

Fitting the Three-Parameter Weibull Distribution: Review and Evaluation of Existing and New Methods

Denis Cousineau

Université de Montréal

C. P. 6128, succ. Centre-ville

Montréal, Québec, H3C 3J7, Canada

ABSTRACT

The three-parameter Weibull distribution is a commonly-used distribution for the study of reliability and breakage data. However, given a data set, it is difficult to estimate the parameters of the distribution and that, for many reasons: (1) the equations of the maximum likelihood estimators are not all available in closed form. These equations can be estimated using iterative methods. However, (2) they return biased estimators and the exact amount of bias is not known. (3) The Weibull distribution does not meet the regularity conditions so that in addition to being biased, the maximum likelihood estimators may also be highly variable from one sample to another (weak efficiency). The methods to estimate parameters of a distribution can be divided into three classes: a) the maximizing approaches, such as the maximum likelihood method, possibly followed by a bias-correction operation; b) the methods of moments; and c) a mixture of the previous two classes of methods. We found using Monte Carlo simulations that a mixed method was the most accurate to estimate the parameters of the Weibull distribution across many shapes and sample sizes, followed by the weighted Maximum Likelihood estimation method. If the shape parameter is known to be larger than 1, the Maximum Product of Spacing method is the most accurate whereas in the opposite case, the mixed method is to be preferred. A test that can detect if the shape parameter is smaller than 1 is discussed and evaluated. Overall, the maximum likelihood estimation method was the worst, with errors of estimation almost twice as large as those of the best methods.

Index Terms — Weibull distribution, parameter estimation, maximum likelihood.

1 INTRODUCTION

THE Weibull distribution is used in reliability studies, for example to study the voltage breakage of electric circuits [1, 2]. It is also commonly used in other disciplines such as physics for the study of crystallization, in climatology for the study of tides and in cognitive psychology to study the time to complete a task [3-5]. The Weibull distribution with non-zero shift has three parameters, denoted in the following: $\gamma > 0$, the shape parameter responsible for the skew of the distribution; $\beta > 0$, the scale parameter; finally α , the shift parameter which is also a lower bound. Its probability density function is given by

$$f(x | \gamma, \beta, \alpha) = \gamma \beta^{-\gamma} (x - \alpha)^{\gamma-1} \exp\left[-\left(\frac{x - \alpha}{\beta}\right)^\gamma\right] \quad (1)$$

Three basic statistics of this distribution, the mean, the variance

and the Fisher skew are given by

$$E(x) = \alpha + \beta \Gamma(1 + 1/\gamma)$$

$$Var(x) = \beta^2 \left(\Gamma(1 + 2/\gamma) - [\Gamma(1 + 1/\gamma)]^2 \right) \quad (2)$$

$$Sk(x) = \frac{[\Gamma(1 + 1/\gamma)]^3 - 3\Gamma(1 + 1/\gamma)\Gamma(1 + 2/\gamma) + \Gamma(1 + 3/\gamma)}{\left(\Gamma(1 + 2/\gamma) - [\Gamma(1 + 1/\gamma)]^2 \right)^{3/2}}$$

in which Γ represents the Gamma function. Its shape can vary from an hyperexponential (whenever $\gamma < 1$) to a near symmetrical (when $\gamma \approx 3.6$) to a negatively skewed distribution (as $\gamma \rightarrow \infty$).

In general, the best method to estimate the parameters of a distribution is the Maximum Likelihood Estimation method (MLE). However, its application to the Weibull distribution is problematic for three reasons:

1) In order for MLE to be the most efficient method (as sample size increases to infinity), the distribution must satisfy the so-called regularity conditions [6]. However, the Weibull

distribution does not satisfy one of the conditions because the domain of the random variable depends on the position of the lower bound (the parameter α). Three cases occurs, identified in [7]: $\gamma \leq 1$, in which case the MLE method is not consistent as there may be more than one solution [8]; $1 < \gamma \leq 2$, where the distribution of the estimates does not follow a normal distribution; finally, $\gamma > 2$, for which the weak regularity conditions are satisfied so that the properties of MLE (consistency, normality, maximum efficiency) are effective as the sample size tends towards infinity.

2) The MLE solutions are biased and it is not known by what amount. The bias strongly depends on the parameter γ and the sample size n [9]. Figure 1 illustrates the mean estimated parameters $\hat{\gamma}$, $\hat{\beta}$ and $\hat{\alpha}$ as a function of the population γ and sample size n . It was suggested to model the bias surfaces of Figure 1 using polynomials so that a correction could be applied on the estimates [10]. However, we found this approach to be very sensitive to the specifics of the implementation (for example, changing the optimization procedure). Likewise, a review of some bias correction procedures was performed in [1]. However, they are all based on the two-parameter Weibull distribution and do not generalize to the three-parameter Weibull distribution.

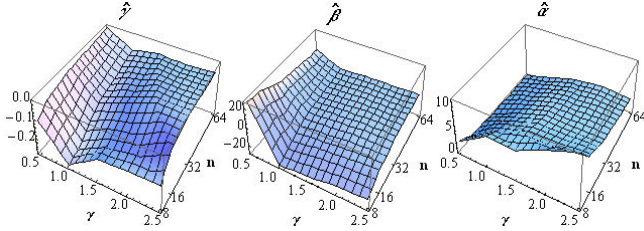


Figure 1. Biases of the MLE method as a function of sample size and true population γ . Left panel: estimated shape parameter; middle panel: estimated scale parameter; right panel: estimated shift parameter. A quantity below zero means an underestimation of the parameter.

