Unified Confidence Bounds for Censored Weibull Data With Covariates

Fritz Scholz
The Boeing Company *

April 1, 2002

Abstract

Data from a 2-parameter Weibull distribution (with covariates) is traditionally analyzed on a logarithmic scale to take advantage of the resulting location-scale nature of the transformed data $Y_1, \ldots, Y_n$. Quantities of interest are the regression parameters $\boldsymbol{\beta}$, the scale parameter σ, the p-quantile $y_p(\boldsymbol{u}) = \boldsymbol{u}'\boldsymbol{\beta} + \sigma \log(-\log(1-p))$, the tail probability $p(y|\boldsymbol{u}) = P(Y \le y|\boldsymbol{u}) = 1 - \exp(-\exp((y - \boldsymbol{u}'\boldsymbol{\beta})/\sigma)))$, and the failure rate function $r(y, \boldsymbol{u}) = [\exp((y - \boldsymbol{u}'\boldsymbol{\beta})/\sigma)]/\sigma$ for a given p-dimensional covariate vector $\boldsymbol{u}$. Usually such data is partially obscured by some sort of censoring which (aside from the case of type II censoring) does not allow exact confidence bounds for these quantities. Thus one resorts to large sample approximations from maximum likelihood theory. Unfortunately this has led to different types of approximations depending on the quantity of interest, see Meeker and Escobar (1998). For example, confidence bounds for $y_p(\boldsymbol{u})$ and $p(y|\boldsymbol{u})$ are not always monotone in p or y and thus are not constructed as inverses to each other as would be the case when exact methods are possible. Also, one usually invokes approximate normality of the m.l.e.'s $\widehat{y}_p(\boldsymbol{u})$ and $\log(\widehat{\sigma})$ (the latter for producing better results) with the apparent inconsistency that $\widehat{y}_p(\mathbf{0}) = \widehat{\sigma} \log(-\log(1-p))$ invokes the inferior approximation for $\widehat{\sigma}$. We resolve these problems by invoking either the approximate $(p+1)$-variate normal approximation for $((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}, \log(\widehat{\sigma}))$ or its bootstrapped approximating distribution. This resolves all the above problems in a clean fashion and in the bootstrap case it leads back to the approach by Robinson (1983).

*Mathematics & Computing Technology, P.O. Box 3707, MS 7L-22,
Seattle WA 98124-2207, e-mail: fritz.scholz@boeing.com

1 Weibull Regression Model for Right Censored Data

Consider the following linear model:

$$y_i = \sum_{j=1}^{p} u_{ij}\beta_j + \sigma\epsilon_i = \boldsymbol{u}_i'\boldsymbol{\beta} + \sigma\epsilon_i \qquad i = 1, \ldots, n$$

where $\epsilon_1, \ldots, \epsilon_n$ are independent random errors, identically distributed according to the extreme value or Gumbel distribution with density $g(x) = \exp[x - \exp(x)]$ and cumulative distribution function $G(x) = 1 - \exp[-\exp(x)]$. The p-quantile of G is denoted by w_p, i.e., $w_p = \log(-\log(1-p))$.

The $n \times p$ matrix $\boldsymbol{U} = (u_{ij})$ of constant regression covariates is assumed to be of full rank p, with $n > p$. The unknown parameters $\sigma, \beta_1, \ldots, \beta_p$ are estimated by the method of maximum likelihood, obtained as the solution to the likelihood equations, provided it exists.

The above model can also arise from the following Weibull regression model:

$$P(T_i \leq t) = 1 - \exp\left(-\left[\frac{t}{\alpha(\boldsymbol{u}_i)}\right]^{\eta}\right)$$

which, after using the log-transformation $Y_i = \log(T_i)$, results in

$$P(Y_i \leq y) = 1 - \exp\left[-\exp\left(\frac{y - \log[\alpha(\boldsymbol{u}_i)]}{(1/\eta)}\right)\right] = 1 - \exp\left[-\exp\left(\frac{y - \mu(\boldsymbol{u}_i)}{\sigma}\right)\right] .$$

Using the identifications $\sigma = 1/\eta$ and $\mu(\boldsymbol{u}_i) = \log[\alpha(\boldsymbol{u}_i)] = \boldsymbol{u}_i'\boldsymbol{\beta}$ this reduces to the previous linear model with the density g.

