Predictive Weibull Models with Applications to Decision-Making in Aircraft Service

N. A. Nechval [1], K. N. Nechval [2], G. Berzins [1], M. Purgailis [1], K. Rozite [1], N. Zolova [1]
[1] Mathematical Statistics Department, University of Latvia
Raina Blvd 19, LV-1050, Riga, Latvia
e-mail: nechval@junik.lv
[2] Applied Mathematics Department, Transport and Telecommunication Institute
Lomonosov Street 1, LV-1019, Riga, Latvia
e-mail: konstan@tsi.lv

Abstract
Based on a random sample from the Weibull distribution with unknown shape and scale parameters, lower and upper prediction limits on a set of m future observations from the same distribution are constructed. The procedures, which arise from considering the distribution of future observations given the observed value of an ancillary statistic, do not require the construction of any tables, and are applicable whether the data are complete or Type II censored. The results have direct application in reliability theory, where the time until the first failure in a group of m items in service provides a measure regarding the operation of the items, as well as in service of fatigue-sensitive aircraft structures to construct strategies of inspections of these structures; examples of applications are given.
Keywords: Aircraft, Fatigue crack, Initiation, Weibull model, Prediction limit.

1 Introduction

Prediction of an unobserved random variable is a fundamental problem in statistics. Patel [1] provides an extensive survey of literature on this topic. In the areas of reliability and lifetesting, this problem translates to obtaining prediction intervals for life distributions such as the Exponential and the Weibull. For the Weibull case, several authors have addressed this issue as well the more complicated problem of deriving prediction limits for order statistics from a future sample. These include Mann and Saunders [2], Mann [3], Antle and Rademaker [4], Lawless [5], Mann, et al [6], Mann [7], Engelhardt and Bain [8], and Fertig, et al [9].

One of the earlier works on prediction for the Weibull distribution is by Mann and Saunders [2]. They considered prediction intervals for the smallest of a set of future observations, based on a small (two or three) preliminary sample of past observations. An expression for the warranty period (time before the failure of the first ordered observation from a set of future observations or a lot) was derived as a function of the ordered past observations.

Mann [3] extended the results for lot sizes n = 10 (5) 25 and sample sizes m = 2 (1) n–3 for a specified assurance level of 0.95. This method requires numerical integration.

In addition, the tables provided are limited to sample sizes less than 25 and are given only for the assurance level of 0.95.

Antle and Rademaker [4] provided a method of obtaining a prediction bound for the largest observation from a future sample of the Type I extreme value distribution, based on the maximum likelihood estimates of the parameters. They used Monte Carlo simulations to obtain the prediction intervals. Using the well-known relationship between the Weibull distribution and the Type I extreme value distribution one can use their method to construct an upper prediction limit for the largest among a set of future Weibull observations. However this method is valid only for complete samples and limited to constructing an upper prediction limit for the largest among a set of future observations.

The distribution theory for estimators of unknown parameters in Weibull models is complicate and cannot be described in explicit forms. Nevertheless, using a conditional method, many problems become analytically manageable. The conditional method used in this paper is the one conditioned on ancillary statistics, which was first suggested by Fisher [10] and promoted further by a number of others (Cox [11], Buehler [12]). Lawless [5] applied this conditional method to different problems relating to the Weibull and extreme value distributions. In the conditional method, quantiles for constructing prediction intervals depend on ancillary statistics of observed data. This procedure, where the results are based on the conditional distribution of the maximum likelihood estimates given a set of ancillary statistics, is exact, but it requires numerical integration, for each new sample obtained, to determine the prediction limits.

We consider in this paper the problem of estimating the minimum time to crack initiation (warranty period or time to a first inspection) for a number of aircraft structure components, before which no cracks (that may be detected) in materials occur, based on the results of previous warranty period tests on the structure components in question. If in a fleet of k aircraft there are km of the same individual structure components, operating independently, the length of time until the first crack initially formed in any of these components is of basic interest, and provides a measure of assurance concerning the operation of the components in question. This leads to the consideration of the following problem. Suppose we have observations X_1, ..., X_n as the results of tests conducted on the components; suppose also that there are km components of the same kind to be put into future use, with times to crack initiation Y_1, ..., Y_{km}. Then we want to be able to estimate, on the basis of X_1, ..., X_n, the shortest time to crack initiation $Y_{(1,km)}$ among the times to crack initiation Y_1, ..., Y_{km}. In other words, it is desirable to construct lower simultaneous prediction limit, $H_{(\gamma)}$, that is exceeded with probability γ by observations or functions of observations of all k future samples, each consisting of m units. In this paper, the problem of estimating $Y_{(1,km)}$, the smallest of all k future samples of m observations from the underlying distribution, based on an observed sample of n observations from the same distribution, is considered. A solution is proposed for constructing a lower simultaneous prediction limit, $H_{(\gamma)}$, for $Y_{(1,km)}$. Various properties of these solutions are derived, and illustrations are given for some important special cases. The results have direct application in reliability theory, where the time until the first failure in a group of m items in service provides a measure of assurance regarding the operation of the items.

2 Equation for Constructing Simultaneous Prediction Limit

An equation, which shows how to construct lower simultaneous one-sided prediction limit for the order statistics in all of future samples when one-sided prediction limit for a single future sample is available, is given by the following theorem.