3) The MLE solutions are not available in a direct form for two of the three parameters. The likelihood function of the Weibull parameters for a sample \mathbf{X} of size n is given by

$$\mathcal{L}(\gamma, \beta, \alpha | \mathbf{X}) = \prod_{i=1}^n f(x_i | \gamma, \beta, \alpha)$$

so that the log of the likelihood function simplifies to:

$$\begin{aligned} \log(\mathcal{L}(\gamma, \beta, \alpha | \mathbf{X})) &= -n\gamma \log(\beta) + n \log(\gamma) \\ &+ (\gamma - 1) \sum_{i=1}^n \log(x_i - \alpha) - \beta^{-\gamma} \sum_{i=1}^n (x_i - \alpha)^\gamma \end{aligned}$$

Maximizing the above is achieved by computing the derivative with respect to the scale parameter β and setting it to zero. The resulting equation can be reorganized into:

$$\beta = \gamma \sqrt[\gamma]{\frac{1}{n} \sum_{i=1}^n (x_i - \alpha)^\gamma} \quad (3)$$

It is the only parameter which can be isolated on the left-hand side of an equation. Note that it requires the knowledge of the other two parameters to be computed.

The derivative with respect to γ yields, replacing $\beta^{-\gamma}$ with $1 / (\frac{1}{n} \sum_{i=1}^n (x_i - \alpha)^\gamma)$ and dividing by n :

$$\frac{1}{\gamma} + \frac{1}{n} \sum_{i=1}^n \log(x_i - \alpha) - \frac{\sum_{i=1}^n \log(x_i - \alpha) (x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^\gamma} = 0. \quad (4)$$

Similarly for α , we get:

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i - \alpha} \times \frac{\sum_{i=1}^n (x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^{\gamma-1}} - \frac{\gamma}{\gamma - 1} = 0 \quad (5)$$

Equations (4) and (5) define implicitly the parameters γ and α , which in turn can be used to derive β .

One way to estimate γ and α is to square equations (4) and (5) and search for the minimum so that the full MLE solution is given by $\{\hat{\gamma}, \hat{\alpha}\}_{\text{MLE}}$ which satisfies the following two constraints:

$$\left\{ \begin{array}{l} \text{Min}_{\gamma, \alpha} \left[\frac{1}{\gamma} + \frac{1}{n} \sum_{i=1}^n \log(x_i - \alpha) - \frac{\sum_{i=1}^n \log(x_i - \alpha) (x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^\gamma} \right]^2 \\ \text{Min}_{\gamma, \alpha} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i - \alpha} \times \frac{\sum_{i=1}^n (x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^{\gamma-1}} - \frac{\gamma}{\gamma - 1} \right]^2 \end{array} \right\} \quad (6)$$

and by
$$\hat{\beta}_{\text{MLE}} = \gamma \sqrt[\gamma]{\frac{1}{n} \sum_{i=1}^n (x_i - \alpha)^\gamma}$$

In the following, we will review alternatives to maximizing the likelihood function, namely the maximum product of spacing (MPS) [11] and the weighted Maximum Likelihood estimation method (w-MLE) [12]. We will also review method of moment estimators [13] and propose new such estimators. We will also consider mixed methods where a subset of the parameters is estimated with one method and the remaining parameters with another method. Finally, we will assess the capabilities of the methods using Monte Carlo simulations. Because the parameters α and β are scaling parameters, they will not be varied in the simulations; only the shape parameter γ and the sample size n will be manipulated.

2 A REVIEW OF EXISTING METHODS TO ESTIMATE THE PARAMETERS OF A THREE-PARAMETER WEIBULL DISTRIBUTION

The methods will be divided into three classes, one that maximizes a measure of adequacy (e.g. the likelihood), one that isolate the parameters from the equations of basic statistics of

equations (2) and a class which mixes methods from the previous two.

2.1 MAXIMIZING ADEQUACY

Maximum Likelihood estimation (MLE). This method was already reviewed above. With this method, the parameter β is a function of the other two parameters. It can be shown that if the true parameter α and γ are known, $\hat{\beta}^\gamma$ follows a Gamma distribution with parameters n , β^γ/n , which tends to a normal distribution as n get large. However, the true parameters α and γ being unknown, this result is of little interest in practical application.

Maximum Product of Spacing (MPS). To avoid inconsistent solutions when $\gamma < 1$, it was proposed to replace the likelihood function with a related function, the spacing function [11]. This function returns the probability that the data are spaced the way they are. To do so, the following quantity is maximized:

$$\begin{aligned} S(\gamma, \beta, \alpha | \mathbf{X}) &= \prod_{i=1}^{n+1} \left(\int_{x_{i-1}}^{x_i} f(x | \gamma, \beta, \alpha) dx \right) \\ &= \prod_{i=1}^{n+1} (F(x_i | \gamma, \beta, \alpha) - F(x_{i-1} | \gamma, \beta, \alpha)) \end{aligned} \quad (7)$$

in which F is the cumulative distribution function of the Weibull distribution $F(x | \gamma, \beta, \alpha) = 1 - \exp(-(x - \alpha)/\beta)^\gamma$ and where $F(x_0)$ and $F(x_{n+1})$ are set to 0 and 1 respectively. This function is consistent for all γ , avoiding the first difficulty with MLE. In addition, it was shown to be hyper-efficient, i.e. to produce less variable estimates than the MLE alternative, a characteristic that will be present in the subsequent simulations.

