldots,N} x_j$ and $w > 1$, the first derivative of $\hat{v}$ with respect to u, is equal to:

$$\frac{\partial \hat{v}}{\partial u}(u, w) = \frac{-1}{N} \sum_{j=1}^{N} w(x_j - u)^{(w-1)} ,$$

which is a non-positive function, and the second derivative of $\hat{v}$ with respect to w, is equal to:

$$\frac{\partial^2 \hat{v}}{\partial w^2}(u, w) = \frac{1}{N} \sum_{j=1}^{N} (x_j - u)^w \ln^2(x_j - u) ,$$

which is a non-negative function. As $\hat{v}$ is continuous on $U \times W$, the result follows. □

PROPOSITION 7. *There exists an open set* Ω *such that*:

(i) $[0, u_0] \subset \Omega$

(ii) L^{**} *is of class* C^1 *on* Ω.

Proof. We proved in Proposition 2 that:

$$\forall u \in [0, u_0] \qquad L^{**} = L^*(u, \hat{w})$$

where $\hat{w}$ is the implicit function defined on $[0, u_0]$ by $f\Big(u, \hat{w}(u)\Big) = 0$.

We proved in Proposition 5 that there exists an open set Ω_1 containing $[0, u_0]$ and such that $\hat{w}$ is of class C^1 on Ω_1.

Moreover, the function L^* defined explicitly by

$$L^*(u, w) = N \ln(Nw) - N - N \ln \sum_{j=1}^{N} (x_j - u)^w + (w - 1) \sum_{j=1}^{N} \ln(x_j - u)$$

is of class $\mathcal{C}^1$ for $u < \min_{j=1,2,\ldots,N} x_j$ and $w > 1$.

Hence, if we denote by Ω, the open set $\Omega_1 \cap]-\infty, \min x_j[$, the function L^{**}, as the composition of the two functions L^* and $\hat{w}$, of class $\mathcal{C}^1$ respectively on $\Omega \times W$ and on Ω, is also of class $\mathcal{C}^1$ on Ω. □

4. An Exact Algorithm

We present in this section an algorithm, called MLEW, to find the maximum log-likelihood estimators for the three-parameter Weibull distribution. As shown in Section 3, this problem reduces to maximization of the function L^{**} where:

$$L^{**}(u) = \max_{w \in W} L^*(u, w) = L^*(u, \hat{w}) = \max_{v \in V} L(u, v, \hat{w}) = L(u, \hat{v}, \hat{w}) .$$

The proposed algorithm is an outer-approximation method, based on Piyavskii's algorithm [23, 24]. A piecewise linear upper-bounding function $\bar{L}^{**}$ of L^{**} is constructed and updated at each iteration. The new evaluation point p_k at iteration k corresponds to the maximum of $\bar{L}^{**}$ at that iteration. This last function is updated using lower and upper bounds K^L and K^U on the slope of $L^{**}(u)$ first between the closest evaluation point u_k^L on the left of p_k (or 0 if none exists) and p_k, and then between p_k and the closest evaluation point u_k^U on the right of p_k (which, due to the initial steps always exists, see below). These bounds are obtained by using results of Section 3 and range inclusion techniques from interval analysis (Moore [20], Ratschek and Rokne [25, 26]). The algorithm stops when the difference between the maximum of function $\bar{L}^{**}$ and the incumbent value L_{opt} does not exceed a given tolerance ε. An alternate stopping rule, described later, considers the optimal location of parameters u, v and w.

Thus, we first present, in Subsection 4.1, an algorithm for finding an ε-optimal value for L. Extensions for finding ε-optimal values for the parameters are discussed in Subsection 4.2.

4.1. Finding an ε-Optimal Value for L

As shown by Horst and Tuy ([13, 12]) for general outer-approximation methods in global optimization, algorithm MLEW can also be viewed as a branch-and-bound method. Branching is made by separating the interval $[u_k^L, u_k^U]$ into two subintervals $[u_k^L, p_k]$ and $[p_k, u_k^U]$ which have only point p_k in common. Bounding $L(u)$ on the interval $[u_k^L, u_k^U]$ is done by finding the intersection point $\left(p_k, \bar{L}^{**}(p_k)\right)$ of lines with slopes K_k^U and K_k^L going through the points $\left(u_k^L, L^{**}(u_k^L)\right)$ and $\left(u_k^U, L^{**}(u_k^U)\right)$ respectively (see Figure 1). The point $\left(p_k, \bar{L}^{**}(p_k)\right)$ is a local maximum of the current upper bounding function and is called a *peak point*. Each

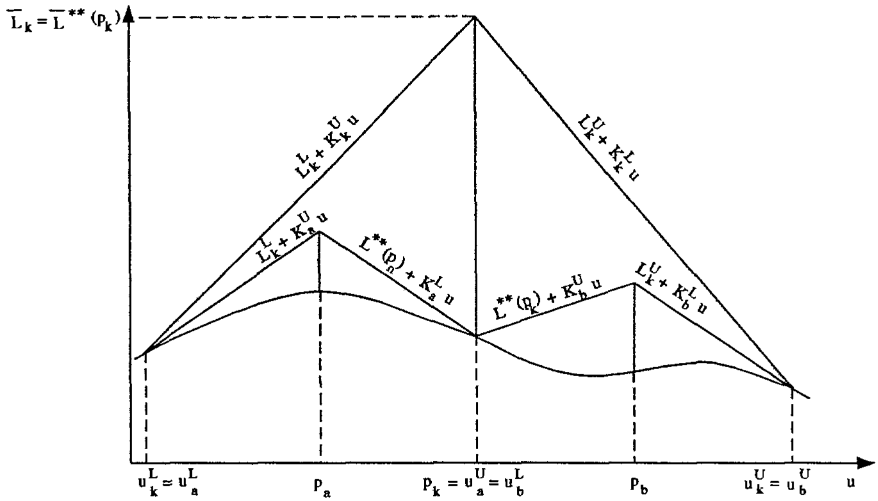

Fig. 1. Illustration of MLEW algorithm.

subproblem is characterized by the quintuplet $(p_k, \bar{L}_k, [u_k^L, u_k^U], L_k^L, L_k^U)$ where $\bar{L}_k = \bar{L}^{**}(p_k)$, $L_k^L = L^{**}(u_k^L)$ and $L_k^U = L^{**}(u_k^U)$. Subproblems are stored in a list, accessed through an adequate data structure for the operations MAX, MIN, INSERT and DELETE, i.e., a double-ended priority queue such as a min-max heap (see Atkinson *et al.* [3], Hasham and Sack [11]).