Theorem 1 (*Lower simultaneous one-sided prediction limit*). Let $(X_1, ..., X_n)$ be a random sample from the cdf $F(.)$, and let $(Y_{1_j},...,Y_{m_j})$ be the jth random sample of m_j "future" observations from the same cdf, $j\in\{1, ..., k\}$. Assume that $(k+1)$ samples are independent. Let $H=H(X_1, ..., X_n)$ be any statistic based on the preliminary sample and let $Y_{(r_j,m_j)}$ denote the r_jth order statistic in the jth sample of size m_j. Then

$$\Pr\left(Y_{(r_1,m_1)} \geq H,...,Y_{(r_j,m_j)} \geq H,...,Y_{(r_k,m_k)} \geq H\right) = \sum_{i_1=0}^{r_1-1}...\sum_{i_j=0}^{r_j-1}...\sum_{i_k=0}^{r_k-1}\binom{m_1}{i_1}...\binom{m_j}{i_j}$$

$$...\binom{m_k}{i_k}\frac{\Pr\left(Y_{(i_\Sigma+1,m_\Sigma)} \geq H\right)-\Pr\left(Y_{(i_\Sigma,m_\Sigma)} \geq H\right)}{\binom{m_\Sigma}{i_\Sigma}}, \tag{1}$$

where

$$i_\Sigma=\sum_{j=1}^{k} i_j, \quad m_\Sigma=\sum_{j=1}^{k} m_j. \tag{2}$$

Proof.

$$\Pr\left(Y_{(r_1,m_1)} \geq H,...,Y_{(r_j,m_j)} \geq H,...,Y_{(r_k,m_k)} \geq H\right) = \prod_{j=1}^{k}\Pr\left(Y_{(r_j,m_j)} \geq H\right)$$

$$= \mathrm{E}\left\{\prod_{j=1}^{k}\sum_{i_j=0}^{r_j-1}\binom{m_j}{i_j}[F(H)]^{i_j}[1-F(H)]^{m_j-i_j}\right\} = \sum_{i_1=0}^{r_1-1}...\sum_{i_j=0}^{r_j-1}...\sum_{i_k=0}^{r_k-1}\binom{m_1}{i_1}...\binom{m_j}{i_j}$$

$$...\binom{m_k}{i_k}\mathrm{E}\left\{[F(H)]^{i_\Sigma}[1-F(H)]^{m_\Sigma-i_\Sigma}\right\}. \tag{3}$$

Since

$$\mathrm{E}\left\{[F(H)]^{i_\Sigma}[1-F(H)]^{m_\Sigma-i_\Sigma}\right\}$$

$$=\binom{m_\Sigma}{i_\Sigma}^{-1}\left[\mathrm{E}\left\{\sum_{i=0}^{i_\Sigma}\binom{m_\Sigma}{i}[F(H)]^i[1-F(H)]^{m_\Sigma-i}-\sum_{i=0}^{i_\Sigma-1}\binom{m_\Sigma}{i}[F(H)]^i[1-F(H)]^{m_\Sigma-i}\right\}\right]$$

$$=\frac{\Pr\left(Y_{(i_\Sigma+1,m_\Sigma)}\geq H\right)-\Pr\left(Y_{(i_\Sigma,m_\Sigma)}\geq H\right)}{\binom{m_\Sigma}{i_\Sigma}}, \tag{4}$$

the joint probability can be written as

$$\Pr\left(Y_{(r_1,m_1)}\geq H,\ldots,Y_{(r_j,m_j)}\geq H,\ldots,Y_{(r_k,m_k)}\geq H\right)=\sum_{i_1=0}^{r_1-1}\ldots\sum_{i_j=0}^{r_j-1}\ldots\sum_{i_k=0}^{r_k-1}\binom{m_1}{i_1}\ldots\binom{m_j}{i_j}$$

$$\ldots\binom{m_k}{i_k}\frac{\Pr\left(\mathrm{Y}_{(i_\Sigma+1,m_\Sigma)}\geq H\right)-\Pr\left(\mathrm{Y}_{(i_\Sigma,m_\Sigma)}\geq H\right)}{\binom{m_\Sigma}{i_\Sigma}}. \tag{5}$$

This ends the proof. □

Corollary 1.1. If $r_j = 1$, $\forall j=1(1)k$, then

$$\Pr\left(Y_{(1,m_1)}\geq H,\ldots,Y_{(1,m_j)}\geq H,\ldots,Y_{(1,m_k)}\geq H\right)=\Pr\left(Y_{(1,m_\Sigma)}\geq H\right). \tag{6}$$

3 Lower One-sided Prediction Limit for Weibull Order Statistic

The Weibull distribution is a powerful modelling tool used in reliability analyses to predict failure rates and to provide a description of the failure of parts and equipment. The Weibull distribution has been widely used in the empirical modelling of economic models. Applications include the modelling of unemployment spells, strike durations, income distributions, the length of a firm's innovation period, and the size of research and development budgets. Depending on the particular problem, the variable under consideration may not be fully observed, requiring censoring procedures for estimation.