A generalization of MPS was proposed, the Quantile Maximum Product Estimation (QMPE) [14]-[17] and explored in [9]. In this generalization, $m \leq n$ quantiles are estimated from the data (with boundaries $[q_i, q_{i+1}]$, each quantile regrouping n_i data ($\sum n_i = n$)). Then, the function to maximize is

$$\begin{aligned} Q(\gamma, \beta, \alpha | \mathbf{X}) &= \prod_{i=1}^{n+1} \left(\int_{q_{i-1}}^{q_i} f(x | \gamma, \beta, \alpha) dx \right)^{n_i} \\ &= \prod_{i=1}^{n+1} (F(q_i | \gamma, \beta, \alpha) - F(q_{i-1} | \gamma, \beta, \alpha))^{n_i} \end{aligned}$$

in which $F(q_0) = 0$ and $F(q_{n+1}) = 1$. The advantage of QMPE is that it is little affected by outliers (suspiciously small or large data) commonly found in experimental psychology. We chose not to explore QMPE here as the optimal number of quantile m is not known.

Weighted Maximum Likelihood estimation (w-MLE). First proposed in [18] for the two-parameter Weibull distribution and extended to the three-parameter case in [12], this approach proposes equations similar to the MLE's but incorporating three weights such that $\{\hat{\gamma}, \hat{\alpha}\}_{w-MLE}$ satisfy the system of equations:

$$\left\{ \begin{array}{l} \text{Min}_{\gamma, \alpha} \left(\frac{w_2}{\gamma} + \frac{1}{n} \sum_{i=1}^n \log(x_i - \alpha) - \frac{\sum_{i=1}^n \log(x_i - \alpha)(x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^\gamma} \right)^2 \\ \text{Min}_{\gamma, \alpha} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i - \alpha} \times \frac{\sum_{i=1}^n (x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^{\gamma-1}} - w_3 \right)^2 \end{array} \right.$$

from which we derive

$$\hat{\beta}_{w-MLE} = \sqrt{\frac{1}{nw_1} \sum_{i=1}^n (x_i - \alpha)^\gamma} \quad (8)$$

The weights are meant to nullify the biases and two of them (w_1 and w_2) depends only on the sample size whereas the third depends on the sample size and the hypothetical shape parameter γ . The weights are not available in closed form; [12] provides tables in which the weights were estimated using Monte Carlo simulations. Contrary to MLE, w-MLE properties exist for all sample sizes, not just for the asymptotic cases.

Related techniques. A nested MLE method was proposed in which the search for the best-fitting parameters is performed by subdividing the parameter space [19]. This technique however returns estimates that are only marginally better than those of the regular MLE technique [20].

Reference [21] supplemented the regular MLE method with an alternative method when the true shape parameter is smaller than 1 to avoid inconsistent solutions. However, the method returned estimates that are poorer than regular MLE (inverting the bias). They also proposed a test to determine whether the population shape is smaller or equal to 1. We will discuss this test in the last section.

Finally, Bayesian estimation techniques and prior-informed MLE are generalizations of the regular MLE method [22], [23]. Because they are based on the likelihood functions, the inherent biases of MLE transfer to these methods as well. Their advantage is that they incorporate priors, if such priors are available in a given theoretical context.

2.2 METHOD OF MOMENTS

The method of moments (MoM) consists in using equations of summary statistics to isolate the parameters to be estimated. Statistics of higher moments (e. g. skewness as opposed to mean) are more variable and more likely to result in a poor estimate. A few MoM estimators were proposed for the two-parameter Weibull distributions which cannot be generalized to a three-parameter Weibull distribution [13]. Harter and Moore suggested to estimate γ using the coefficient of variation (ratio of the mean onto the standard deviation). If there is no shift ($\alpha = 0$), this ratio is equal to an expression involving γ only (that expression cannot be manipulated to isolate γ but using a simple binary search, its value can be estimated). In the case where α is not zero and unknown, their estimate

$$\hat{\gamma}_{HM} \text{ such that } \frac{\bar{X} - \alpha}{\bar{X}} = \frac{\Gamma(1+1/\hat{\gamma}_{HM})}{\sqrt{\Gamma(1+2/\hat{\gamma}_{HM}) - [\Gamma(1+1/\hat{\gamma}_{HM})]^2}}$$

(in which \bar{X} is the mean of the sample X and \bar{X} is the sample standard deviation) requires the knowledge of α to be obtained. Likewise, [13] proposed a MoM estimator for β which requires both α and γ : $\hat{\beta}_{HM} = (\bar{X} - \alpha) / \Gamma(1+1/\gamma)$ but this method will not be examined further.