Formal rules of the algorithm are given next, followed by their justification, proof of the algorithm's finite ε-convergence and estimation of its rate of convergence.

ALGORITHM MLEW

1. Initialization of the location, scale and shape parameters. Sufficient condition for a corner solution.

$u_{opt} \leftarrow \min_{j=1,2,\ldots,N} x_j$;

$w_{opt} \leftarrow 1$;

$v_{opt} \leftarrow \frac{1}{N} \sum_{j=1}^{N} (x_j - u_{opt})$;

$L_{opt} \leftarrow -N \ln v_{opt} - N$;

If $\frac{\partial L^*}{\partial w}(0,1) < 0$ **then**

stop: $(u_{opt}, v_{opt}, w_{opt})$ is an optimal solution with value L_{opt}

Endif ;

$u_0 \leftarrow$ root of $\dfrac{\partial L^*}{\partial w}(u,1) = 0$ on $[0, \min_{j=1,2,\ldots,N} x_j[$;

2. Initialization of the upper bounding function.

$[K_0^L, K_0^U] \leftarrow$ interval containing the range of the slope of $L^{**}(u)$ for $0 \le u \le u_0$;

$u_1 \leftarrow 0$;

$[K_w^L, K_w^U] \leftarrow$ interval containing the range of the slope of $\dfrac{\partial L^*}{\partial w}(0,w)$ for $w \ge 1$;

$K_w \leftarrow \max\{|K_w^L|, |K_w^U|\}$;

$w_1 \leftarrow$ root of $\dfrac{\partial L^*}{\partial w}(0,w) = 0$ over $[1, +\infty[$ with a precision $K_w \dfrac{\varepsilon}{2}$;

$v_1 \leftarrow \dfrac{1}{N} \sum_{j=1}^{N} x_j^{w_1}$;

If $L(u_1, v_1, w_1) > L_{opt}$ **then**

$L_{opt} \leftarrow L(u_1, v_1, w_1)$;

$(u_{opt}, v_{opt}, w_{opt}) \leftarrow (u_1, v_1, w_1)$

Endif ;

If $K_0^U \le 0$ or $K_0^L \ge 0$ **then**

stop: $(u_{opt}, v_{opt}, w_{opt})$ is an optimal solution with value L_{opt}

Endif ;

$\bar{L}_2^{**}(u) \leftarrow \min\{L^{**}(0) + K_0^U u, L^{**}(u_0) + K_0^L(u - u_0)\}$ for $0 \le u \le u_0$;

$p_2 \leftarrow \arg \max_{u \in [0,u_0]} \bar{L}_2^{**}(u)$;

$\bar{L}_2 \leftarrow \bar{L}_2^{**}(p_2)$;

Insert $P_2 = \left(p_2, \bar{L}_2, [0, u_0], L^{**}(0), L^{**}(u_0)\right)$ in *list* ;

$k \leftarrow 2$;

3. Optimality (ε-optimal value) and first range reduction tests.

Extract from *list* the subproblem $P_l \leftarrow \left(p_l, \bar{L}_l, [u_l^L, u_l^U], L^{**}(u_l^L), L^{**}(u_l^U)\right)$ for which $\bar{L}_l = \max_i \bar{L}_i$;

$[K_w^L, K_w^U] \leftarrow$ interval containing the range of the slope of $\dfrac{\partial L^*}{\partial w}(p_l, w)$ for $w \ge 1$;

$K_w \leftarrow \max\{|K_w^L|, |K_w^U|\}$;

$w_l \leftarrow$ root of $\dfrac{\partial L^*}{\partial w}(p_l, w) = 0$ over $[1, +\infty[$ with a precision $K_w \dfrac{\varepsilon}{2}$;

$v_l \leftarrow \dfrac{1}{N} \sum_{j=1}^{N} (x_j - p_l)^{w_l}$;

If $L(p_l, v_l, w_l) > L_{opt}$ **then**

$L_{opt} \leftarrow L(p_l, v_l, w_l)$;

$(u_{opt}, v_{opt}, w_{opt}) \leftarrow (p_l, v_l, w_l)$;

delete from *list* all subproblems P_s with $\bar{L}_s \leq L_{opt}$

Endif ;

If $\bar{L}_l - L_{opt} \leq \frac{\varepsilon}{2}$ **then**

stop: L_{opt} is an ε-optimal value

Endif ;

4. Branching (new subproblems) and monotonicity range reduction test.

$[\underline{u}_a, \bar{u}_a] \leftarrow [u_l^L, p_l]$;

$[\underline{u}_b, \bar{u}_b] \leftarrow [p_l, u_l^U]$;

For $i = a, b$ **do**

$[K^L, K^U] \leftarrow$ interval containing the range of the slope of L^{**} on $[\underline{u}_i, \bar{u}_i]$;

if ($K^L \leq 0$ **and** $K^U \geq 0$) **then**

$\bar{L}^{**}_{k+1}(u) \leftarrow \min\{L^{**}(\underline{u}_i) + K^U(u - \underline{u}_i), L^{**}(\bar{u}_i) + K^L(u - \bar{u}_i)\}$;

$p_{k+1} \leftarrow \arg \max_{u \in [\underline{u}_i, \bar{u}_i]} \bar{L}^{**}_{k+1}(u)$;

$\bar{L}_{k+1} \leftarrow \bar{L}^{**}_{k+1}(p_{k+1})$;

insert in *list* $P_{k+1} = \Big(p_{k+1}, \bar{L}_{k+1}, [\underline{u}_i, \bar{u}_i], L^{**}(\underline{u}_i), L^{**}(\bar{u}_i)\Big)$;

$k \leftarrow k + 1$

endif

End For ;

Return to Step 3.

Step 1 of algorithm MLEW computes, for the largest possible value of u, i.e., $\min_{j=1,2,\ldots,N} x_j$, the optimal values of the parameters v and w, i.e., $\hat{v}$ which is given in Proposition 1, and $\hat{w}$, which, according to Theorem 1, is equal to 1. It then checks if the sufficient condition of Proposition 2 for this point to be optimal is satisfied and if it is the case the procedure stops.

