Statistique et analyse des données

Mohamed M. T. Limam

Tolerance bounds for Weibull regression models

Statistique et analyse des données, tome 16, nº 1 (1991), p. 43-54.
<http://www.numdam.org/item?id=SAD_1991__16_1_43_0>

Statistique et Analyse des Données
1991 - Vol. 16 n° 1 p. 43 - 54

TOLERANCE BOUNDS FOR WEIBULL REGRESSION MODELS

Mohamed M.T. LIMAM

Statistics Department
King Saud University
Saudi Arabia

Résumé.

Récemment, la construction des intervalles simultanés de tolérance a reçu une attention dans les publications statistiques. Cet article étend un travail précédent de l'auteur sur les modèles de régression linéaire, aux modèles de régression Weibull. Quelques distributions approximatives de l'estimateur du maximum de vraisemblance de l'écart-type sont discutées. Le niveau de confiance actuel des limites de tolérance est évalué à l'aide d'une étude de simulation. Cette étude montre que les limites sont légèrement conservatrices pour des échantillons modérés ou larges, mais en général elles sont performantes. Enfin, un exemple numérique est discuté pour illustrer l'usage des limites de tolérance suggérées.

Mots clés : intervalles simultanés de tolérance, inférence statistique, distribution de valeurs extrêmes.

Classification AMS : 62 G 15
Classification STMA : 04 064

Reçu le 19 octobre 1990, révisé le 1er août 1991.

Abstract.

The construction of simultaneous tolerance intervals for linear models has recently received attention in the literature. This article extends earlier work by this author in dealing with the linear normal regression models by considering extreme value and Weibull regression models. Alternative approximate distributions for the maximum likelihood estimator of the scale parameter are discussed. A simulation study is used to estimate the actual confidence level of the bounds. This study showed that the bounds are slightly conservative for moderate to large sample sizes, but actual confidence levels are fairly close to the nominal ones. Finally, a numerical example is given to illustrate the use of the suggested one-sided tolerance bounds.

Keywords : simultaneous tolerance intervals, statistical inference, extreme value distribution.

1. INTRODUCTION.

In this article, a procedure for constructing lower tolerance bounds in the Weibull regression models is discussed. The Weibull distribution is widely used in modeling lifetime data and failure strength of ceramic materials. Also, Weibull regression models are assumed in many lifetime data statistical analyses. These models are usually applied for regressing the lifetime of items on the regressor variables $x_1, x_2, \ldots, x_{q-1}$, where the scale parameter of the lifetime distribution, and not the shape parameter, depends on the x's. Lawless (1982) describes experiments in which Weibull regression models would be appropriate.

Let $W_1, W_2, \ldots, W_n$ denote independent lifetimes from the Weibull distribution, and define $Y = \ln W$ as the log of lifetime. The probability density function (p.d.f.) of Y given $\underline{x}$ is in the extreme value form :

$$f(y|\underline{x}) = \exp\{(y - \underline{x}'\underline{\beta})/\sigma - \exp(y - \underline{x}'\underline{\beta}/\sigma]/\sigma, \quad -\infty < y < \infty .$$

Then, the data $Y_i = \ln(W_i)$, $\underline{x}_i' = (1, x_{i1}, x_{i2}, \ldots, x_{i,q-1})$, $i = 1, \ldots, n$ can be written in the linear model form

$$Y_i = \underline{x}_i' \underline{\beta} + \sigma Z_i\ , \quad i = 1,\dots, n$$

$$= \beta_0 + \beta_1 x_{i1} + \dots + \beta_{q-1} x_{i,q-1} + \sigma Z_i\ , \qquad (1.1)$$

where $\underline{\beta}$ is a $q \times 1$ vector of unknown regressor coefficients and Z_i has a standard extreme value distribution with p.d.f. $g(z) = \exp(z - e^z)$, where $-\infty < z < \infty$.