Based on engineering and macroscopic viewpoints, the mechanical properties of metallic materials are often considered homogeneous. However, a considerable amount of scatter has been observed in fatigue data even under the same loading condition. It may be attributed to the inhomogeneous material properties. As a result, probabilistic approaches for the fatigue crack initiation and growth have received great attention in recent years. Along with the development of fracture mechanics for the past three decades and the need of reliability or risk assessment for some important structures or components such as

- Transportation Systems and Vehicles – aircraft, space vehicles, trains, ships

- Civil Structures – bridges, dams, tunnels
- Power Generation – nuclear, fossil fuel and hydroelectric plants
- High-Value Manufactured Products – launch systems, satellites, semiconductor and electronic equipment
- Industrial Equipment – oil and gas exploration, production and processing equipment, chemical process facilities, pulp and paper

the so-called 'probabilistic fracture mechanics' has thus arisen [13]. One of the important issues in the probabilistic fracture mechanics analysis lies in the probabilistic modelling of fatigue crack initiation (or growth) phenomenon. Many probabilistic models have been proposed to capture the scatter of the fatigue crack growth data. Some of these models are based on the two-parameter Weibull distribution. It exhibits a wide range of shapes for the density and hazard functions that makes this distribution suitable for modelling complex failure data sets. Many authors have considered the problem of constructing prediction limits for the extreme value and Weibull distributions. References [14] and [6] contain good discussions of available procedures. As a rule, the better procedures involve the use of tables generated by Monte Carlo methods.

In this section our focus is on prediction limits for future samples of observations from the two-parameter Weibull distribution and the purpose is to present a technique for constructing the prediction limits which can be used very generally, for Type II censored as well as complete data. The procedures should in particular be useful in situations not handled by the tables in the aforementioned references.

The proposed technique may be useful when we consider, for example, the reliability problem associated with fatigue damage that arises from the initiation of fatigue cracks originating from rivet holes along the top longitudinal row of the outer skin of the fuselage (Figure 1).

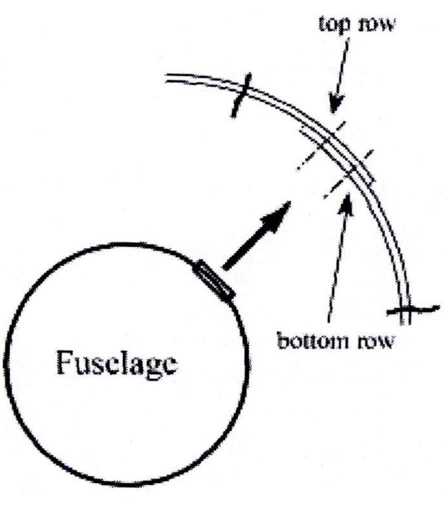

Figure 1: Rivet row in consideration.

It is assumed that a fatigue crack can initiate randomly at either side of a hole with diameter d. Experiments show that the number of flight cycles at which an initial crack will appear at one side with respect to a particular rivet follows the two-parameter Weibull distribution.

A post-failure photograph of one of the F-16 479 bulkhead test components (Figure 2) indicates the location of fatigue crack initiation at the radius between the bulkhead and one of the two vertical tail attach pads. Experiments show that the time to fatigue crack initiation follows the two-parameter Weibull distribution.

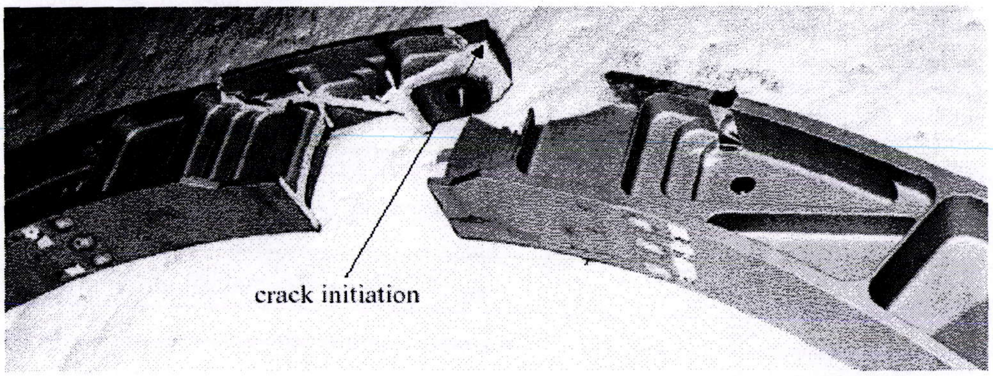

Figure 2: F-16 479-bulkhead test specimen number -7B, post-failure crack initiation.

The probability density function for the random variable X of the two-parameter Weibull distribution is given by

$$f(x;\sigma,\delta)=\frac{\delta}{\beta}\left(\frac{x}{\beta}\right)^{\delta-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\delta}\right]\ (x>0), \tag{7}$$

where $\delta>0$ and $\beta>0$ are the shape and scale parameters, respectively. Writing

$$S=\mu+\sigma Z, \tag{8}$$

where Z is a random variable with standardized extreme value density,

$$f(z)=\exp(z-e^{z}),\quad -\infty<z<\infty, \tag{9}$$

then the density of S can be obtained as

$$f(s;\mu,\sigma)=\frac{1}{\sigma}\exp\left(\frac{s-\mu}{\sigma}-\exp\left(\frac{s-\mu}{\sigma}\right)\right), \quad -\infty<s<\infty. \tag{10}$$

The distribution of S is known as the smallest extreme value (SEV) distribution. If S=ln$X$, so that, X=e^S, then

$$f(x;\mu,\sigma)=\frac{1}{x}\frac{1}{\sigma}(xe^{-\mu})^{1/\sigma}\exp[-(xe^{-\mu})^{1/\sigma}]. \tag{11}$$

With $\sigma=1/\delta$ and $\mu=\ln\beta$, X is distributed as Weibull with shape parameter δ and scale parameter β. Given this, for analytical and computational convenience, this paper works in the S=ln$X$ scale, the results, however, are reported directly for the Weibull observations.