Step 2 first computes u_0 (using Newton's method) the largest value of u for which $\hat{w} > 1$. According to Proposition 4, the open interval $]u_0, \min_{j=1,2,\ldots,N} x_j[$ cannot contain an optimal value for u, and hence is discarded. A subroutine, using the natural interval extension of interval arithmetic (see, e.g., Ratschek and Rokne [25, 26]), is then used to bound the slope of $L^{**}(u)$ on the interval $[0, u_0]$. This subroutine provides an interval $[K_0^L, K_0^U]$ which is an inclusion for the range of the derivative of $L^{**}(u)$ over a given interval $[a, b]$. The upper bounding function $\bar{L}^{**}(u)$ is then initialized as the lower envelope of lines with slopes K_0^U and K_0^L

through points $\left(0, L^{**}(0)\right)$ and $\left(u_0, L^{**}(u_0)\right)$ respectively. The first subproblem, P_2, is then obtained by computing the maximum $\bar{L}_2$ of this upper bounding function and its argument p_2.

Current iterations are described from Step 3 onwards: the subproblem with the largest upper bound is selected and the function $L^{**}(u) = L\left(u, \hat{v}(u, \hat{w}), \hat{w}(u)\right)$ evaluated at its peak point. Note that $\hat{w}(u)$ can be computed using, e.g., Newton's method as $\partial L^*/\partial w\,(p_l, w) = 0$ has a unique root over $[1, +\infty[$ (see Theorem 1). If the new evaluation point is better than the incumbent one it replaces it, the incumbent value is updated and the list of subproblems trimmed by deleting those with an upper bound not larger than the incumbent one (first range reduction test). The stopping criterion is then applied: the algorithm ends if the difference between the upper bound and the incumbent value is sufficiently small.

Branching takes place in Step 4, where the current problem is bipartitioned. To this effect the corresponding interval of values of u is split at the peak point of $\bar{L}^{**}(u)$ and the range of its derivative is evaluated on each of the subintervals so obtained. If the lower and upper bounds on the derivative of $L^{**}(u)$ are respectively non-positive and non-negative, the upper bounding function $\bar{L}^{**}$ is updated on this subinterval, its maximum and corresponding peak point found and the data quintuplet inserted in the list of subproblems. If a lower bound on the derivative of $L^{**}(u)$ is positive or an upper bound negative the corresponding subproblem is discarded, i.e., not inserted in the list (monotonicity range reduction test). Indeed in such a case a globally optimal value for a subproblem can only occur at one of the extreme point of the corresponding subinterval (assuming L^{**} to be continuous, which has been shown in Proposition 7) and $L^{**}(u)$ has already been evaluated at both of these extreme points.

We now prove finite ε-convergence of Algorithm MLEW.

THEOREM 2. *Algorithm* MLEW *finds an ε-optimal value for the log-likelihood function $L(u, v, w)$ of the three-parameter Weibull distribution in finite time for any given positive constant ε.*

Proof. We first show that each step of algorithm MLEW takes a finite time. All steps involve only simple calculations except for computation of the root of $\partial L^*/\partial w\,(u, 1) = 0$ in Step 1, of the roots of $\partial L^*/\partial w\,(0, w) = 0$ in Step 2 and of $\partial L^*/\partial w\,(p_l, w) = 0$ in Step 3. From Theorem 1, the root u_0 of $\partial L^*/\partial w\,(u, 1) = 0$ is unique; it can be found in finite time, to any desired precision ε_{u_0}, by dichotomous search. Let u_0^U be an ε_{u_0}-optimal overestimate of u_0. We eliminate the interval $]u_0^U, \min_{j=1,2,\ldots,N} x_i[$ and go on with the optimization of L^{**} on $[0, u_0^U]$. From Proposition 4, the function $\hat{w}(u)$ defined as the solution of $\partial L^*/\partial w\,(u, w) = 0$ is non-increasing, and from Proposition 5, it has a continuous derivative. Therefore $\hat{w}(u)$ can be found in finite time for fixed u, to any desired precision $\varepsilon_{\hat{w}}$ on w, by using Newton's method (or dichotomous search in case of non convergence of Newton's method). In order to reach an $\varepsilon/2$-optimal value of L^{**}, $\varepsilon_{\hat{w}}$ can be chosen to be equal to $\varepsilon K_w/2$, where K_w is an upper-bound on the absolute value of $\partial L^*/\partial w\,(u, w)$.

We next note that while Steps 1 and 2 occur once, Steps 3 and 4 are repeated until the stopping condition $\bar{L}_l - L_{opt} \le \varepsilon/2$ holds. At any iteration k, the upper bounding function $\bar{L}_k^{**}$ is updated into $\bar{L}_{k+1}^{**}$: the peak point p_l disappears and is replaced by at most two new ones, as one or both of the subintervals obtained may be deleted due to the monotonicity test. Therefore we are left with subproblems such that $K^L < 0$ and $K^U > 0$ on each interval $[\underline{u}_l, \bar{u}_l]$. Moreover the width of the interval $[\underline{u}_l, \bar{u}_l]$ where $\underline{u}_l$ and $\bar{u}_l$ are evaluation points, respectively to the left and the right of peak point p_l is at least $(1/K^U - 1/K^L)\varepsilon/2 \ge (1/K_0^U - 1/K_0^L)\varepsilon/2 \ge \varepsilon/K$ where $K = \max\{-K_0^L, K_0^U\}$ (as $L^{**}(p_l) - L_{opt} \ge \varepsilon/2$ implies $L^{**}(p_l) - \max\{L^{**}(\underline{u}_l), L^{**}(\bar{u}_l)\} \ge \varepsilon/2$) if branching is to occur on $[\underline{u}_l, \bar{u}_l]$.

Therefore, the range of values of u for any subproblem obtained by branching on $[\underline{u}_l, \bar{u}_l]$ is at least of width ε/K. Comparing this value with the width of $[0, u_0]$ shows that the number of iterations of algorithm MLEW is bounded by $u_0 K/\varepsilon + 1$, which is finite for any strictly positive ε.

Due to the precision $\varepsilon_w = \varepsilon K_w/2$ used to compute $\hat{w}(u)$, the resulting error on $L^{**}(u)$ does not exceed $\varepsilon/2$. Moreover, as the stopping criterion is performed with $\varepsilon/2$, this yields an ε-optimal value. □

As with other global optimization algorithms, the rate of convergence of algorithm MLEW is not large, in the worst case.

