Maximum Likelihood Parameter Estimation in the Extended Weibull Distribution and its Applications to Breakdown Voltage Estimation

Hideo Hirose
Faculty of Computer Science and Systems Engineering
Department of Control and Engineering Science
Kyushu Institute of Technology
Iizuka, Fukuoka 820-8502, Japan

ABSTRACT

Although the Weibull distribution is widely used in a variety of reliability applications, difficulties in its treatment, particularly in three parameter cases in the maximum likelihood estimation, hinder us from using the distribution. The extended Weibull distribution proposed by Marshall and Olkin (1997) can avoid the difficulties which appear in the conventional Weibull distribution models. This paper shows the maximum likelihood estimation method in the extended Weibull distribution model. The paper also illustrates some typical applications for breakdown voltage estimation in which the extended models are superior to the conventional Weibull models. The central discussion is whether the shape parameters in the extended model accomplish the mass shifting effect of the distribution.

1 INTRODUCTION

PARAMETER estimation in the Weibull distribution is widely used in a variety of reliability applications, e.g., in civil engineering, mechanical engineering, electrical engineering, and medical and pharmaceutical fields. However, in three parameter models in particular, it is known that computing the maximum likelihood estimates (MLEs) is extremely difficult. This difficulty originates with the inclusion of the location parameter γ_w of the distribution when a conventional three-parameter Weibull cumulative distribution (W3P) is expressed by

$$F_{W3P} = 1 - \exp\left\{ -\left(\frac{x-\gamma_w}{\eta_w} \right)^{\beta_w} \right\}, \ (\eta_w > 0, \beta_w > 0, x \geq \gamma_w). \tag{1}$$

Then, the likelihood function $L_{W3P}(\eta_w, \beta_w, \gamma_w)$, in complete continuous data cases,

$$L_{W3P}(\eta_w, \beta_w, \gamma_w) = \prod_{i=1}^{n} \frac{\beta_w}{\eta_w} \left(\frac{x_i - \gamma_w}{\eta_w} \right)^{\beta_w - 1} \exp\left\{ -\left(\frac{x_i - \gamma_w}{\eta_w} \right)^{\beta_w} \right\}, \tag{2}$$

is unbounded for $\beta_w < 1$, where η_w and β_w are the scale and shape parameters, respectively and n is the sample size. In the case $\beta_w \geq 1$, the likelihood equation $\nabla \log L_{W3P} = 0$ need not have a solution and the global maximum of $L_{W3P}(\eta_w, \beta_w, \gamma_w)$ may occur on the boundary $\beta_w = 1$ [1]. The case $\beta_w \leq 2$ corresponds to the so-called non-regular model, under which the asymptotic joint distribution of the MLEs is not normal and the asymptotic variances are not inversely proportional to the sample size [2]. Moreover, an estimate of β_w goes to infinity in some data case, which corresponds to the case of Gumbel or Fréchet distribution [3]. This is called the divergent case here. These two (non-regular and divergent problems) are fundamental problems which occur in the W3P model. Thus, such issues lead us to use alternative methods; using the moment estimators [4], applying the Bayesian method [5], and extending the Weibull model to a generalized extreme-value distribution model [6] are among them. However, properties of the MLE, e.g., asymptotic efficiency and etc., are still attractive.

Marshall and Olkin have recently proposed a new distribution family which has the property of geometric-extreme stability which means that the minimum or the maximum of a geometric number of independent random variables with common distribution in the family has a distribution again in the family [7]. They indicate that the particular case that the distribution is an exponential distribution yields a new two-parameter family of distributions which may sometimes be a competitor to the Weibull and gamma families. Similarly, a two-parameter Weibull model yields a new three-parameter distribution which

Manuscript received on 14 March 2001, in final form 25 April 2002.

may be a competitor to the conventional Weibull models. We abbreviate these extended distributions for the exponential and Weibull models as the EE2P and EW3P, respectively, but we mainly deal with the case of EW3P in this paper.

In some data cases which appear in electrical engineering, e.g., breakdown voltage data, the extended models fit quite well to the actual data, and therefore they may provide smaller confidence intervals for certain percentile point estimators; electrical engineers often want to know the reliability of lower percentile points. This tendency appears in both the EE2P and EW3P for breakdown voltage data. More important point in the maximum likelihood parameter estimation, in three parameter cases, is the exclusion of non-regular cases caused by inclusion of the location parameter. The problem of the blowup of the likelihood function in the W3P model will vanish in the EW3P model.

Since computing the MLEs of the extended models is not straightforward, a method to estimate the MLEs using the predictor-corrector method is briefly introduced.

2 EXTENDED MODELS FOR THE EXPONENTIAL AND WEIBULL DISTRIBUTIONS

If a survival function corresponding to a cumulative distribution function $F(x;\theta)$ is denoted by $\bar{F}(x;\theta)$, a new family of distributions G can be generated as,

$$G(x;\alpha,\theta)=1-\frac{\alpha\bar{F}(x;\theta)}{1-\bar{\alpha}\bar{F}(x;\theta)}, \tag{3}$$

where $\bar{\alpha}=1-\alpha(\alpha>0)$ and θ is the parameter vector in the distribution F [7]. Then, the density function of the extended model is

$$g(x;\alpha,\theta)=\frac{\alpha f(x;\theta)}{\left\{1-\bar{\alpha}\bar{F}(x;\theta)\right\}^2} \tag{4}$$

where $f(x;\theta)$ is the density of $F(x;\theta)$.

2.1 EXTENDED TWO-PARAMETER EXPONENTIAL (EE2P)

If the cumulative distribution function for the exponential model is expressed by

$$F(x;\eta_e)=1-\exp\left(-\frac{x}{\eta_e}\right),\ (\eta_e>0), \tag{5}$$

then the extended cumulative distribution function and the density function are

$$G(x;\alpha,\eta_e)=1-\frac{\alpha\nu}{1-(1-\alpha)\nu}, \tag{6}$$

and

$$g(x;\alpha,\eta_e)=\frac{\alpha\nu}{\eta_e\{1-(1-\alpha)\nu\}^2}, \tag{7}$$

where $\nu=\exp(-y)$, $y=x/\eta_e$; η_e is the scale parameter.

2.2 EXTENDED THREE-PARAMETER WEIBULL MODEL (EW3P)

The cumulative two-parameter Weibull distribution model can be obtained by setting $\gamma_w=0$ in (1), which is abridged by W2P here. Then, the extended cumulative distribution function and the density function for the W2P are

$$G(x;\alpha,\eta_e,\beta_e)=1-\frac{\alpha w}{1-(1-\alpha)w}, \tag{8}$$

$$g(x;\alpha,\eta_e,\beta_e)=\frac{\alpha\beta_e y^{\beta_e-1}w}{\eta_e\{1-(1-\alpha)w\}^2}, \tag{9}$$

where $w=\exp(-z)$, $z=y^{\beta_e}$, $y=x/\eta_e$; η_e and β_e are the scale and shape parameters.

2.3 IMPORTANT PROPERTIES OF THE EW3P

The problem in the W3P model that the log-likelihood becomes unbounded is caused by inclusion of the location parameter, and such a problem will vanish in the EW3P model if (9) is bounded to the above. This is easily obtained because

$$\{1-(1-\alpha)w_i\}^2>0,\ (\alpha>0,\ 0<w_i<1),$$
$$0<y_i^{\beta_e-1}<\infty,\ (x_i>0), \tag{10}$$

to each observed value x_i. If the regularity conditions [8] hold, we can obtain the MLEs and their confidence intervals by using the observed Fisher information matrix. Here, we use the log-likelihood function in the EW3P as:

$$\log L(\alpha,\eta_e,\beta_e)=\sum_{i=1}^{n}\log g(x_i;\alpha,\eta_e,\beta_e). \tag{11}$$

We have not experienced the divergent cases so far in the extended models in simulation studies, and thus we do not believe that the divergent problem is serious in using the extended models.

