

# MAXIMUM-LIKELIHOOD ESTIMATION OF THE PARAMETERS OF A FOUR-PARAMETER CLASS OF PROBABILITY DISTRIBUTIONS

by SIEGFRIED H. LEHNIGK

(Received 5th November 1986)

## 1. Introduction

We shall concern ourselves with the class of continuous, four-parameter, one-sided probability distributions which can be characterized by the probability density function (*pdf*) class

$$f(x) = f(x; c, b, p, \beta) = \begin{cases} \frac{\beta}{\Gamma((1-p)\beta^{-1})} b^{-1} \xi^{-p} \exp -\xi^\beta, \xi = (x-c)b^{-1}, x > c, \\ 0, x < c. \end{cases} \quad (1.1)$$

It depends on the four parameters: shift  $c \in \mathbb{R}$ , scale  $b > 0$ , initial shape  $p < 1$ , and terminal shape  $\beta > 0$ . For  $p \leq 0$ , the definition of  $f(x)$  can be completed by setting  $f(c) = \beta/b\Gamma(\beta^{-1}) > 0$  if  $p = 0$ , and  $f(c) = 0$  if  $p < 0$ . For  $0 < p < 1$ ,  $f(x)$  remains undefined at  $x = c$ ;  $f(x) \uparrow +\infty$  as  $x \downarrow c$ .

Apparently, the distribution class (1.1) has been introduced by Amoroso [1]. It is closely related to the class of generalized Feller differential equations, Lehnigk [10], which are of Fokker–Planck type. Through this connection with analysis the distribution class (1.1) can be traced back in some of its special cases to the heat equation ( $p = 0, \beta = 2$ , Fourier, Gauss) and to the Kepinski and Feller equations ( $p < 1, \beta = 1$ ), Lehnigk [9]. The function theoretical properties of the characteristic function associated with the *pdf* class (1.1) have been discussed by Lehnigk [11].

From a purely statistical point of view the distribution class (1.1), which has been designated a generalized Gamma distribution by various authors, has received considerable attention in more recent publications, Stacy [14], Stacy and Mihram [15], Parr and Webster [12], Harter [5], and Essenwanger [3], some of which concentrate on the maximum-likelihood estimation problem. In the present paper we shall pick up this problem once more in an attempt to make this estimation technique more easily accessible to the statistical analyst.

Although serious efforts have been made, (Parr and Webster [12], Harter [5], Essenwanger [3], also Harter and Moore [6] for the special cases of Gamma ( $p < 1, \beta = 1$ ) and Weibull ( $p = 1 - \beta < 1$ ) distributions) to formalize and standardize the maximum-likelihood approach to the population parameter estimation problem

associated with the class (1.1) it seems that its actual implementation has been rather limited. In fact, the distribution class (1.1) has not been used as extensively in everyday statistical practice as it should have been. The main reason for this state of affairs is most likely to be attributed to computational intensity and possibly to convergence problems associated with the solution of the resulting system of transcendental equations as they appear in the literature.

It is our objective to demonstrate that the maximum-likelihood method for the estimation of the parameters  $c$ ,  $b$ ,  $p$ , and  $\beta$  of the distribution class (1.1) can actually be handled in a computationally economical and interactive fashion on a *PC*. As a matter of fact, we shall show in Section 2 that it is only necessary to solve a system of two equations,  $g(\beta, c) = 0$ ,  $h(\beta, c) = 0$ , if the shift parameter  $c$  is considered as unknown, and to solve only one equation,  $g(\beta) = 0$ , if  $c$  is known (Section 4).

If there exists a solution of the system  $g(\beta, c) = 0$ ,  $h(\beta, c) = 0$  ( $c$  unknown), or of the single equation  $g(\beta) = 0$  ( $c$  known) which are estimates  $\hat{\beta}$  (and  $\hat{c}$ ) for the terminal shape parameter  $\beta$  (and for the shift parameter  $c$ ), then estimates  $\hat{b}$  and  $\hat{p}$  for the scale parameter  $b$  and for the initial shape parameter  $p$  can easily be determined from simple auxiliary formulas.

In the present paper we shall restrict ourselves to the discussion of general aspects of the likelihood function approach. In Section 2 we shall derive the appropriate equations. In Section 3 we shall discuss the properties of the functions  $g(\beta, c)$  and  $h(\beta, c)$ . The case that the shift parameter  $c$  is known will briefly be summarized in Section 4. Some major special cases will be mentioned in Section 5. Another paper to be published elsewhere will be devoted to the implementation of the likelihood function technique, to the computational approach to the solution of the equations  $g(\beta, c) = 0$ ,  $h(\beta, c) = 0$ , or simply  $g(\beta) = 0$  if  $c$  is known, and to the presentation of empirical examples.

## 2. The maximum-likelihood equations

Let  $P$  be a parameter vector. Relative to a shift parameter  $c$  contained in  $P$  and located on a coordinate axis of  $x$ , let  $X_v - c$  ( $v = 1, \dots, N$ ) represent i.i.d. random sample values of a random variable  $X$  which is assumed to be distributed according to the density function  $f(x; P)$ . The likelihood function is then defined as

$$L(P) = \prod_{v=1}^N f(X_v - c; P)$$

(see, e.g., Roussas [13, Section 12.5], and Law and Kelton [7, Section 5.4]). For the distributions characterized by the *pdf* class (1.1) this likelihood function takes the particular form

$$L(P) = \left[ \frac{\beta}{b\Gamma((1-p)\beta^{-1})} \right]^N b^{Np} \left( \prod_{v=1}^N (X_v - c) \right)^{-p} \exp - \left( b^{-\beta} \sum_{v=1}^N (X_v - c)^\beta \right) \quad (2.1)$$

with  $P = (c, b, p, \beta)$ .