Rather than observing the responses y_i completely, the data are allowed to be censored, i.e., for each response y_i one either observes it or some censoring time c_i. The response y_i is observed whenever $c_i \geq y_i$ and otherwise one observes c_i, and one knows whether the observation is a y_i or a c_i. One will also always know the corresponding covariates $u_{ij}, j = 1, \ldots, p$ for $i = 1, \ldots, n$. Such censoring is called multiple right censoring. Thus the data consist of

$$\boldsymbol{S} = \{(x_1, \delta_1, \boldsymbol{u}_1), \ldots, (x_n, \delta_n, \boldsymbol{u}_n)\} ,$$

where $x_i = y_i$ and $\delta_i = 1$ when $y_i \leq c_i$, and $x_i = c_i$ and $\delta_i = 0$ when $y_i > c_i$. The number of uncensored observations is denoted by $r = \sum_{i=1}^{n} \delta_i$ and the index sets of uncensored and censored observations by $\mathcal{D}$ and $\mathcal{C}$, respectively, i.e.,

$$\mathcal{D} = \{i : \delta_i = 1, i = 1, \ldots, n\} = \{i_1, \ldots, i_r\} \quad \text{and} \quad \mathcal{C} = \{i : \delta_i = 0, i = 1, \ldots, n\} .$$

The likelihood function of the data $\boldsymbol{S}$ is

$$L(\boldsymbol{\beta},\sigma)=\prod_{i\in\mathcal{D}}\frac{1}{\sigma}\exp\left[\frac{x_i-\boldsymbol{u}_i'\boldsymbol{\beta}}{\sigma}-\exp\left(\frac{x_i-\boldsymbol{u}_i'\boldsymbol{\beta}}{\sigma}\right)\right]\prod_{i\in\mathcal{C}}\exp\left[-\exp\left(\frac{x_i-\boldsymbol{u}_i'\boldsymbol{\beta}}{\sigma}\right)\right]$$

and the corresponding log-likelihood is

$$\begin{aligned}\ell(\boldsymbol{\beta},\sigma) &= \log[L(\boldsymbol{\beta},\sigma)]\\ &= \sum_{i\in\mathcal{D}}\left[\frac{x_i-\boldsymbol{u}_i'\boldsymbol{\beta}}{\sigma}-\exp\left(\frac{x_i-\boldsymbol{u}_i'\boldsymbol{\beta}}{\sigma}\right)\right]-\sum_{i\in\mathcal{D}}\log\sigma-\sum_{i\in\mathcal{C}}\exp\left(\frac{x_i-\boldsymbol{u}_i'\boldsymbol{\beta}}{\sigma}\right)\ .\end{aligned}$$

In [10] conditions were given under which maximum likelihood estimates or m.l.e.'s of the parameters $(\beta_1,\ldots,\beta_p,\sigma)$ exist and are the unique solution of the likelihood equations

$$\frac{\partial\ell(\boldsymbol{\beta},\sigma)}{\partial\sigma}=0 \qquad \text{and} \qquad \frac{\partial\ell(\boldsymbol{\beta},\sigma)}{\partial\beta_j}=0\ , \qquad \text{for} \quad j=1,\ldots,p\ .$$

Also given were algorithmic details for obtaining these estimates.