THEOREM 3. *Algorithm* MLEW *has a logarithmic rate of convergence.*

Proof. Consider algorithm MLEW at the n^{th} iteration. Let:

$$F_n^* = \max_{u \in [0, u_0]} \bar{L}^{**}(u)\,,$$

$$L_{opt}^n = \max_{k=1,2,\ldots,n} L^{**}(u_k)\,,$$

$$\varepsilon_n = F_n^* - L_{opt}^n\,.$$

After n iterations, the upper-bounding function $\bar{L}^{**}$ has, at most, $n-1$ peak points, which we call *first generation* peak points. After at most $n-1$ additional iterations, the highest peak point is not one of these, i.e., it is a second generation one. Let m be the iteration at which this is the case for the first time ($m \le 2n-1$). Let i be the iteration at which the peak point selected at iteration m has been obtained.

The upper bound

$$F_m^* = \bar{L}^{**}(p_m) = \max\{\bar{L}^{**}(p_a), \bar{L}^{**}(p_b)\}$$

where p_a and p_b are the peak points obtained when updating $\bar{L}^{**}(u)$ on $[u_i^L, u_i^U]$ (see Figure 2). Bounds on the slope of $L^{**}(u)$ on the intervals $[u_i^L, u_i = p_i]$ and $[u_i, u_i^U]$ have been used for that purpose. If the less precise bounds defined on $[u_i^L, u_i^U]$ had

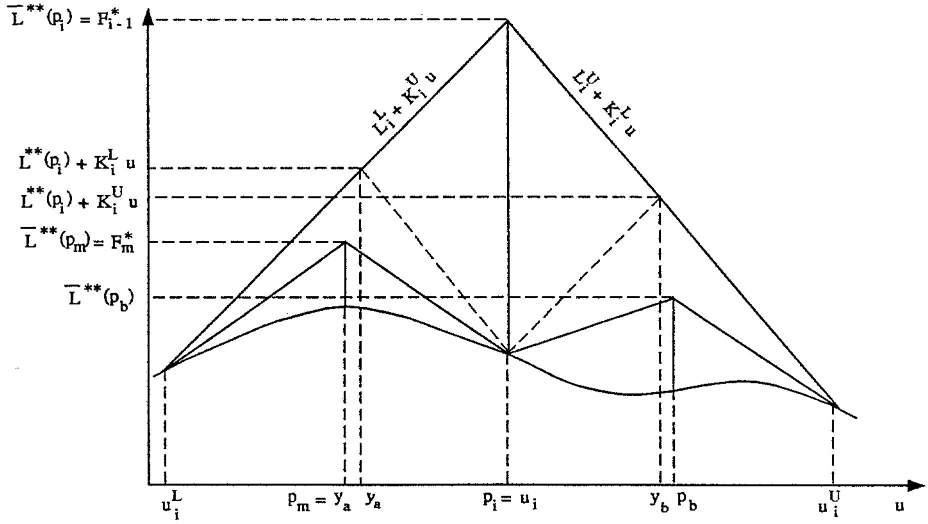

Fig. 2. Current iteration of MLEW algorithm.

been used instead one would have obtained peak points with ordinates y_a (at the intersection of lines $F_{i-1}^* + K_i^U(u - u_i)$ and $L^{**}(u_i) + K_i^L(u - u_i)$) and y_b (at the intersection of lines $F_{i-1}^* + K_i^L(u - u_i)$ and $L^{**}(u_i) + K_i^U(u - u_i)$, see again Figure 2). Then, from elementary geometry

$$F_m^* \leq \max\Big(\bar{L}^{**}(y_a), \bar{L}^{**}(y_b)\Big) = \begin{cases} \dfrac{K_i^U\, L^{**}(u_i) - K_i^L\, F_{i-1}^*}{K_i^U - K_i^L} & \text{if } K_i^U + K_i^L \leq 0 \\[2ex] \dfrac{K_i^U\, F_{i-1}^* - K_i^L\, L^{**}(u_i)}{K_i^U - K_i^L} & \text{if } K_i^U + K_i^L > 0\ . \end{cases}$$

Consider now $F_m^* - L_{opt}^n$ as a function of ε_n; as ε_n does not increase with n, $\varepsilon_{2n-1} = F_{2n-1}^* - L_{opt}^{2n-1} \leq F_m^* - L_{opt}^m$.

Two cases may arise:

(i) if $L^{**}(u_i) > L_{opt}^{i-1}$ then $L_{opt}^i = L^{**}(u_i)$;

(ii) if $L^{**}(u_i) \leq L_{opt}^{i-1}$ then $L_{opt}^i = L_{opt}^{i-1}$.

In case (i), $L_{opt}^m \geq L_{opt}^i = L^{**}(u_i)$. Setting $K^{sup} = \max(K_i^U, -K_i^L)$ and $K^{inf} = \min(K_i^U, -K_i^L)$, as $F_{i-1}^* > L^{**}(u_i)$ one has:

$$F_m^* - L_{opt}^m \leq \frac{K_i^{sup}\, F_{i-1}^* + K_i^{inf}\, L^{**}(u_i)}{K_i^U - K_i^L} - L_{opt}^i$$

$$\leq \frac{K_i^{sup} F_{i-1}^* + K_i^{inf} L_{opt}^i}{K_i^U - K_i^L} - \frac{(K_i^{sup} + K_i^{inf}) L_{opt}^i}{K_i^U - K_i^L}$$

$$= \frac{K_i^{sup}}{K_i^U - K_i^L} (F_{i-1}^* - L_{opt}^i) \leq \frac{K^{sup}}{K_i^U - K_i^L} (F_{i-1}^* - L_{opt}^{i-1})$$

$$\leq \frac{K_i^{sup}}{K_i^U - K_i^L} \varepsilon_n \, .$$

Similarly, in case (ii), as $L_{opt}^m \geq L_{opt}^{i-1} = L^{**}(u_i)$,

$$F_m^* - L_{opt}^m \leq \frac{K_i^{sup}}{K_i^U - K_i^L} (F_{i-1}^* - L_{opt}^{i-1}) \leq \frac{K^{sup}}{K_i^U - K_i^L} \varepsilon_n \, .$$

Hence,

$$\varepsilon_{2n-1} \leq \frac{K_i^{sup}}{K_i^U - K_i^L} \varepsilon_n \, .$$

Moreover,

$$\frac{K_i^{sup}}{K_i^U - K_i^L} = \frac{K_i^{sup}}{K_i^{sup} + K_i^{inf}} = K < 1 \text{ as } K_i^{inf} \geq \frac{\varepsilon_n}{u_i^U - u_i^L} \geq \frac{\varepsilon_n}{u_0} \, .$$

Therefore as doubling the number of iterations decreases the error by a positive percentage, algorithm MLEW has a logarithmic rate of convergence. □

4.2. Finding an ε-Optimal Parameter Vector

In this subsection, we discuss an alternate stopping rule that guarantees the ε-optimality of the parameter vector. We suppose a tolerance vector $(\varepsilon_u, \varepsilon_v, \varepsilon_w)$ on the parameter vector (u, v, w) is given. We proceed in three steps. First we obtain an ε_u-optimal value for the location parameter, then an ε_w-optimal value for the shape parameter and finally an ε_v-optimal value for the scale parameter.