It is possible to derive a new MoM estimate for γ which does not require information of the other two parameters, but this estimator requires an estimate of the Fisher skew, which is a moment-3 statistic: $\hat{\gamma}_{Sk}$ such that:

$$\hat{X} = \frac{[\Gamma(1+1/\hat{\gamma}_{Sk})]^3 - 3\Gamma(1+1/\hat{\gamma}_{Sk})\Gamma(1+2/\hat{\gamma}_{Sk}) + \Gamma(1+3/\hat{\gamma}_{Sk})}{(\Gamma(1+2/\hat{\gamma}_{Sk}) - [\Gamma(1+1/\hat{\gamma}_{Sk})]^2)^{3/2}} \quad (9)$$

in which \hat{X} represents the unbiased sample skew [24] measured by

$$\hat{X} = \frac{n^2}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{X})^3.$$

Regarding the other two parameters, it seems difficult to produce an estimate not based on the other parameters, but we found a few estimators based only on one parameter, the shape parameter γ .

One was suggested in [13] and uses the standard deviation (moment-2 statistic):

$$\hat{\beta}_{SD} = \frac{\bar{X}}{\sqrt{\Gamma(1+2/\gamma) - [\Gamma(1+1/\gamma)]^2}} \quad (10)$$

In the following, two new MoM estimators will be proposed, based on the notion of expected minimum of a sample, $E(\text{Min}(X))$. The minimum of a sample of size n is smaller than x with probability:

$$Pr(\text{Min}(X) \leq x) = 1 - Pr(x_1 \geq x \ \& \ x_2 \geq x \ \& \ \dots \ \& \ x_n \geq x)$$

which, if the data are independent, equals:

$$Pr(\text{Min}(X) \leq x) = 1 - \prod_{i=1}^n Pr(x_i \geq x)$$

In the case of a Weibull distribution, this simplifies into:

$$Pr(\text{Min}(X) \leq x) = 1 - \exp\left(-\left(\frac{x - \alpha}{\beta / \sqrt[n]{n}}\right)^\gamma\right)$$

which is a Weibull distribution with the same shape and shift parameter than the sample population but with scale reduced by a factor $\sqrt[n]{n}$. Hence, the expected minimum of a sample is given by $E(\text{Min}(X)) = \alpha + (\beta / \sqrt[n]{n}) \Gamma(1+1/\gamma)$. Using this statistic, we have that $E(X) - E(\text{Min}(X)) = \beta \Gamma(1+1/\gamma) (1 - 1/\sqrt[n]{n})$ such that a new MoM estimate for β is given by:

$$\hat{\beta}_{Min} = \frac{\bar{X} - X_{(1)}}{\Gamma(1+1/\gamma) (1 - 1/\sqrt[n]{n})} \quad (11)$$

in which $X_{(1)}$ is the minimum of the sample. Likewise, $\sqrt[n]{n} E(\text{Min}(X)) - E(X) = (\sqrt[n]{n} - 1)\alpha$ such that a MoM estimator of α is

$$\hat{\alpha}_{Min} = \frac{\sqrt[n]{n} X_{(1)} - \bar{X}}{\sqrt[n]{n} - 1} \quad (12)$$

2.3 MIXED METHODS

We have found two MoM estimators which are dependent on γ . Because they involve only moment-1 statistics (expected value and expected minimum value) which are not biased statistics, they have the potential to be unbiased estimators if the shape parameter γ is itself without bias. However, the MoM estimator for γ is based on skew, a moment-3 statistic which is very likely to be a poor estimator (as will be confirmed in the next section). This opens the possibility to have a mixed method in which γ is estimated using a maximum-adequacy method and the other two parameters are then estimated using the MoM estimators. This is one approach that will be examined in the subsequent section.

3 ESTIMATING THE SHAPE PARAMETER γ

In this section, we concentrate only on methods to estimate the shape parameters. We identified in the previous section four methods that can be used: MLE (equation 6), MPS (equation 7), w-MLE (equation 8) and MoM using skew (equation 9). These methods were tested using Monte Carlo simulations (methodology described in the appendix) varying the sample size n {8, 16, 32, 64} from very small to moderately large sample size and the true shape parameter γ_T {0.5, 1.0, 1.5, 2.0, 2.5} from regions with inconsistent solutions to regions satisfying the weak regularity conditions.

In this section, we screened the estimated $\hat{\gamma}$ for possible extreme estimates. Indeed, negatively skewed Weibull distributions occur for γ in the infinite range [3.6 .. ∞]. As a result, a Weibull distribution with shape 10 has an undistinguishable shape from another Weibull distribution with shape 10,000. A very small number of such extreme estimates would increase the measures of bias, efficiency and root mean square error (RMSE). Following [25], we rejected estimates of γ which were above 5.

There was a total of 2% of extreme estimates. The extreme estimates were not randomly distributed. 70% of them were located in the $\gamma_T = 2.5$ conditions. This is understandable as the true shape is closer to the cut-off point. Similarly, the methods differ regarding the number of outlying estimates they produced. The MLE method produced the largest amount of extreme estimates (41%), followed closely by MoM(Sk) and MPS (35% and 24% respectively). The w-MLE method never produced estimates above 5.

Figure 2 illustrates the biases, the efficiencies and the RMSE for the four sample sizes and the four methods divided according to the population shape.