2.4 SHAPES OF THE DENSITY FUNCTIONS OF THE EXTENDED MODELS

Typical shapes of the density functions of the EW3P are shown in Figure 1. We can find a variety of shape patterns of the density functions, which suggests the appropriateness of fit to actual data cases with a small number of free parameters. Figure 2 shows a modified Pear-

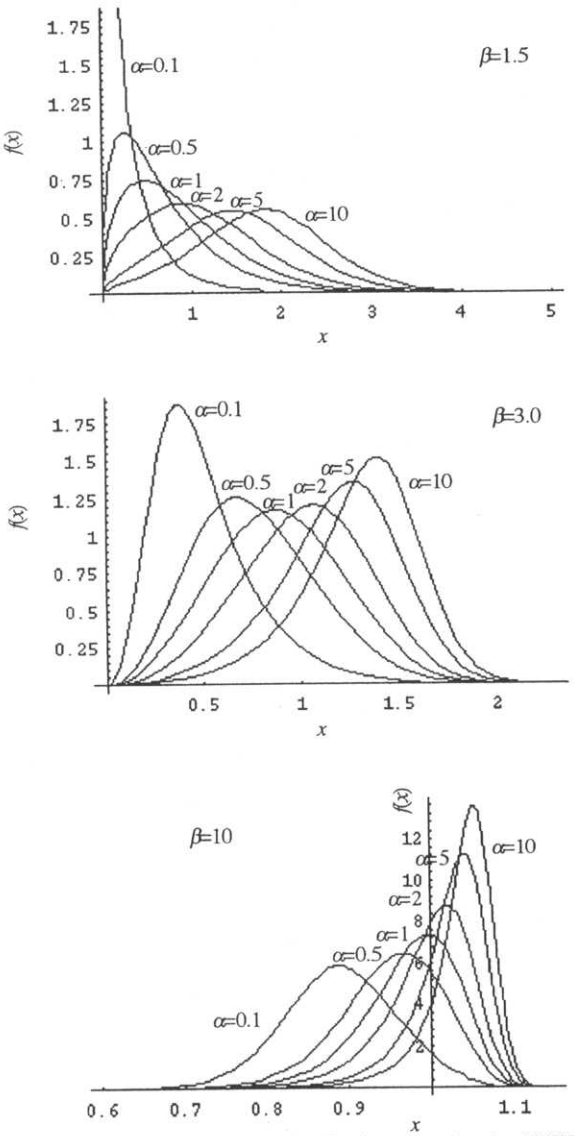


Figure 1. Typical shapes of the density functions for the EW3P.

son's diagram where the ordinate and abscissa express the skewness (*skw*) and kurtosis (*krt*), respectively, where

$$skw = \mu_3/\mu_2^{3/2},$$

$$krt = \mu_4/\mu_2^2, \tag{12}$$

$$\mu_2 = \mu_2' - (\mu_1')^2,$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3,$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4,$$

$$\mu_j' = \int_0^\infty x^j g(x)dx \ (g(x): \text{density}, j = 1,2,3,4). \tag{13}$$

The figure reveals a wider applicability of the EW3P to the field data than that of the W3P.

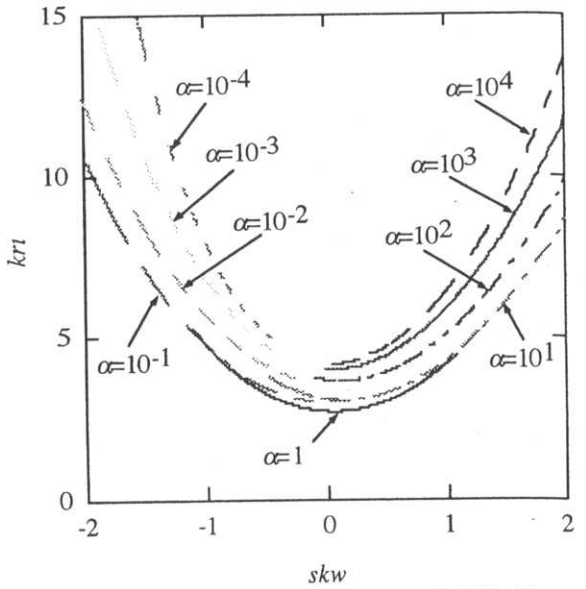


Figure 2. Modified Pearson's diagram for the EW3P. The curve when $\alpha = 1$ expresses the W3P.

In the EE2P, it is interesting that the larger the shape parameter α, the more the distribution moves to the right; the shape parameter α works as if it were a location parameter (or a shift parameter). This is because we can obtain

$$\mu_{\text{EE2P}} = -\frac{\eta_e \alpha \log \alpha}{(1-\alpha)},$$

$$\approx \eta_e \log \alpha \ (\alpha \text{ is large}), \tag{14}$$

and

$$\sigma_{\text{EE2P}} \to \text{constant} \ (\alpha \text{ is large}), \tag{15}$$

where μ and σ denote the mean and standard deviation, respectively; (14) is obtained analytically but (15) is computed by numerical integration. This phenomenon is called here *the mass shifting effect* of the distribution. Figure 3

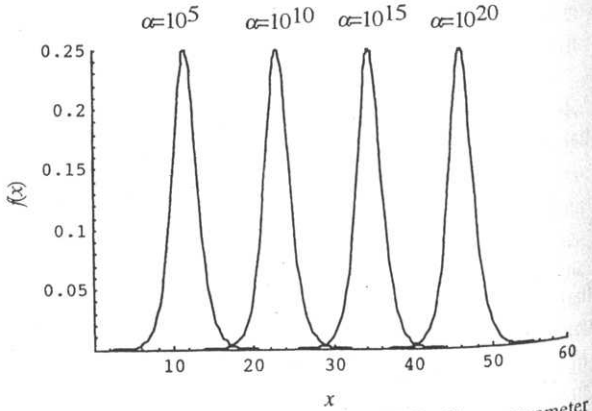


Figure 3. Typical density functions for the EE2P. Shape parameter α varies exponentially with $\eta_e = 1$.

shows typical density functions when α varies exponentially with the fixed scale parameter $\eta_e = 1$.

In the case of EW3P, this phenomenon is not exactly the same; however, the mass shifting effect of the distribution in the W3P model by the location parameter γ_w can roughly be realized by using the two shape parameters and scale parameter in the EW3P model as shown later.

3 APPLICATIONS TO BREAKDOWN VOLTAGE ESTIMATION

Typical examples to breakdown voltage estimations are introduced here to explain how applicable the extended models are. In 3.1, an example of the mass shifting effect in 2 parameter case (EE2P) is demonstrated, and the EE2P result is compared with the W2P result. In 3.2, EW3P fitting of five data sets are shown, which suggests a typical shape pattern observed in breakdown voltage. Since the data case is the same as in 3.1, this subsection extends the discussion in 3.1 to the three-parameter cases. In 3.3, a non-regular case in W3P is used, comparing the results of the W3P with those of the EW3P. In 3.4, a divergent case in W3P is examined in the EW3P. The parameter estimation method is given in Appendix 6.1.

3.1 EPOXY RESIN BREAKDOWN VOLTAGES A

A data case of 100 dielectric breakdown voltages of epoxy resin test pieces in [9] is used to investigate the two-parameter models: EE2P and W2P. Figure 4 shows the histogram of the data; the distribution is easily seen to be shifted away from the origin, which suggests the use of the mass shifting effect of the EE2P. Table 1 shows the fitted MLEs of the parameters and percentile points x_p along with the log-likelihood values, where

$$x_p = F^{-1}(p), \tag{16}$$

denotes the breakdown voltage with breakdown probability p. The large difference of the two log-likelihood values (larger than two, which corresponds to two additional free parameter inclusion in a model [10]),

$$\log L_{\mathrm{EE2P}} - \log L_{\mathrm{W2P}} = -157.040 - (-159.126) = 2.086 \tag{17}$$

suggests that the EE2P model is a better fit to the data than the W2P model. Although the mass of the distribution in the W2P can be shifted away from the origin by using the larger value of β_w, the shape of the W2P tends to that of the Gumbel distribution, i.e. skewed to the left. The histogram of the data reminds us the large value of kurtosis and symmetric distribution to the mean value, which suggests that the EE2P model is a better fit.

