THE INSTITUTE
OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

LIKELIHOOD AND MODIFIED LIKELIHOOD ESTIMATION
FOR DISTRIBUTIONS WITH UNKNOWN ENDPOINTS

by

Richard L. Smith

September 1994

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

MIM EO SERIES #2327	Richard L. Smith LIKELIHOOD AND MODIFIED LIKELIHOOD ESTIMATION FOR DISTRIBUTION WITH UNKNOWN ENDPOINTS.	
NAME		DATE

LIKELIHOOD AND MODIFIED LIKELIHOOD ESTIMATION FOR DISTRIBUTIONS WITH UNKNOWN ENDPOINTS

by

Richard L. Smith

University of North Carolina

1. Introduction

A recurrent theme in reliability research is the analysis of distributions with unknown endpoints. Specific examples are the three-parameter Weibull, lognormal, gamma and inverse Gaussian distributions. Each of these distributions has the property of being concentrated on an interval (γ, ∞), where γ, typically called an endpoint or threshold parameter, is unknown. In most cases this is non-negative and is interpreted as the minimum possible lifetime of the system, a parameter of obvious importance in reliability analysis. In such families it has been known for a long time that there are difficulties with maximum likelihood estimation.

Clifford Cohen was one of the first to consider maximum likelihood estimation (MLE) for distributions of this type, and over the years has made many contributions. However, he has also been one of the pioneers in considering alternative methods. He has developed a number of variants of a method which he called modified maximum likelihood estimation (MMLE), and in recent years has also worked on modified moments estimators (MME). The main motivation for developing these alternative methods has apparently been ease of numerical implementation. However, it has also been suggested in a number of simulation studies that they may have superior statistical properties as well, in the sense of having smaller bias or variance in small or moderate-sized samples. In this paper I want to explore these aspects further, with particular attention to a theoretical comparison of the MLE and MMLE techniques.

The paper is organised as follows. In Section 2, I review the fields of MLE and MMLE for distributions with unknown endpoints, with particular attention to Cohen's own work on these techniques. Then in Section 3, I review some known results on asymptotics for MLE, and develop some new approximations for MMLE. Finally, Section 4 presents some numerical and simulation results.

1

2. Estimation of distributions with unknown endpoints

Historically, the first of these distributions to be studied was the three-parameter lognormal distribution. Cohen's (1951) paper was one of the first papers to consider maximum likelihood estimation for this family (though not the first: the paper by Wilson and Worcester (1945) preceded him). Starting with the density

$$f(x; \gamma, \delta, \beta) = \frac{1}{\sqrt{2\pi}\delta(x - \gamma)} \exp\left[-\frac{\log^2\{(x - \gamma)/\beta\}}{2\delta^2} \right], \quad x > \gamma, \tag{2.1}$$

he developed the maximum likelihood estimating equations and proposed a graphical method of numerical solution. He also calculated the theoretical Fisher information matrix for the unknown parameter vector (γ, δ, β).

However, in this paper he also introduced modified maximum likelihood. In this, the likelihood equation for γ was replaced by the equation

$$x_0 - \gamma = \beta e^{\delta t_0}. \tag{2.2}$$

Here, $x_0 = X_1 + \frac{1}{2}\eta$, X_1 being the smallest observation in the sample, η is the precision or measurement accuracy of the observations, and t_0 is given by

$$\Phi(t_0) = \frac{k}{n} \tag{2.3}$$

where Φ is the standard normal distribution function, n the total sample size and k the number of observations recorded as being equal to X_1. This formulation allows for the possibility of rounded data; if the data were recorded to a very high precision so that there were effectively no ties, then one would presumably take $x_0 = X_1$ and $k = 1$. Thus it can be seen that the maximum likelihood equation for γ is replaced by a moments equation based on the approximate mean of the smallest order statistic. The usual maximum likelihood estimating equations for the other parameters δ and β are retained, thus giving a system of three nonlinear equations in the three unknown parameters. His main argument for presenting this approach was computational simplicity, but he also hinted that it may have superior statistical properties as well.

In work subsequent to Cohen's original paper on the lognormal distribution, Hill (1963) pointed out that the solutions to the likelihood equations – that is, setting the derivatives of the likelihood function equal to 0 – in fact gave only a local and not a global maximum of the likelihood function, the global maximum being $+\infty$ attained as $\alpha \uparrow x_1$. Nevertheless, he argued in favor of the local maximum likelihood estimator, essentially by analogy with the corresponding Bayesian solution. Harter and Moore (1966) also advocated using the local maximum, extending the discussion to include censored data, and proposed an alternative modified scheme based on treating X_1 as a censored value when the local maximum does not exist. Griffiths (1980) expanded further on the theme of local maximum likelihood estimation by proposing interval estimation procedures

based on the maximized likelihood. Meanwhile Cohen and Whitten (1980) gave a more detailed review of the problem presenting a number of different forms of modified maximum likelihood and moments procedures.

Possibly to an even greater extent than the lognormal distribution, the three-parameter Weibull is widely studied and hotly debated. Cohen (1975) defined this family by the density f and distribution function F given by

$$f(x; \gamma, \delta, \theta) = \frac{\delta}{\theta}(x - \gamma)^{\delta-1} \exp\left\{-\frac{(x - \gamma)^\delta}{\theta}\right\},$$

$$F(x; \gamma, \delta, \theta) = 1 - \exp\left\{-\frac{(x - \gamma)^\delta}{\theta}\right\}, \tag{2.4}$$

both expressions being valid on $x > \gamma$, and considered the multi-censored likelihood function

$$L = \prod_{i=1}^{n} f(X_i; \gamma, \delta, \theta) \prod_{j=1}^{k} \{1 - F(T_j; \gamma, \delta, \theta)\}^{r_j}$$

which is based on n uncensored observations at $X_1, ..., X_n$ together with a group of r_j right-censored observations at T_j for each of k censoring points $T_1, ..., T_k$. An earlier development for the two-parameter Weibull (in which γ is known) was due to Cohen (1965).

For the three-parameter case with $\delta > 1$, he advocated direct application of the maximum likelihood method. The above parametrization in fact allows one to eliminate θ analytically, so the equations actually reduce to a set of two nonlinear equations in two unknowns. For $\delta < 1$, these equations have no solution since the likelihood function becomes unbounded when $\gamma \uparrow X_1$, where we assume $X_1 = \min(X_1, ..., X_n)$. In this case he advocated replacing the likelihood equation for the endpoint γ by $\hat{\gamma} = X_1 - \frac{1}{2}\eta$, where η is the sampling precision, or else by the more sophisticated estimating equation

$$X_1 = \gamma + \left(\frac{\theta}{N}\right)^{1/\delta} \Gamma\left(1 + \frac{1}{\delta}\right) \tag{2.5}$$

where N is the total sample size $n + r_1 + ... + r_k$ and $\Gamma(\cdot)$ the gamma function. Equation (2.5) amounts to equating the smallest order statistic X_1 to its expected value under the assumed model, and thus is very much in the same spirit as Cohen's (1951) earlier suggestion for the lognormal distribution.

Cohen and Whitten (1982b) proposed a number of alternative forms for the modified maximum likelihood idea. In this paper they used the parametrization

$$f(x; \gamma, \delta, \beta) = \frac{\delta}{\beta^\delta}(x - \gamma)^{\delta-1} \exp\left\{-\left(\frac{x - \gamma}{\beta}\right)^\delta\right\}, \quad x > \gamma. \tag{2.6}$$

3

Based on simulations of the bias and standard deviation of the estimators in the complete sample case, they actually advocated maximum likelihood only when $\delta > 2.2$, and discussed five forms of modified maximum likelihood estimator (MMLE) for use in other cases:

I: Set $F(X_r) = r/(n+1)$, where X_r is the r'th smallest order statistic in a sample of size n and F is the cumulative distribution function. Detailed discussion was restricted to $r = 1$ though they also pointed out that $r > 1$ might be desirable for robustness purposes.

II: Set $X_1 = \gamma + n^{-1/\delta}\beta\Gamma(1 + 1/\delta)$, as in (2.5) with $N = n$.

III: Equate the sample and population means, i.e. $\bar{X} = \gamma + \beta\Gamma(1 + 1/\delta)$.

IV: Equate the sample and population variances, i.e. $s_X^2 = \beta^2\{\Gamma(1 + 2/\delta) - \Gamma^2(1 + 1/\delta)\}$.