As seen, the w-MLE method systematically outperforms the other methods. The worst method is the MoM(Sk) which is not

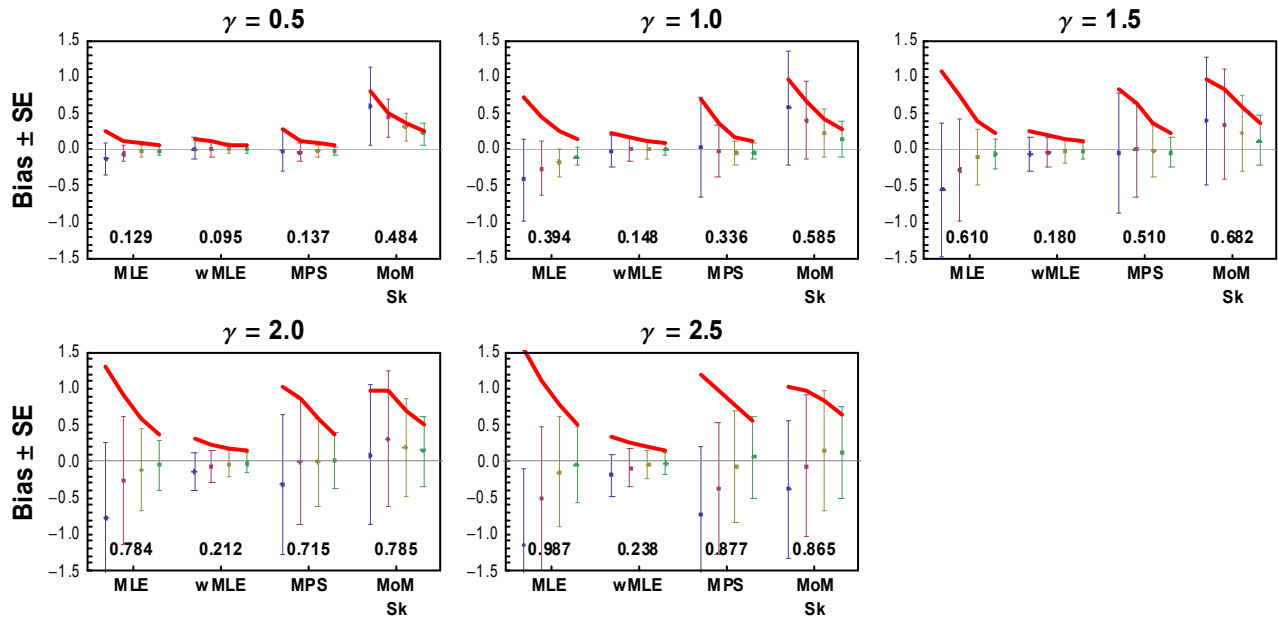


Figure 2. Bias, efficiency and RMSE for the shape estimates as a function of the true population shape (graphs), methods (abscissa) and sample sizes (in the order 8, 16, 32, 64). The dots shows the bias, the error bars, the efficiency, and the thick line, the RMSE. The number in the bottom of the graphs is the mean RMSE across the four sample sizes.

so surprising since it is based on a moment-3 statistic. Its poor performance are mostly apparent for small γ ; for large γ , its RMSE gets closer to MLE's RMSE.

MLE are also generally bad but surprisingly, not for the cases where $\gamma < 1$. In those situations, there is a second inconsistent solution but as shown in [8], this second solution is always located on one of the boundaries in the parameter space (e.g. when $\alpha = X_{(1)}$). Because all the boundaries were excluded from the starting values, it seems that the optimization procedure generally found the (correct) local maximum of the likelihood surface.

Table 1 summarizes the RMSE and the biases across all **Table 1.** Mean estimated RMSE and Bias on the parameter γ across sample sizes as a function of the true parameter γ for the four methods of estimation explored in the text

True γ	MLE	w-MLE	MPS	MoM(Sk)
RMSE				
0.5	0.129	0.095	0.137	0.484
1.0	0.394	0.148	0.336	0.585
1.5	0.610	0.180	0.510	0.682
2.0	0.784	0.212	0.715	0.785
2.5	0.987	0.238	0.877	0.865
Unsigned mean	0.581	0.175	0.515	0.680
Bias				
0.5	-0.052	0.011	-0.020	0.393
1.0	-0.238	-0.005	-0.015	0.339
1.5	-0.249	-0.035	-0.026	0.274
2.0	-0.304	-0.068	-0.081	0.180
2.5	-0.461	-0.087	-0.279	-0.044
Unsigned mean	0.261	0.041	0.084	0.246

sample sizes. Regarding biases, both w-MLE and MPS have smallest bias so that they are potential contenders for providing the γ estimate in a mixed method.

4 ESTIMATING ALL THREE PARAMETERS SIMULTANEOUSLY

We chose to examine three maximum-adequacy methods (namely MLE, (equation 6), MPS (equation 7), and w-MLE, (equation 8)) and three mixed methods. In all three mixed methods, β and α are estimated using the mean and the expected minimum (equations 11 and 12). They differ on the origin of $\hat{\gamma}$: it was taken from the previous three methods (the estimates are thus noted Mixed($\gamma \leftarrow$ MLE), Mixed($\gamma \leftarrow$ MPS), and Mixed($\gamma \leftarrow$ w-MLE)). In light of the results in the previous section, we expect MLE to provide poor estimates but it is nonetheless provided as a baseline and because it is still the most commonly used method.

The quality of the estimates are again reported in terms of RMSE, bias and efficiency, but this time computed on the vector composed of the three parameters $\bar{\theta} = \{\gamma, \beta, \alpha\}$ and distance computed using the Euclidian distance (the norm). The results are presented in Figure 3 and summarized in Table 2.