The standard errors using the observed Fisher information matrix shown in parentheses in Table 1 also suggest

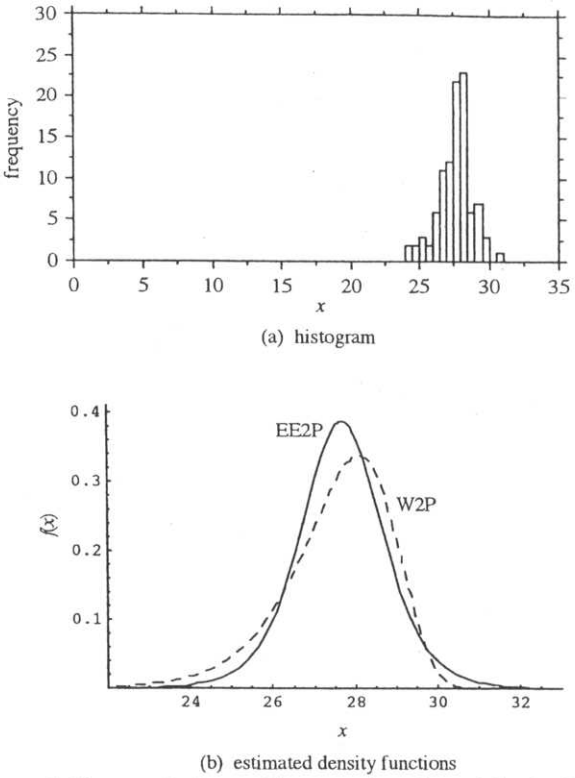


Figure 4. Epoxy resin data. a, histogram; b, estimated density functions.

Table 1. MLEs of parameters and percentile points for epoxy resin breakdown voltages A (100 sampled data).

Model	Estimates						
	α	η	β	$x_{.005}$	$x_{.01}$	$x_{.05}$	$\log L$
EE2P	4.609×10^{18}	.6441		24.27	24.72	25.78	−157.040
	(1.683×10^{19})	(.05453)		(.314)	(.279)	(.200)	
				[.353]	[.316]	[.230]	
W2P		28.14	25.80	22.92	23.54	25.08	−159.126
		(.1152)	(1.850)	(.378)	(.344)	(.259)	
				[.515]	[.458]	[.315]	

Standard errors using the observed Fisher information matrix are in parentheses.
Standard errors using the bootstrap method are in brackets.

the superiority of the EE2P over the W2P to this data case. However, the approximate standard errors shown in Table 1 are derived from each population model, e.g., EE2P errors are obtained from EE2P distribution; we cannot straightforwardly compare their values to each other. Therefore, a bootstrap method which mimics the non-parametric bootstrap [11] is used here to compare the standard errors of the percentile points; 100 data are sampled from the original 100 data with equal probability (.01) with replacement for 1000 trails, and estimates are obtained from each distribution model. In Table 1, the standard errors using the bootstrap are shown in brackets, and the results again support the superiority of the EE2P over the W2P. Figure 4 shows the density functions of the EE2P and W2P models fitted to the data.

This is only just one case comparison. It should be mentioned that we cannot simply address the superiority of the EE2P model over the W2P model. However, typical data cases observed in electrical breakdown voltage tests reveal that such a phenomenon is not rare, as will be shown later. It is worth noting that a remarkable difference between the two log-likelihood values (with the same number of free parameters) is not so common.

3.2 EPOXY RESIN BREAKDOWN VOLTAGES B

Five data sets of twenty samples [9] obtained by the epoxy resin insulation breakdown voltage tests are used here. First, we look at the shapes of the empirical distributions. Table 2 shows the mean, standard deviation, coefficient of variation, skewness, and kurtosis for five data sets as well as for 100 samples of all data sets. Typically, the table shows that the values of coefficient of variation (ratio of standard deviation to mean) are located around 0.05 and that distributions are negatively skewed and sharp. Table 3 shows the likelihood values fitted by the EW3P and W3P models. The table indicates that the extended model shows superior fit to the data.

Figure 5 shows the cumulative distribution function (cdf) for 100 samples with the superimposition of the empirical distribution. It appears that the cdfs of the extended models are closer to the empirical distribution. The values of

Table 2. Shape of the empirical distribution for epoxy resin breakdown voltages.

Case	Mean	s.d.	*cv*	*skw*	*krt*
1	27.27	1.205	0.04417	−1.032	3.294
2	27.80	1.462	0.05260	−0.2488	3.210
3	27.57	1.057	0.03833	−0.3111	2.580
4	27.51	1.174	0.04269	−1.501	5.075
5	27.85	1.057	0.03795	−0.2377	2.830
all	27.60	1.194	0.04327	−0.6040	3.845

Table 3. Log-likelihood values of EW3P and W3P fitted to epoxy resin breakdown voltages.

Case	EW3P	W3P
1	−28.8501*	> −28.9799
2	−35.2909	> −35.3747
3	−28.7608	< −28.6517
4	−26.9227*	> −27.6132
5	−28.7234	> −28.8236
all	−155.096	> −157.073

*Limiting values are used because of the divergent cases.

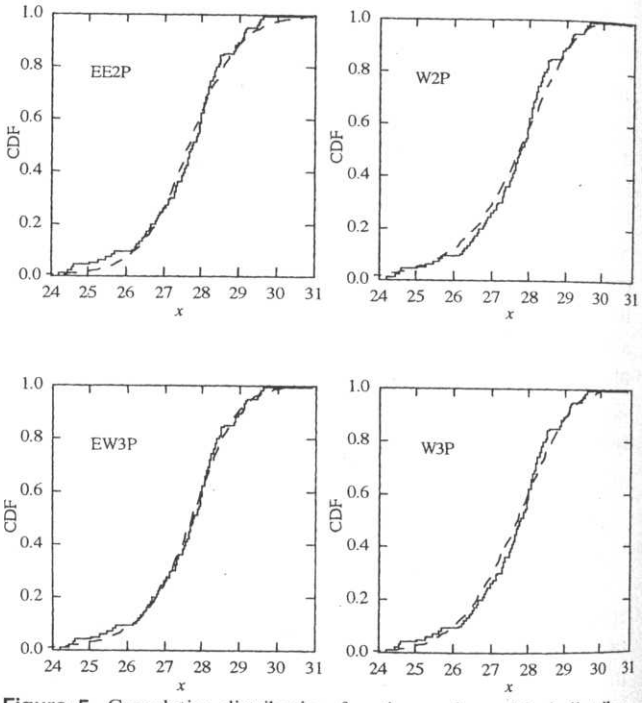


Figure 5. Cumulative distribution function and empirical distribution function. Solid line, empirical distribution; dashed line, CDF.

the Kolmogorov statistic [11], D_n, defined by

$$D_n = \sup_{x \geq 0} \left| \hat{F}_n(x) - F(x) \right| \tag{18}$$

for EW3P and W3P as well as EE2P and W2P are 0.049, 0.072, 0.058, 0.091, respectively, where $\hat{F}_n$ expresses the empirical distribution. Since the critical value D_{100} for the significance level 0.05 for 100 samples is 0.134, we cannot reject all the hypothetical models, but we should note that the values of D_{100} in the extended models are smaller than those in the W3P or W2P.

3.3 OIL INSULATION BREAKDOWN VOLTAGES

Table 4 represents a data set of 20 breakdown voltages when increasing voltages were applied to transformer oil in a test tank at the Hitachi Research Laboratory until insulation broke down. This corresponds to Test #4 in

Table 4. Oil insulation breakdown voltages.