V: Equate the sample and population medians, i.e. $X_{\mathrm{med}} = \gamma + \beta(\ln 2)^{1/\delta}$.

In each case the new moment equation replaced the likelihood equation for γ, $d\ln L/d\gamma = 0$, the likelihood equations for β and δ being retained in their usual form. MMLE of types I and II had been considered in the lognormal case by Cohen and Whitten (1980). Other related papers were Cohen and Whitten (1982a) for the three-parameter gamma distribution, and Chan, Cohen and Whitten (1984) for the three-parameter inverse Gaussian distribution. They also considered a family of "modified moment estimators" in which the usual equations for the first two moments, III and IV above, are supplemented by one of I, II or V instead of the more traditional third-moment equation.

For the three-parameter Weibull case, Cohen and Whitten (1982b) carried out a simulation study based on samples of sizes $n = 10$, 25 and 100 and six different values of δ. They suggested that ordinary maximum likelihood estimation works well when $\delta > 2.2$ but that versions I and II of MMLE would work well over the whole range of δ. They also computed the asymptotic variance of the maximum likelihood estimates when $\delta > 2$ but pointed out that for $\delta \leq 2$ the standard asymptotic theory fails – we consider this aspect in more detail in Section 3. They also gave some attention to the MME approach and this was developed further in subsequent papers: Cohen, Whitten and Ding (1984), Cohen, Whitten and Ding (1985), Cohen and Whitten (1985), Cohen and Whitten (1986), respectively for the Weibull, lognormal, inverse Gaussian and gamma cases.

Cohen has also studied truncated distributions. In Charernkavanich and Cohen (1984), a truncated two-parameter Weibull distribution was considered, of the form

$$f_{LT}(x; \alpha, \delta, \theta) = \frac{f(x; \delta, \theta)}{1 - F(\alpha; \delta, \theta)}, \qquad x > \alpha, \tag{2.7}$$

4

f and F denoting the density and distribution function of the two-parameter Weibull distribution,

$$f(x;\delta,\theta) = \frac{\delta}{\theta}x^{\delta-1}\exp\left(-\frac{x^\delta}{\theta}\right), \qquad F(x;\delta,\theta) = 1 - \exp\left(-\frac{x^\delta}{\theta}\right), \qquad x > 0. \qquad (2.8)$$

In this case maximum likelihood estimation results in $\widehat{\alpha} = x_1$ but modifications similar to I and II above may also be considered to reduce the obvious bias in this. Right-truncation was also considered.

Cohen's books (Cohen and Whitten 1988, Cohen 1991) have reviewed all of these estimation techniques, and have extended them to numerous parametric families beyond the ones discussed here. Throughout Cohen's work, there has been a strong emphasis on efficient computational implementation of the procedures, and it would appear that this aspect was his own motivation for developing alternatives to MLE. Nevertheless, a recurring theme developed in simulations of these estimators is that in small samples MMLE or MME may have superior statistical properties as well. I believe this aspect is ultimately more important than issues of computational convenience, and this is what I want to discuss in the remainder of the paper. I concentrate on MMLE as I believe this is the more interesting and general technique – for example, it extends easily to the case of censored data – but the general approach taken may be relevant for MME as well.

3. Theory of MLE and MMLE

In this section I review the known maximum likelihood theory for three-parameter Weibull and related distributions, and develop some new results for modified maximum likelihood estimation. The MMLEs will always be version II in the classification of Section 1, in which the smallest order statistic X_1 is equated to its expected value under the model, together with the usual likelihood estimating equations for β and δ. I restrict detailed discussion to the three-parameter Weibull distribution with density defined by (2.6). However, the general concepts and qualitative results are also applicable to many of the other estimators and distributions considered in Section 1.

Maximum likelihood estimation for the three-parameter Weibull distribution was considered by a number of authors including Wingo (1973), Rockette, Antle and Klimko (1974), Lemon (1975) and of course Cohen (1975). When $\delta > 2$ the Fisher information is finite and the maximum likelihood estimators obey all the usual asymptotic properties, such as asymptotic normality and asymptotic efficiency, of regular parametric problems. Moreover, it is appropriate to seek a local maximum of the log likelihood function and to ignore the singularity in the likelihood that occurs as $\gamma \uparrow X_1$ when $\delta < 1$. Rigorous proofs of these statements were provided by Smith (1985). There is no guarantee that a local maximum of the likelihood function exists, but Rockette, Antle and Klimko (1974) conjectured that when a local maximum does exist, it is unique. As far as I know, this

conjecture has never been proved, but I have never encountered a counterexample and believe the conjecture to be true for all practical purposes.

The Fisher information matrix has been given by numerous authors; here I follow Section 3.8 of Cohen and Whitten (1988) by writing the Fisher information matrix $A = A(\gamma, \delta, \beta) = (a_{ij})$, $i, j = 1, 2, 3$ (so $a_{11} = \mathrm{E}\{-\partial^2 \log f(X; \gamma, \delta, \beta)/\partial\gamma^2\}$, etc.) in the form

$$a_{11} = \frac{C}{\beta^2}, \qquad a_{22} = \frac{K}{\delta^2}, \qquad a_{33} = \frac{\delta^2}{\beta^2},$$

$$a_{12} = a_{21} = \frac{J}{\beta}, \qquad a_{13} = a_{31} = \frac{\delta^2}{\beta^2}\Gamma\left(2 - \frac{1}{\delta}\right), \qquad a_{23} = a_{32} = -\frac{\Psi(2)}{\beta}, \tag{3.1}$$

where

$$C = (\delta - 1)\left\{\Gamma\left(1 - \frac{2}{\delta}\right) + \delta\Gamma\left(2 - \frac{2}{\delta}\right)\right\} = (\delta - 1)^2\Gamma\left(1 - \frac{2}{\delta}\right)$$

$$J = \Gamma\left(1 - \frac{1}{\delta}\right) - \left\{1 + \Psi\left(2 - \frac{1}{\delta}\right)\right\}\Gamma\left(2 - \frac{1}{\delta}\right),$$

$$K = \Psi'(1) + \Psi^2(2).$$

Here $\Psi(\cdot)$ is the digamma function (the derivative of $\log\Gamma(\cdot)$) and $\Psi'(\cdot)$, the derivative of $\Psi(\cdot)$, is the trigamma function. These functions are tabulated in Abramowitz and Stegun (1964).

Provided $\delta > 2$, the (local) maximum likelihood estimates are asymptotically unbiased and normally distributed, with variance-covariance matrix given by $n^{-1}A^{-1}$, where n is the sample size. However, as $\delta \downarrow 2$, $C \uparrow \infty$, so this theory breaks down when $\delta \leq 2$.

The case $\delta \leq 2$ was considered in detail by Smith (1985). If δ and β are known, so that the only unknown parameter is γ itself, then there are several different cases depending on the exact value of δ. For $\delta = 2$ the maximum likelihood estimator $\hat{\gamma}$ is asymptotically normal but $\hat{\gamma} - \gamma$ is $O_p((n \log n)^{-1/2})$, instead of the usual $O_p(n^{-1/2})$ (Woodroofe 1972). For $1 < \delta < 2$, $\hat{\gamma}$ has an extremely complicated non-normal distribution obtained by Woodroofe (1974); here $\hat{\gamma} - \gamma = O_p(n^{-1/\delta})$. For $\delta \leq 1$ no local maximum of the likelihood function exists. However, for $\delta < 2$, Akahira (1975) showed that $O_p(n^{-1/\delta})$ is the optimal rate of convergence for any estimator and Ibragimov and Has'minskii (1981) obtained asymptotically optimal estimates, though these optimality results are also very complicated and not easy to apply in practice. A different approach is based on the smallest order statistic X_1, since if we define $Z_1 = n^{1/\delta}\beta^{-1}(X_1 - \gamma)$, we have

$$\Pr\{Z_1 \leq z\} = 1 - \exp(-z^\delta), \qquad z \geq 0, \tag{3.2}$$

so if X_1 is treated as an estimator of γ, it has an extremely simple distributional form with an error of $O_p(n^{-1/\delta})$. Akahira's result shows that this is the optimal *rate* of convergence whenever $\delta < 2$, though not for $\delta > 2$ (since in this case the MLE converges at rate $O_p(n^{-1/2})$).