Let the set of the sample values of the  $X_v$ 's contain  $m \leq N$  ( $m \geq 2$ ) distinct elements  $x_v$ ,  $x_1 < x_2 < \dots < x_m$ . Then  $x_v$  occurs with absolute (integral valued) frequency  $f_{av} \geq 0$ ,  $\sum_{v=1}^m f_{av} = N$ , and

$$\prod_{v=1}^N (X_v - c) = \prod_{v=1}^m (x_v - c)^{f_{av}} = \prod_{v=1}^m \exp(f_{av} \log(x_v - c)),$$

$$\sum_{v=1}^N (X_v - c)^\beta = \sum_{v=1}^m f_{av}(x_v - c)^\beta = \sum_{v=1}^m f_{av} \exp(\beta \log(x_v - c)).$$

(We have used here the general definition  $b^a = \exp(a \log b)$ ,  $\log$  designating the principal value.) Clearly, without loss of generality, we may assume that the absolute frequencies  $f_{a1}$  and  $f_{am}$  of  $x_1$  and  $x_m$ , respectively, are positive. This implies that the shift parameter  $c$  is restricted to  $c < x_1$ . If the shift parameter is known it suffices to assume that, in addition to  $f_{am}$ , there is at least one other positive frequency  $f_{ak}$  with  $1 \leq k < m$ . Setting  $\log(x_v - c) = \rho_v$  we obtain from (2.1)

$$L(P) = \beta^N \Gamma^{-N} ((1-p)\beta^{-1}) b^{-(1-p)N} \left( \prod_{v=1}^m e^{f_{av}\rho_v} \right)^{-p} \exp \left( - \left( b^{-\beta} \sum_{v=1}^m f_{av} e^{\beta\rho_v} \right) \right). \tag{2.2}$$

The objective of the likelihood function approach is to determine those values of the parameters  $c$ ,  $b$ ,  $p$ , and  $\beta$ , subject to the conditions  $c < x_1$ ,  $b > 0$ ,  $p < 1$ ,  $\beta > 0$  (if such a set exists), which maximizes the function  $L(P)$ . It is more convenient in our situation to use the equivalent function  $R(P) = \log L(P)$ . Therefore, taking logarithms in (2.2), we obtain

$$R(P) = N \log \beta - N \log \Gamma((1-p)\beta^{-1}) - (1-p)N \log b - p \sum_{v=1}^m f_{av}\rho_v - b^{-\beta} \sum_{v=1}^m f_{av} e^{\beta\rho_v}. \tag{2.3}$$

It is useful at this point to change from absolute frequencies  $f_{av}$  to relative frequencies  $f_v = N^{-1} f_{av} \geq 0$  with  $\sum_{v=1}^m f_v = 1$ , and  $f_1 > 0$ ,  $f_m > 0$ . We introduce the abbreviations

$$B = B(\beta, c) = N^{-1} \sum_{v=1}^m f_{av} e^{\beta\rho_v} = \sum_{v=1}^m f_v e^{\beta\rho_v}, \tag{2.4a}$$

$$C = C(c) = N^{-1} \sum_{v=1}^m f_{av}\rho_v = \sum_{v=1}^m f_v\rho_v. \tag{2.4b}$$

Then (2.3) can be rewritten in the form

$$N^{-1}R(P) = \log \beta - \log \Gamma((1-p)\beta^{-1}) - (1-p) \log b - pC - b^{-\beta}B. \tag{2.5}$$

A necessary condition for  $R(P)$  to take a maximum in the interior of the open parameter domain  $\mathcal{P}: c < x_1$ ,  $b > 0$ ,  $p < 1$ ,  $\beta > 0$ , is that the four simultaneous equations  $\partial R/\partial c = 0$ ,  $\partial R/\partial b = 0$ ,  $\partial R/\partial p = 0$ , and  $\partial R/\partial \beta = 0$  be satisfied. Under the assumption that

all four parameters are unknown these equations are of the form

$$pE + \beta b^{-\beta} F = 0, \tag{2.6}$$

$$-(1-p)b^{-1} + \beta b^{-\beta-1} B = 0, \tag{2.7}$$

$$\beta^{-1} \psi((1-p)\beta^{-1}) + \log b - C = 0, \tag{2.8}$$

$$\beta^{-1} + (1-p)\beta^{-2} \psi((1-p)\beta^{-1}) + b^{-\beta} B \log b - b^{-\beta} D = 0, \tag{2.9}$$

with

$$D = D(\beta, c) = \sum_{v=1}^m f_v \rho_v e^{\beta \rho_v}, E = E(c) = \sum_{v=1}^m f_v e^{-\rho_v}, F = F(\beta, c) = \sum_{v=1}^m f_v e^{(\beta-1)\rho_v}, \tag{2.10}$$

and  $\psi(y)$  being the psi function,  $\psi(y) = d \log \Gamma(y) / dy$  (see, for example, Gradshteyn and Ryzhik [4, formula 8.360]).