Data Case	Data									
4	3.0	4.2	3.8	3.8	3.3	2.7	2.7	3.3	3.2	3.5
	3.5	2.6	3.1	3.4	2.8	2.9	3.1	3.1	3.6	3.4

(×100 kV)

Table 1 in [13]. Table 5 shows the fitted MLEs of the parameters, percentile points x_p and the log-likelihood values. As seen in Table 5, this data case corresponds to the non-regular case in the W3P model (since the estimate of the shape parameter is less than two in the W3P model, the asymptotic joint distribution of the MLEs is not normal and the asymptotic variances are not inversely proportional to the sample size); we cannot obtain the standard errors by using the observed Fisher information matrix if we adopt the W3P model as the underlying distribution. Provided that the regularity conditions hold in the EW3P, the standard errors using the observed Fisher information matrix are shown in the Table.

Although the log-likelihood value in the EW3P is slightly smaller than that in the W3P, standard errors in the EW3P obtained by using the mimicked non-parametric bootstrap show superiority over those in the W3P when percentile points are $x_{.005}$ and $x_{.01}$.

The values of the Kolmogorov statistic for the EW3P and W3P are 0.083 and 0.088 respectively, which means that the EW3P is a little bit closer to the empirical distribution. Both the models are not rejected because the critical value of D_{20} is 0.294.

3.4 EPOXY RESIN BREAKDOWN VOLTAGES C

Table 6 shows the fitted MLEs of the parameters, percentile points x_p, and the log-likelihood values to the data case 1 in Table 1 in [9]. As is explained in [9], this data case corresponds to the divergent case in the W3P model; we can neither obtain the estimates of the parameters nor their standard errors by using the observed Fisher information matrix if we adopt the W3P model as the underlying distribution. Only the percentile points can be obtained by adopting the extended maximum likelihood estimates [14]. Provided that the regularity conditions hold in the EW3P model, the standard errors using the observed Fisher information matrix are computed as shown in the table. The log-likelihood value in the EW3P is slightly larger than that in the W3P; the standard errors in the EW3P obtained by using the mimicked non-parametric bootstrap show superiority over those in the W3P; note that the extended MLEs are used in this bootstrap procedure in the divergent cases in the W3P model.

The values of the Kolmogorov statistic for the EW3P and the limiting distribution for the W3P are 0.088 and 0.126 respectively.

Table 5. MLEs of parameters and percentile points for oil insulation breakdown voltages.

Model	Estimates							
	α	η	β	γ	$x_{.005}$	$x_{.01}$	$x_{.05}$	$\log L$
EW3P	.05039	4.064	13.11		2.160	2.279	2.584	−10.2972
	(.1109)	(.5193)	(.2940)		(.196)	(.181)	(.141)	
					[.148]	[.138]	[.114]	
W3P		.8351	1.877	2.507	2.557	2.579	2.679	−9.45101
		—	—	—	—	—	—	
					[.200]	[.160]	[.093]	

Standard errors using the observed Fisher information matrix are in parentheses.
Standard errors using the bootstrap method are in brackets.

Table 6. MLEs of parameters and percentile points for epoxy resin breakdown voltages B (20 sampled data).

Model	Estimates							
	α	η	β	γ	$x_{.005}$	$x_{.01}$	$x_{.05}$	$\log L$
EW3P	3.314	27.12	24.98		23.01	23.66	25.21	−28.8501
	(5.825)	(1.130)	(11.44)		(1.29)	(1.06)	(.599)	
					[.908]	[.815]	[.598]	
W3P		—	—		23.30	23.89	25.27	−28.9799
		—	—	—	—	—	—	
					[1.07]	[.930]	[.625]	

Standard errors using the observed Fisher information matrix are in parentheses.
Standard errors using the bootstrap method are in brackets.

4 HOW EW3P COMPETE WITH W3P

As seen in the previous section, there are cases that the EW3P fits better to the breakdown voltage data than the W3P does from a view-point of the log-likelihood, confidence interval, or the Kolmogorov statistic. In this section, we show when and how the EW3P behaves better than the W3P does, and discuss whether the EW3P can be a competitor for the W3P. In the discussion, we limit the application field to the breakdown voltage phenomena. By experience in breakdown voltage tests, we know *a priori* that the coefficient of variation (cv) of the breakdown voltage data is typically around 0.05 (more specifically, from 0.03 to 0.1); e.g. see Table 2. Thus, we first assume that the cv is 0.05.

The central discussion here is whether the shape parameters in the extended model accomplish the mass shifting effect of the distribution, because the *raison d'être* of the W3P is that it has the positive endpoint to the left in engineering applications, while the EW3P does not.

If the cv is fixed and the endpoint γ_w is determined, then the shape of the W3P can completely be established (i.e., β_w is defined uniquely) by solving the equation,

$$cv_w = \frac{\sigma_w}{\mu_w} = \frac{\eta_w \cdot \left[\Gamma(1+2/\beta_w) - \{\Gamma(1+1/\beta_w)\}^2\right]^{1/2}}{\gamma_w + \eta_w \cdot \Gamma(1+1/\beta_w)}$$

$$= \frac{\left[\Gamma(1+2/\beta_w) - \{\Gamma(1+1/\beta_w)\}^2\right]^{1/2}}{\gamma_w/\eta_w + \Gamma(1+1/\beta_w)}. \quad (19)$$

Thus, β_w is a function of γ_w/η_w. We can determine the shape of the distribution as:

$$\gamma_w/\eta_w \Rightarrow (\gamma_w/\eta_w, cv_w) \Rightarrow \beta_w \Rightarrow (\gamma_w/\eta_w, \beta_w) \Rightarrow \mu_w/\eta_w, \sigma_w/\eta_w.$$

Similarly, the shape of the EW3P can also be confirmed by solving the simultaneous equations,

$$cv_e = \frac{\sigma_e}{\mu_e} = \frac{\sigma_w}{\mu_w} = cv_w,$$

$$\mu_e = \mu_w, \text{ (or equivalently, } \sigma_e = \sigma_w) \quad (20)$$

where μ_e and σ_e are obtained numerically as

$$\mu_e = \int_0^\infty x g(x)\,dx,$$

$$\sigma_e = \left\{ \int_0^\infty x^2 g(x)\,dx - \left(\int_0^\infty x g(x)\,dx \right)^2 \right\}^{1/2}. \quad (21)$$

Then, α and β_e are a function of γ_w/η_w. We can determine the shape of the distribution as:

$$\gamma_w/\eta_w \Rightarrow (\mu_w/\eta_w, \sigma_w/\eta_w) \Rightarrow \mu_e, \sigma_e \Rightarrow \alpha, \beta_e.$$

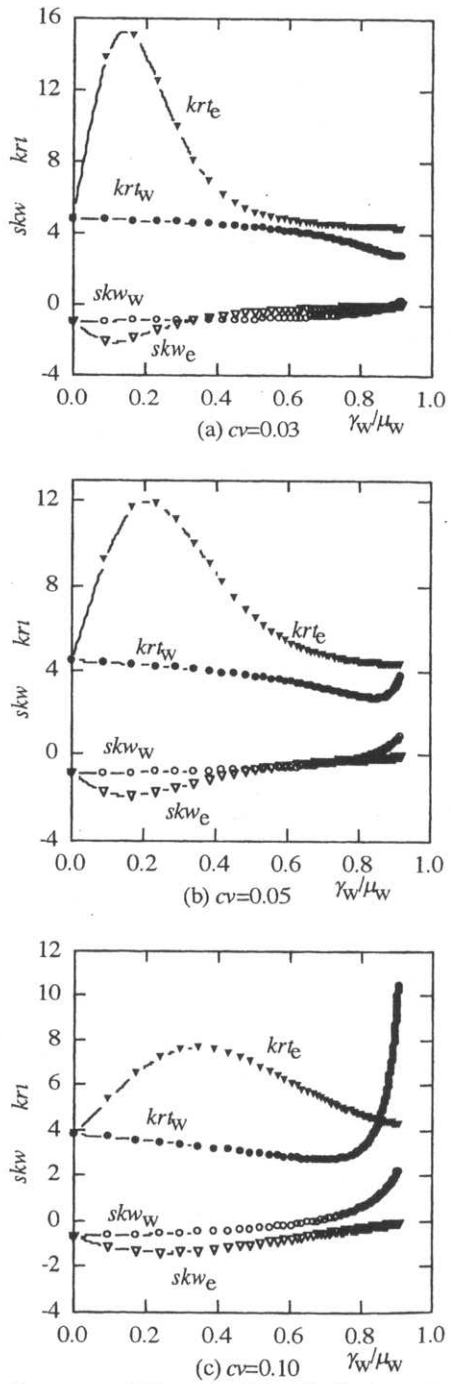


Figure 6. Skewness and Kurtosis of the distributions. A, $cv = 0.03$; b, $cv = 0.05$; c, $cv = 0.10$.