6

In this context, the MMLE proposed by Cohen reduces to

$$\bar{\gamma} = X_1 - n^{-1/\delta}\beta\Gamma(1 + 1/\delta) \qquad (3.3)$$

and we see that equation (3.2) gives the exact distribution of this estimator; we still have $\bar{\gamma} - \gamma = O_p(n^{-1/\delta})$ but the subtracted term in (3.3) amounts to a bias correction which improves its properties as a point estimator. The estimator therefore appears to be a good estimator when $\delta < 2$, made more attractive by the simplicity of (3.2). However for $\delta > 2$ the maximum likelihood estimator is asymptotically more efficient.

Now let us turn to the case where all three parameters are unknown, and $\delta < 2$. In this case, I showed (in Smith 1985) that there is an asymptotic independence property between estimates of γ and those of the other parameters: in large samples, ignorance about δ and β does not affect one's ability to estimate γ, and conversely, if one tries to estimate δ and β using an estimated value of γ, then one gets exactly the same asymptotic distribution as if γ were known. The intuitive reason for this is that efficient estimates of γ converge at rate $O_p(n^{-1/\delta})$, which is smaller than the error $O_p(n^{-1/2})$ in $\widehat{\delta}$ and $\widehat{\beta}$, so asymptotically, the error in estimating γ has no influence on the solution.

In pursuit of a specific procedure with these properties, I proposed estimating γ by the sample minimum X_1, and then constructing the usual likelihood equations for δ and β based on the differences $X_2 - X_1, ..., X_n - X_1$. When $\delta < 2$, these estimators have the same asymptotic properties as the regular two-parameter Weibull estimates when γ is known, but when $\delta > 2$, they converge only at the rate $O_p(n^{-1/\delta})$ and are therefore inferior to the three-parameter maximum likelihood estimates (Theorem 4 of Smith, 1985).

Although this procedure has the right asymptotic properties, it is questionable whether it is in fact the best way to proceed in small samples. In a recent paper Lockhart and Stephens (1994) have proposed an "iterative bias reduction" procedure very similar to Cohen's MMLE, and claimed on the basis of Monte Carlo studies that it gave a better fit to the Weibull distribution than estimates based on X_1. It therefore seems worthwhile to re-examine the whole idea of MMLE estimation in the light of these asymptotic results.

First, an elementary but important result:

Theorem 1. Suppose an estimator $\tilde{\gamma}$ is defined by the right hand side of (3.3), but with estimators $\tilde{\delta}$, $\tilde{\beta}$ substituted for the true values δ and β. Suppose these estimators satisfy

$$\tilde{\beta} - \beta \to_p 0, \qquad \log n \ (\tilde{\delta} - \delta) \to_p 0 \qquad \text{as } n \to \infty. \qquad (3.4)$$

Here $\to_p$ denotes convergence in probability. Then $\tilde{\gamma}$ has the same asymptotic properties as $\bar{\gamma}$; in particular, the distribution of

$$\tilde{Z}_1 = n^{-1/\delta}\frac{\tilde{\gamma} - \gamma}{\beta} - \Gamma(1 + \frac{1}{\delta})$$

7

converges to the Weibull random variable Z_1 given by (3.2).

We should also note that $\Gamma(1+1/\delta)$ is the mean of Z_1, so this result also makes precise the claim that $\tilde{\gamma}$ is asymptotically unbiased. We note also that this result applies for all values of δ, not just $\delta < 2$.

Proof. Taylor expansion shows that

$$\tilde{\gamma} - \bar{\gamma} \approx (\tilde{\delta} - \delta)\frac{\partial\bar{\gamma}}{\partial\delta} + (\tilde{\beta} - \beta)\frac{\partial\bar{\gamma}}{\partial\beta}$$

$$= (\tilde{\delta} - \delta)n^{-1/\delta}\frac{\beta}{\delta^2}\left\{\Gamma'\left(1 + \frac{1}{\delta}\right) - \log n\ \Gamma\left(1 + \frac{1}{\delta}\right)\right\} - (\tilde{\beta} - \beta)n^{-1/\delta}\Gamma\left(1 + \frac{1}{\delta}\right).$$

$$(3.5)$$

Under the assumption (3.4), the right hand side of (3.5) is $o_p(n^{-1/\delta})$, so $\tilde{Z}_1 - Z_1 = o_p(1)$, and the result follows.

Now, let us consider a corresponding result for the estimation of δ and β:

Theorem 2. Suppose γ is estimated by an estimator $\tilde{\gamma}$ such that $\tilde{\gamma} - \gamma = O_p(n^{-1/\delta})$. Suppose the estimators $\tilde{\delta}$ and $\tilde{\beta}$ are obtained by maximizing the two-parameter log likelihood for (δ, β), but with $\tilde{\gamma}$ substituted for γ. Also let $(\bar{\delta}, \bar{\beta})$ denote the corresponding estimators computed from the same data using the true value of γ. Then $\tilde{\delta} - \bar{\delta}$ and $\tilde{\beta} - \bar{\beta}$ are each of $O_p(n^{-1/\delta})$ if $\delta > 1$, $O_p(n^{-1}\log n)$ if $\delta \le 1$.

The idea of the proof is the same as Theorem 4 of Smith (1985); see also Theorem 3 of Smith (1994). We omit the details.

For $\delta < 2$, $\tilde{\delta} - \bar{\delta}$ and $\tilde{\beta} - \bar{\beta}$ are of smaller order of magnitude than $\bar{\delta} - \delta$ and $\bar{\beta} - \beta$, the latter being of $O_p(n^{-1/2})$, so in this case Theorem 2 makes precise the notion that $\tilde{\delta}$ and $\tilde{\beta}$ are asymptotically unaffected by γ being unknown. For $\delta > 2$, however, the $O_p(n^{-1/\delta})$ error dominates, so this implies that the MMLE procedure is inefficient in this case.

So far, I have confined my discussion entirely to asymptotic theory. However, there are a number of respects in which these results do not ring true for finite samples. For example, consider Figure 1. In this figure I have computed n times the asymptotic variance of $\hat{\gamma}$, assuming $\beta = 1$ and for a range of values of $\delta > 2$, where (a) $\hat{\gamma}$ is the MLE for γ assuming δ and β unknown (this is curve I on Figure 1), (b) $\hat{\gamma}$ is the MLE for γ assuming δ and β known (curve II), (c) $\hat{\gamma}$ is the MMLE $\tilde{\gamma}$ (curve III). In this case I assume sample size $n = 100$, and the plotted curves are based on (a) the top left hand entry of A^{-1}, where the matrix $A = (a_{ij})$ is defined by (3.1), (b) $1/a_{11}$, again from (3.1), (c) $n^{1-2/\delta}\text{Var}(Z_1) = n^{1-2/\delta}\{\Gamma(1 + 2/\delta) - \Gamma^2(1 + 1/\delta)\}$ where Z_1 has the Weibull distribution (3.2). It can be seen that curve III always lies above curve II but, for $\delta > 2.45$, it lies below curve I. The interpretation, taken literally, is that for $\delta > 2.45$ we would prefer to use MMLE, despite the asymptotic inefficiency as $n \to \infty$. The comparison is not so good for larger n – for $n = 1000$, the crossover between curves I and III occurs at $\delta = 2.96$

– but in most reliability applications we have $n < 100$ so comparisons based on $n = 100$ seem meaningful in practice. However, the sharp contrast between curves I and II for MLE suggests there should be a similar contrast for MMLE, but our asymptotic results do not capture that. Similarly, for $\delta < 2$ and the estimation of δ and β, the asymptotic results say that one can ignore the effect of estimating γ, but this does not seem very realistic either.

For the remainder of this section, then, I discuss an alternative approach to characterizing the variance-covariance matrix of the MMLE, with more attention to practical approximation than theoretical asymptotics.