Before we proceed we observe the following facts concerning the coefficients  $B, C, D, E,$  and  $F$  in the equations (2.6), ..., (2.9), which will be of importance later.  $B(\beta, c)$  is a positive function of  $\beta \in (0, +\infty)$  and  $c \in (-\infty, x_1)$ ;  $B(1, c) = \sum_{v=1}^m f_v (x_v - c)$  is the sample mean relative to the point  $x = c$ ; and, for every  $c < x_1, B(\beta, c) \rightarrow \sum_{v=1}^m f_v = 1$  as  $\beta \downarrow 0$ . Since  $\rho_v = \log(x_v - c) < \rho_m = \log(x_m - c) (v = 1, \dots, m-1)$  it follows that

$$C = \sum_{v=1}^m f_v \rho_v < \rho_m \sum_{v=1}^m f_v = \rho_m. \tag{2.11}$$

Furthermore,

$$D(\beta, c) = \frac{\partial B(\beta, c)}{\partial \beta}, D(\beta, c) \rightarrow C(c) \text{ as } \beta \downarrow 0, E(c) = -\frac{dC(c)}{dc} > 0,$$

$$F(\beta, c) = -\beta^{-1} \frac{\partial B(\beta, c)}{\partial c} > 0, F(1, c) = \sum_{v=1}^m f_v = 1, F(\beta, c) \rightarrow E(c) \text{ as } \beta \downarrow 0.$$

We now return to the equations (2.6), ..., (2.9). Our first essential observation is that equation (2.6) can be satisfied only if  $p < 0$ . In other words, if all four parameters are considered unknown the distribution *pdf* must be of hump-shaped type ( $p < 0$ ), i.e.,  $f(x, P) \downarrow 0, x \downarrow c$ . Whether or not this is the case must be ascertained from the given data.

If there is clear indication that the *pdf* should not go to zero as  $x \downarrow c$ , i.e., if  $f(x, P)$  has to approach either a finite limit or  $+\infty$  as  $x \downarrow c$ , then the initial shape parameter  $p$  must be in the interval  $0 \leq p < 1$ . (In this situation the distribution is either of purely exponential type ( $p = 0$ ) or of  $J$ -shaped type ( $0 < p < 1$ .) If  $0 \leq p < 1$ , the left-hand side of equation (2.6) is positive which means that the function  $R(P)$  does not take a maximum in the interior of the open parameter domain  $\mathcal{P}: c < x_1, b > 0, \beta > 0$  (if  $p = 0$ ) or  $\mathcal{P}: c < x_1, b > 0, \beta > 0, 0 < p < 1$ , otherwise. Let  $x_0$  be the infimum of the set of all possible observations  $x_1$ . Then (2.5) shows that

$$\sup_{P \in \mathcal{P}} N^{-1}R(P) = \sup_{c \leq x_0} N^{-1}R(P) = \log \beta - \log \Gamma((1-p)\beta^{-1}) - (1-p) \log b$$

$$-p \inf_{c \leq x_0} C(c) - b^{-\beta} \inf_{c \leq x_0} B(\beta, c)$$

with  $\inf C(c) = C(x_0)$ ,  $\inf B(\beta, c) = B(\beta, x_0)$ . Therefore, if  $0 \leq p < 1$ , the maximum-likelihood estimate  $\hat{c}$  for the shift parameter  $c$  is given by  $\hat{c} = x_0$ . Then, with  $\hat{c}$  known (as in any case of a known shift parameter), one can return to (2.5) in which now only  $b$ ,  $\beta$ , and  $p$  (if  $p \neq 0$ ) are left as parameters. Consequently, the equation  $\partial R / \partial c = 0$  becomes superfluous, i.e., (2.6) becomes vacuous, and one is left with the equations (2.7), (2.8), and (2.9). The special case of a known shift parameter will be discussed in Section 4.

Our next observation is that by means of equation (2.7) the scale parameter  $b$  can be eliminated from the other three. Indeed, equation (2.7) immediately leads to

$$b^\beta = (1-p)^{-1} \beta B. \tag{2.12}$$

Then, upon substitution, we obtain from (2.6)

$$pE + (1-p)B^{-1}F = 0, \tag{2.13}$$

and from (2.8) and (2.9), after simplifying multiplications by appropriate factors,

$$\psi((1-p)\beta^{-1}) + \log [(1-p)^{-1} \beta B] - \beta C = 0, \tag{2.14}$$

$$\psi((1-p)\beta^{-1}) + \log [(1-p)^{-1} \beta B] + (1-p)^{-1} \beta - \beta B^{-1}D = 0. \tag{2.15}$$

Finally, we observe the obvious but essential fact that the initial shape parameter  $p$  can be expressed in terms of the shift and terminal shape parameters  $c$  and  $\beta$ . Subtraction of (2.14) from (2.15) leads to

$$(1-p)^{-1} \beta = AB^{-1} \tag{2.16}$$

with

$$A = A(\beta, c) = \beta(D - BC) = \beta \sum_{v=1}^m f_v(\rho_v - C) e^{\beta \rho_v}, \tag{2.17}$$

$B$ ,  $C$ , and  $D$  as given in (2.4) and (2.10), respectively. If  $A \neq 0$ , (2.16) can also be written in the forms

$$1-p = \beta A^{-1}B \quad \text{or} \quad p = 1 - \beta A^{-1}B. \tag{2.18}$$

Using these identities in (2.13) we obtain

$$(1 - \beta A^{-1}B)E + \beta A^{-1}F = 0$$

or, after multiplying by  $\beta^{-1}A$ ,

$$h(\beta, c) = (\beta^{-1}A - B)E + F = 0. \tag{2.19a}$$

If we observe (2.17) we can also write

$$h(\beta, c) = (D - BC - B)E + F = 0. \tag{2.19b}$$

Next, using (2.16) in (2.14), we arrive at a second equation,

$$g(\beta, c) = \psi(A^{-1}B) + \log A - \beta C = 0. \tag{2.20}$$

If the simultaneous equations (2.19) and (2.20) have a solution  $(\hat{\beta}, \hat{c})$  which is an estimate of the parameters  $\beta$  and  $c$  we return to the formula (2.18) to obtain an estimate  $\hat{p}$  for the initial shape parameter  $p$ ,