Since the solutions of equations (19) and (20) cannot be expressed in closed forms, they have to be solved numerically. A method to find the solutions is shown in Appendix 6.2.

Suppose that $cv = 0.05$. Figure 6 shows the *skw* and *krt* (equation (12)) for the W3P and EW3P parameterized by γ_w/μ_w (γ_w/μ_w becomes a function of γ_w/η_w). Figure 7

shows the probability of $\int_0^{\gamma_w} g(x)dx$ which is expected to be very small.

At a point of γ_w/μ_w where *skw* and *krt* in the W3P differ most from those in the EW3P, the shapes of the two distributions can be considered to differ most from each other. This point is $\gamma_w/\mu_w = 0.211$ ($\gamma_w/\eta_w = 0.26$), and

$$skw_w = -0.862, \quad krt_w = 4.25,$$
$$skw_e = -1.84, \; krt_e = 12.0 \qquad (22)$$

The density functions at the point are shown in Figure 8; at first glance we cannot distinguish the difference between the two distributions around the endpoint of the W3P. Actually, the probability of $\int_0^{\gamma_w} g(x)dx$ is very small as Figure 8 indicates; numerical integration tells us that it is 1.18×10^{-5}. The EW3P seems to accomplish the mass shifting effect of the distribution by the two shape parameters in the EW3P as the location parameter in the W3P shifts the distribution. We will see if this is true by using Monte Carlo simulations from two viewpoints: the likelihood value and the percentile point estimator.

4.1 COMPARISON BY LIKELIHOOD

Preserving the *cv* value as 0.05, we use two kinds of distribution function as a mother distribution for random number generation; one is the EW3P ($\alpha_e = 225$, $\eta_e = 1$, $\beta_e = 7.82$) and the other is the W3P ($\eta_w = 1$, $\beta_w = 19.5$, $\gamma_w = 0.211$); these parameter values are determined when the shapes of the two distributions differ from each other most, i.e., $\gamma_w/\eta_w = 0.211$. Then, we fit the distribution functions of the EW3P and W3P to the generated data. The sample size is 100 and the number of trials is 1000.

Figure 9 shows the difference of the two likelihood values between the EW3P and the W3P. The random numbers are generated from the EW3P and from the W3P. Figure 9(a) suggests the high superiority of the EW3P over the W3P, while Figure 9(b) indicates that the EW3P is much the same to the W3P. In short, the EW3P can be a formidable competitor to the W3P.

4.2 COMPARISON BY PERCENTILE POINT

The major flaw of the W3P in maximum likelihood estimation occurs when the location parameter estimator is used: 1. non-regular problem may prohibit to exist the maximum likelihood estimates in some cases, 2. the location parameter estimate may go to minus infinity, 3. the standard deviation of the estimate of the location parameter is usually very large (it may include negative value). Thus, Hirose and Lai [3] proposed to use the low percentile point as a substitution of the endpoint of the distribution. Therefore, using the low percentile point in the EW3P also seems to be natural.

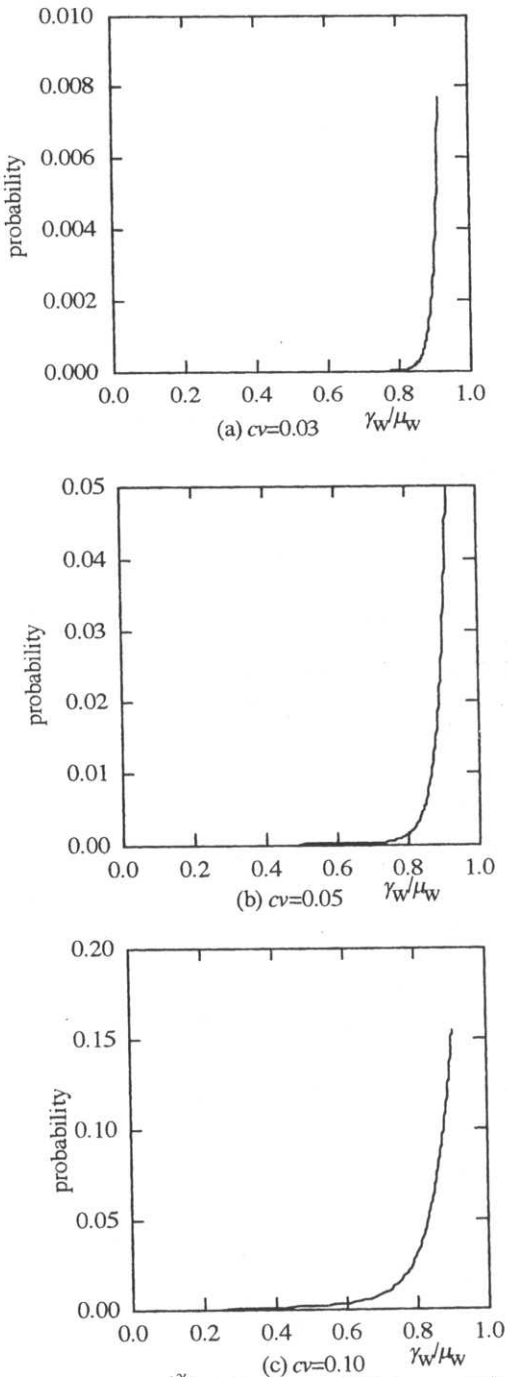


Figure 7. Probability $\int_0^{\gamma_w} g(x)dx$. a, $cv = 0.03$; b, $cv = 0.05$; c, $cv = 0.10$.

Using the same random numbers generated in 4.1, we compared the estimates of the low percentile points, $x_{0.01}$ and $x_{0.005}$, in the EW3P, along with those in the W3P. Figure 10, the scatter points of these estimates by each distribution fitting, suggests that there is no remarkable difference of the estimates between the EW3P and W3P. The bias and RMSE (root mean square error) are shown in Table 7, which confirms that the EW3P is much the

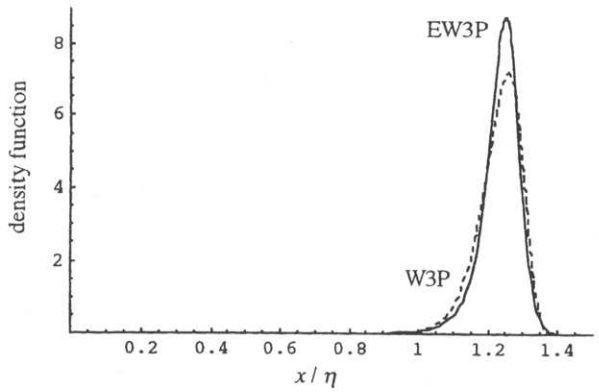
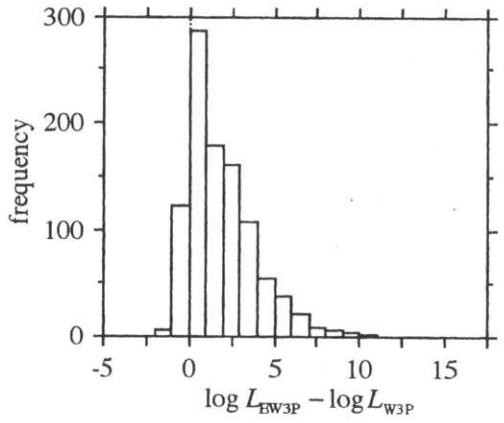


Figure 8. Density functions when $cv = 0.05$.