The estimating equations for MMLE may be written in the form $h_n(\gamma, \delta, \beta) = 0$, where the suffix n is written to make explicit the dependence on the sample size n, and

$$
h_n = \begin{bmatrix} h_n^{(1)} \\ h_n^{(2)} \\ h_n^{(3)} \end{bmatrix}
$$

$$
= \begin{bmatrix} -(X_1 - \gamma) + n^{-1/\delta}\beta\Gamma\left(1 + \frac{1}{\delta}\right) \\ -\frac{1}{\delta} - \frac{1}{n}\sum \log\left(\frac{X_i - \gamma}{\beta}\right) + \frac{1}{n}\sum \left(\frac{X_i - \gamma}{\beta}\right)^\delta \log\left(\frac{X_i - \gamma}{\beta}\right) \\ \frac{\delta}{\beta} - \frac{\delta}{\beta}\frac{1}{n}\sum \left(\frac{X_i - \gamma}{\beta}\right)^\delta \end{bmatrix}
$$

Define the matrix

$$
H_n = \begin{bmatrix} \frac{\partial h_n^{(1)}}{\partial \gamma} & \frac{\partial h_n^{(1)}}{\partial \delta} & \frac{\partial h_n^{(1)}}{\partial \beta} \\ \frac{\partial h_n^{(2)}}{\partial \gamma} & \frac{\partial h_n^{(2)}}{\partial \delta} & \frac{\partial h_n^{(2)}}{\partial \beta} \\ \frac{\partial h_n^{(3)}}{\partial \gamma} & \frac{\partial h_n^{(3)}}{\partial \delta} & \frac{\partial h_n^{(3)}}{\partial \beta} \end{bmatrix}. \tag{3.6}
$$

Assuming consistent solutions $(\tilde{\gamma}_n, \tilde{\delta}_n, \tilde{\beta}_n)$ exist, we will have

$$
h_n(\tilde{\gamma}_n, \tilde{\delta}_n, \tilde{\beta}_n) - h_n(\gamma, \delta, \beta) = H_n^* \cdot \begin{bmatrix} \tilde{\gamma}_n - \gamma \\ \tilde{\delta}_n - \delta \\ \tilde{\beta}_n - \beta \end{bmatrix}
$$

where H_n^* is H_n evaluated at some $(\gamma_n^*, \delta_n^*, \beta_n^*)$ on the line joining $(\tilde{\gamma}_n, \tilde{\delta}_n, \tilde{\beta}_n)$ to (γ, δ, β). Noting that $h_n(\tilde{\gamma}_n, \tilde{\delta}_n, \tilde{\beta}_n) = 0$, we then have

$$
\begin{bmatrix} \tilde{\gamma}_n - \gamma \\ \tilde{\delta}_n - \delta \\ \tilde{\beta}_n - \beta \end{bmatrix} = -(H_n^*)^{-1} \cdot h_n \tag{3.7}
$$

where h_n (here and in all subsequent appearances) is evaluated at the true parameter values (γ, δ, β).

9

For large n and for $\delta > 1$, the entries of H_n converge to asymptotic values given by the law of large numbers. (There is a difficulty when $\delta \leq 1$, which is explained below). Moreover, the entries are continuous functions of the parameters (γ, δ, β). Thus for approximate calculations we may replace H_n^* by a matrix $\tilde{H}_n$, defined as the limiting form of H_n at the true parameter values. This gives the approximation

$$\begin{bmatrix} \tilde{\gamma}_n - \gamma \\ \tilde{\delta}_n - \delta \\ \tilde{\beta}_n - \beta \end{bmatrix} \approx -(\tilde{H}_n)^{-1} \cdot h_n, \tag{3.8}$$

which is easier to handle than (3.7) because $\tilde{H}_n$ is now a fixed (non-random) matrix.

Using this approximation, if the variance-covariance matrix of h_n is denoted V_n, we have that the limiting variance-covariance matrix of $(\tilde{\gamma}_n, \tilde{\delta}_n, \tilde{\beta}_n)$ is given by

$$W_n = (\tilde{H}_n)^{-1} V_n (\tilde{H}_n^T)^{-1}.$$

Now let us look more closely at the structure of $\tilde{H}_n$ and V_n. We may write

$$\tilde{H}_n = \begin{bmatrix} 1 & b_n^T \\ c & A_1 \end{bmatrix}, \qquad V_n = \begin{bmatrix} d_n & f_n^T \\ f_n & n^{-1}A_1 \end{bmatrix},$$

where b_n and c are (column) vectors of constants, $d_n = n^{-2/\delta}\{\Gamma(1 + 2/\delta) - \Gamma^2(1 + 1/\delta)\}$ and A_1 is the lower right 2×2 submatrix of A, where A was defined in (3.1). The vector f_n will be explained further below. It should be pointed out that if we had a more general p-dimensional problem, in which the first parameter was an endpoint parameter estimated by an equation similar to (3.3) and the remaining $p - 1$ parameters were estimated from their likelihood equations, then the same general structure, and the manipulations to follow, would be valid. Thus the theory being outlined here could be taken as a first attempt at a general theory for MMLE in problems involving endpoint estimation, though I am confining my detailed calculations to the three-parameter Weibull case.

If we define

$$U_n = (A_1 - cb_n^T)^{-1},$$

then we readily check that

$$\tilde{H}_n^{-1} = \begin{bmatrix} v_n & -b_n^T U_n \\ -U_n c & U_n \end{bmatrix}$$

where $v_n = 1 + b_n^T U_n c$. Consequently,

$$W_n = \begin{bmatrix} w_{n1} & w_{n2}^T \\ w_{n2} & W_{n3} \end{bmatrix}$$

10

where the scalar w_{n1}, the 2×1 vector w_{n2} and the 2×2 matrix W_{n3} are given by

$$w_{n1} = d_n v_n^2 - 2 v_n f_n^T U_n^T b_n + \frac{1}{n} b_n^T U_n A_1 U_n^T b_n,$$

$$w_{n2} = -d_n v_n U_n c + U_n c f_n^T U_n^T b_n + v_n U_n f_n - \frac{1}{n} U_n A_1 U_n^T b_n,$$

$$W_{n3} = d_n U_n c c^T U_n^T - U_n (c f_n^T + f_n c^T) U_n^T + \frac{1}{n} U_n A_1 U_n^T.$$

These expressions may be expressed directly in terms of the elements of $\tilde{H}_n$ and V_n if we write

$$U_n = A_1^{-1} + v_n A_1^{-1} c b_n^T A_1^{-1}, \quad v_n = (1 - b_n^T A_1^{-1} c)^{-1}$$

from which it also follows that

$$b_n^T U_n = v_n b_n^T A_1^{-1}, \quad U_n c = v_n A_1^{-1} c.$$

We then have

$$w_{n1} = v_n^2 \left(d_n - 2 f_n^T A_1^{-1} b_n + \frac{b_n^T A_1^{-1} b_n}{n} \right),$$

$$w_{n2} = -v_n^2 \left(d_n - 2 f_n^T A_1^{-1} b_n + \frac{b_n^T A_1^{-1} b_n}{n} \right) A_1^{-1} c - v_n A_1^{-1} \left(\frac{b_n}{n} - f_n \right),$$

$$W_{n3} = v_n^2 \left(d_n - 2 f_n^T A_1^{-1} b_n + \frac{b_n^T A_1^{-1} b_n}{n} \right) A_1^{-1} c c^T A_1^{-1}$$
$$+ v_n A_1^{-1} \left\{ c \left(\frac{b_n}{n} - f_n \right)^T + \left(\frac{b_n}{n} - f_n \right) c^T \right\} A_1^{-1} + \frac{1}{n} A_1^{-1}. \tag{3.9}$$

To evaluate these expressions, we still need to know b_n, c and f_n. However, from (3.6) we have $c = (a_{12} \ a_{13})^T$ where a_{12} and a_{13} are given by (3.1). Also,

$$b_n = \begin{bmatrix} n^{-1/\delta} \left(\frac{\beta}{\delta^2} \right) \{ \Gamma \left(1 + \frac{1}{\delta} \right) \log n - \Gamma' \left(1 + \frac{1}{\delta} \right) \} \\ n^{-1/\delta} \Gamma \left(1 + \frac{1}{\delta} \right) \end{bmatrix}.$$

To evaluate f_n, we need to calculate

$$\text{Cov} \left\{ X_1, \frac{1}{n} \sum \left(\frac{X_j - \gamma}{\beta} \right)^\theta \right\}$$

for arbitrary index $\theta \geq 0$.