Verhagen (1961) showed that the maximum likelihood estimators (MLE's), $\hat{\underline{\beta}} = \hat{\underline{\beta}}(y)$ and $\hat{\sigma} = \hat{\sigma}(y)$ are equivariant estimators (see Lawless 1982, pp. 538). For this case, the distribution of the pivotal quantities $\underline{b} = (\hat{\underline{\beta}} - \underline{\beta})/\sigma$, and $u = \hat{\sigma}/\sigma$ depend only on the distribution of Z and not on $\hat{\underline{\beta}}$ and $\hat{\sigma}$.

The form of the model (1.1) is particularly attractive because its parameters are in a location-scale form. Furthermore, either model, the Weibull and the extreme value, can be of interest in its own right and procedures developed for one model can be applied to the other. Harter (1978) provides an interesting bibliography on the applications of the extreme value distribution. Also, Johnson and Kotz (1970) give references of applications for extreme value and Weibull regression models.

Lieberman and Miller (1963), Wilson (1967), and Limam and Thomas (1988a), proposed simultaneous tolerance intervals for normal regression models. Earlier, Jones et al. (1985) considered the problem of constructing tolerance bounds for log gamma regression models. In this article we extend the development of simultaneous tolerance intervals, suggested by Limam and Thomas (1988a), for the linear regression model, to include Weibull regression models defined in (1.1).

Usually, in lifetime analyses lower tolerance bounds are of interest because they give us information about the percentage of items with lifetimes exceeding these bounds. Thus, the problem is to find a tolerance factor δ such that the probability is $1 - \alpha$ that at least a given proportion p of the population being sampled, the extreme value distribution with location parameter $\underline{x}'\underline{\beta}$, is above the quantity $\underline{x}'\hat{\underline{\beta}} - \delta\hat{\sigma}$, for every $\underline{x}$ and p. That is

$$Pr_{\hat{\beta},\hat{\sigma}}\{P_Y[Y > \underline{x}'\hat{\underline{\beta}} - \delta\hat{\sigma}, \text{ for all } \underline{x} \mid \hat{\beta},\hat{\sigma}] \geq p\} = 1 - \alpha\ ,$$

where Y has the extreme value distribution defined in (1.1).

2. THE CONFIDENCE SET PROCEDURE.

The suggested simultaneous lower one-sided tolerance bounds are based on the confidence set procedure credited to Wilson (1967), where tolerance bands for the normal linear regression model are constructed. Later Limam and Thomas (1988a) used a similar method for their simultaneous tolerance intervals procedure, to produce narrower bands than those suggested by Wilson (1967). This method uses confidence sets for the parameters to put a lower bound on the content of the bands.

The content of the extreme value distribution above the lower one-sided tolerance bound $\underline{x}'\hat{\underline{\beta}} - \delta\hat{\sigma}$ is

$$1 - G[(\underline{x}'\hat{\underline{\beta}} - \delta\hat{\sigma} - \underline{x}'\beta)/\sigma],$$

where G denotes the standard extreme value cumulative distribution function, and $\hat{\underline{\beta}}$ and $\hat{\sigma}$ are MLE's. This content can be expressed in terms of the pivotal quantities $\underline{b}$ and u as

$$C(\underline{x}'\underline{b}, \delta u) = 1 - G[\underline{x}'\underline{b} - \delta u)] .$$

Then, tolerance factors are needed such that

$$\Pr\{C(\underline{x}'\underline{b}, \delta u) \geq p \quad \text{for all } \underline{x}\} = 1 - \alpha .$$