Theorem 2 (*Lower one-sided prediction limit for Weibull order statistic*). Let $X_1<\ldots<X_r$ be the first r ordered past observations from a sample of size n from the distribution (7). Then a lower one-sided conditional $(1-\alpha)$ prediction limit $h_{(1-\alpha)}$ on the lth order statistic Y_l of a set of m future ordered observations $Y_1<\ldots<Y_m$ is given by

$$\Pr\{Y_l>h_{(1-\alpha)};\mathbf{z}\}=\Pr\left\{\hat{\delta}\ln\left(\frac{Y_l}{\hat{\beta}}\right)>\hat{\delta}\ln\left(\frac{h_{(1-\alpha)}}{\hat{\beta}}\right);\mathbf{z}\right\}=\Pr\{W_l>w_{(1-\alpha)};\mathbf{z}\}$$

$$=\frac{\sum_{j=0}^{l-1}\left[\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^{r}\ln(x_i/\hat{\beta})}\left((m-j)e^{vw_{(1-\alpha)}}+\sum_{i=1}^{r}e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv\right]}{\sum_{j=0}^{l-1}\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^{r}\ln(x_i/\hat{\beta})}\left(\sum_{i=1}^{r}e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv}$$

$$=1-\alpha, \tag{12}$$

where $\hat{\beta}$ and $\hat{\delta}$ are the maximum likelihood estimators of β and δ based on the first r ordered past observations (X_1, …, X_r) from a sample of size n from the Weibull distribution, which can be found from solution of

$$\hat{\beta}=\left(\frac{\sum_{i=1}^{r}x_i^{\hat{\delta}}+(n-r)x_r^{\hat{\delta}}}{r}\right)^{1/\hat{\delta}}, \tag{13}$$

and

$$\widehat{\delta} = \left[\left(\sum_{i=1}^{r} x_i^{\widehat{\delta}} \ln x_i + (n-r) x_r^{\widehat{\delta}} \ln x_r \right) \left(\sum_{i=1}^{r} x_i^{\widehat{\delta}} + (n-r) x_r^{\widehat{\delta}} \right)^{-1} - \frac{1}{r} \sum_{i=1}^{r} \ln x_i \right]^{-1}, \tag{14}$$

$$\mathbf{z} = (z_1, z_2, ..., z_{r-2}), \tag{15}$$

$$Z_i = \widehat{\delta} \ln\left(\frac{X_i}{\widehat{\beta}} \right), \; i = 1, ..., r-2, \tag{16}$$

$$W_l = \widehat{\delta} \ln\left(\frac{Y_l}{\widehat{\beta}} \right), \; w_{(1-\alpha)} = \widehat{\delta} \ln\left(\frac{h_{(1-\alpha)}}{\widehat{\beta}} \right). \tag{17}$$

Proof. The joint density of S_1=ln(X_1), …, S_r=ln(X_r) is given by

$$f(s_1, ..., s_r; \mu, \sigma) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} \frac{1}{\sigma} \exp\left(\frac{s_i - \mu}{\sigma} - \exp\left(\frac{s_i - \mu}{\sigma} \right) \right) \exp\left(-(n-r) \exp\left(\frac{s_r - \mu}{\sigma} \right) \right). \tag{18}$$

Let $\widehat{\mu}$, $\widehat{\sigma}$ be the maximum likelihood estimators (estimates) of μ, σ based on S_1, …, S_r and let

$$V_1 = \frac{\widehat{\mu} - \mu}{\widehat{\sigma}}, \tag{19}$$

$$V = \frac{\widehat{\sigma}}{\sigma}, \tag{20}$$

and

$$Z_i = \frac{S_i - \widehat{\mu}}{\widehat{\sigma}}, \; i = 1(1)r. \tag{21}$$

Parameters μ and σ in (18) are location and scale parameters, respectively, and it is well known that if $\widehat{\mu}$ and $\widehat{\sigma}$ are estimates of μ and σ, possessing certain invariance properties, then the quantities V_1 and V are parameter-free. Most, if not all, proposed estimates of μ and σ possess the necessary properties; these include the maximum likelihood estimates and various linear estimates. Z_i, i=1(1)r, are ancillary statistics, any r-2 of which form a functionally independent set. We then find in a straightforward manner that the joint density of V_1, V, conditional on fixed $\mathbf{z}=(z_1, z_2, \ldots, z_{r-2})$, is

$$f(v_1,v;\mathbf{z}) = \vartheta(\mathbf{z})v^{r-1}\exp\left(\sum_{i=1}^{r}(z_i+v_1)v - \sum_{i=1}^{r}\exp[(z_i+v_1)v] - (n-r)\exp[(z_r+v_1)v]\right),$$

$$v_1\in(-\infty,\infty),\quad v\in(0,\infty). \tag{22}$$

where

$$\vartheta(\mathbf{z}) = \left(\int_0^{\infty}\int_{-\infty}^{\infty} v^{r-1}\exp\left(\sum_{i=1}^{r}(z_i+v_1)v - \sum_{i=1}^{r}\exp[(z_i+v_1)v] - (n-r)\exp[(z_r+v_1)v]\right)dv_1 dv\right)^{-1}$$

$$\tag{23}$$

is the normalizing constant. For notational convenience we include all of z_1, …, z_r in (21); z_{r-1} and z_r can be expressed as function of z_1, …, z_r only.