$$\hat{p} = 1 - \hat{\beta}A^{-1}(\hat{\beta}, \hat{c})B(\hat{\beta}, \hat{c}) = 1 - B(\hat{\beta}, \hat{c})[D(\hat{\beta}, \hat{c}) - B(\hat{\beta}, \hat{c})C(\hat{c})]^{-1}, \tag{2.21}$$

provided  $A(\hat{\beta}, \hat{c}) \neq 0$ . We note that  $\hat{p}$  is uniquely determined for every fixed vector  $(\hat{\beta}, \hat{c})$ . Finally, with  $\hat{\beta}, \hat{c}$ , and  $\hat{p}$  specified, we obtain a unique estimate  $\hat{b}$  for the scale parameter  $b$  from (2.12),

$$\hat{b} = \exp \{ \hat{\beta}^{-1} \log [\hat{\beta}B(\hat{\beta}, \hat{c})(1 - \hat{p})^{-1}] \}, \tag{2.22a}$$

which, by means of (2.18), can also be written in the form

$$\hat{b} = \exp \{ \hat{\beta}^{-1} \log A(\hat{\beta}, \hat{c}) \}. \tag{2.22b}$$

### 3. Properties of the functions $g(\beta, c)$ and $h(\beta, c)$

We have seen in the previous section that the derivation of the equations  $g(\beta, c) = 0$  and  $h(\beta, c) = 0$  given in (2.19) and (2.20) requires the assumption that  $A = A(\beta, c) \neq 0$ . Therefore, let us start our discussion by showing that  $A$  given in (2.17) is always positive for  $\beta > 0$  and  $c < x_1$ .

Setting  $\rho_v - C = \alpha_v$  we can express  $D - BC$  as

$$D - BC = e^{c\beta} \sum_{v=1}^m f_v \alpha_v e^{\alpha_v \beta} \tag{3.1}$$

(see (2.4) and (2.10)). As  $\beta \downarrow 0$  this shows that, for any fixed  $c < x_1$ ,

$$D - BC \rightarrow \sum_{v=1}^m f_v \alpha_v = \sum_{v=1}^m f_v \rho_v - C \sum_{v=1}^m f_v = 0, \beta \downarrow 0, \tag{3.2}$$

since  $\sum_{v=1}^m f_v = 1$ . We consider now the derivative

$$\frac{\partial}{\partial \beta} (D - BC) = C e^{c\beta} \sum_{v=1}^m f_v \alpha_v e^{\alpha_v \beta} + e^{c\beta} \sum_{v=1}^m f_v \alpha_v^2 e^{\alpha_v \beta}.$$

As  $\beta \downarrow 0$ , the first term approaches  $C \sum_{v=1}^m f_v \alpha_v = 0$  according to (3.2), and the second term approaches  $\sum_{v=1}^m f_v \alpha_v^2$ . Since all  $x_v$  are distinct it follows that all  $\rho_v = \log(x_v - c)$  are distinct ( $v = 1, \dots, m$ ). Therefore, at most one of the numbers  $\alpha_v = \rho_v - C$  can be zero. Furthermore, since, by assumption, the elements  $f_1$  and  $f_m$  of the set  $\{f_v\}$  are different from zero, at least one of the products  $f_1 \alpha_1^2$  and  $f_m \alpha_m^2$  is different from zero. This implies  $\sum_{v=1}^m f_v \alpha_v^2 > 0$ . Consequently, we arrive at

$$\frac{\partial}{\partial \beta}(D - BC) \rightarrow \sum_{v=1}^m f_v \alpha_v^2 > 0, \beta \downarrow 0 \quad (\alpha_v = \rho_v - C). \tag{3.3}$$

This together with the limit relation (3.2) implies that the function  $D(\beta, c) - B(\beta, c)C(c) > 0$  for any fixed  $c < x_1$  at least for sufficiently small values of  $\beta > 0$ . Therefore, according to (2.17),  $A(\beta, c) > 0$  for any  $c < x_1$  and  $\beta > 0$  sufficiently small.

We can now establish the fact that  $A(\beta, c) > 0$  for any  $c < x_1$  and for every  $\beta > 0$ . According to (2.11) we have  $\rho_m - C > 0$ . Therefore, there exists an integer  $n, 1 \leq n \leq m$ , such that  $\alpha_v = \rho_v - C > 0$  for  $v \geq n$  and  $\alpha_v \leq 0$  for  $v < n$ . If  $n \geq 2$ , at least one of these  $\alpha_v$ 's is strictly negative. We break up the sum at the right-hand side of (3.1) as follows,

$$D - BC = e^{C\beta} \left[ \sum_{v=1}^{n-1} f_v \alpha_v e^{\alpha_v \beta} + \sum_{v=n}^m f_v \alpha_v e^{\alpha_v \beta} \right].$$

The second sum is a positive strictly monotonically increasing function of  $\beta$ . In the terms of the first sum the factors  $\alpha_v e^{\alpha_v \beta} \leq 0$  are certainly not decreasing. They are strictly monotonically increasing if  $\alpha_v < 0$ . Therefore, the first sum is not decreasing. Consequently, since  $D - BC > 0$  for sufficiently small values of  $\beta > 0$ , it is always positive. This implies that  $A(\beta, c) = \beta [D(\beta, c) - B(\beta, c)C(c)] > 0$  for any fixed  $c < x_1$  and for every  $\beta > 0$ . As  $\beta \downarrow 0, A(\beta, c) \downarrow 0$  according to (3.2).