Obtaining an ε_u-optimal value for the shape parameter

Algorithm MLEW iteratively reduces the interval $[0, \min_{j=1,\ldots,N} x_j]$ by discarding subintervals that cannot contain a globally optimal parameter u^*. Let $\mathcal{A}$ be the set of indices of active subproblems, i.e., subproblems that have not been discarded once an ε-optimal value for L^{**} has been reached. The remaining active subspace is a union of subintervals:

$$\bigcup_{l \in \mathcal{A}} [u_l^L, u_l^U] \, .$$

Denote by u^L and u^U the leftmost and rightmost endpoints of the remaining subintervals, respectively. If $u^U - u^L \leq \varepsilon_u$, and ε_u-optimal value has been reached for u. Assuming there is only one global optimum point, after further iterations of algorithm MLEW, the condition $u^U - u^L \leq \varepsilon_u$ can be satisfied. Otherwise, one could adapt the algorithm of Hansen, Jaumard and Lu [8, 9] for global optimization of univariate Lipschitz functions. This algorithm allows the localization of each global optimum of such a function within an interval containing only ε-optimal points, provided any two successive local optima differ in value by at least ε. In such a case, denote by u^L and u^U the extreme points of one of the remaining intervals. This, however, has not been needed as for all tested samples, algorithm MLEW has always provided solutions satisfying the tolerance condition on u.

Obtaining an ε_v-optimal value for the shape parameter

It follows from Proposition 4, which states that $\tilde{w}(u)$ is non-increasing, that the remaining w-subspace is included in $[w^L, w^U]$ where $w^L = \tilde{w}(u^U)$ and $w^U = \tilde{w}(u^L)$. After the first condition $u^U - u^L \leq \varepsilon_u$ has been fulfilled, algorithm MLEW will carry on until $(w^U - w^L) < \varepsilon_w$, is satisfied.

Obtaining an ε_w-optimal value for the scale parameter

Proposition 6 enables us to find bounds on the remaining v-subspace: as $\hat{v}$ is non-increasing with respect to u, the lower bound in v is necessarily reached when u is equal to u^U and the upper bound when u is equal to u^L. Moreover, as $\hat{v}$ is convex with respect to w, the lower bound in v is reached at the minimum of a convex function over a closed set, and the upper bound when w is equal to w^L or w is equal to w^U.

Combining these results, it is easy to deduce the value of the upper bound in v (see Figure 3):

$$v^U = \sup\left(\hat{v}(u^L, w^L), \hat{v}(u^L, w^U)\right) .$$

To find the lower bound in v, one first has to check whether the minimum of the convex function $\hat{v}(u^U, w)$ occurs within the bounds w^L and w^U or not. This can easily be done by computing the first derivative of v with respect to w at $u = u^L$ and $u = u^U$. In case the minimum occurs within these bounds, rather than computing its exact value, which is too time consuming for a stopping test performed at each iteration, algorithm MLEW computes a linear approximation of this minimum. The exact value or the approximation of the lower bound in v is obtained by the following algorithm:

If $\dfrac{\partial v}{\partial w}(u^U, w^L) \geq 0$ **then**

$\quad v^L \leftarrow \hat{v}(u^U, w^L)$

else

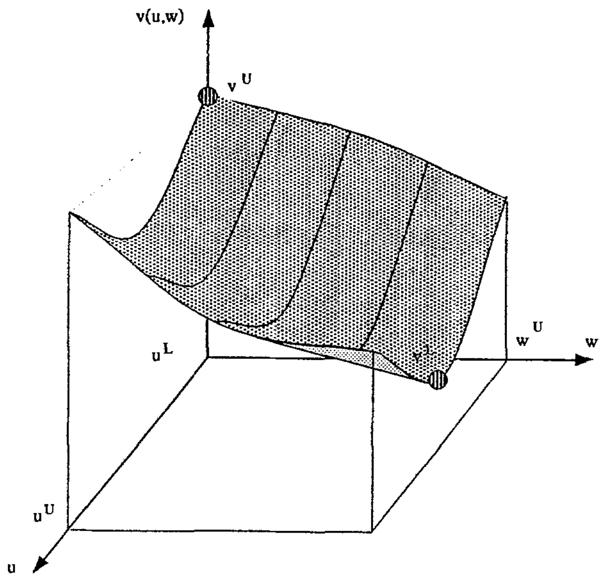

Fig. 3. Shape of the implicit function $v(u, w)$.

if $\dfrac{\partial v}{\partial w}(u^U, w^U) \leq 0$ **then**

$v^L \leftarrow \hat{v}(u^U, w^U)$

else

$$v^L \leftarrow \frac{\left(\frac{\partial v}{\partial w}(u^U,w^U)\hat{v}(v^U,w^L)(1-\frac{\partial v}{\partial w}(u^U,w^L))-\frac{\partial v}{\partial w}(u^U,w^L)\hat{v}(v^U,w^U)(1-\frac{\partial v}{\partial w}(u^U,w^U))\right)}{\left(\frac{\partial v}{\partial w}(u^U,w^U)-\frac{\partial v}{\partial w}(u^U,w^L)\right)}$$

endif

Endif

When the third condition $v^U - v^L < \varepsilon_v$ is also satisfied, as the two previous criteria are still holding, the algorithm stops and the parameter vector provided is then $(\varepsilon_u, \varepsilon_v, \varepsilon_w)$-optimal.