Writing

$$W = \frac{\ln Y_l - \mu}{\sigma}, \tag{24}$$

where Y_l is the lth order statistic from an independent second sample of m observations also from the distribution (7), and noting that exp(W) is the lth order statistic in a sample of m observations from the standard exponential distribution, we have the density of W as

$$f(w) = \frac{m!}{(l-1)!(m-l)!}[1-\exp(-e^w)]^{l-1}[\exp(-e^w)]^{m-l}e^w\exp(-e^w)$$

$$= \frac{m!}{(l-1)!(m-l)!}e^w\sum_{j=0}^{l-1}\binom{l-1}{j}(-1)^{l-1-j}\exp[-(m-j)e^w],\quad w\in(-\infty,\infty). \tag{25}$$

Since W is distributed independently of v_1, v we find the joint density of w, v_1, v, conditional on $\mathbf{z}$, as the product of (22) and (25),

$$f(w,v_1,v;\mathbf{z}) = f(w)f(v_1,v;\mathbf{z}). \tag{26}$$

Note that

$$W_l = \frac{\ln Y_l - \widehat{\mu}}{\widehat{\sigma}} = \frac{W - V_1 V}{V}; \tag{27}$$

making the transformation $w_l=(w-v_1v)/v$, $v_1=v_1$, $v=v$ we find the joint density of w_l, v_1, v, conditional on $\mathbf{z}$, as

$$f(w_l,v_1,v;\mathbf{z})=\frac{m!}{(l-1)!(m-l)!}\vartheta(\mathbf{z})\sum_{j=0}^{l-1}\left[\binom{l-1}{j}(-1)^{l-1-j}v^r\exp\left((r+1)v_1v+\left(w_l+\sum_{i=1}^{r}z_i\right)v\right)\right.$$

$$\left.\times\exp\left(-(m-j)\exp[(w_l+v_1)v]\right)\exp\left(-\exp[v_1v]\left(\sum_{i=1}^{r}\exp[z_iv]+(n-r)\exp[z_rv]\right)\right)\right],$$

$$w_l\in(-\infty,\infty),\quad v_1\in(-\infty,\infty),\quad v\in(0,\infty). \tag{28}$$

Now v_1 can be integrated out of (28) in a straightforward way to give

$$f(w_l,v;\mathbf{z})=\frac{m!}{(l-1)!(m-l)!}\vartheta(\mathbf{z})\Gamma(r+1)$$

$$\times\sum_{j=0}^{l-1}\left[\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\frac{v^{r-2}\exp\left(v\sum_{i=1}^{r}z_i\right)(m-j)v\exp[w_lv]}{\left((m-j)\exp[w_lv]+\sum_{i=1}^{r}\exp[z_iv]+(n-r)\exp[z_rv]\right)^{r+1}}\right]$$

$$=\frac{\sum_{j=0}^{l-1}\left[\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\frac{rv^{r-2}\exp\left(v\sum_{i=1}^{r}z_i\right)(m-j)v\exp[w_lv]}{\left((m-j)\exp[w_lv]+\sum_{i=1}^{r}\exp[z_iv]+(n-r)\exp[z_rv]\right)^{r+1}}\right]}{\sum_{j=0}^{l-1}\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}\exp\left(v\sum_{i=1}^{r}z_i\right)\left(\sum_{i=1}^{r}\exp[z_iv]+(n-r)\exp[z_rv]\right)^{-r}dv},$$

$$w_l\in(-\infty,\infty),\quad v\in(0,\infty). \tag{29}$$

Thus, for fixed $w_{(1-\alpha)}$ $(-\infty<w_{(1-\alpha)}<\infty)$,

$$\Pr\{W_l>w_{(1-\alpha)};\mathbf{z}\}=\int_0^\infty\int_{w_{(1-\alpha)}}^\infty f(w_l,v;\mathbf{z})dw_ldv$$

$$
=\frac{\sum_{j=0}^{l-1}\left[\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}\exp\left(v\sum_{i=1}^r z_i\right)\left((m-j)e^{vw_{(1-\alpha)}}+\sum_{i=1}^r \exp[z_i v]+(n-r)\exp[z_r v]\right)^{-r}dv\right]}{\sum_{j=0}^{l-1}\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}\exp\left(v\sum_{i=1}^r z_i\right)\left(\sum_{i=1}^r \exp[z_i v]+(n-r)\exp[z_r v]\right)^{-r}dv}
$$

$$
=\frac{\sum_{j=0}^{l-1}\left[\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^r \ln(x_i/\hat{\beta})}\left((m-j)e^{vw_{(1-\alpha)}}+\sum_{i=1}^r e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv\right]}{\sum_{j=0}^{l-1}\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^r \ln(x_i/\hat{\beta})}\left(\sum_{i=1}^r e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv}.
\tag{30}
$$

This completes the proof. □

Corollary 2.1. A lower one-sided conditional $(1-\alpha)$ prediction limit $h_{(1-\alpha)}$ on the minimum Y_1 of a set of m future ordered observations $Y_1< \ldots <Y_m$ is given by

$$
\Pr\{Y_1>h_{(1-\alpha)};\mathbf{z}\}=\Pr\left\{\hat{\delta}\ln\left(\frac{Y_1}{\hat{\beta}}\right)>\hat{\delta}\ln\left(\frac{h_{(1-\alpha)}}{\hat{\beta}}\right);\mathbf{z}\right\}=\Pr\{W_1>w_{(1-\alpha)};\mathbf{z}\}
$$

$$
=\frac{\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^r \ln(x_i/\hat{\beta})}\left(me^{vw_{(1-\alpha)}}+\sum_{i=1}^r e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv}{\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^r \ln(x_i/\hat{\beta})}\left(\sum_{i=1}^r e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv}=1-\alpha. \tag{31}
$$

Thus, when $l = 1$, (12) reduces to formula (31).