2 Normal Approximations

Under the conditions of Theorem 2 of [10] the likelihood equations have a unique solution, the m.l.e. $(\widehat{\boldsymbol{\beta}},\widehat{\sigma})$, and the matrix $\widehat{\boldsymbol{H}}$ of second partial derivatives of ℓ at this solution is

$$\widehat{\boldsymbol{H}}=-\frac{1}{\widehat{\sigma}^2}\begin{pmatrix}\sum_{i=1}^n\exp(\hat{z}_i)\boldsymbol{u}_i\boldsymbol{u}_i' & \sum_{i=1}^n\hat{z}_i\exp(\hat{z}_i)\boldsymbol{u}_i\\ \sum_{i=1}^n\hat{z}_i\exp(\hat{z}_i)\boldsymbol{u}_i' & r+\sum_{i=1}^n\hat{z}_i^2\exp(\hat{z}_i)\end{pmatrix}=-\frac{1}{\widehat{\sigma}^2}\,\widehat{\boldsymbol{B}}\ ,$$

where $\hat{z}_i=(x_i-\boldsymbol{u}_i'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$. The negative of $\widehat{\boldsymbol{H}}$ is also known as the observed Fisher information matrix. Denoting $\hat{w}_i=\exp(\hat{z}_i)/\sum_{j=1}^n\exp(\hat{z}_j)$ and $\widehat{W}=\sum_{j=1}^n\exp(\hat{z}_j)$ one can write

$$\widehat{\boldsymbol{B}}=\widehat{W}\begin{pmatrix}\sum_{i=1}^n\hat{w}_i\boldsymbol{u}_i\boldsymbol{u}_i' & \sum_{i=1}^n\hat{w}_i\hat{z}_i\boldsymbol{u}_i\\ \sum_{i=1}^n\hat{w}_i\hat{z}_i\boldsymbol{u}_i' & r/\widehat{W}+\sum_{i=1}^n\hat{w}_i\hat{z}_i^2\end{pmatrix}=\widehat{W}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{b}\\ \boldsymbol{b}' & c\end{pmatrix}\ .$$

In [10] this matrix $\widehat{\boldsymbol{B}}$ was shown to be positive definite.

It is conventional practice to invoke large sample maximum likelihood theory to claim that

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\sigma} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta} \\ \sigma \end{pmatrix} \;\sim\; \mathcal{N}_{p+1}\left(\mathbf{0}, -\widehat{\boldsymbol{H}}^{-1}\right) = \mathcal{N}_{p+1}\left(\mathbf{0}, \widehat{\sigma}^2 \widehat{\boldsymbol{B}}^{-1}\right) \;, \tag{1}$$

where the $p+1$ dimensional normal distribution with mean vector $\mathbf{0}$ and covariance matrix $-\widehat{\boldsymbol{H}}^{-1}$ serves as an approximation for the distribution of the left-hand side. The fact that the observed Fisher information is used in expressing the covariance matrix is a matter of convenience, since the estimated Fisher information matrix is usually difficult to evaluate in the context of censoring.

The estimated Fisher information matrix is obtained by computing the Fisher information matrix $I(\boldsymbol{\beta}, \sigma) = -E(\boldsymbol{H})$, where $\boldsymbol{H}$ is the matrix of second partial derivatives of ℓ evaluated at $(\boldsymbol{\beta}, \sigma)$ and the expectation E is evaluated under $(\boldsymbol{\beta}, \sigma)$ as well, and then estimating $I(\boldsymbol{\beta}, \sigma)$ by $I(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})$. Note that the lack of censoring times for any of the n cases may even make it impossible to evaluate the Fisher information since the evaluation of $\ell(\boldsymbol{\beta}, \sigma)$ (and thus its derivatives) involves the determination of $i \in \mathcal{C}$, which can only be done when the censoring time is known for case i.

This substitution of $-\widehat{\boldsymbol{H}}^{-1}$ for $I(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})^{-1}$ is usually justified by the consistency properties of m.l.e.'s and continuity considerations. We will not dwell on conditions that support the validity of approximate normality and the above use of the observed Fisher information in the presence of right censoring and covariates.