Confidence sets for the parameters $(\underline{\beta},\sigma)$ will be described in terms of a product set S, for pivotal quantities $\underline{b}$ and u, such that applying the Bonferroni inequality to the events $\underline{b} \in E_1$ and $u \in E_2$ gives

$$\Pr\{(\underline{b}, u) \in S\} \geq 1 - 2\alpha . \qquad (2.1)$$

For a pivotal set S and a specified p content, which can depend on $\underline{x}$, let

$$\delta^* = \delta^*(p, \underline{x}) = \min\{\delta : C(\underline{x}'\underline{b}, \delta u) \geq p, \text{ for all } (\underline{b}, u) \in S\}$$

denote the optimal tolerance factor. Then, $(\underline{b}, u) \in S$ implies that $C(\underline{x}'\underline{b}, \delta^* u) \geq p$, for all $\underline{x}$. Hence

$$\Pr\{C(\underline{x}'\underline{b}, \delta^* u) \geq p, \text{ for all } \underline{x}\}$$

$$\geq \Pr\{(\underline{b}, u) \in S\} \geq 1 - 2\alpha. \qquad (2.2)$$

In (2.1) the Bonferroni inequality is used to obtain the lower bound of $1 - 2\alpha$. Williams (1962) has shown that the probability in (2.1) has an upper bound of $1 - \alpha$, such that

$$1 - 2\alpha \leq \Pr\{(\underline{b}, u) \in S\} \leq 1 - \alpha .$$

In this next section two confidence sets, E_1 and E_2 are developed for $\underline{b}$ and u, respectively. The pivotal set for $\underline{b}$, E_1, depends on u which implies that the probability in (2.1) is equal to its upper bound $1 - \alpha$. Actually, this is a similar situation to the one encountered by Limam and Thomas (1988b). Then the probability in (2.2) has a lower bound of $1 - \alpha$. This fact suggests that by using $1 - \alpha$ level sets for β and σ we obtain $1 - \alpha$ tolerance bounds.

In the following section we adopt this procedure to derive tolerance bounds for the Weibull regression model. It is clear from the previous development that the suggested procedure will be a conservative one. Improvements on this method are difficult and complicate the derivation.

3. DEVELOPMENT OF THE TOLERANCE BOUNDS.

Tolerance bounds, developed in this article, employ large sample approximations for the distribution of the MLE's : $(\hat{\underline{\beta}}, \hat{\sigma}) \sim N[(\underline{\beta}, \sigma) ; I_0^{-1}]$, where I_0 is the observed information matrix. For uncensored data the expected information matrix is very simple, and a normal approximation employing the expected information matrix can be used : $(\hat{\underline{\beta}}, \hat{\sigma}) \sim N[(\underline{\beta}, \sigma), I^{-1}]$ (see Lawless 1982, pp. 301).

Letting the inverse of I_0 be

$$I_0^{-1} = \hat{\sigma}^2 \begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{bmatrix} ,$$

then $\hat{\sigma}^2 C_{11}$ is the q×q asymptotic covariance matrix for $\hat{\underline{\beta}}$. Under $H_0 : \underline{\beta} = \underline{\beta}_0$, the quadratic form $(\hat{\underline{\beta}} - \underline{\beta}_0)C_{11}^{-1}(\hat{\underline{\beta}} - \underline{\beta}_0)/\sigma^2$ is approximately χ_q^2 in large samples. Also, $\hat{\sigma}$ is distributed $N(\sigma, \hat{\sigma}^2 C_{22})$, where $\hat{\sigma}^2 C_{22}$ is an approximation to the asymptotic variance of $\hat{\sigma}$. By using normal approximation theory, the $1 - \alpha$ confidence set for $\underline{\beta}$ is

$$E_1 = \{\underline{\beta} : (\hat{\underline{\beta}} - \underline{\beta})C_{11}(\hat{\underline{\beta}} - \underline{\beta}) \leq \sigma^2 k_1^2 \}$$

where

$$k_1^2 = \chi^2_{(q,1-\alpha)} \tag{3.1}$$

is the upper $1 - \alpha$ percentile of the chi-square distribution with q degrees of freedom. Also, the $1 - \alpha$ confidence set for σ is

$$E_2 = \{\sigma : 0 < \sigma \leq \hat{\sigma}[1 - Z_\alpha C_{22}^{1/2}]\},$$

where Z_α is the lower α percentile of the standard normal distribution. In terms of the pivotal quantities, both sets can be written as

$$E_1 = \{\underline{b} : \underline{b}'C_{11}^{-1}\underline{b} \leq u^2k_1^2\},$$

$$E_2 = \{u : u \geq k_2\}, \text{ where } k_2 = [1 - Z_\alpha C_{22}^{1/2}]^{-1}. \tag{3.2}$$