11

To simplify the calculations, we assume without loss of generality that $\gamma = 0$, $\beta = 1$. Consider the following scheme for generating the sample $(X_1, ..., X_n)$: set $X_1 = n^{-1/\delta}Z_1$, where Z_1 has c.d.f. (3.2), then generate $X_2, ..., X_n$ conditionally independently from the density

$$\exp(X_1^\delta)\delta x^{\delta-1}\exp(-x^\delta), \qquad x > X_1.$$

This is equivalent to drawing an independent sample $(X_1, ..., X_n)$ and reordering so that X_1 is the smallest value (with the others ordered randomly).

Now, for $j > 1$,

$$
\begin{aligned}
\mathrm{E}\{X_j^\theta | X_1\} &= \exp(X_1^\delta)\int_{X_1}^\infty \delta x^{\theta+\delta-1}\exp(-x^\delta)dx \\
&= \exp(X_1^\delta)\Gamma\left(1+\frac{\theta}{\delta}\right)\left\{1 - \int_0^{X_1}\delta x^{\theta+\delta-1}\exp(-x^\delta)dx\right\} \\
&= \Gamma\left(1+\frac{\theta}{\delta}\right)\left(1+X_1^\delta+\frac{1}{2}X_1^{2\delta}+...\right)\left(1-\frac{\delta}{\theta+\delta}X_1^{\theta+\delta}+\frac{\delta}{\theta+2\delta}X_1^{\theta+2\delta}-...\right) \\
&= \Gamma\left(1+\frac{\theta}{\delta}\right)\left(1+X_1^\delta-\frac{\delta}{\theta+\delta}X_1^{\theta+\delta}+...\right)
\end{aligned}
$$

where we have retained only the three leading terms and dropped the rest. It follows that

$$
\begin{aligned}
\mathrm{E}\{X_1 X_j^\theta\} &= \Gamma\left(1+\frac{\theta}{\delta}\right)\mathrm{E}\left\{X_1+X_1^{\delta+1}-\frac{\delta}{\theta+\delta}X_1^{\theta+\delta+1}+...\right\} \\
&= n^{-1/\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(1+\frac{1}{\delta}\right)+n^{-1-1/\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(2+\frac{1}{\delta}\right) \\
&\quad - n^{-1-1/\delta-\theta/\delta}\frac{\delta}{\theta+\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(2+\frac{1}{\delta}+\frac{\theta}{\delta}\right)+...
\end{aligned}
$$

Hence

$$\mathrm{E}\left\{X_1\cdot\frac{1}{n}\sum_2^n X_j^\theta\right\} = \left(1-\frac{1}{n}\right)\left\{n^{-1/\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(1+\frac{1}{\delta}\right)+...\right\}.$$

We also have

$$\mathrm{E}\left\{\frac{1}{n}X_1^{\theta+1}\right\} = n^{-1-1/\delta-\theta/\delta}\Gamma\left(1+\frac{1}{\delta}+\frac{\theta}{\delta}\right)$$

and that

$$\mathrm{E}\{X_1\} = n^{-1/\delta}\Gamma\left(1+\frac{1}{\delta}\right), \qquad \mathrm{E}\left\{\frac{1}{n}\sum_1^n X_j^\theta\right\} = \Gamma\left(1+\frac{\theta}{\delta}\right).$$

12

Hence we deduce

$$\operatorname{Cov}\left\{X_1, \frac{1}{n}\sum_1^n X_j^\theta\right\} \approx$$

$$-n^{-1-1/\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(1+\frac{1}{\delta}\right)+n^{-1-1/\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(2+\frac{1}{\delta}\right)$$

$$-n^{-1-1/\delta-\theta/\delta}\frac{\delta}{\theta+\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(2+\frac{1}{\delta}+\frac{\theta}{\delta}\right)+n^{-1-1/\delta-\theta/\delta}\Gamma\left(1+\frac{1}{\delta}+\frac{\theta}{\delta}\right)$$

$$=n^{-1-1/\delta}\cdot\frac{1}{\delta}\Gamma\left(1+\frac{\theta}{\delta}\right)\Gamma\left(1+\frac{1}{\delta}\right)$$

$$-n^{-1-1/\delta-\theta/\delta}\Gamma\left(1+\frac{1}{\delta}+\frac{\theta}{\delta}\right)\left\{\left(1+\frac{1}{\theta+\delta}\right)\Gamma\left(1+\frac{\theta}{\delta}\right)-1\right\}$$

$$=R_n(\theta)\qquad\text{say.}$$

$$(3.10)$$

We may also formally differentiate with respect to θ in (3.10) to deduce

$$\operatorname{Cov}\left\{X_1, \frac{1}{n}\sum_1^n X_j^\theta\log X_j\right\}\approx R_n'(\theta)$$

$$=n^{-1-1/\delta}\cdot\frac{1}{\delta^2}\Gamma'\left(1+\frac{\theta}{\delta}\right)\Gamma\left(1+\frac{1}{\delta}\right)$$

$$+n^{-1-1/\delta-\theta/\delta}\frac{\log n}{\delta}\Gamma\left(1+\frac{1}{\delta}+\frac{\theta}{\delta}\right)\left\{\left(1+\frac{1}{\theta+\delta}\right)\Gamma\left(1+\frac{\theta}{\delta}\right)-1\right\}$$

$$-n^{-1-1/\delta-\theta/\delta}\frac{1}{\delta}\Gamma'\left(1+\frac{1}{\delta}+\frac{\theta}{\delta}\right)\left\{\left(1+\frac{1}{\theta+\delta}\right)\Gamma\left(1+\frac{\theta}{\delta}\right)-1\right\}$$

$$-n^{-1-1/\delta-\theta/\delta}\Gamma\left(1+\frac{1}{\delta}+\frac{\theta}{\delta}\right)\left\{\frac{1}{\delta}\left(1+\frac{1}{\theta+\delta}\right)\Gamma'\left(1+\frac{\theta}{\delta}\right)-\frac{1}{(\theta+\delta)^2}\Gamma\left(1+\frac{\theta}{\delta}\right)\right\}$$

Based on these approximations, and reinserting the scale parameter β, we deduce that

$$f_n \approx \left[\frac{\beta\{R_n'(0)-R_n'(\delta)\}}{\delta R_n(\delta)}\right].\tag{3.11}$$

However $f_n = O(n^{-1-1/\delta}\log n)$, so it seems safe to drop terms in which the power of n is $-2-1/\delta$. With this further simplification for $R_n(\delta)$ and $R_n'(\delta)$, we have

$$R_n(\delta) = n^{-1-1/\delta}\cdot\frac{1}{\delta}\Gamma\left(1+\frac{1}{\delta}\right),$$

$$R_n'(0)-R_n'(\delta) = n^{-1-1/\delta}\cdot\frac{1}{\delta^2}\Gamma\left(1+\frac{1}{\delta}\right)\left\{\log n-\Psi\left(1+\frac{1}{\delta}\right)-(\delta+1)\Psi(1)\right\}.$$

$$(3.12)$$

13

Equations (3.11) and (3.12) define our final approximation for f_n, and complete our derivation of the approximate variances and covariances in (3.9).

We are now in a position to consider the implications of these results. For $\delta < 2$ we have $w_{n1} \sim d_n$ and $W_{n3} \sim n^{-1} A_1^{-1}$. Also $w_{n2} = o(n^{-1/\delta - 1/2} \log n)$, so that the correlations between $\tilde{\gamma}_n$ and $(\tilde{\delta}_n, \tilde{\beta}_n)$ are asymptotically negligible in this case. Thus for $\delta < 2$, the asymptotic form of W_n is

$$\begin{bmatrix} d_n & 0 \\ 0 & n^{-1} A_1^{-1} \end{bmatrix}$$

confirming our earlier result that estimation of γ and the pair (δ, β) are effectively independent in this case. On the other hand, for $\delta > 2$ the matrix W_n is asymptotic to

$$d_n \begin{bmatrix} 1 & -c^T A_1^{-1} \\ -A_1^{-1} c & A_1^{-1} c c^T A_1^{-1} \end{bmatrix}$$

which is of rank 1; in other words, in this case the variability of $\tilde{\gamma}$ effectively dominates the whole problem.

It is of more interest, however, to examine the individual terms in relation to their asymptotic values. In the case when δ and β are known, the variance of $\tilde{\gamma}$ is d_n; in w_{n1}, the main change is to multiply this by a factor v_n^2 which tends to 1 at rate $O(n^{-1/\delta} \log n)$ as $n \to \infty$. In fact we often have $v_n > 2$; for instance, at $n = 100$ this is true for all $\delta > 2.5$. Thus in practice the difference between the two variances is not negligible and the calculations we have made are essential to obtain realistic approximations. This will be supported by simulations in Section 4.