5. Experimental Results

The algorithms described in the previous section are now applied to several samples from the literature to illustrate their efficiency. A comparison with the method of Panchang and Gupta [21] is also made. Both versions of algorithm MLEW are implemented in Fortran 77 and run on a Sun SPARC (with a 16 Mips processor). The data are presented in Table I, with their reference in brackets. Table II provides, for each sample, the value of the maximum of the log-likelihood function, obtained

TABLE I. Test samples of a Weibull distribution.

sample	observations					
1 [27]	3.1	4.6	5.6	6.8		
2 [22]	35.0	34.7	30.9	29.0	28.1	24.9
	24.0	23.9	23.3	22.6	22.4	19.8
	19.8	19.4	19.0	17.6	16.5	15.9
	13.3	12.3	12.0	12.0		
3 [1]	10.0805	10.0990	10.2757	10.6545	10.6883	11.0666
	11.2083	11.2558	11.8761	12.2103		
4 [15]	143	164	188	188	190	192
	206	209	213	216	220	227
	230	234	246	265	304	216*
	244*					
51 [4]	15	20	27	42	42	43
	44	46	64	65		
52 [4]	sample 51 +					
	65	68	68	71	74	75
	75	76	77	78		
61 [17]	540	700	800	900	980	1060
	1120	1180	1250	1340	1400	1500
	1500*	1500*	1500*	1500*		
62 [17]	sample 61 +					
	1500*	1500*	1500*	1500*		
7 [16]	17.88	28.92	33.00	41.52	42.12	45.60
	48.48	51.84	51.96	54.12	55.56	67.80
	68.64	68.64	68.88	84.12	93.12	98.64
	105.12	105.84	127.92	128.04	173.40	
81 [19]	93.4	98.7	116.6	117.8	132.7	136.6
	140.3	158.0	164.8	183.9		
82 [19]	152.7	172.0	172.5	173.3	193.0	204.7
	216.5	234.9	262.6	422.6		
9 [6]	0.265	0.269	0.297	0.315	0.3235	0.338
	0.379	0.379	0.392	0.402	0.412	0.416
	0.418	0.423	0.449	0.484	0.494	0.613
	0.654	0.740				

by algorithm MLEW with a precision $\varepsilon = 10^{-6}$, and the corresponding vector of parameter values. Characteristics of the solution process are presented in Table III and are the following: ε_{value} is the precision required, $\bar{L} - L_{opt}$ is the precision obtained at the end of the solution process, *tcpu* is the computing time (user time) in seconds, *iter* is the number of iterations (parameter k in algorithm MLEW), *disc*1 is the number of intervals discarded by the first range reduction test of algorithm MLEW, *disc*2 is the number of intervals discarded by the monotonicity range reduction test. It appears that: *(i)* solutions for all samples are obtained in very moderate computing time; *(ii)* the number of iterations is also moderate; *(iii)* both range reduction tests are effective, the first one more than the second one on average, but their relative usefulness varies from sample to sample; *(iv)* increasing precision from $\varepsilon = 10^{-1}$ to $\varepsilon = 10^{-6}$ moderately augments computation times and number of iterations (by not more than a factor of 4).

TABLE II. Optimal values and vectors for the test sample sets.

sample	bounds/value	u_{opt}	v_{opt}	w_{opt}	L_{opt}
1	value	3.100000	1.925000	1.000000	-6.619704
2	value	10.98784	41.86285	1.515528	-71.75690
3	value	10.08050	0.861010	1.000000	-8.503508
4	value	122.0259	329418.6	2.711477	-87.32424
51	lower	1.052543	43749.96	2.811973	-41.54652
	upper	1.052635	43751.49	2.811981	-41.54652
52	value	0.000000	1604933.	3.445769	-87.74309
61	value	425.6769	646838.5	2.000065	-91.96168
62	value	474.8136	55280.13	1.633255	-96.30728
7	value	14.87591	755.6003	1.594299	-112.8502
81	value	86.57383	1001.132	1.737348	-46.8779
82	value	152.7000	67.78000	1.000000	-52.1627
9	value	0.261144	0.112443	1.244496	17.0683

In the last column of Table III, *passiv* is the number of iterations required by a passive algorithm (grid search) to provide an ε-optimal value for $L^{**}(u)$. This last parameter allows an indirect empirical comparison with Panchang and Gupta's method [21] modified to guarantee an ε-optimal value. Indeed, consider a function f of class C^1 and assume one wants to evaluate its optimal value on an interval $[\underline{u}, \bar{u}]$ (this is the aim of Panchang and Gupta's method for function $L^{**}(u)$). Denote by K the Lipschitz constant of f on the interval $[\underline{u}, \bar{u}]$, i.e., the best possible bound on the absolute value of its first order derivative on $[\underline{u}, \bar{u}]$. Then the number of iterations required to guarantee an ε-optimal value when applying a passive algorithm (or grid search) is $[K(\bar{u} - \underline{u})]/\varepsilon$. This is therefore the minimum number of iterations needed by Panchang and Gupta's algorithm to guarantee a precision of ε. An iteration of Panchang and Gupta's algorithm requires an amount of computations in the same order of magnitude as an iteration of algorithm MLEW. Therefore the latter algorithm is much faster than the former. Indeed, for some test problems (samples 3, 4 and 81) Panchang and Gupta's method requires an inordinate number of iterations to guarantee even a low precision of $\varepsilon = 10^{-1}$. The maximum log-likelihood function of sample 2 has been represented on Figure 4.

Results obtained with the algorithm MLEW modified in order to obtain an ε-optimal vector are described in Table IV. The characteristic parameters are the following: ε_{vector} is the precision required on each parameter (here the same for each of the three parameters), Δ_u (resp. Δ_v, Δ_w) is the relative precision obtained for parameter u (resp. v, w) when the algorithm stops. When the global optimum is reached at the local optimum vector characterized by $u = \min_{j=1,2,\ldots,N} x_j$, its value can be computed analytically, hence with $\Delta_u = \Delta_v = \Delta_w = 0$. The parameters *tcpu*, *iter*, *disc*1 and *disc*2 have the same meaning as in Table III. When the required precision cannot be reached in a reasonable amount of computing time,

TABLE III. Performance of Algorithm MLEW (ε-optimal value).