We can now discuss the properties of the functions  $g(\beta, c)$  and  $h(\beta, c)$ . We shall start with the function

$$g(\beta, c) = \psi(A^{-1}(\beta, c)B(\beta, c)) + \log A(\beta, c) - \beta C(c). \tag{3.4}$$

It is the most important one in the sense that it is the only one to be considered in the special case that the shift parameter  $c$  is known (Section 4).

According to its definition (2.4a) we have  $B(\beta, c) > 0$  for any  $c < x_1$  and for every  $\beta > 0$ . Furthermore, we have seen that  $B(\beta, c) \rightarrow 1$  as  $\beta \downarrow 0$ . Since, as we just saw,  $A(\beta, c) > 0$  and  $A(\beta, c) \downarrow 0$  as  $\beta \downarrow 0$ , it follows that the argument  $A^{-1}B$  of the psi function in (3.4) is positive and approaches  $+\infty$  as  $\beta \downarrow 0$ . For large values of its argument the psi function has the asymptotic representation  $\psi(y) \sim \log y, y \uparrow + \infty$  (see, e.g., Gradshteyn and Ryzhik [4, formula 8.362.2]). Therefore  $\psi(A^{-1}B) \sim \log B - \log A$ . Consequently, according to (3.4),  $g(\beta, c) \sim \log B(\beta, c) - \beta C(c)$  as  $\beta \downarrow 0$  for any fixed  $c < x_1$ , i.e.,  $g(\beta, c) \rightarrow 0$  as  $\beta \downarrow 0$ .

Next we look at the derivative  $\partial g / \partial \beta$ ,

$$g' = A^{-2} [AA' + (AB' - A'B)\psi'(A^{-1}B) - A^2C] \tag{3.5}$$

where the prime signifies differentiation with respect to  $\beta$ . From (2.7) we obtain

$$A' = (D - BC) + \beta(D' - DC)$$

so that

$$AA' = \beta(D - BC)^2 + \beta^2(D - BC)(D' - DC)$$

since  $B' = D$  as indicated right after (2.11). Furthermore,

$$AB' = \beta(D - BC)D,$$

$$A'B = (D - BC)B + \beta(D' - DC)B.$$

The function  $\psi(y)$  is strictly monotonically increasing. Therefore, its derivative has the asymptotic representation  $\psi'(y) \sim d \log y / dy = y^{-1}$  as  $y \uparrow +\infty$ , i.e.,  $\psi'(A^{-1}B) \sim AB^{-1} = \beta(D - BC)B^{-1}$  as  $\beta \downarrow 0$ . Substituting this asymptotic relation and the expressions for  $AA'$ ,  $AB'$ , and  $A'B$  into the right-hand side of (3.5) we arrive, after appropriate cancellations, at

$$g'(\beta, c) \sim A^{-2}[\beta^2(D - BC)^2DB^{-1} - A^2C] = (D - BC)B^{-1}, \beta \downarrow 0.$$

This shows that, for any  $c < x_1$ ,  $g'(\beta, c) > 0$  if  $\beta > 0$  is sufficiently small and that, according to (3.2) and since  $B \rightarrow 1$ ,  $g'(\beta, c) \downarrow 0$  as  $\beta \downarrow 0$ . Since  $g(\beta, c) \rightarrow 0$  as  $\beta \downarrow 0$ , it follows that  $g(\beta, c) > 0$  for  $\beta > 0$  sufficiently small and for any  $c < x_1$ .

Finally, we investigate the behaviour of  $g(\beta, c)$  as  $\beta \uparrow +\infty$ . For this purpose we want to express  $A(\beta, c)$  and  $A^{-1}(\beta, c)B(\beta, c)$  in more convenient forms. Using the definitions of  $A$  and  $B$  as given in (2.17) and (2.4a), respectively, together with the definitions of  $C(c)$  and  $D(\beta, c)$  in (2.4b) and (2.10), we can write

$$A(\beta, c) = \beta e^{\beta \rho_m} \sum_{v=1}^m f_v(\rho_v - C) e^{-(\rho_m - \rho_v)\beta},$$

$$B(\beta, c) = e^{\beta \rho_m} \sum_{v=1}^m f_v e^{-(\rho_m - \rho_v)\beta},$$

$$A^{-1}(\beta, c)B(\beta, c) = \frac{\sum_{v=1}^m f_v e^{-(\rho_m - \rho_v)\beta}}{\beta \sum_{v=1}^m f_v(\rho_v - C) e^{-(\rho_m - \rho_v)\beta}}.$$

Since  $\rho_m - \rho_v > 0$  for  $v = 1, \dots, m - 1$ , the identities for  $A$  and  $A^{-1}B$  provide us with the asymptotic representations

$$A(\beta, c) \sim \beta e^{\beta \rho_m} f_m(\rho_m - C), \beta \uparrow +\infty,$$

$$A^{-1}(\beta, c)B(\beta, c) \sim \beta^{-1}(\rho_m - C)^{-1}, \beta \uparrow +\infty.$$

From the first one we deduce

$$\log A(\beta, c) \sim \log \beta + \rho_m \beta + \log f_m(\rho_m - C).$$

The second one shows that, as  $\beta \uparrow + \infty$ , the argument of  $\psi(A^{-1}B)$  approaches zero. Since  $\psi(y) \sim -\gamma - y^{-1}, y \downarrow 0, \gamma = -\psi(1)$  being Euler's constant (see e.g., Gradshteyn and Ryzhik [4, formulas 8.362.1 and 8.367.1]), we see that  $\psi(A^{-1}B) \sim -\gamma - (\rho_m - C)\beta, \beta \uparrow + \infty$ . Consequently, according to (3.4),

$$g(\beta, c) \sim \log \beta - \gamma + \log f_m(\rho_m - C), \beta \uparrow + \infty,$$

which implies the important fact that  $g(\beta, c) > 0$  if  $\beta > 0$  is sufficiently large. The ultimate and initial positivity of  $g(\beta, c)$  means that, for any fixed  $c < x_1$ , the equation  $g(\beta, c) = 0$  has an even number of positive roots.