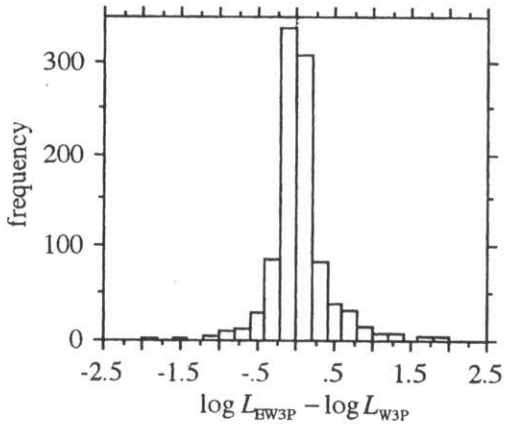
same to the W3P. In short, the EW3P can be a competitor to the W3P.

4.3 OTHER CASE STUDIES

We have been discussing so far the difference of the two distributions only on the case that cv is 0.05 and γ_w/η_w is 0.211 where the shapes of the two distributions differ from each other most. The properties obtained in that case are also almost true in other cases as far as $\gamma_w/\mu_w \leq 0.5$. If we assume that the value of cv varies from 0.03 to 0.1 in electrical insulation breakdown voltage data as stated before, then $(\mu - 5\sigma)/\mu$ which would cover almost all the breakdown voltages varies from 0.5 to 0.85. Even if the condition is worst for the EW3P (this would be the case that $cv = 0.1$ and $\gamma_w/\mu_w = 0.5$) in percentile point estimations, the results of the comparison are much the same as discussed in subsections 4.1 and 4.2. From Figures 6 and 7, we can easily imagine that the EW3P behaves better than the W3P does even when $cv \leq 0.03$. Figure 11 and Table 8 show the simulation results when $cv = 0.1$ and $\gamma_w/\mu_w = 0.5$. As a result, the EW3P can be a competitor to the W3P as far as $cv \leq 0.1$.



(a) Random Numbers are Generated from the EW3P



(b) Random Numbers are Generated from the W3P

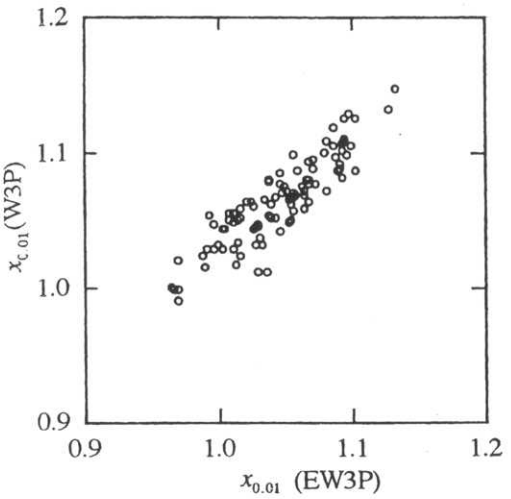
Figure 9. Difference of the two log-likelihood values between the EW3P and the W3P. $cv = 0.05$, $\gamma_w/\mu_w = 0.211$. a, Random numbers are generated from the EW3P; b, random numbers are generated from the W3P.

5 CONCLUDING REMARKS

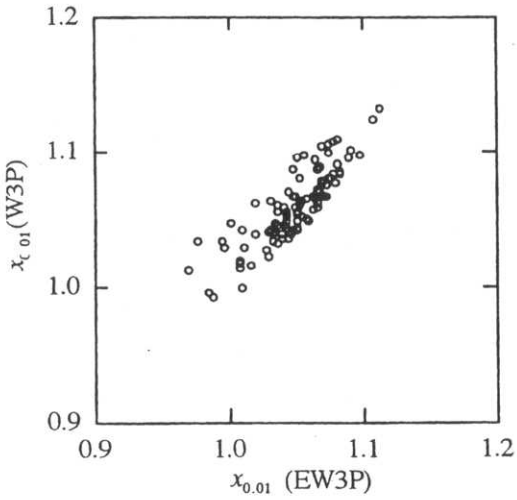
Marshall and Olkin predict that the extended exponential and Weibull distribution may be competitors to the family of Weibull and gamma models (see [7]). In electrical engineering fields, the extended model sometimes is a better fit to the actual breakdown voltage data than the conventional Weibull model in the sense of the likelihood and confidence intervals of the percentile point estimates. Since the three-parameter Weibull distribution has mainly

Table 7. Bias and RMSE of x_p fitted by the EW3P and W3P. $cv = 0.05$ and $\gamma_w/\mu_w = 0.211$ ($\gamma_w/\eta_w = 0.26$).

Random number is generated from the EW3P				
Distribution	p	x_p	Bias	RMSE
EW3P	0.01	1.022	0.01998	0.04264
	0.005	0.9650	0.03448	0.06053
W3P	0.01	1.022	0.03971	0.04943
	0.005	0.9650	0.07036	0.07912
Random number is generated from the W3P				
Distribution	p	x_p	Bias	RMSE
EW3P	0.01	1.050	−0.001008	0.02643
	0.005	1.023	−0.003625	0.03323
W3P	0.01	1.050	0.06102	0.02611
	0.005	1.023	0.07238	0.03244

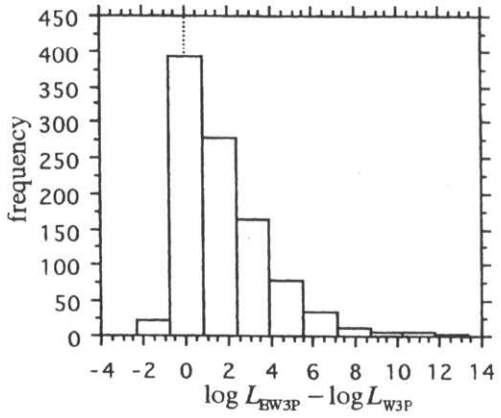


(a) Random Numbers are Generated from the EW3P

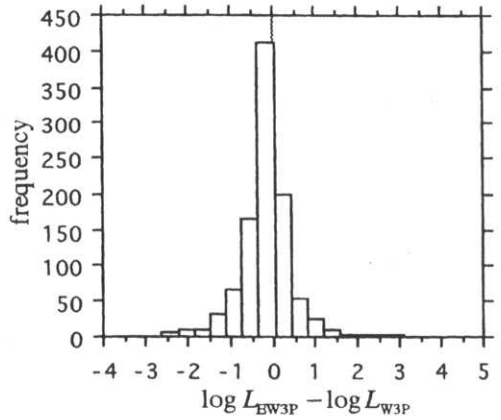


(b) Random Numbers are Generated from the W3P

Figure 10. Comparison of $x_{0.01}$ between the EW3P and W3P. cv = 0.05, $\gamma_w/\mu_w = 0.211$. a, random numbers are generated from the EW3P; b, random numbers are generated from the W3P.