As seen, MLE and Mixed($\gamma \leftarrow$ MLE) are the worst methods (with one exception: if $\gamma = 0.5$, MPS is the second worst method). Yet, MPS is the most reliable method for $\gamma > 1$. When $\gamma \leq 1$, the Mixed($\gamma \leftarrow$ w-MLE) method is superior to all other methods. If you have to adopt only one method and you expect to have γ s that are above and below 1, the recommended methods are either Mixed($\gamma \leftarrow$ w-MLE) or w-MLE.

Because the reliability of the methods depends on whether $\gamma \leq 1$ or $\gamma > 1$, it would be nice to have a test to decide what

Table 2. Mean estimated RMSE and Bias on the parameter vector across sample sizes as a function of the true parameter γ for the six methods of estimation explored in the text.

True γ	MLE	w-MLE	MPS	Mixed ($\gamma \leftarrow$ MLE)	Mixed ($\gamma \leftarrow$ w-MLE)	Mixed ($\gamma \leftarrow$ MPS)
RMSE						
0.5	62.080	24.735	38.091	21.632	19.875	23.544
1.0	11.137	9.539	10.032	10.201	9.029	9.223
1.5	10.309	8.057	7.759	9.922	8.312	8.171
2.0	13.770	8.698	7.783	13.251	9.834	9.378
2.5	16.930	9.484	8.838	16.498	11.430	11.309
Unsigned mean	22.845	12.103	14.500	14.301	11.696	12.325
Bias						
0.5	16.515	6.912	11.106	-2.742	2.286	5.241
1.0	0.705	2.098	0.662	-2.356	0.560	0.131
1.5	3.497	0.835	-0.570	0.509	0.079	0.078
2.0	4.722	0.207	-0.156	1.262	-0.060	1.205
2.5	5.734	-0.467	0.981	1.942	-0.167	2.682
Unsigned mean	6.235	2.104	2.695	1.762	0.630	1.867

method to use. Such a test was proposed in [21], hereafter called the Lockhart and Stephens test. It consists in minimizing the following:

$$\left(\frac{1}{\gamma} + \frac{1}{n} \sum_{i=1}^n \log(x_i - \alpha) - \frac{\sum_{i=1}^n \log(x_i - \alpha)(x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^\gamma} \right)^2 + \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i - \alpha} \times \frac{\sum_{i=1}^n (x_i - \alpha)^\gamma}{\sum_{i=1}^n (x_i - \alpha)^{\gamma-1}} - \frac{\gamma}{\gamma-1} \right)^2 \quad (13).$$

If the minimized value is not zero, we are in a region with inconsistent solutions and therefore, γ must be smaller or equal to 1. We verified this test with Monte Carlo simulations using a very small sample size ($n = 8$) and found it to be very sensitive, as seen in Table 3.

By first using the test followed by either MPS if equation (13) is zero or Mixed($\gamma \leftarrow$ w-MLE) otherwise, the global RMSE is lowered to 10.656, a significant improvement over any one method used exclusively.

5 CONCLUSION

In this report, we examined methods to estimate the parameters of the three-parameter Weibull distribution. The results showed that the regular MLE is the worst method and should be avoided at all cost unless the sample size is very large (much larger than the sizes tested here). Any other method tested in the third section performed almost 40% better than MLE.

However, there is another reason why MLE should not be used: the precision in measurements. In all the empirical sciences, the data are measured with a precision of $\pm \Delta x$ (in the Monte Carlo simulations reported above, we generated random deviates with 40 digits of precision so that Δx can be ignored). As such, the likelihood of a data set is unfounded since we only have access to approximate measures and this fact must have repercussions on the likelihood measure:

$$\begin{aligned} \mathcal{L}(\gamma, \beta, \alpha | \mathbf{X} \pm \Delta x) &= \prod_{i=1}^n \Pr(x - \Delta x \leq x_i \leq x + \Delta x) \\ &= \prod_{i=1}^n (F(x_i + \Delta x | \gamma, \beta, \alpha) - F(x_i - \Delta x | \gamma, \beta, \alpha)) \end{aligned} \quad (14)$$

As seen, equation (14) is much more similar to MPS (equation 7) in which the spacing to the subsequent datum provided a form of measurement precision. As such, this method should be preferred over the standard MLE method, more so if γ is known to be larger than 1.

APPENDIX

In all the Monte Carlo simulations reported in the text, we used the following method. In one simulation, a random sample of size n was obtained from a population following a Weibull distribution with true parameters $\{\gamma, \beta = 100, \alpha = 300\}$. Because α and β are scaling parameters, it is not necessary to vary them systematically. The sample was then analysed using the methods described in the text and the resulting estimates were saved. This simulation process was repeated a thousand times for a given condition of n and γ .