Then, the pivotal set S is

$$S = \{(\underline{b}, u) : \underline{b}'C_{11}^{-1}\underline{b} \leq u^2k_1^2 \text{ and } u \geq k_2\}.$$

By applying the Scheffé projection result (Miller 1981, p. 16) to the set E_1, we obtain the following upper bounds for $|\underline{x}'\underline{b}|$

$$|\underline{x}'\underline{b}| \leq uk_1A(\underline{x}) \text{ for all } \underline{x} \text{ and } (\underline{b}, u) \in S, \tag{3.3}$$

where $A(\underline{x}) = [\underline{x}'C_{11}\underline{x}]^{1/2}$.

Development of the tolerance bounds requires the following result : given a desired p content, and $u = k$, by simple differentiation we see that $C(\underline{x}'\underline{b}, \delta u)$ is decreasing in $|\underline{x}'\underline{b}|$. Using this result with (3.3) gives

$$C(\underline{x}'\underline{b}, \delta u) \geq C(uk_1A(\underline{x}), \delta u) \text{ for all } (\underline{b}, u) \in S.$$

Note also that $C(uk_1A(\underline{x}), \delta u)$ is an increasing function of u for a given $\underline{x}$. This result yields

$$C(uk_1A(\underline{x}), \delta u) \geq C(k_2k_1A(\underline{x}), \delta k_2), \text{ for all } u \geq k_2 .$$

Finally, solving $C(k_2k_1A(\underline{x}), \delta k_2) = p$ for the tolerance factor $\delta = \delta(p, \underline{x})$ yields

$$\delta(p, \underline{x}) = k_1A(\underline{x}) - \text{Log}(-\text{Log } p)/k_2 . \tag{3.4}$$

These tolerance factors are developped on the extreme value scale, Y = ln(W). Then, to obtain tolerance bounds on the Weibull scale we take exponentials of the extreme value tolerance bounds.

Note that standard regression programs such as SAS lifereg procedure provide all estimates needed to compute the suggested tolerance bounds.

4. DIFFERENT DISTRIBUTIONS FOR THE MLE OF σ.

Bain and Engelhardt (1981) considered different approximate distributions for the MLE of σ. Their first approximation is $n\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-1}$, which is considered to be quite adequate for practical applications, and a more reasonable approximation than the normal, in (3.2), since it gives a desired skewness to the distribution of $\hat{\sigma}$. This approximate distribution yields the following lower bound on u

$$k_2' = [\chi^2_{\alpha,n-1}/n]^{1/2} .$$

As an improvement on the previous approximation, Bain and Engelhardt (1981) suggested, for the complete sample case, the following approximate distributions : $cn\hat{\sigma}^2/\sigma^2 \sim \chi^2_{c(n-1)}$, where c = .822 is chosen to make this distribution exact asymptotically. For the complete sample case, this approximation yields $1 - \alpha$ lower confidence bounds for u of the form

$$k_2'' = [\chi^2_{\alpha,\ .822(n-1)}/.822n]^{1/2} . \tag{4.1}$$

Bain and Engelhardt (1986) showed that a chi-square distribution provides a useful approximation for both complete and censored sample. They suggested the following simple approximation :

$$cfn(\hat{\sigma}/\sigma)^{(1+f^2)} \sim \chi^2_{c(fn-1)} ,$$

where f is the censoring fraction, and they showed that it gives a good accuracy for all cases. Values of c are tabulated in Bain and Engelhardt (1986) for censoring fractions f = 0.1(.1)1. Constants c go from .822 to 2.0 as f goes from 1 to zero. For the simulation study in the next section, we have a complete sample data, which means an uncensored data, and the upper bound k_2'' defined in (4.1) is used.