If  $(\beta^*, c^*)$  is a point such that  $g(\beta^*, c^*) = 0$  and if the conditions of the implicit function theorem (see, e.g., Courant [2, vol. 1, p. 482]) are satisfied at  $(\beta^*, c^*)$ , then the equation  $g(\beta, c) = 0$  defines implicitly a single-valued, differentiable function  $\beta = y(c)$  in some neighbourhood of  $\beta^*$  such that  $y(c^*) = \beta^*$ .

We turn now to the discussion of the function (2.19),

$$h(\beta, c) = [D(\beta, c) - B(\beta, c)C(c) - B(\beta, c)]E(c) + F(\beta, c). \tag{3.6}$$

It is easy to see that, for fixed  $c < x_1, h(\beta, c) \rightarrow 0$  as  $\beta \downarrow 0$  as a consequence of the limit relations  $D - BC \rightarrow 0, B \rightarrow 1,$  and  $F \rightarrow E$  (see (3.2) and the remarks made on  $B$  and  $F$  in Section 2).

For the derivative  $\partial h / \partial \beta$  we have

$$h'(\beta, c) = [(D - BC)' - B']E + F',$$

the primes signifying differentiation with respect to  $\beta$ . According to the limit relations established earlier,  $(D - BC)' \rightarrow \sum_{v=1}^m f_v(\rho_v - C)^2, B' = D \rightarrow C = \sum_{v=1}^m f_v \rho_v$ . Furthermore  $F' = \sum_{v=1}^m f_v \rho_v e^{(\beta-1)\rho_v} \rightarrow \sum_{v=1}^m f_v \rho_v e^{-\rho_v}$  as  $\beta \downarrow 0$ . Therefore, with  $E = \sum_{v=1}^m f_v e^{-\rho_v}$  we obtain

$$h'(\beta, c) \rightarrow \left[ \sum_{v=1}^m f_v(\rho_v - C)^2 - \sum_{v=1}^m f_v \rho_v \right] \sum_{v=1}^m f_v e^{-\rho_v} + \sum_{v=1}^m f_v \rho_v e^{-\rho_v}, \beta \downarrow 0. \tag{3.7}$$

This expression for the limit can also be written in the form

$$\sum_{v=1}^m f_v e^{-\rho_v} (K + \rho_v) \tag{3.8}$$

with

$$K + \rho_v = \sum_{\mu=1}^m f_\mu \rho_\mu^2 - \left( \sum_{\mu=1}^m f_\mu \rho_\mu \right)^2 - \sum_{\mu=1}^m f_\mu \rho_\mu + \rho_v \quad (v = 1, \dots, m). \tag{3.9}$$

To evaluate the limit (3.8) it is necessary to take the shift parameter  $c$  into account. If we let  $c \uparrow x_1$ , then  $\rho_v = \log(x_v - c) \downarrow \log(x_v - x_1) > -\infty (v = 2, \dots, m)$ , and  $\rho_1 = \log(x_1 - c) \downarrow -\infty$ . We now investigate the factors (3.9) which appear in (3.8). Since  $\rho_1 < \rho_v (v = 2, \dots, m)$  it is sufficient to look only at  $K + \rho_1$ . All terms in (3.9) which do not depend on  $\rho_1$  approach finite constant values as  $c \uparrow x_1$ . Therefore, we have

$$\begin{aligned}
 K + \rho_1 &= f_1 \rho_1^2 - f_1^2 \rho_1^2 - (\text{bounded terms}) \rho_1 - (\text{bounded terms}) \\
 &= f_1(1 - f_1) \rho_1^2 - (\text{bounded terms}) \rho_1 - (\text{bounded terms}).
 \end{aligned}$$

This shows that  $K + \rho_v > K + \rho_1 \uparrow + \infty$  as  $c \uparrow x_1 (v=2, \dots, m)$ . Thus, all terms in (3.8) are positive if  $x_1 - c > 0$  is sufficiently small. Consequently, according to (3.7),  $h'(\beta, c)$  approaches a positive limit as  $\beta \downarrow 0$  if  $x_1 - c > 0$  is sufficiently small. This together with the limit relation  $h(\beta, c) \rightarrow 0$  as  $\beta \downarrow 0$  implies that  $h(\beta, c) > 0$  for sufficiently small  $\beta > 0$  and sufficiently small  $x_1 - c > 0$ .

If we let  $c \downarrow -\infty$ , then  $\rho_v = \log(x_v - c) \uparrow + \infty$  and  $\rho_v / \rho_m \uparrow 1 (v=1, \dots, m)$ . Therefore,

$$K + \rho_v = \rho_m^2 \sum_{\mu=1}^m f_{\mu} \left( \frac{\rho_{\mu}}{\rho_m} \right)^2 - \rho_m^2 \left( \sum_{\mu=1}^m f_{\mu} \frac{\rho_{\mu}}{\rho_m} \right)^2 - \rho_m \sum_{\mu=1}^m f_{\mu} \frac{\rho_{\mu}}{\rho_m} + \rho_m \frac{\rho_v}{\rho_m} \rightarrow 0$$

( $v=1, \dots, m$ ) as  $c \downarrow -\infty$ . Consequently, according to (3.8) and (3.7),  $h' \rightarrow 0$  as  $\beta \downarrow 0$ . Unfortunately, because of the complexity of the terms which appear in (3.7), nothing can be said in general about the limit in (3.7) for  $c$  away from  $-\infty$  and not close to  $x_1$ . We may add, however, that, from a phenomenological point of view, the difference  $x_1 - c$  cannot be arbitrarily large for the class of distributions characterized by the pdf class (1.1).