Even if the large sample approximation is theoretically justified, one still has to explore via simulation to what extent the sample is large enough for the approximation to be reasonable. Sample size alone will not decide the issue. There is also the question of how much censoring can be tolerated. It is unclear whether the number of uncensored cases needs to approach infinity in order for asymptotic results to hold. As Le Cam[5] says in his *Principle 7:* "If you need to use asymptotic arguments, don't forget to let your number of observations tend to infinity."

Although the above focusses on multiple right censoring we point out that the idea of the new method also applies to other types of censoring, such as interval censoring.

When using the observed Fisher information in the normal approximation we have

$$\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \alpha\widehat{\sigma} \;\sim\; \mathcal{N}\left(\boldsymbol{u}'\boldsymbol{\beta} + \alpha\sigma, \tau^2(\boldsymbol{u}, \alpha)\right) \;, \tag{2}$$

where

$$\tau^2(\boldsymbol{u}, \alpha) \;=\; \widehat{\sigma}^2 \,(\boldsymbol{u}', \alpha)\, \widehat{\boldsymbol{B}}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \alpha \end{pmatrix} \tag{3}$$

For later computational reference it is useful to develop an alternate expression for the quadratic form in (3). To this end we use the partitioned form of $\widehat{\boldsymbol{B}}$ and note that its inverse can be expressed as

$$\widehat{\boldsymbol{B}}^{-1} = \frac{1}{\widehat{W}} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}' & c \end{pmatrix}^{-1} = \frac{1}{\widehat{W}\rho} \begin{pmatrix} \rho\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{b}\boldsymbol{b}'\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{b} \\ -\boldsymbol{b}'\boldsymbol{A}^{-1} & 1 \end{pmatrix}, \tag{4}$$

where $\rho = c - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} > 0$. In fact, it was shown previously in [10] (when establishing that $\widehat{\boldsymbol{B}}$ is positive definite) that

$$\rho = \frac{r}{\widehat{W}} + \sum_{i=1}^{n} \hat{w}_i \left(\hat{z}_i - \boldsymbol{u}_i'\boldsymbol{A}^{-1} \sum_{j=1}^{n} \hat{w}_j \hat{z}_j \boldsymbol{u}_j \right)^2 \geq \frac{r}{\widehat{W}} . \tag{5}$$

From the above form of the partitioned inverse matrix one can express the quadratic form in (3) as

$$(\boldsymbol{u}', \alpha)\, \widehat{\boldsymbol{B}}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \alpha \end{pmatrix} = \frac{1}{\widehat{W}\rho} \left[\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + (\alpha - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b})^2 \right] . \tag{6}$$

One construction of upper confidence bounds for the cumulative distribution function $G((y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma)$ for given threshold y_0 and covariate vector $\boldsymbol{u}$ involves a normal approximation for $(y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$. Again this can be obtained by the delta method from (1). One arrives then at the following approximation:

$$\frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} \sim \mathcal{N}\left(\frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}, \frac{1}{\rho\widehat{W}} \left[\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left(\frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b} \right)^2 \right] \right) . \tag{7}$$

Experience has shown that the normal approximation for $\widehat{\sigma}$ is not good in small samples or in samples involving only few failures. Partly this is due to the fact that $\widehat{\sigma}$ is naturally bounded from below by zero while the normal distribution has an infinite range. Thus one often invokes instead the approximate normality of $\widehat{\lambda} = \log(\widehat{\sigma})$. Using the delta method one finds as resulting approximation

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \log(\widehat{\sigma}) \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta} \\ \log(\sigma) \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\lambda} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta} \\ \lambda \end{pmatrix} \sim \mathcal{N}_{p+1}\left(\boldsymbol{0}, \widehat{\sigma}^2 \widehat{\boldsymbol{D}}\,\widehat{\boldsymbol{B}}^{-1}\widehat{\boldsymbol{D}}' \right) ,$$

where

$$\widehat{\boldsymbol{D}} = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{0}' & 1/\widehat{\sigma} \end{pmatrix} \quad \text{and } \boldsymbol{I}_p \text{ is a } p \times p \text{ identity matrix.}$$