If we look instead at W_{n3}, it can be seen that this is written as the sum of three terms, the first of which measures the influence of the variation in $\tilde{\gamma}$, the last is the asymptotic covariance matrix when γ is known, and the second represents a cross-correlation between these two effects. The first term is of $O(n^{-2/\delta})$ and the third of $O(n^{-1})$; although asymptotically one of these terms always dominates the other (except when $\delta = 2$), from the point of view of practical calculation it would seem much more sensible to retain both of them.

Before leaving this section we return briefly to the case $\delta \leq 1$, for which the results are not valid. The difficulty is that in this case $\partial h_n^{(2)} / \partial \gamma$ does not obey the law of large numbers, so that c is infinite. In fact $\partial h_n^{(2)} / \partial \gamma = O_p(n^{1/\delta - 1})$ (Smith 1985, Lemma 4.II(iii)) so we could bound c by an increasing sequence c_n of this (or very slightly larger) order of magnitude. However in this case the direct interpretability of the expression for W_{n3} is lost.

Another point to make is that, although I have calculated an approximate covariance matrix for the estimators, it does not follow that they are approximately normally distributed. Indeed from (3.8), in which $h_n^{(2)}$ and $h_n^{(3)}$ are asymptotically bivariate normal but $h_n^{(1)}$ has a Weibull distribution, it will follow that the true limiting distribution

14

is a complicated mixture of Weibull and normal random variables. However, I have not attempted to work out the details of this.

4. Numerical results and simulations

Curve IV on Figure 1 plots n times the approximate variance of $\tilde{\gamma}$, when $n = 100$, based on (3.9). It can be seen that it lies well above the first-order approximation based on Theorem 1, and still somewhat above curve I, which is for three-parameter maximum likelihood estimation. So we do not actually claim that MMLE is more efficient than MLE, though the difference is not too great when based on the new approximations for MMLE. In fact, if the curves in Figure 1 are continued for larger values of δ, they cross over (curve IV lies below curve I) at about $\delta = 15$, but this is most likely a numerical artifact; we cannot expect the approximations in Section 3 to work well for very large δ.

In Figures 2 and 3, similar curves are plotted for n times the variances of the estimators of δ and β respectively. Here curve I is for the three-parameter MLE, curve II for the two-parameter MLE when γ is known, and curve III is the new approximation based on (3.9). When $\delta >> 2$, curve I lies well above curve II; again, curve III is above both of them, but the difference between curve I and curve III is not too great. For $\delta \leq 2$, based on the results of Smith (1985), curves I and II merge into a single curve, but still curve III lies considerably above them, reflecting the real influence of γ being unknown. Only as δ approaches 1 does the new approximation appear to break down; I conjecture that this is because of the instability of the parameter a_{12}, which becomes infinite as $\delta \downarrow 1$.

To examine how well these results reflect the true variances of the estimators, a simulation study was performed. The simulation was based on samples of size 100 for each of four values of δ, namely 4.0, 2.5, 1.75 and 1.2, and 1000 replications were performed of each experiment. For each sample, the MLE $(\hat{\gamma}, \hat{\delta}, \hat{\beta})$ was calculated using the optimization routine DFPMIN from Press et al. (1986). This routine assumes that function values and first-order derivatives are available, but I have approximated the first-order derivatives of the log likelihood by using a simple differencing procedure, so effectively converting it into a derivative-free optimizer. For the MMLE, the objective is to find a value $\tilde{\gamma}$ such that (3.3) holds when δ and β are replaced by their MLEs $\tilde{\delta}$ and $\tilde{\beta}$ under the two-parameter model with $\gamma = \tilde{\gamma}$ fixed. This was done by starting with two trial values of $\tilde{\gamma}$ and linearly interpolating to obtain a third value, the process being iterated to convergence. This procedure, which has been advocated by Cohen in a number of his papers, usually converges in fewer than ten iterations. In contrast, Lockhart and Stephens (1994) advocated direct iteration between (3.3) and the maximum likelihood solution for (δ, γ); this can be a very slow procedure. There are two known circumstances under which MLE can fail to converge. The first has already been pointed out, and arises when $\delta \leq 1$ and $\gamma \uparrow X_1$. The second occurs for large δ and has been called *the embedded model problem* by Cheng and Iles (1989). This occurs because, under some conditions involving $\delta \rightarrow \infty$, $\beta \rightarrow \infty$ and $\gamma \rightarrow -\infty$, the three-parameter Weibull distribution converges to a quite different model (the Gumbel or two-parameter extreme value distribution) which may fit the data better

than any finite case of the three-parameter Weibull. If this happens, then the estimates will appear to diverge. In principle, this could happen with MMLE as well, though in fact in only one of our 4000 simulations, involving the MLE with $\delta = 4.0$, did this actually happen. A more serious point, however, is that there were several instances in which both $\tilde{\delta}$ and $X_1 - \tilde{\gamma}$ were very large, and this distorted the variances of the estimators. Some trial runs with $n = 25$ and $\delta = 4.0$ suggested that for this smaller value of n, there were many more instances when the procedures did not converge at all, but these are not reproduced here.

As an assessment of the performance of the estimators, both variances and mean squared errors (MSEs) of the estimators were calculated, ignoring any instances where the estimator did not converge. In most cases the MSE was not much larger than the variance, showing that bias is not a serious issue at this sample size. However, as a robust alternative to the standard deviation, I also computed 0.741 times the interquartile range (IQR) of the simulated estimates. In this case, non-converging cases were not deleted as these may legitimately be regarded as outliers of the sampling distribution. The multiplier 0.741 is based on the fact that the standard deviation is 0.741 IQR for a normal distribution, though as previously pointed out, the estimators are not exactly normally distributed (even asymptotically) for this problem. In all cases, the quantity actually tabulated is n times the estimated variance of the distribution.

For each value of δ, the four rows of Table 1 represent (i) the sample variance of the simulations, (ii) the MSE, (iii) the IQR-based method just described, (iv) the theoretical value obtained from the information matrix in the case of MLE, and (3.9) for MMLE. With MLE for $\delta < 2$, no asymptotic value is shown for $\hat{\gamma}$, while the estimates for $\hat{\delta}$ and $\hat{\beta}$ are those for the two-parameter Weibull, since this is the right asymptotic result when $\delta < 2$. Also shown in parentheses under the MMLE for γ is nd_n, which is the value that would be obtained under the first-order approximation given by Theorem 1 of Section 3.

For $\delta = 4$, the differences between the variance (or MSE) and the IQR-based estimates are drastic; however the variance is dominated by the effect of a few outlying terms, and for this reason I believe that the IQR is a much more realistic estimate of the variability of this procedure. For the other values of δ, this difference does not appear to matter.

In comparing theoretical and simulated (IQR-based) variances, it can be seen that with $\delta = 4$ or 2.5, the theoretical results for both MLE and MMLE somewhat underestimate the true value, but the difference is not too bad, and certainly much better than for the uncorrected estimated for the variance of $\tilde{\gamma}$, which is out by a factor of up to 10. When $\delta = 1.75$, the approximation for the MMLE is very good, and in the case of $\tilde{\delta}$ and $\tilde{\beta}$ certainly better than the uncorrected approximations, which are the same as for MLE. Only for $\delta = 1.2$ are the new results somewhat unconvincing, but I have already given a plausible explanation of this.

In an attempt to see just how well the results are capable of doing, the whole experiment was repeated with 500 replications of sample size 500, with results in Table 2. This

16

is too large a sample size for most reliability experiments, but the simulation is included mainly to demonstrate that for a value of n that is large but not extremely large, the new approximations do perform well and much better than the first-order asymptotics. In this case, such a conclusion seems substantiated for all cases expect $\delta = 1.2$; here there is not much to choose between the two methods of approximation.