5. SIMULATION STUDY.

In this section, the accuracy of the suggested tolerance bounds is examined by a simulation study, for different p contents, confidence levels, and sample sizes. We assume model (1.1) with one independent variable,

$$y_i = \beta_0 + \beta_1 x_i + \sigma Z_i , \qquad (5.1)$$

where $Z_i \sim EV(0,1)$. Without loss of generality we take $\beta_0 = 0, \beta_1 = 1, \sigma = 1$, and assume that $x \sim N(\mu_x, \sigma_x^2)$. Tolerance factor, δ, and $\underline{x}'\underline{\beta}$ are invariant under linear transformations of x. Then, we let $x \sim N(0,1)$. IMSL (1987) is used to generate Z_i's, and x_i's, for a given sample size n. Then, Y_i are calculated according to (5.1). MLE's for β_0, β_1, σ, and the information matrix are computed through SAS (1985) Lifereg procedure. Then for specified α, and p, the smallest content $C[k_2'' k_1 A(x), \delta(p,x)k_2'']$, over the range of $x : (-\infty, \infty)$, is checked with the desired content p.

Independent sets of 5000 samples were generated on an IBM 3083 computer for different sample sizes, α, and p. The percentage of samples satisfying the condition $C[k_2'', k_1 A(x), \delta(p,x)k_2''] \geq p$, is recorded as the actual confidence level. The standard error of the estimated confidence coefficient would be approximately .003. Table 1 gives the empirical confidence levels of the suggested tolerance bounds, for $\alpha = .05$, p = .80, .90, .95, .99, and n = 10, 15, 20, 30,100.

It is easy to see that there are three factors affecting the tolerance bounds confidence level : the asymptotic approximate distributions of $\hat{\beta}$ and $\hat{\sigma}$, the desired p content, and the conservative nature of the suggested tolerance bounds. Table 1 reveals that for small sample sizes the asymptotic approximation does not work and thus the confidence levels tend to be less than the nominal ones. But although the actual confidence levels are not conservative for $n = 10$ they may provide a reasonable approximation for some applications. As the sample size increases, the small sample size effect decreases and the approximate distribution is improved. Then the conservative behaviour of the bounds is exhibited. For $n = 100$, the small sample size effect disappears, and the bounds are slightly conservative. Thus, for moderate or large sample sizes the lower p-content tolerance bound has a true confidence level sufficiently close to the nominal level.

6. NUMERICAL EXAMPLE.

We illustrate the suggested lower one-sided simultaneous tolerance bounds, by employing the example discussed by Nelson (1970), and used by Lawless (1982, p. 185). Results of an accelerated life test experiment on a type of electrical insulating fluid were presented. The uncensored data are breakdown times for seven groups of specimens, each group involving a different voltage level (kvolts). For a fixed voltage level, the model suggests that breakdown times have a Weibull distribution. Also, the distribution for each voltage is assumed to differ only in its scale parameter. In terms of log lifetime the model is of the form

$$y_{ij} = \beta_0 + \beta_1 x_i + \sigma Z_{ij}, \quad i = 1, \ldots, 7, \quad j = 1, \ldots, n_j,$$

where x_i is the log of voltage level v_i, and $\Sigma n_i = n = 76$.
The MLE's of β_0, β_1, and σ are found by iteration to be $\hat{\beta}_0 = 64.842$, $\hat{\beta}_1 = -17.728$ and $\hat{\sigma}^2 = 1.659$. The inverse of the observed information matrix is

$$I_0^{-1} = 1.659 \begin{bmatrix} 19.0436 & -5.4429 & -.00541 \\ -5.4429 & 1.5569 & .00057 \\ -.00541 & .00057 & .00775 \end{bmatrix}.$$