We investigate next the behaviour of  $h(\beta, c)$  as  $\beta \uparrow + \infty$ . Using the definitions of the functions  $B, D$ , and  $F$  as given in (2.4a) and (2.10) in (3.6) we can write

$$h(\beta, c) = e^{\beta \rho_m} \left[ \sum_{v=1}^m f_v (\rho_v - C - 1) e^{-(\rho_m - \rho_v)\beta} \right] E + e^{\beta \rho_m} \sum_{v=1}^m f_v e^{-(\rho_m - \rho_v)\beta} e^{-\rho_v}.$$

Therefore, we obtain the asymptotic representation

$$h(\beta, c) \sim e^{\beta \rho_m} f_m [(\rho_m - C)E - E + e^{-\rho_m}] = M, \beta \uparrow + \infty.$$

Since  $E = \sum_{v=1}^m f_v e^{-\rho_v} > f_m e^{-\rho_m} > 0$ , we have  $-E < -f_m e^{-\rho_m}$ . Therefore,  $-E + e^{-\rho_m} < -f_m e^{-\rho_m} + e^{-\rho_m} = e^{-\rho_m}(1 - f_m) > 0$ . Furthermore, since  $\rho_m - C > 0$ , it follows that the asymptotic comparison function  $M$  of  $h$  is positive. In other words,  $h(\beta, c) > 0$  for  $c < x_1$  if  $\beta$  is sufficiently large. This result together with the fact that  $h(\beta, c) > 0$  for  $x_1 - c > 0$  sufficiently small if  $\beta > 0$  is sufficiently small leads us to conclude that, for fixed  $c$  such that  $x_1 - c > 0$  is sufficiently small, the equation  $h(\beta, c) = 0$  has an *even number of positive roots*.

If the shift parameter  $c$  is away from  $-\infty$  and not close to  $x_1$  we cannot say anything about the number of roots of the equation  $h(\beta, c) = 0$  except that this number is *odd* if  $h(\beta, c) < 0$  for small values of  $\beta > 0$ . In any case, if  $(\beta^*, c^*)$  is a point such that  $h(\beta^*, c^*) = 0$  and if the conditions of the implicit function theorem are satisfied at  $(\beta^*, c^*)$ , then  $h(\beta, c) = 0$  defines implicitly a single-valued, differentiable function  $\beta = z(c)$  in some neighbourhood of  $\beta^*$  such that  $z(c^*) = \beta^*$ .

As a result of these considerations we may finally conclude that, if  $p < 0$  and if there exists a number  $\hat{c} < x_1$  such that  $z(\hat{c}) = y(\hat{c}) = \hat{\beta}$  ( $y(c)$  being the function implicitly defined

by  $g(\beta, c)=0$ , the numbers  $\hat{\beta}$  and  $\hat{c}$  are estimates for the terminal shape and shift parameters  $\beta$  and  $c$ , respectively. The estimates  $\hat{p}$  and  $\hat{b}$  for the initial shape and scale parameters  $p$  and  $b$  can then quickly be calculated from the formulas (2.21) and (2.22). It is important to observe, however, that the estimate vector  $\hat{P}=(\hat{c}, \hat{b}, \hat{p}, \hat{\beta})$  need not be unique so that, if there is more than one such vector, it may be necessary to verify which one actually maximizes the logarithmic likelihood function  $R(P)$  given in (2.5).

It is clear from the structure of the functions  $g(\beta, c)$  and  $h(\beta, c)$  that it is not easy to ascertain whether or not the equations  $g=0, h=0$ , can have more than one solution. However, it seems that, if different solutions  $(\hat{\beta}_v, \hat{c}_v)$  exist, they can be ordered such that  $\hat{\beta}_v < \hat{\beta}_{v+1}, \hat{c}_v < \hat{c}_{v+1}$ , and that the pair  $(\hat{\beta}_1, \hat{c}_1)$  maximizes  $R(P)$  together with the corresponding values  $\hat{p}$  and  $\hat{b}$  obtained from (2.21) and (2.22). This topic will be discussed elsewhere.

**4. The case of a known shift parameter**

Frequently, the shift parameter of a distribution can be determined from physical or equivalent evidence. It is also known if  $0 \leq p < 1$  as mentioned earlier. Without loss of generality, we may assume in such a situation that  $c=0$ . In these circumstances the logarithmic likelihood function (2.5) reduces to a function of the (in general) three unknown parameters  $b, p$ , and  $\beta$ . Consequently, equation (2.6), i.e.,  $\partial R/\partial c=0$ , becomes vacuous so that we are left with equations (2.7), (2.8), and (2.9) or (2.12), (2.14), and (2.15), respectively. In different notations they appear, for example, in Parr and Webster [12] and Essenwanger [3, p. 117]. The same reasoning as applied in Section 2 now leads us to the equation

$$g(\beta) = \psi(A^{-1}B) + \log A - \beta C = 0.$$

It is the only equation that must actually be solved. It agrees with (2.20) for  $c=0$ , the coefficients  $A$  and  $B$  now being functions of  $\beta$  only,  $A=A(\beta), B=B(\beta)$ . They are determined in (2.17) and (2.4a), respectively, for  $c=0$ , i.e., with  $\rho_v = \log x_v$ . The coefficient  $C$  in  $g(\beta)$  is now a constant.