As far as the comparison between MLE and MMLE is concerned — well, there are no cases when MMLE does clearly better than MLE on the simulations, except when judged by raw variances in the $n = 100$, $\delta = 4$ case. It may be that such a result would be seen for smaller sample sizes, but in that case even our new approximations are not going to be good enough to explain what is going on theoretically. The computational simplicity of MMLE, which Cohen has demonstrated on numerous occasions, remains one feature in its favor. Apart from that, I believe the main advantage of MMLE is that its behavior is reasonably well understood, and certainly better understood than the multi-parameter MLE, for the regime when δ is near to or less than 2, which is when the standard asymptotics of MLE start to break down. This is very much consistent with the conclusions Cohen himself has reached from his extensive numerical studies. It is hoped that the results given here will serve to stimulate further theoretical research into these issues.

References

Abramowitz, M. and Stegun, I.A. (1964), *Handbook of Mathematical Functions.* National Bureau of Standards, Washington D.C., reprinted by Dover, New York.

Akahira, M. (1975), Asymptotic theory for estimation of location in non-regular cases. I: Order of convergence of consistent estimators. II: Bounds of asymptotic distributions of consistent estimators. *Rep. Statist. Appl. Res. Union Jap. Sci. Eng.* **22**, 8-26 and 99-115.

Chan, M., Cohen, A.C. and Whitten, B.J. (1984), Modified maximum likelihood and modified moment estimators for the three-parameter inverse Gaussian distribution. *Comm. Stat. Simul. Comp.* **13**, 47-68.

Cheng, R.C.H. and Iles, T.C. (1989), Embedded models in three parameter distributions and their estimation. *J.R. Statist. Soc. B.* **52**, 135-149.

Charernkavanich, D. and Cohen, A.C. (1984), Estimation in the singly truncated Weibull distribution with unknown truncation point. *Comm. Stat. Theor. Meth.* **A13**, 843-857.

Cohen, A.C. (1951), Estimating parameters of logarithmic-normal distributions by the method of maximum likelihood. *J. Amer. Statist. Assoc.* **46**, 206-212.

Cohen, A.C. (1965), Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics* **7**, 579-588.

17

Cohen, A.C. (1975), Multi-censored sampling in the three parameter Weibull distribution. *Technometrics* **17**, 347-351.

Cohen, A.C. (1991), *Truncated and Censored Samples: Theory and Applications.* Marcel Dekker, New York.

Cohen, A.C. and Whitten, B.J. (1980), Estimation in the three-parameter lognormal distribution. *J. Amer. Statist. Assoc.* **75**, 399-404.

Cohen, A.C. and Whitten, B.J. (1982a), Modified moment and maximum likelihood estimators for parameters of the three-parameter gamma distribution. *Comm. Stat. Simul. Comp.* **11**, 197-214.

Cohen, A.C. and Whitten, B.J. (1982b), Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution. *Comm. Stat. Theor. Meth.* **11**, 2631-2656.

Cohen, A.C. and Whitten, B.J. (1985), Modified moment estimation for the three-parameter inverse Gaussian distribution. *J. Qual. Tech.* **17**, 147-154.

Cohen, A.C. and Whitten, B.J. (1986), Modified moment estimation for the three-parameter gamma distribution. *J. Qual. Tech.* **18**, 53-62.

Cohen, A.C. and Whitten, B.J. (1988), *Parameter Estimation in Reliability and Life Span Models.* Marcel Dekker, New York.

Cohen, A.C., Whitten, B.J. and Ding, Y. (1984), Modified moment estimation for the three-parameter Weibull distribution. *J. Qual. Tech.* **16**, 159-167.

Cohen, A.C., Whitten, B.J. and Ding, Y. (1985), Modified moment estimation for the three-parameter lognormal distribution. *J. Qual. Tech.* **17**, 92-99.

Griffiths, D.A. (1980), Interval estimation for the three-parameter lognormal distributions via the likelihood equations. *Applied Statistics* **29**, 58-68.

Harter, H.L. and Moore, A.M. (1966), Local maximum likelihood estimation of the parameters of three-parameter log-normal populations from complete and censored samples. *J. Amer. Statist. Assoc.* **61**, 842-851.

Hill, B.M. (1963), The three-parameter log-normal distribution and Bayesian analysis of a point-source epidemic. *J. Amer. Statist. Assoc.* **58**, 72-84.

Ibragimov, I.A. and Has'minskii, R.Z. (1981), *Statistical Estimation.* Springer Verlag, Berlin.

Lemon, G.H. (1975), Maximum likelihood estimation for the three-parameter Weibull distribution based on censored samples. *Technometrics* **17**, 247-254.

Lockhart, R.A. and Stephens, M.A. (1994), Estimation and tests of fit for the three-parameter Weibull distribution. *J.R. Statist. Soc. B* **56**, 491-500.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986), *Numerical Recipes: The Art of Scientific Computation.* Cambridge University Press, Cambridge.

Rockette, H., Antle, C. and Klimko, L.A. (1974), Maximum likelihood estimation with the Weibull model. *J. Amer. Statist. Assoc.* **69**, 246-249.

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

Smith, R.L. (1994), Nonregular regression. *Biometrika* **81**, 173-183.

Wilson, E.B. and Worcester, J. (1945), The normal logarithmic transform. *Rev. Econ. Statist.* **27**, 17-22.

Wingo, D.R. (1973), Solution of the three-parameter Weibull equations by constrained modified quasi-linearization (progressively censored samples). *IEEE Trans. Reliability* **R22**, 96-102.

Woodroofe, M. (1972), Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Math. Stat.* **43**, 113-122.

Woodroofe, M. (1974), Maximum likelihood estimation of translation parameter of truncated distribution II. *Ann. Statist.* **2**, 474-488.

Table 1: Simulation results and approximations for n=100

	MLE				MMLE		
Number	$\widehat{\gamma}$	$\widehat{\delta}$	$\widehat{\beta}$	Number	$\tilde{\gamma}$	$\tilde{\delta}$	$\tilde{\beta}$
(a) $\delta = 4$							
999	35.7	825	36.4	1000	17.7	461	18.5
	35.7	828	36.4		18.4	479	19.1
	3.85	88.2	4.29		6.68	143	7.33
	3.21	76.1	3.54		4.84	95.0	5.21
					(0.647)		
(b) $\delta = 2.5$							
1000	1.18	18.9	1.67	1000	1.40	21.0	1.95
	1.25	19.1	1.77		1.43	22.0	1.99
	0.890	13.1	1.26		1.31	19.0	2.00
	0.396	8.04	0.670		1.08	13.1	1.47
					(0.362)		
(c) $\delta = 1.75$							
1000	0.254	3.78	0.722	1000	0.292	4.17	0.794
	0.332	4.10	0.870		0.294	4.29	0.795
	0.219	3.26	0.753		0.282	4.11	0.894
	–	1.86	0.362		0.267	3.62	0.774
					(0.143)		
(d) $\delta = 1.2$							
921	0.0305	1.01	0.725	1000	0.0407	1.31	0.829
	0.0556	1.07	0.810		0.0412	1.38	0.829
	0.0207	1.17	0.695		0.0296	1.25	0.753
	–	0.875	0.770		0.050	1.75	1.00
					(0.029)		

This table lists n times the variance of the MLEs and MMLEs for simulated samples of size $n = 100$ and four values of δ (together with $\gamma = 0$, $\beta = 1$). The results are based on 1000 replications and "Number" denotes the number of those replications for which the estimators were successfully computed. For each δ, the four rows of numbers are based on 1. the sample variance of the simulations, 2. the mean squared deviation from the true value (sample variance plus squared bias), 3. variance estimator computed from the IQR, 4. theoretical result as obtained from the Fisher information matrix in the case of MLE or (3.9) in the case of MMLE. The fifth number in parentheses in the column for $\tilde{\gamma}$ is nd_n, i.e. the crude variance approximation based on direct application of Theorem 1.