Now, we compute lower one-sided tolerance bound for time to breaking, with 95 % confidence level, a content $p = .80$, and a voltage stress level $v = 30$. Equations (3.1), (3.3) and (4.1) yield $k_1 = 2.447$, $A(\underline{x}) = .1711$, and $k_2'' = .85$, respectively.
Then, $k_2''k_1A(\underline{x}) = .356$, and using equation (3.4) with $p = .80$ yields a tolerance factor $\delta = 2.183$. Therefore, the lower one-sided tolerance bound on the extreme value scale is $\underline{x}'\hat{\underline{\beta}} - \delta\hat{\sigma} = 1.733$. Taking the exponential of this lower bound, to make the inference on the Weibull scale, yields 5.660. The interpretation of this result is that we are 95 % sure that at least 80 % of times to breaking of the electrical insulating fluid exceed 5.660 minutes, for a voltage level $v = 30$ kvolts.

ACKNOWLEDGMENTS.

Part of this work was completed while the author is visiting the Department of Statistics at Oregon State University, during Summer 1988. The author is indebted to Professor David R. Thomas for his valuable suggestions, and to Badreddine Sofiane for helping with the computations. The author thanks the editor and two referees for useful suggestions.

REFERENCES.

Bain, L.J. and Engelhardt, M., (1981), Simple approximate distributional results for confidence and tolerance limits for the Weibull distribution based on maximum likelihood estimators, *Technometrics*, 23, pp. 15-19.

Bain, L.J. and Engelhardt, M., (1986), Approximate distributional results based on the maximum likelihood estimators for the Weibull distribution, *Journal of Quality Technology*, 18, pp. 174-181.

Harter, L.H., (1978), A bibliography of extreme value theory, International Statistical Review, 46, pp. 279-036.

International Mathematical and Statistical Libraries, Inc., (1987), IMSL library Reference Manual, Houston.

Johnson, N.L. and Kotz, S., (1970), Distributions in Statistics : Continuous Univariate Distributions - 1, Houghton Mifflin Co., Boston.

Jones, R.A., Scholz, F.W. Ossiander, M. and Shorack, G.R., (1985), Tolerance bounds for log gamma regression models, *Technometrics*, 27, pp. 109-118.

Lawless, J.F., (1982), Statistical Models and Methods for Lifetime Data, First Edition, John Wiley and Sons, Inc.

Lieberman, G. and Miller, R., (1963), Simultaneous tolerance intervals in regression, Biometrika, 50, pp. 155-168.

Limam, M.M.T. and Thomas, D.R., (1988a), Simultaneous tolerance intervals for the linear regression model, *Journal of the American Statistical Association*, 83, 403, pp. 801-804.

Limam, M.M.T. and Thomas, D.R., (1988b), Simultaneous tolerance intervals in the random one-way model with covariates, Communications in Statistics, Simulation and Computation, 17, 3, pp. 1007-1019.

Miller, R.G. Jr, (1981), Simultaneous Statistical Inference, Second Edition, Springer-Verlag, New York.

Nelson, W.B., (1970), Statistical methods for accelerated lifetest data - the inverse power law model, General Electric Co. Technical Report 71-C-011, Schenectady, New York.

Statistical Analysis System, (1985), SAS/Stat Guide, SAS Institute Inc,. Box 8000, Cary, North Carolina.

Verhagen, A.M., (1961), The estimation of regression and error scale parameter when the joint distribution of the errors is of any continuous form and known apart from a scale parameter, Biometrika, 48, pp. 125-132.

Williams, J.S., (1962), A confidence interval for variance components, Biometrika, 49, pp. 278-281.

Wilson, A.L., (1967), An approach to simultaneous tolerance intervals in regression, Annals of Statistics, 30, pp. 939-959.

Table 1. Empirical confidence levels of the tolerance bounds.

	P			
n	.80	.90	.95	.99
10	.935	.936	.935	.934
15	.945	.947	.946	.949
20	.953	.954	.952	.951
30	.956	.955	.954	.955
100	.955	.954	.954	.955