Corollary 2.2. An upper one-sided conditional $(1-\alpha)$ prediction limit $h^{(1-\alpha)}$ on the maximum Y_m of a set of m future ordered observations $Y_1< \ldots <Y_m$ is given by

$$
\Pr\{Y_m<h^{(1-\alpha)};\mathbf{z}\}=\Pr\left\{\hat{\delta}\ln\left(\frac{Y_m}{\hat{\beta}}\right)<\hat{\delta}\ln\left(\frac{h^{(1-\alpha)}}{\hat{\beta}}\right);\mathbf{z}\right\}=\Pr\{W_m<w^{(1-\alpha)};\mathbf{z}\}
$$

$$
=1-\frac{\sum_{j=0}^{m-1}\left[\binom{m-1}{j}\frac{(-1)^{m-1-j}}{m-j}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^r \ln(x_i/\hat{\beta})}\left(\frac{m-j}{e^{-vw^{(1-\alpha)}}}+\sum_{i=1}^r e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv\right]}{\sum_{j=0}^{m-1}\binom{m-1}{j}\frac{(-1)^{m-1-j}}{m-j}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^r \ln(x_i/\hat{\beta})}\left(\sum_{i=1}^r e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv}
$$

$$=1-\frac{\sum_{j=0}^{m-1}\left[\binom{m-1}{j}\frac{(-1)^j}{j+1}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^{r}\ln(x_i/\hat{\beta})}\left((j+1)e^{vw_{(1-\alpha)}}+\sum_{i=1}^{r}e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv\right]}{\sum_{j=0}^{m-1}\binom{m-1}{j}\frac{(-1)^j}{j+1}\int_0^\infty v^{r-2}e^{v\hat{\delta}\sum_{i=1}^{r}\ln(x_i/\hat{\beta})}\left(\sum_{i=1}^{r}e^{v\hat{\delta}\ln(x_i/\hat{\beta})}+(n-r)e^{v\hat{\delta}\ln(x_r/\hat{\beta})}\right)^{-r}dv}$$

$$=1-\alpha. \tag{32}$$

Corollary 2.3. If $r=n$ and n is large, (12) should be more or less independent of **Z**. Also, Z_1, …, Z_n will be nearly independent and approximately distributed as standard extreme values, with pdf

$$f_{\circ}({}^{\bullet}w)=e^{{}^{\bullet}w}\exp(-e^{{}^{\bullet}w}),\quad {}^{\bullet}w\in(-\infty,\infty). \tag{33}$$

Our first step is to replace z_1, …, z_n in the numerator of (12) by $n\mathrm{E}\{{}^{\bullet}W\}=-n\gamma$, where γ=0.577215… is the Euler constant. We now suppose that $(1/n)\sum_{i=1}^{n}\exp[z_i v]$ will be approximately equal to the moment generating function for (33), with dummy variable v. Since $\mathrm{E}\{\exp[{}^{\bullet}w\eta]\}=\Gamma(1+\eta)$, we approximate the above sum in the denominator of (12) by $n\Gamma(1+v)$. We thus arrive at the following approximation to (12), where $r=n$,

$$\Pr\{Y_l>h_{(1-\alpha)};\mathbf{z}\}=\Pr\left\{\hat{\delta}\ln\left(\frac{Y_l}{\hat{\beta}}\right)>\hat{\delta}\ln\left(\frac{h_{(1-\alpha)}}{\hat{\beta}}\right);\mathbf{z}\right\}=\Pr\{W_l>w_{(1-\alpha)};\mathbf{z}\}$$

$$=\frac{\sum_{j=0}^{l-1}\left[\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{n-2}e^{-vn\gamma}\left(\frac{m-j}{n}e^{vw_{(1-\alpha)}}+\Gamma(1+v)\right)^{-n}dv\right]}{\sum_{j=0}^{l-1}\binom{l-1}{j}\frac{(-1)^{l-1-j}}{m-j}\int_0^\infty v^{n-2}e^{-vn\gamma}\left[\Gamma(1+v)\right]^{-n}dv}=1-\alpha. \tag{34}$$

4 Examples of Applications

4.1 Assigning Warranty Period

For instance, consider the data of fatigue tests on a particular type of structural components (stringer) of aircraft IL-86. The data are for a complete sample of size $r=n=5$, with observations (Table 1).

Table 1: The data of fatigue tests.

Observations	Time to crack initiation (in number of 10^4 flight hours)
x_1	5
x_2	6.25
x_3	7.5
x_4	7.9
x_5	8.1

and results being expressed here in number of 10^4 flight-hours. On the basis of these data it is wished to estimate a lower 0.95 prediction limit on Y_1 in a group of $m = 5$ identical components (for a fleet of m =5 aircraft IL-86) which are to be put into service.