Table 2: Simulation results and approximations for n=500

	MLE				MMLE		
Number	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\beta}$	Number	$\tilde{\gamma}$	$\tilde{\delta}$	$\tilde{\beta}$
(a) $\delta = 4$							
500	4.11	91.8	4.39	500	6.92	148	7.32
	4.11	91.8	4.40		6.98	150	7.38
	3.69	85.6	4.13		7.22	152	7.42
	3.21	76.1	3.54		6.53	138	7.09
					(1.45)		
(b) $\delta = 2.5$							
500	0.790	12.1	1.14	500	1.06	14.8	1.49
	0.808	12.3	1.16		1.07	14.9	1.50
	0.695	11.1	1.10		1.16	16.0	1.66
	0.396	8.04	0.670		1.02	13.7	1.41
					(0.500)		
(c) $\delta = 1.75$							
500	0.126	2.97	0.611	500	0.146	3.17	0.641
	0.167	3.24	0.695		0.146	3.17	0.642
	0.108	3.23	0.612		0.134	3.24	0.604
	–	1.86	0.362		0.158	3.06	0.615
					(0.114)		
(d) $\delta = 1.2$							
500	0.0099	0.974	0.883	500	0.0109	1.01	0.904
	0.0181	1.04	0.933		0.0109	1.02	0.905
	0.0069	0.883	0.946		0.0080	0.880	0.943
	–	0.875	0.770		0.0117	1.09	0.825
					(0.0098)		

This table is computed in the same way as for Table 1, except that it is for 500 replications based on samples of size $n = 500$.

21

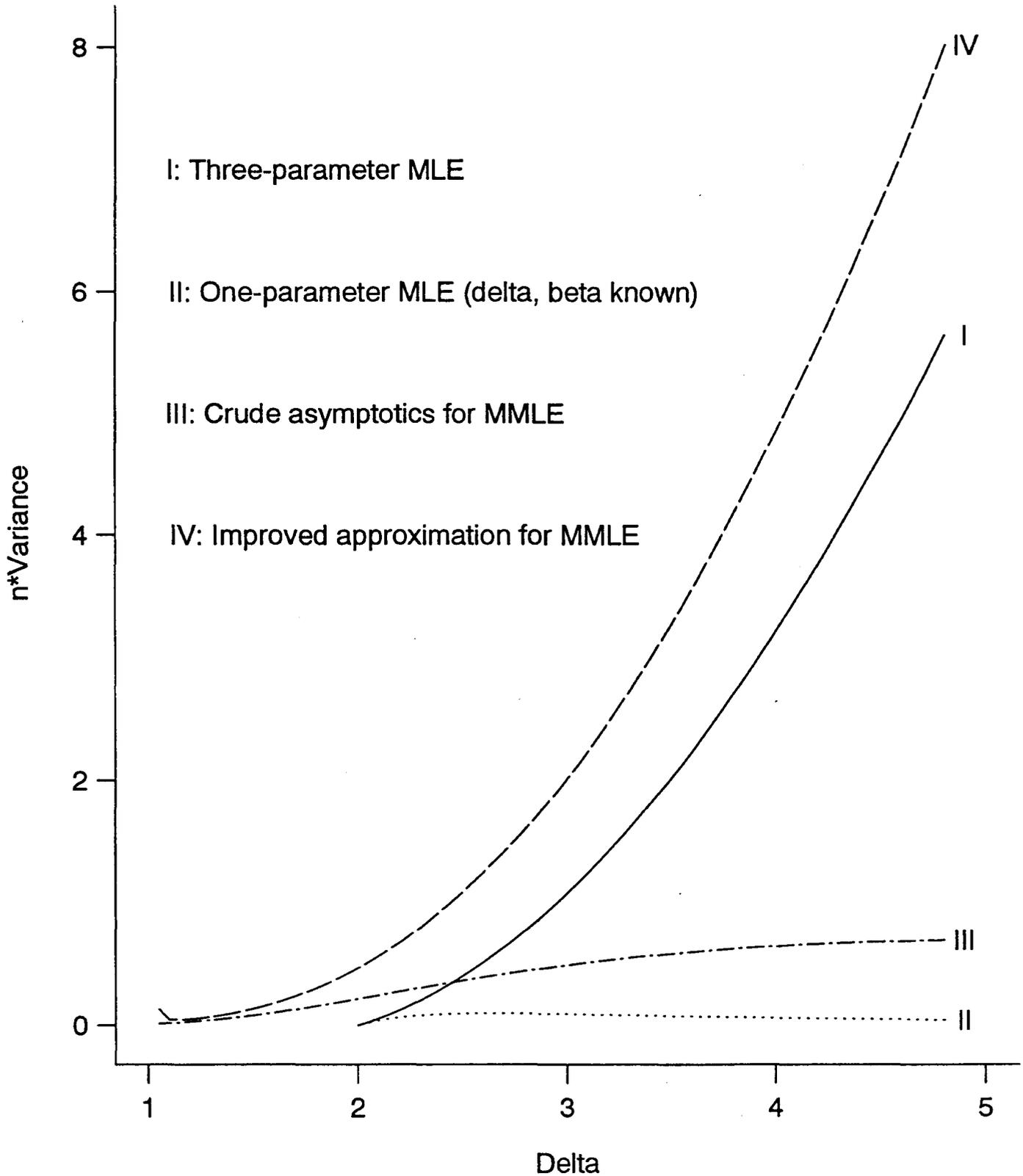Figure 1: Variance approximations for estimating Gamma (sample size n=100)

I: Three-parameter MLE

II: One-parameter MLE (delta, beta known)

III: Crude asymptotics for MMLE

IV: Improved approximation for MMLE

Figure 2: Variance approximations for estimating Delta (sample size n=100)

I: Three-parameter MLE

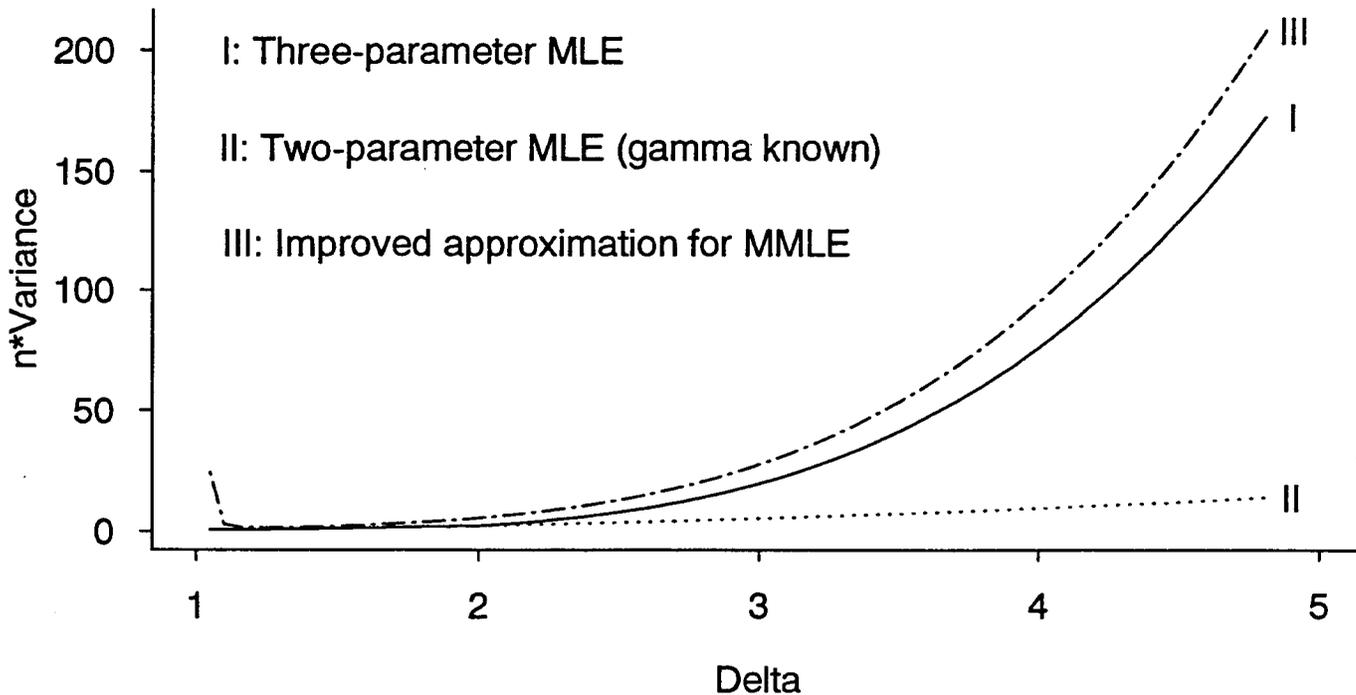II: Two-parameter MLE (gamma known)

III: Improved approximation for MMLE

Figure 3: Variance approximations for estimating Beta (sample size n=100)

I: Three-parameter MLE

II: Two-parameter MLE (gamma known)

III: Improved approximation for MMLE