A simple linear transformation yields

$$\begin{pmatrix} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma} \\ \widehat{\lambda} - \lambda \end{pmatrix} \sim \mathcal{N}_{p+1}\left(\boldsymbol{0}, \widehat{\boldsymbol{B}}^{-1} \right) \tag{8}$$

and using the partitioned form of $\widehat{\boldsymbol{B}}^{-1}$ on the right side of (4) results in

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma} \\ \widehat{\lambda} - \lambda \end{pmatrix} \sim \mathcal{N}_2\left(\mathbf{0}, \frac{1}{\widehat{W}\rho}\begin{pmatrix} c_{11} & c_{12} \\ c_{12} & 1 \end{pmatrix}\right), \tag{9}$$

where $c_{11} = \rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left(\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b}\right)^2$ and $c_{12} = -\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b}$.
To obtain the conventional confidence bounds for the failure rate functions in the Weibull or Gumbel model we need to invoke the following normal approximation. From (1) one obtains

$$\begin{aligned} \begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}'\widehat{\boldsymbol{\beta}} \\ \widehat{\sigma} \end{pmatrix} - \begin{pmatrix} \boldsymbol{u}'\boldsymbol{\beta} \\ \sigma \end{pmatrix} \\ \sim \mathcal{N}_2\left(\mathbf{0}, \widehat{\sigma}^2 \begin{pmatrix} \boldsymbol{u}' & 0 \\ \mathbf{0}' & 1 \end{pmatrix} \widehat{\boldsymbol{B}}^{-1} \begin{pmatrix} \boldsymbol{u} & \mathbf{0} \\ 0 & 1 \end{pmatrix}\right). \end{aligned} \tag{10}$$

For a function $f(\mu, \sigma)$ with gradient vector

$$(d_1, d_2) = \left(\frac{\partial f(\mu,\sigma)}{\partial \mu}, \frac{\partial f(\mu,\sigma)}{\partial \sigma}\right)\Bigg|_{\mu=\widehat{\mu},\sigma=\widehat{\sigma}}$$

the delta method yields

$$\begin{aligned} f(\widehat{\mu}, \widehat{\sigma}) - f(\mu, \sigma) &\sim \mathcal{N}\left(0, \widehat{\sigma}^2(d_1, d_2)\begin{pmatrix} \boldsymbol{u}' & 0 \\ \mathbf{0}' & 1 \end{pmatrix} \widehat{\boldsymbol{B}}^{-1} \begin{pmatrix} \boldsymbol{u} & \mathbf{0} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}\right) \\ &\sim \mathcal{N}\left(0, \widehat{\sigma}^2(d_1\boldsymbol{u}', d_2)\widehat{\boldsymbol{B}}^{-1}\begin{pmatrix} d_1\boldsymbol{u}' \\ d_2 \end{pmatrix}\right) \end{aligned} \tag{11}$$

and using the expression in (6) one can express the variance of this normal approximation as

$$\frac{\widehat{\sigma}^2}{\widehat{W}\rho}\left[\rho d_1^2\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + (d_2 - d_1\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b})^2\right].$$

For the Gumbel model the failure rate function is

$$r_G(y_0, \boldsymbol{u}) = \frac{\frac{1}{\sigma}g\left(\frac{y_0-\boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\right)}{1 - G\left(\frac{y_0-\boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\right)} = \frac{1}{\sigma}\exp\left(\frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\right)$$

and its logarithm is

$$\log\left(r_G(y_0, \boldsymbol{u})\right) = -\log(\sigma) + \frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}.$$