Goodness-of-fit testing. We assess the statistical significance of departures from the Weibull model by performing empirical distribution function goodness-of-fit test. We use the $_\bullet S$ statistic [15]. For censoring (or complete) datasets, the $_\bullet S$ statistic is given by

$$_\bullet S = \frac{\sum_{i=[r/2]+1}^{r-1} \left(\frac{\ln(x_{i+1}/x_i)}{M_i} \right)}{\sum_{i=1}^{r-1} \left(\frac{\ln(x_{i+1}/x_i)}{M_i} \right)} = \frac{\sum_{i=3}^{4} \left(\frac{\ln(x_{i+1}/x_i)}{M_i} \right)}{\sum_{i=1}^{4} \left(\frac{\ln(x_{i+1}/x_i)}{M_i} \right)} = 0.184, \tag{35}$$

where $[r/2]$ is a largest integer $\le r/2$, the values of M_i are given in Table 13 [15]. The reject region for the α level of significance is $\{_\bullet S > _\bullet S_{n;1-\alpha}\}$. The percentage points for $_\bullet S_{n;1-\alpha}$ were given by Kapur and Lamberson [15]. For this example,

$$_\bullet S = 0.184 < {}_\bullet S_{n=5;\,1-\alpha=0.95} = 0.86. \tag{36}$$

Thus, there is not evidence to rule out the Weibull model. The maximum likelihood estimates of unknown parameters β and δ are $\widehat{\beta} = 7.42603$ and $\widehat{\delta} = 7.9081$, respectively. It follows from (31) that

$$\Pr\{Y_1 > h_{(1-\alpha)};\mathbf{z}\} = \Pr\left\{ \widehat{\delta} \ln\left(\frac{Y_1}{\widehat{\beta}} \right) > \widehat{\delta} \ln\left(\frac{h_{(1-\alpha)}}{\widehat{\beta}} \right);\mathbf{z} \right\} = \Pr\{W_1 > w_{(1-\alpha)};\mathbf{z}\}$$

$$= \Pr\{W_1 > -8.4378;\mathbf{z}\} = \frac{0.0000141389}{0.0000148830} = 0.95 \tag{37}$$

and a lower 0.95 prediction limit for Y_1 is $h_{(1-\alpha)}$=2.5549 (×10^4) flight hours, i.e., we have obtained the warranty period (or the time to the first inspection) equal to 25549 flight hours with confidence level $\gamma = 1-\alpha = 0.95$.

4.2 Planning In-service Inspections for Detection of Initial Crack after Warranty Period

Let us assume that in a fleet of m aircraft there are m of the same individual structure components, operating independently. Suppose an inspection is carried out at time τ_j, and this shows that initial crack (which may be detected) has not yet occurred. We now have to schedule the next inspection. Let Y_1 be the minimum time to crack initiation in the above components. In other words, let Y_1 be the smallest observation from an independent second sample of m observations also from the distribution (7). Then the inspection times can be calculated recursively as

$$\tau_j = \widehat{\beta}\exp(w_{j,(1-\alpha)}/\widehat{\delta}), \quad j\geq 2, \tag{38}$$

where it is assumed that τ_0=0, τ_1 is a time of the first inspection (warranty period), w_j is determined from

$$\Pr\{Y_1 > \tau_j; Y_1 > \tau_{j-1}, \mathbf{z}\} = \Pr\left\{\widehat{\delta}\ln\left(\frac{Y_1}{\widehat{\beta}}\right) > \widehat{\delta}\ln\left(\frac{\tau_j}{\widehat{\beta}}\right); \widehat{\delta}\ln\left(\frac{Y_1}{\widehat{\beta}}\right) > \widehat{\delta}\ln\left(\frac{\tau_{j-1}}{\widehat{\beta}}\right), \mathbf{z}\right\}$$

$$= \Pr\{W_1 > w_{j,(1-\alpha)}; W_1 > w_{j-1,(1-\alpha)}, \mathbf{z}\} = \frac{\Pr\{W_1 > w_{j,(1-\alpha)}; \mathbf{z}\}}{\Pr\{W_1 > w_{j-1,(1-\alpha)}; \mathbf{z}\}} = 1-\alpha \tag{39}$$

or, equivalently, from

$$\Pr\{Y_1 > \tau_j; \mathbf{z}\} = \Pr\left\{\widehat{\delta}\ln\left(\frac{Y_1}{\widehat{\beta}}\right) > \widehat{\delta}\ln\left(\frac{\tau_j}{\widehat{\beta}}\right); \mathbf{z}\right\} = \Pr\{W_1 > w_{j,(1-\alpha)}; \mathbf{z}\}$$

$$= \frac{\int_0^\infty v^{r-2} e^{v\widehat{\delta}\sum_{i=1}^{r}\ln(x_i/\widehat{\beta})}\left(me^{vw_{j,(1-\alpha)}} + \sum_{i=1}^{r} e^{v\widehat{\delta}\ln(x_i/\widehat{\beta})} + (n-r)e^{v\widehat{\delta}\ln(x_r/\widehat{\beta})}\right)^{-r} dv}{\int_0^\infty v^{r-2} e^{v\widehat{\delta}\sum_{i=1}^{r}\ln(x_i/\widehat{\beta})}\left(\sum_{i=1}^{r} e^{v\widehat{\delta}\ln(x_i/\widehat{\beta})} + (n-r)e^{v\widehat{\delta}\ln(x_r/\widehat{\beta})}\right)^{-r} dv} = (1-\alpha)^j, \quad j\geq 1, \tag{40}$$

$\widehat{\beta}$ and $\widehat{\delta}$ are the MLE's of β and δ, respectively, and can be found from solution of (13) and (14), respectively.