sample	ε_{value}	$\bar{L} - L_{opt}$	tcpu	iter	disc1	disc2	passiv
1	1e-01	5e-02	7	12	9	0	7e+04
	1e-03	0	7	13	11	0	7e+06
	1e-06	0	7	13	11	0	7e+09
2	1e-01	8e-02	24	17	9	0	2e+05
	1e-03	9e-04	42	30	23	0	2e+07
	1e-06	8e-07	75	50	43	3	2e+10
3	1e-01	5e-02	13	21	17	0	5e+22
	1e-03	0	17	23	21	0	5e+24
	1e-06	0	17	23	21	0	5e+27
4	1e-01	5e-02	30	25	12	6	1e+19
	1e-03	8e-04	48	38	25	6	1e+21
	1e-06	8e-07	71	59	42	10	1e+24
51	1e-01	3e-01	7	12	4	1	9e+05
	1e-03	7e-04	20	27	14	1	9e+07
	1e-06	9e-07	45	61	50	4	9e+10
52	1e-01	0	15	13	1	10	7e+09
	1e-03	0	15	13	1	10	7e+11
	1e-06	0	15	13	1	10	7e+14
61	1e-01	9e-02	15	16	5	2	2e+11
	1e-03	8e-04	35	32	21	3	2e+13
	1e-06	9e-07	67	57	44	6	2e+16
62	1e-01	8e-02	19	16	6	2	9e+09
	1e-03	8e-04	39	29	17	4	9e+11
	1e-06	7e-07	69	48	32	7	9e+14
7	1e-01	9e-02	17	13	2	5	5e+07
	1e-03	5e-04	30	20	6	9	5e+09
	1e-06	4e-07	45	31	11	16	5e+12
81	1e-01	5e-02	16	24	14	0	3e+13
	1e-03	7e-04	29	41	30	0	3e+15
	1e-06	8e-07	50	70	60	0	3e+18
82	1e-01	0	5	8	4	2	3e+07
	1e-03	0	5	8	4	2	3e+09
	1e-06	0	5	8	4	2	3e+12
9	1e-01	6e-02	16	13	6	1	4e+06
	1e-03	5e-04	31	23	11	6	4e+08
	1e-06	4e-07	45	36	22	10	4e+11

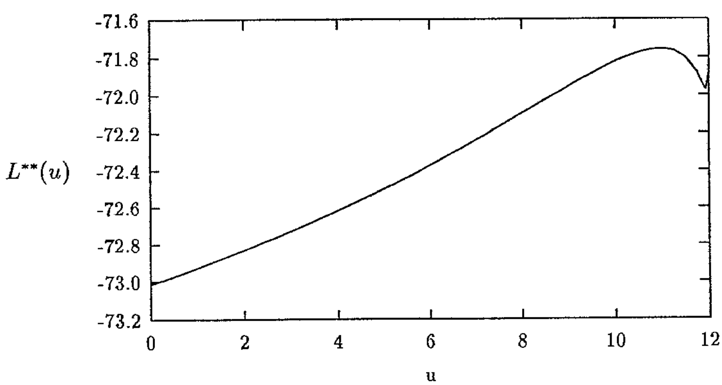

Fig. 4. Maximum log-likelihood of sample 2 ($w = \hat{w} \geq 1$).

TABLE IV. Performance of Algorithm MLEW (ε-optimal vector).

sample	ε_{vector}	Δ_u	Δ_v	Δ_w	tcpu	iter	disc1	disc2
1	1e-01	0	0	0	4	13	11	0
	1e-03	0	0	0	4	13	11	0
	1e-06	0	0	0	4	13	11	0
2	1e-01	8e-03	8e-02	2e-02	47	35	28	1
	1e-03	6e-05	6e-04	1e-04	88	62	52	4
	1e-06	1e-07	1e-06	2e-07	153	110	87	16
3	1e-01	0	0	0	14	23	21	0
	1e-03	0	0	0	14	23	21	0
	1e-06	0	0	0	14	23	21	0
4	1e-01	5e-03	9e-02	7e-03	55	47	32	9
	1e-03	3e-05	6e-04	5e-05	95	78	58	13
	1e-06	0	0	0	128	94	72	20
51	1e-01	5e-02	2e-02	1e-03	42	65	55	1
	1e-03	0	0	0	43	66	59	5
	1e-06	0	0	0	43	66	59	5
52	1e-01	0	0	0	15	13	1	10
	1e-03	0	0	0	15	13	1	10
	1e-06	0	0	0	15	13	1	10
61	1e-01	6e-03	6e-02	6e-03	53	51	40	3
	1e-03	8e-05	8e-04	7e-05	84	79	65	6
	1e-06	5e-09	5e-08	4e-09	139	128	114	10
62	1e-01	4e-04	4e-02	6e-03	48	39	27	6
	1e-03	3e-05	3e-04	5e-05	82	64	48	11
	1e-06	0	0	0	122	67	51	14
7	1e-01	3e-02	9e-02	1e-02	27	20	6	9
	1e-03	1e-04	5e-04	6e-05	51	35	13	18
	1e-06	4e-07	7e-07	9e-08	95	62	17	37
81	1e-01	3e-03	8e-02	9e-03	38	60	50	0
	1e-03	3e-05	8e-04	9e-05	63	93	85	0
	1e-06	0	0	0	83	106	99	5
82	1e-01	0	0	0	5	8	4	2
	1e-03	0	0	0	5	8	4	2
	1e-06	0	0	0	5	8	4	2
9	1e-01	8e-03	9e-02	4e-02	24	20	11	4
	1e-03	6e-05	7e-04	3e-04	48	36	22	10
	1e-06	0	0	0	63	45	31	12

a limit of 1000 iterations has been imposed (this is the case for $\varepsilon_{vector} = 1e - 06$ for samples 4, 51, 62, 81 and 9).

It appears that: *(i)* solutions for all problems are obtained in reasonable computing times; they vary greatly from problem to problem and sometimes augment sharply when high precision is required (this appears to be due to a flat optimum); *(ii)* the number of iterations remains moderate for low precision and for some samples augments sharply when high precision is required; both reduction tests are effective, the monotonicity test being the most useful one for the more difficult to solve problems.

Algorithm MLEW was also extended to the case of type I right censored samples (simple or progressive). Recall that, a sample $(x_i)_{1 \leq i \leq N}$ is said to be simply right censored (type I) if all observations higher than a given bound X_0 are set to X_0. The sample is said to be progressively right censored (type I) if all observations higher than a given bound X_0 are set to different values X_K lower than the expected observed value x_K. In both cases, only R of the values $(x_i)_{1 \leq i \leq R}$, on a total number of N, are based on real observations. The $N - R$ remaining ones are censored.

Some minor changes in the log-likelihood allow us to handle the case of right censored sample (type I):

$$L_C(u, v, w) = R \ln\left(\frac{w}{v}\right) - \frac{1}{v} \sum (x_j - u)^w + (w - 1) \sum_{j=1}^{R} \ln(x_j - u) .$$

All results of Section 3 and the algorithm MLEW described in Section 4 were adapted and 2 problems based on right-censored (type I) observations (namely, problems 61 and 62) were solved. Computation times and other characteristics appear to be similar to those obtained in the case of uncensored samples.