If  $\hat{\beta}$  is an estimate of the terminal shape parameter  $\beta$  (i.e., a solution of  $g(\beta)=0$ , in which case there are at least two), the corresponding estimates  $\hat{p}$  and  $\hat{b}$  for the initial shape and scale parameters are then determined by the formulas

$$\hat{p} = 1 - \hat{\beta} A^{-1}(\hat{\beta}) B(\hat{\beta})$$

and

$$\hat{b} = \exp \{ \hat{\beta}^{-1} \log [ \hat{\beta} B(\hat{\beta}) (1 - \hat{p})^{-1} ] \}$$

which correspond to (2.21) and (2.22). It is our experience that the smallest positive root  $\hat{\beta}$  of  $g(\beta)=0$  provides the maximum for  $R(P)$ .

**5. Special cases**

For particular values of the two shape parameters the class of probability distributions characterized by the *pdf* class (1.1) contains a number of special cases well-known in statistics and statistical physics. The major ones are: (one-sided) Gauss

( $p=0, \beta=2$ ), Weibull ( $p=1-\beta < 1$ ), exponential ( $p=1-\beta=0$ ), Rayleigh ( $p=1-\beta=-1$ ), Gamma ( $p < 1, \beta=1$ ), chi-square ( $p=(2-\nu)/2 < 1, \beta=1$ ), Maxwell ( $p=-2, \beta=2; x=vt_0, b=(2kT/m)^{1/2}t_0$ ), and Wien ( $p=-3, \beta=1; x=2\pi c\omega_0^{-2}\omega, b=2\pi c\omega_0^{-2}kT\hbar^{-1}$ ). These special cases of (1.1) have been recognized earlier, for example in Stacy and Mihram [15], Lehnigk [8], and Lehnigk [9].

The frequent practice of hypothesizing a particular distribution from among the class (1.1) is equivalent to assigning particular values to at least one of the two shape parameters or to establishing a particular relationship between them with the objective to single out one of the above mentioned special cases in an attempt to reduce the number of unknown parameters. Our experience has shown, however, that the equations  $g(\beta, c)=0$  and  $h(\beta, c)=0$  for the four-parameter-unknown case, or just the equation  $g(\beta)=0$  for the three-parameter-unknown case, can easily be solved so that there is really no need to hypothesize a special distribution. (As mentioned in the introduction this claim will be substantiated in a separate publication.)

The reduced maximum-likelihood equations for the special cases are well known. The reader is referred to the literature, e.g., Law and Kelton [7, Chap. 5.2.2], and Essenwanger [3, Chap. 2.11.2] for the three-parameter Weibull distribution.

For applications in the area of statistical physics it may be of interest to consider the details for the Maxwell distribution for which  $p=-2, \beta=2, b>0, c < x_1$ . (2.5) becomes

$$N^{-1}R = \log 2 - \log \Gamma(3/2) - 3 \log b + 2C - b^{-2}B.$$

Therefore,

$$\partial R / \partial c = -2E + 2b^{-2}F = 0,$$

$$\partial R / \partial b = -3b^{-1} + 2b^{-3}B = 0.$$

The resulting equations are then  $b^2=2B/3$  and  $2BE-3F=0$ ,  $B=B(2, c)$ ,  $E=E(c)$ ,  $F=F(2, c)$ . If  $c=0$  we are left with  $b^2=2B(2, 0)/3$ .

#### REFERENCES

1. L. AMOROSO, Ricerche intorno alla curva dei redditi, *Ann. Mat. Pura. Appl.* (4) 2 (1924), 123–157.
2. R. COURANT, *Differential and Integral Calculus* (2nd ed., vol. 1, Interscience Publishers, New York, 1937).
3. O. M. ESSENWANGER, *Applied Statistics in Atmospheric Sciences* (Elsevier, Amsterdam, 1976).
4. I. S. GRADSHTEYN and I. M. RYZHIK, *Tables of Integrals, Series, and Products* (4th ed., Academic Press, New York, 1965).
5. H. L. HARTEK, Maximum-likelihood estimation of the parameters of a four-parameter generalized Gamma population from complete and censored samples, *Technometrics* 9 (1967), 159–165.
6. H. L. HARTEK and A. H. MOORE, Maximum-likelihood estimation of Gamma and Weibull populations from complete and censored samples, *Technometrics* 7 (1965), 639–643.

7. A. M. LAW and W. D. KELTON, *Simulation Modeling and Analysis* (McGraw-Hill, New York, 1982).
8. S. H. LEHNIGK, Maxwell and Wien processes as special cases of the generalized Feller diffusion process, *J. Math. Phys.* **18** (1977), 104–105.
9. S. H. LEHNIGK, Initial condition solutions of the generalized Feller equations, *J. Appl. Math. Phys. (ZAMP)* **29** (1978), 273–294.
10. S. H. LEHNIGK, On a class of probability distributions, *Math. Methods Appl. Sci.* **9** (1987), 210–219.
11. S. H. LEHNIGK, Characteristic functions of a class of probability distributions, *Complex Variables Theory Appl.* (1987), to appear.
12. B. VAN PARR and J. T. WEBSTER, A method for discriminating between failure density functions used in reliability predictions, *Technometrics* **7** (1965), 1–10.
13. G. G. ROUSSAS, *A First Course in Mathematical Statistics* (Addison-Wesley, Reading, 1973).
14. E. W. STACY, A generalization of the Gamma distribution, *Ann. Math. Stat.* **33** (1962), 1187–1192.
15. E. W. STACY and G. A. MIHRAM, Parameter estimation for a generalized Gamma distribution, *Technometrics* **7** (1965), 349–358.

MICOM, AMSMI-RD-RE-OP  
REDSTONE ARSENAL  
AL 35898-5248  
USA