But again, for instance, consider the data of fatigue tests on a particular type of structural components of aircraft IL-86: x_1=5, x_2=6.25, x_3=7.5, x_4=7.9, x_5=8.1 (in number of 10^4 flight hours) given in Table 1, where $r=n=5$ and the maximum likelihood estimates of unknown parameters β and δ are $\widehat{\beta} = 7.42603$ and $\widehat{\delta} = 7.9081$, respectively. Thus, using (38) with τ_1=2.5549 (×10^4 flight hours) (the time of the first inspection), we have obtained the following inspection time sequence (see Table 2).

Table 2: The inspection time sequence.

w_j	Inspection time τ_j (×10^4 flight hours)	Interval $\tau_{j+1}-\tau_j$ (flight hours)
–	$\tau_0 = 0$	–
$w_1 = -8.4378$	$\tau_1 = 2.5549$	25549
$w_2 = -6.5181$	$\tau_2 = 3.2569$	7020
$w_3 = -5.5145$	$\tau_3 = 3.6975$	4406
$w_4 = -4.8509$	$\tau_4 = 4.0212$	3237
$w_5 = -4.3623$	$\tau_5 = 4.2775$	2563
$w_6 = -3.9793$	$\tau_6 = 4.4898$	2123
$w_7 = -3.6666$	$\tau_7 = 4.6708$	1810
$w_8 = -3.4038$	$\tau_8 = 4.8287$	1579
$w_9 = -3.1780$	$\tau_9 = 4.9685$	1398
⋮	⋮	⋮

5 Conclusions

The method of constructing prediction limits for future samples from a Weibull distribution introduced in this paper utilizes all the information in a sample, but since it involves the use of numerical integration, many may prefer to use this technique only in situations not readily handled by other of the methods described earlier. With modern computing, however, the conditional prediction limits are not difficult to calculate and should be recommended when the ability to do computations is available.

Although the results of this paper can be obtained through simulation, the simulation results are unstable; they vary from one to another. From theoretical as well as practical points of view, analytical solutions should be used if they are available. The results of this paper provide such analytical solutions. Furthermore, the techniques used in this paper can be applied to obtaining explicit formulae for computing prediction limits for any other location-scale distributions.

Acknowledgements

This research was supported in part by Grant No. 06.1936 and Grant No. 07.2036 from the Latvian Council of Science and the National Institute of Mathematics and Informatics of Latvia.

References

[1] Patel, J. K. (1989). Prediction Intervals - a Review. Communications in Statistics – Theory and Methods, Vol. 13, pp. 2393-2465.

[2] Mann, N. R. and Saunders, S. C. (1969). On Evaluation of Warranty Assurance when Life Has a Weibull Distribution. Biometrika, Vol. 56, pp. 615-625.

[3] Mann, N. R. (1970). Warranty Periods Based on Three Ordered Sample Observations from a Weibull Population. IEEE Transactions on Reliability, Vol. R-19, pp. 167-171.

[4] Antle, C. E. and Rademaker, F. (1972). An Upper Confidence Limit on the Maximum of *m* Future Observations from a Type 1 Extreme-Value Distribution. Biometrika, Vol. 59, pp. 475-477.

[5] Lawless, J. F. (1982). Statistical Models and Methods for Lifetime Data. John Wiley, New York.

[6] Mann N.R., Schafer R.E. and Singpurwalla N.D. (1974). Methods for Statistical Analysis of Reliability and Life Data. New York: John Wiley and Sons.

[7] Mann, N. R. (1976). Warranty Periods for Production Lots Based on Fatigue-Test Data. Engineering Fracture Mechanics, Vol. 8, pp. 123-130.

[8] Engelhardt, M. and Bain, L. J. (1982). On Prediction Limits for Samples from a Weibull or Extreme-Value Distribution. Technometrics, Vol. 24, pp. 147-150.

[9] Fertig, K. W., Mayer, M. and Mann, N. R. (1980). On Constructing Prediction Intervals for Samples from a Weibull or Extreme-Value Distribution. Technometrics, Vol. 22, pp. 567-573.

[10] Fisher, R. A. (1934). Two New Properties of Mathematical Likelihood. Proc. Roy. Statist. Soc., Ser. A 144, pp. 285-307.

[11] Cox, D. R. (1958). Some Problems Connected with Statistical Inference. Ann. Math. Statist., Vol. 29, pp. 357-372.

[12] Buehler, R. J. (1959). Some Validity Criteria for Statistical Inferences. Ann. Math. Statist., Vol. 30, pp. 845-863.

[13] Blischke, W.R. and Murthy, D.N.P. (2000). Reliability. New York: Wiley.

[14] Mann, N.R. and Fertig, K.W. (1973). Tables for Obtaining Weibull Confidence Bounds and Tolerance Bounds Based on Best Linear Invariant Estimates of Parameters of Extreme-Value Distribution. Technometrics, Vol. 15, pp. 87-101.

[15] Kapur, K.C. and Lamberson, L.R. (1977). Reliability in Engineering Design. New York: John Wiley and Sons.