(a) Random Numbers are Generated from the EW3P



(b) Random Numbers are Generated from the W3P

Figure 11. Difference of the two log-likelihood values between the EW3P and the W3P. $cv = 0.1$, $\gamma_w/\mu_w = 0.5$. a, random numbers are generated from the EW3P; b, random numbers are generated from the W3P.

two major problems due to the inclusion of the location parameter, many researchers propose alternative estimation methods to circumvent these problems. However, these problems may vanish if we adopt the extended three-parameter Weibull model. Thus, the Marshall and

Table 8. Bias and RMSE of x_p fitted by the EW3P and W3P. $cv = 0.1$ and $\gamma_w/\mu_w = 0.5$.

Random number is generated from the EW3P				
Distribution	p	x_p	Bias	RMSE
EW3P	0.01	1.271	0.03762	0.1089
	0.005	1.139	0.06670	0.1476
W3P	0.01	1.271	0.07027	0.1203
	0.005	1.139	0.1344	0.1793
Random number is generated from the W3P				
Distribution	p	x_p	Bias	RMSE
EW3P	0.01	1.373	−0.02207	0.05693
	0.005	1.322	−0.03911	0.07282
W3P	0.01	1.373	0.009688	0.06077
	0.005	1.322	0.01003	0.07430

Olkin's extended models are particularly useful under such conditions. This paper has discussed whether the shape parameters in the extended model accomplish the mass shifting effect of the distribution, and it has been positively clarified when the value of coefficient of variation is small.

6 APPENDIX

6.1 PARAMETER ESTIMATION METHOD

6.1.1 THE MAIN IDEA

In solving the likelihood equations, iterative methods are used because of the nonlinearity of the equations. The Newton-Raphson method and the continuation method [15–18] are among the iterative methods, and they work well in finding the solutions.

Figure 2 indicates that a solution of the likelihood equations may not be uniquely determined; for example, the curves when $\alpha = 0.1$ and $\alpha = 10$ have an intersection point around $skw = 0$. Actually, the multiple local maximum points are obtained in an epoxy resin breakdown voltage data case, and such a case is not rare. Therefore, we have to search for solutions carefully. Here, a search method using the profile log-likelihood is recommended. This method can be considered as one of the *predictor-corrector* methods; the central idea is similar to the literature [19].

6.1.2 ESTIMATION PROCEDURE

Step 1: Obtain the MLE of the W2P, $\hat{\eta}_{\mathrm{W2P}}$ and $\hat{\beta}_{\mathrm{W2P}}$. This solution is the case of $\alpha = 1$ in the extended model; we set $\hat{\eta}_0 = \hat{\eta}_{\mathrm{W2P}}$ and $\hat{\beta}_0 = \hat{\beta}_{\mathrm{W2P}}$. The solution is the starting point, $k = 0$.

Step 2: Define a multiplier $m_1 > 1$.

Step 3: Set $\alpha_{k+1} = m_1 \times \alpha_k$. Using $\hat{\eta}_k$ and $\hat{\beta}_k$ as a starting point, find the solution of the equations, (25) and (26), by the Newton method, and set $\hat{\eta}_{k+1} = \hat{\eta}$, $\hat{\beta}_{k+1} = \hat{\beta}$, and $\log L_{k+1} = \log L$. Here, $\hat{\eta}_k$ and $\hat{\beta}_k$ are considered to be simple predictors and $\hat{\eta}_{k+1}$ and $\hat{\beta}_{k+1}$ are correctors. Iterate this procedure until $\alpha_{k+1} \geq M_1$.

Step 4: Define a multiplier $0 < m_2 < 1$.

Step 5: Set $\alpha_{k-1} = m_2 \times \alpha_k$. Using $\hat{\eta}_k$ and $\hat{\beta}_k$ as a starting point, find the solution of the equations, (25) and (26), and set $\hat{\eta}_{k-1} = \hat{\eta}$, $\hat{\beta}_{k-1} = \hat{\beta}$, and $\log L_{k-1} = \log L$. Iterate this procedure until $\alpha_{k-1} \leq M_2$.

Step 6: Search k such that $\log L_k$ is maximum. Using α_k, $\hat{\eta}_k$ and $\hat{\beta}_k$ as a starting point, find the solution of the equations, (24)-(26), by the Newton method.

6.1.3 LIKELIHOOD EQUATIONS AND HESSIANS

The concrete formulae of the first and second derivatives of $\log L$ for the extended Weibull model are presented here for readers' convenience in programming.

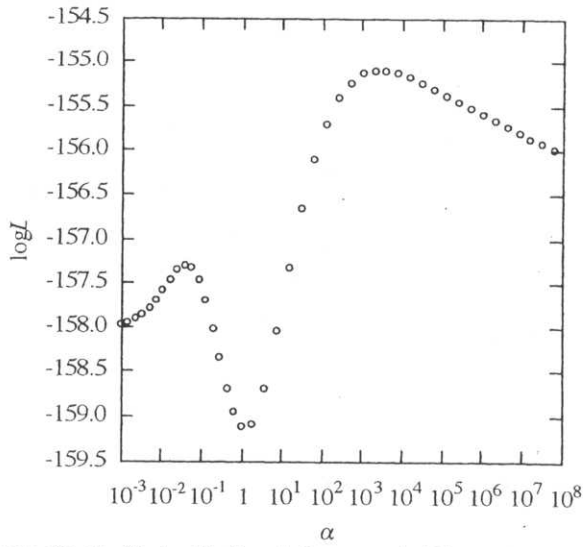


Figure 12. Profile log-likelihood. Data case in [9].

The log-likelihood function for the EW3P is

$$\log L = n(\log\alpha + \log\beta - \log\eta) + \sum_{i=1}^{n}\left[(\beta-1)\log y_i + \log w_i - 2\log\{1-(1-\alpha)w_i\}\right]. \quad (23)$$

The corresponding likelihood equations are

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - 2\sum_{i=1}^{n}\frac{w_i}{1-(1-\alpha)w_i} = 0, \quad (24)$$

$$\frac{\partial \log L}{\partial \eta} = \frac{n\beta}{\eta} - \sum_{i=1}^{n}\left\{1 - z_i\frac{1+(1-\alpha)w_i}{1-(1-\alpha)w_i}\right\} = 0, \quad (25)$$

$$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n}\log y_i\left\{1 - z_i\frac{1+(1-\alpha)w_i}{1-(1-\alpha)w_i}\right\} = 0. \quad (26)$$

The second derivatives of the log-likelihood function (23) are

$$\frac{\partial^2 \log L}{\partial^2 \alpha} = -\frac{n}{\alpha^2} + 2\sum_{i=1}^{n}\frac{w_i^2}{\{1-(1-\alpha)w_i\}^2}, \quad (27)$$

$$\frac{\partial^2 \log L}{\partial\eta\partial\alpha} = \frac{\partial^2 \log L}{\partial\alpha\partial\eta} = -\frac{2\beta}{\eta}\sum_{i=1}^{n}\frac{z_i w_i}{\{1-(1-\alpha)w_i\}^2}, \quad (28)$$

$$\frac{\partial^2 \log L}{\partial\beta\partial\alpha} = \frac{\partial^2 \log L}{\partial\alpha\partial\beta} = 2\sum_{i=1}^{n}\frac{z_i w_i \log y_i}{\{1-(1-\alpha)w_i\}^2}, \quad (29)$$

$$\frac{\partial^2 \log L}{\partial^2 \eta} = \frac{n\beta}{\eta^2} + \frac{\beta}{\eta^2}\sum_{i=1}^{n}\frac{z_i\{(1-\alpha)^2(1+\beta)w_i^2 + 2\beta(1-\alpha)z_i w_i - (1+\beta)\}}{\{1-(1-\alpha)w_i\}^2}, \quad (30)$$

$$\frac{\partial^2 \log L}{\partial\beta\partial\eta} = \frac{\partial^2 \log L}{\partial\eta\partial\beta} = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} \frac{z_i\left\{-(1-\alpha)^2(1+\beta\log y_i)w_i^2 - 2\beta(1-\alpha)z_i w_i \log y_i + (1+\beta\log y_i)\right\}}{\left\{1-(1-\alpha)w_i\right\}^2}, \quad (31)$$

$$\frac{\partial^2 \log L}{\partial^2\beta} = \frac{1}{\beta^2}\sum_{i=1}^{n} \frac{-(1-\alpha)^2\left\{1-\beta z_i(\log y_i)^2\right\}w_i^2 + 2(1-\alpha)\left\{1+\beta^2 z_i^2(\log y_i)^2\right\}w_i - 1 - \beta^2 z_i(\log y_i)^2}{\left\{1-(1-\alpha)w_i\right\}^2} \quad (32)$$

The first and second derivatives of $\log L$ for the EE2P model are treated as the special case in the EW3P model.