In conclusion, the use of global optimization methods, i.e., projection and outer approximations, allows the determination, with high precision and in very moderate computing time, of the globally optimal maximum log-likelihood parameter values for the three-parameter Weibull distribution, in both the uncensored and in the right-censored cases.

Acknowledgements

Research of the second and the third authors has been supported by AFOSR grant 90-0008 to Rutgers University. Research of the second author has been supported by NSERC (Natural Sciences and Engineering Council of Canada) grant GP0105574. Research of the third author has been supported by NSERC grant GP0036426 and FCAR (Fonds pour la Formation des Chercheurs et l'Aide à la Recherche) grants 89EQ4144, 92EQ1048 and 91NC0305. Research of the third author has also been supported by an NSF Professorship for Women in Science at Princeton University. We would also like to thank the referees for their helpful suggestions which improved the conciseness of some of the proofs.

References

1. Adatia, A. and Chan, L.K. (1985), Robust Estimators of the 3-Parameter Weibull Distribution, *IEEE Transactions on Reliability* **34**(4), 347–351.
2. Arnaudies, J.M. and Fraysse, H. (1989), *Cours de Mathématiques – 3: Compléments d'analyse*, Dunod Université, Paris.
3. Atkinson, M.D., Sack, J.-R., Santoro, N. and Strothotte, T. (1986), Min-Max Heaps and Generalized Priority Queues, *Communications of the ACM* **29**(10), 996–1000.
4. Bain, L.J. (1978), Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Statistics: Textbooks and Monographs* **24**, Marcel Dekker Inc., New York.
5. Cohen, A.C., Whitten, B.J. and Ding, Y. (1984), Modified Moment Estimation for the Three-Parameter Weibull Distribution, *Journal of Quality Technology* **16**(3), 159–167.
6. Dumonceaux, R. and Antle, C.E. (1973), Discrimination between the Log-Normal and the Weibull Distribution, *Technometrics* **15**, 923–926.

7. Geoffrion, A.M. (1972), Generalized Benders Decomposition, *Journal of Optimization Theory and Applications* **10**(4), 237–260.
8. Hansen, P., Jaumard, B. and Lu, S.-H. (1992), Global Optimization of Univariate Lipschitz Functions: I. Survey and Properties, *Mathematical Programming* **55**, 251–272.
9. Hansen, P., Jaumard, B. and Lu, S.-H. (1992), Global Optimization of Univariate Lipschitz Functions: II. New Algorithms and Computational Comparison, *Mathematical Programming* **55**, 273–292.
10. Harter, H.L. (1971), Some Optimization Problems in Parameter Estimation, *Optimizing Methods in Statistics*, 33–62.
11. Hasham, A. and Sack, J.-R. (1987), Bounds for Min-Max Heaps, *BIT* **27**, 315–323.
12. Horst, R. and Tuy, H. (1987), On the Convergence of Global Methods in Multiextremal Optimization, *Journal of Optimization Theory and Applications* **54**, 253–271.
13. Horst, R. and Tuy, H. (1990), *Global Optimization – Deterministic Approaches*, Springer-Verlag, New York (2nd edition 1992).
14. Kotz, S. and Johnson, N.L. (1973), Statistical Distributions: Survey of the Literature, Trends and Prospects, *The American Statistician* **27**(1), 15–17.
15. Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, New York, Wiley.
16. Lieblein, J. and Zelen, M. (1956), Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings, *Journal of Research of the National Bureau of Standards* **47**, 273–316.
17. Lochner, H.R., When and How to Use the Weibull Distribution, Research Report.
18. McCool, J.I. (1970), Inference on Weibull Percentiles and Shape Parameters from Maximum Likelihood Estimates, *IEEE Transactions on Reliability* **19**, 2–9.
19. McCool, J.I. (1974), Inferential techniques for Weibull populations, *Aerospace Research Laboratories Report ARL* TR 74-0180, Wright-Patterson AFB, Ohio.
20. Moore, R.E. (1979), *Methods and Applications of Interval Analysis*, SIAM Studies in Applied Mathematics, Philadelphia.
21. Panchang, V.G. and Gupta, R.C. (1989), On the Determination of Three-Parameter Weibull MLE's, *Communications in Statistics and Computation* **18**(3), 1037–1057.
22. Petrauskas, A. and Aagaard, P.M. (1971), Extrapolation of Historical Storm Data for Estimating Design Wave Heights, *Society of Petroleum Engineers Journal* **11**, 23–37.
23. Piyavskii, S.A. (1967), 'An Algorithm for Finding the Absolute Minimum of a Function', *Theory of Optimal Solutions* **2**, Kiev, IK AN USSR (in Russian), pp. 13–24.
24. Piyavskii, S.A. (1972), An Algorithm for Finding the Absolute Extremum of a Function, *USSR Computational Mathematics and Mathematical Physics* **12**, 57–67.
25. Ratschek, H. and Rokne, J. (1984), *Computer Methods for the Range of Functions*, Ellis Horwood Series Mathematics and its Applications, Wiley, New York.
26. Ratschek, H. and Rokne, J. (1988), *New Computer Methods for Global Optimization*, Ellis Horwood, Chichester.
27. Rockette, H., Antle, C. and Klimko, L.A. (1974), Maximum Likelihood Estimation with the Weibull Model, *Journal of the American Statistical Association* **69**(345), 246–249.
28. Wingo, D.R. (1972), Maximum Likelihood Estimation of the Parameters of the Weibull Distribution by Modified Quasilinearization, *IEEE Transactions on Reliability* **21**, 89–93.
29. Wingo, D.R. (1973), Solution of the Three-Parameter Weibull Equations by Constrained Modified Quasilinearization (Progressively Censored Samples), *IEEE Transactions on Reliability* **22**(2), 96–102.
30. Zanakis, S.H. (1977), Computational Experience with Some Nonlinear Optimization Algorithms in Deriving Maximum Likelihood Estimates for the Three-Parameter Weibull Distribution, in *Algorithmic Methods in Probability*, M.F. Neuts (ed.), TIMS Studies in Management Sciences, North-Holland **7**, 63–77.
31. Zanakis, S.H. (1979), Extended Pattern Search with Transformations for the Three-Parameter Weibull MLE Problem, *Management Science* **25**(11), 1149–1161.
32. Zanakis, S.H. and Kyparisis, J. (1986), A Review of Maximum Likelihood Estimation Methods for the Three-Parameter Weibull Distribution, *Journal of Statistical Computation and Simulation* **25**, 53–73.