For all the maximizing adequacy methods, we used a simplex algorithm [27]. Both β and γ were constrained to be larger than zero. The shift parameter was constrained to be smaller than the smallest value of the sample (as α is the lower bound

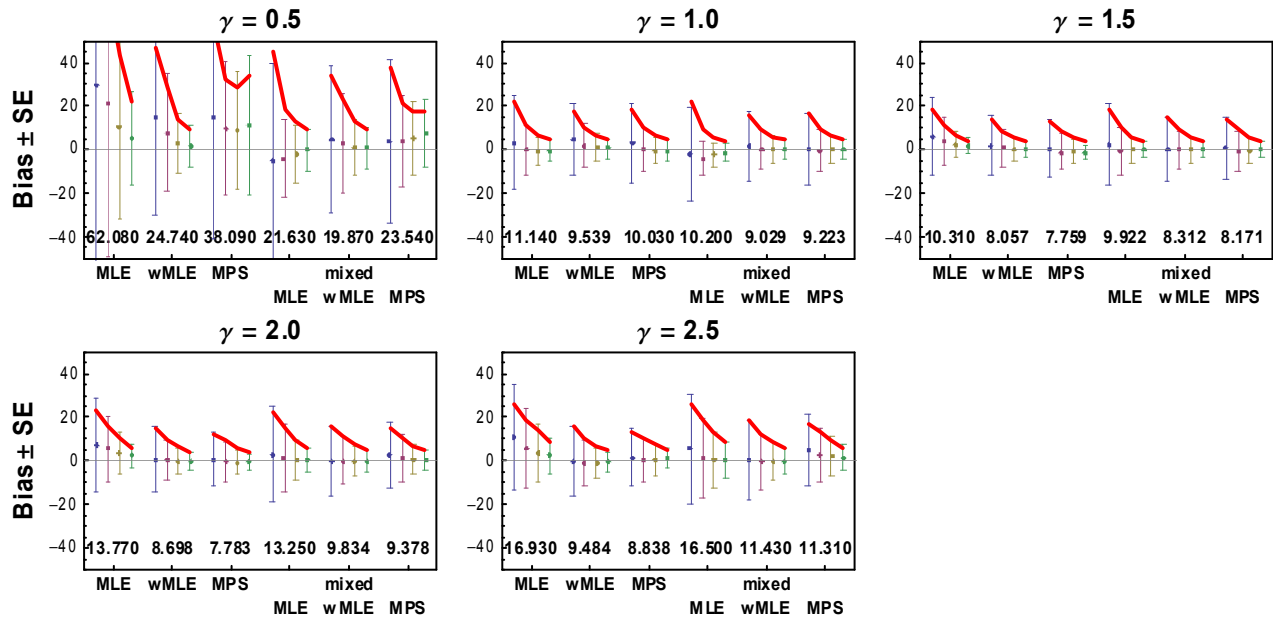


Figure 3. Bias, efficiency and RMSE for the estimated parameters in the same format as Figure 2.

parameter, all the data have to be above it). The search was provided two starting values for each parameter. They were the same in all conditions: For γ , 0.6 and 2.4; for β , 80 and 120; for α , 280 and 320. These quantities for α and β correspond to what can generally be estimated from a visual inspection of the empirical distribution.

Across simulations, we varied the sample size n , with values $\{8, 16, 32, 64\}$. These numbers represent a very small, small, moderate and moderate to large sample sizes which are commonly found in empirical papers. We did not explore larger sample sizes as the problem of parameter estimation is much less acute for $n \geq 100$. We also varied the shape parameter γ , with values $\{0.5, 1.0, 1.5, 2.0, 2.5\}$. We chose one value for each scenario $\{\gamma = 0.5$ implies the presence of inconsistent estimates; $\gamma = 1.5$ implies non normal estimates; $\gamma = 2.5$ implies normal and asymptotically most efficient estimates) as well as the boundaries between the scenarios ($\gamma = 1.0$ and $\gamma = 2.0$). In total, we had 20 conditions each containing 1000

Table 3. Proportion of times the Lockart and Stephens (1994) test detected a γ larger than 1 as a function of the true γ .

γ	Proportion
0.98	0.000
0.99	0.000
1.00	0.000
1.01	0.777
1.02	0.919
1.03	0.962
1.04	0.985
1.05	0.994
1.06	0.996
1.07	0.998
1.08	1.000

simulations.

For each of the conditions, we computed the following statistics: bias (difference between the mean estimated parameter and the true parameter), efficiency (standard deviation of the difference between the individual parameter estimates and the mean estimate) and the root mean square error (based on the squared difference between the individual estimates and the true parameters). They were computed on the difference (in the case of a single parameter) or the Euclidian distance (in the case where all three parameters are considered simultaneously as a vector). In either case, we have the relation

$$RMSE^2 = bias^2 + variance.$$

where *variance* is the square of efficiency illustrated in Figures 2 and 3

For practical considerations, the RMSE is the most relevant measure as it quantifies the ability of a method to estimate the parameters adequately: a method with significantly smaller RMSE is to be preferred whenever possible.

ACKNOWLEDGMENT

This research was supported in part by a research grant from the Conseil pour la Recherche en Sciences Naturelles et en Génie of Canada.

REFERENCES

- [1] M. Cacciari, G. Mazzanti and G. C. Montanari, "Comparison of maximum likelihood unbiasing methods for the estimation of the Weibull parameters". IEEE Trans. Dielectr. Electr. Insul., Vol. 3, pp. 18-27, 1996.
- [2] H. Hirose, "Maximum likelihood estimation in the 3-parameter Weibull distribution: A look through the Generalized Extreme-value distribution". IEEE Trans. Dielectr. Electr. Insul., Vol. 3, pp. 43-55, 1996.