6.1.4 NUMERICAL EXAMPLE IN THE ESTIMATION

A numerical example is demonstrated here for 100 samples in [9]. The MLE of the W2P, $\hat{\eta}_{W2P} = 28.14$ and $\hat{\beta}_{W2P} = 25.80$, are obtained by an iterative method, e.g., by [14]. We define $m_1 = 2$. Then, $\hat{\eta}_1 = 27.63$ and $\hat{\beta}_1 = 22.12$ with $\alpha_1 = 2$ are obtained by the Newton-Raphson method. This procedure is continued until $\alpha_k > 5\times10^{11}$. Similarly, by defining $m_2 = 2/3$, $\hat{\eta}_{-1} = 28.41$ and $\hat{\beta}_{-1} = 28.02$ with $\alpha_{-1} = 2/3$ are obtained, and this procedure is continued until $\alpha_k < 10^{-5}$. Figure 12 shows the profile log-likelihood where local maximum values are seen around $\alpha = 0.04$ and 2000. Using α_{11}, $\hat{\eta}_{11}$, and $\hat{\beta}_{11}$ as a starting point, the solution can be found to be,

$$\hat{\alpha} = 2772,\ \hat{\eta} = 19.18,\ \hat{\beta} = 5.622, \quad (33)$$

with $\log L_{max} = -155.096$ by the Newton method. The estimate for a percentile point, $x_{0.01}$, is 23.80.

The observed information matrix for the EW3P is,

$$V = -\begin{bmatrix} \frac{\partial^2 \log L}{\partial\sigma^2} & \frac{\partial^2 \log L}{\partial\mu\partial\sigma} & \frac{\partial^2 \log L}{\partial k\partial\sigma} \\ \frac{\partial^2 \log L}{\partial\sigma\partial\mu} & \frac{\partial^2 \log L}{\partial\mu^2} & \frac{\partial^2 \log L}{\partial k\partial\mu} \\ \frac{\partial^2 \log L}{\partial\sigma\partial k} & \frac{\partial^2 \log L}{\partial\mu\partial k} & \frac{\partial^2 \log L}{\partial k^2} \end{bmatrix} = \begin{bmatrix} 4.376\times10^{-6} & 0.02832 & -0.03583 \\ 0.02832 & 195.3 & -254.0 \\ -0.03583 & -254.0 & 334.6 \end{bmatrix}, \quad (34)$$

and V^{-1} is,

$$V^{-1} = \begin{bmatrix} 1.045\times10^{8} & -49460 & -26360 \\ -49460 & 23.82 & 12.79 \\ -26360 & 12.79 & 6.893 \end{bmatrix}. \quad (35)$$

If we can use the asymptotic normality and the delta method, approximated 90% confidence intervals for $\hat{x}_{0.01}$ is computed as [22.31, 25.30].

6.2 A METHOD TO SOLVE THE NONLINEAR EQUATIONS

Since (20) is a system of nonlinear equations, we often use a Newton-type iterative method to solve the equations, such as:

$$X_{i+1} = X_i - J_i^{-1}F(X_i), \quad (i = 0,1, \ldots,), \quad (36)$$

where,

$$X = (x_1, x_2, \ldots, x_n)^T,$$
$$F = (f_1(X), f_2(X), \ldots, f_n(X))^T,$$
$$J = (\partial f_s(X)/\partial x_t). \quad (37)$$

However, the computation of the Jacobian is extremely difficult because $f_s(X)$ is expressed by integration forms as in (21) and they cannot be expressed explicitly. The computational difficulty for the quasi-Newton method is much the same. Therefore, a simple optimization method is used here to solve the system of nonlinear equations.

Since

$$f_1 = f_2 = \cdots = f_n = 0 \quad (38)$$

is equivalent to

$$f = f_1^2 + f_2^2 + \cdots + f_n^2 = 0, \quad (39)$$

We need to find the optimum (minimum) point of X such that $f(X) = 0$. The Nelder-Mead simplex method [21] is handy, and the solutions are obtained by examining only the function values.

REFERENCES

[1] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, New York: Wiley, 1982.

[2] S. D. Dubey, "Hyper-Efficient Estimator of the Location Parameter of the Weibull Laws", *Naval Research Logistic Quarterly*, Vol. 13, pp. 253–264, 1966.

[3] 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.

[4] A. C. Cohen and B. J. Whitten, *Parameter Estimation in Reliability and Life Span Models*, New York: Marcel Dekker, 1988.

[5] R. L. Smith and J. C. Naylor, "A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull distribution", Appl. Statistics, Vol. 36, pp. 358–369, 1987.

[6] R. C. H. Cheng and T. C. Iles, "Embedded Models in Three-Parameter Distributions and Their estimation", J. Royal Statistical Soc., Ser. B, Vol. 52, pp. 135–149, 1990.

[7] A. W. Marshall and I. Olkin, "A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families", Biometrika, Vol. 84, pp. 641–652, 1997.

[8] E. L. Lehmann, *Theory of Point Estimation*, New York, Wiley, 1983.

[9] H. Hirose, "Maximum Likelihood Estimation in the 3-Parameter Weibull Distribution: A Look Through the Generalized Extreme-Value Distribution", IEEE Trans. DEI, Vol. 3, pp. 43–55, 1996.

[10] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle", 2nd International Symposium on Information Theory, Akademiai Kiado, Budapest, pp. 267–281, 1973.

[11] B. Efron, "Bootstrap Methods: Another Look at the Jackknife", Annals of Statistics, Vol. 7, pp. 1–26, 1979.

[12] P. J. Bickel and K. Doksum, *Mathematical Statistics*, Prentice Hall. 1977.

[13] Y. Kako, "Estimation of Weibull Parameters for BDV Data of Transformer Oil by Most Likelihood Methods", Conference Internationale des Grands Reseaux Electriques, Working Group 15–01, Task Force 02, 1986.

[14] H. Hirose, "Percentile Point Estimation in the Three Parameter Weibull Distribution by the Extended Maximum Likelihood Estimates", Computational Statistics and Data Analysis, Vol. 11, pp. 309–331, 1991.

[15] A. C. Cohen, "Maximum Likelihood Estimation in The Weibull Distribution Based on Complete and Censored Samples", Technometrics, Vol. 7, pp. 579–588 (1965)

[16] E. L. Allgower and K. Georg, *Numerical Continuation Methods*, Springer-Verlag. 1990.

[17] H. Hirose, "Parameter Estimation in the Extreme-Value Distributions Using the Continuation Method", Trans. Information Processing Soc. Japan, Vol. 35, pp. 1674–1681, 1994.

[18] H. Hirose, "Maximum Likelihood Parameter Estimation in the Three-Parameter Log-Normal Distribution", Computational Statistics and Data Analysis, Vol. 24, pp. 139–152, 1997.

[19] H. Hirose, "Parameter Estimation for the 3-Parameter Gamma Distribution Using the Continuation Method", IEEE Trans. Reliability, Vol. 47, pp. 188–196, 1998.

[20] H. Hirose, "Maximum Likelihood Parameter Estimation in the Three-Parameter Gamma Distribution", Computational Statistics and Data Analysis, Vol. 20, pp. 343–354, 1995.

[21] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization", Computer J., Vol. 7, pp. 308–313, 1965.