- [3] J. Feng, "Generalization errors of the simple perceptron". *J. Phys. A: Mathematical and General*, Vol. 31, pp. 4037-4048, 1998.
- [4] L. de Haan, "Fighting the arch-enemy with mathematics", *Statistica Neerlandica*, Vol. 44, pp. 45-68, 1990.
- [5] D. Cousineau, V. Goodman, and R. M. Shiffrin, "Extending statistics of extremes to distributions varying on position and scale, and implication for race models", *J. Mathematical Psychology*, Vol. 46, pp. 431-454, 2002.
- [6] N. M. Kiefer, Maximum likelihood estimation (MLE), <http://instruct1.cit.cornell.edu/courses/econ620/reviewm5.pdf>, 2007.
- [7] R. L. Smith, "Maximum likelihood estimation in a class of nonregular cases", *Biometrika*, Vol. 72, pp. 67-90, 1985.
- [8] H. Rockette, C. Antle, and L. A. Klimko, "Maximum likelihood estimation with the Weibull model", *J. American Statistical Association*, Vol. 69, pp. 246-249, 1974.
- [9] D. Cousineau, S. Brown, and A. Heathcote, "Fitting distributions using maximum likelihood: Methods and packages", *Behavior Research Methods, Instruments, & Computers*, Vol. 36, pp. 742-756, 2004.
- [10] H. Hirose, "Bias correction for the maximum-likelihood estimates in the two-parameter Weibull distribution", *IEEE Trans. Dielectr. Electr. Insul.*, Vol. 6, pp. 66-68, 1999.
- [11] R. C. Cheng, and N. A. K. Amin, "Estimating parameters in continuous univariate distributions with a shifted origin", *J. Royal Statistical Society B*, Vol. 45, pp. 394-403, 1983.
- [12] D. Cousineau, "Nearly unbiased estimators for the three-parameter Weibull distribution with greater efficiency than the iterative likelihood method", *British J. Mathematical and Statistical Psychology*, in press.
- [13] H. L. Harter, and H. Moore, "Maximum likelihood estimation of the parameters of Gamma and Weibull populations from complete and from censored samples". *Technometrics*, Vol. 7, pp. 639-643, 1965.
- [14] A. Heathcote, S. Brown, and D. J. K. Mewhort, "Quantile maximum likelihood estimation of response time distributions", *Psychonomic Bulletin & Review*, Vol. 9, pp. 394-401, 2002.
- [15] A. Heathcote, S. Brown, and D. Cousineau, "QMPE: Estimating Lognormal, Wald and Weibull RT distributions with a parameter dependent lower bound", *Behavior Research Methods, Instruments, & Computers*, Vol. 36, pp. 277-290, 2004.
- [16] P. L. Speckman, and J. N. Rouder, "A comment on Heathcote, Brown and Mewhort's QMLE estimation method for response time distributions", *Psychonomic Bulletin & Review*, Vol. 11, pp. 574-576, 2004.
- [17] A. Heathcote, and S. Brown, "Reply to Speckman and Rouder: A theoretical basis for QML", *Psychonomic Bulletin & Review*, Vol. 11, pp. 577-578, 2004.
- [18] J. Jacquelin, "Generalization of the method of maximum likelihood", *IEEE Trans. Electr. Insul.*, Vol. 28, pp. 65-72, 1993.
- [19] E. Gourdin, P. Hansen, and B. Jaumard, "Finding maximum likelihood estimators for the three-parameter Weibull distribution", *J. Global Optimization*, Vol. 5, pp. 373-397, 1994.
- [20] D. Cousineau, "Implementing and evaluating the nested maximum likelihood estimation technique", *Tutorials in Quantitative Methods for Psychology*, Vol. 3, pp. 8-13, 2007.
- [21] R. A. Lockhart, and M. A. Stephens, "Estimation and tests of fit for the three-parameter Weibull distribution", *J. Royal Statistical Society B*, Vol. 56, pp. 491-500, 1994.
- [22] J. N. Rouder, D. Sun, P. L. Speckman, J. Lu, and D. Zhou, "A hierarchical bayesian statistical framework for response time distributions", *Psychometrika*, Vol. 68, pp. 589-606, 2003.
- [23] D. Cousineau, and S. Hélie, "Improving maximum likelihood estimation using prior probabilities", *Journal of Mathematical Psychology*, submitted for publication, 2008.
- [24] C. Rose, and M. D. Smith, *Mathematical Statistics with Mathematica*, New York: Springer-Verlag, pp. 253-256, 2001.
- [25] J. N. Rouder, J. Lu, P. L. Speckman, D. Sun, and Y. Jiang, "A hierarchical model for estimating response time distributions", *Psychonomic Bulletin & Review*, Vol. 12, pp. 195-223, 2005.
- [26] H. Hirose and T. L. Lai, "Inference from grouped data in three-parameter Weibull models with applications to breakdown-voltage experiments", *Technometrics*, Vol. 39, pp. 199-210, 1997.
- [27] J. A. Nelder, and R. Mead, "A simplex method for function minimization", *The Computer J.*, Vol. 7, pp. 308-313, 1965.



Denis Cousineau is professor at the Université de Montréal since 2000. He previously was post-doctoral fellow at Indiana University. He holds B.Sc. in computer science, in education and in psychology. His areas of research include response time models, attention and visual search. Starting in 2002, he organizes yearly the Summer School in Advanced Methodological Methods and he founded in 2005 the *Tutorials in Quantitative Methods for Psychology* journal. He just published his first book (in French): *Statistiques: Concepts et applications* with Robert Haccoun.