Using (11) with $f(\mu,\sigma) = -\log(\sigma) + (y_0-\mu)/\sigma$ one gets the following normal approximation for the maximum likelihood estimate

$$\log\left(\widehat{r}_G(y_0,\boldsymbol{u})\right) = -\log(\widehat{\sigma}) + \frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}$$

of $\log\left(r_G(y_0,\boldsymbol{u})\right)$, namely

$$\log\left(\widehat{r}_G(y_0,\boldsymbol{u})/r_G(y_0,\boldsymbol{u})\right) \sim \mathcal{N}\left(0, \widehat{\tau}_G^2(y_0,,\boldsymbol{u})\right) \tag{12}$$

where

$$\widehat{\tau}_G^2(y_0,\boldsymbol{u}) = \frac{1}{\widehat{W}\rho}\left(\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left[\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b} - 1 - \frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right]^2\right)$$

On the other hand, for the corresponding Weibull model with $\log(\alpha) = \boldsymbol{u}'\boldsymbol{\beta}$, $\eta = 1/\sigma$ and $t_0 = \exp(y_0)$ the failure rate function is

$$r_W(t_0,\boldsymbol{u}) = \frac{\eta}{\alpha}\left(\frac{t_0}{\alpha}\right)^{\eta-1}$$

and its logarithm is

$$\log\left(r_W(t_0,\boldsymbol{u})\right) = -\log(\sigma) + \frac{\log(t_0) - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma} - \log(t_0)\ .$$

Note that this form differs from $\log\left(r_G(y_0,\boldsymbol{u})\right)$ by the term $\log(t_0)$. This is caused by the differentiation process that is involved in the definition of the failure rate.
We use

$$\log\left(\widehat{r}_W(t_0,\boldsymbol{u})\right) = -\log(\widehat{\sigma}) + \frac{\log(t_0) - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} - \log(t_0)$$

as maximum likelihood estimate of $\log\left(r_W(t_0)\right)$ and as normal approximation the delta method yields

$$\log\left(\widehat{r}_W(t_0,\boldsymbol{u})/r_W(t_0,\boldsymbol{u})\right) \sim \mathcal{N}\left(0, \widehat{\tau}_W^2(t_0,\boldsymbol{u})\right) \tag{13}$$

where

$$\widehat{\tau}_W^2(t_0,\boldsymbol{u}) = \frac{1}{\widehat{W}\rho}\left(\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left[\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b} - 1 - \frac{\log(t_0) - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right]^2\right)\ .$$

3 Two Quantile Lower Confidence Bounds

In this section we present the traditional lower confidence bound for the p-quantile $y_p(\boldsymbol{u}) = \boldsymbol{u}'\boldsymbol{\beta} + \sigma w_p$ at a given covariate vector $\boldsymbol{u}$ and introduce a new proposal. We focus on lower bounds since a $100\gamma\%$ lower bound is also a $100(1-\gamma)\%$ upper bound for the same target quantity.

Both lower bounds are of the form $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + t\widehat{\sigma}$, but different types of normal approximations are invoked. It is shown that the traditional lower confidence bound may have certain defects in small samples or in samples with few failures. The new lower bound avoids these defects.

3.1 Traditional Quantile Lower Confidence Bound

Since $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma} w_p$ is the m.l.e. of the p-quantile $y_p(\boldsymbol{u})$ and since

$$\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma} w_p \sim \mathcal{N}(\boldsymbol{u}'\boldsymbol{\beta} + \sigma w_p, \tau^2(\boldsymbol{u}, w_p))$$

one common and natural lower confidence bound for $y_p(\boldsymbol{u})$, see [4] or [7] for example, is

$$\widehat{y}_{p,L1}(\gamma, \boldsymbol{u}) = \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma} w_p - z_\gamma \tau(\boldsymbol{u}, w_p) \ ,$$

where z_γ is the γ-quantile of the standard normal distribution function. Since w_p affects $\widehat{y}_{p,L1}(\gamma, \boldsymbol{u})$ both through $\widehat{\sigma} w_p$ and through $\tau(\boldsymbol{u}, w_p)$ it is not immediately clear whether $\widehat{y}_{p,L1}(\gamma, \boldsymbol{u})$ is a monotone function of w_p. Since the target $y_p(\boldsymbol{u}) = \boldsymbol{\beta} + \widehat{\sigma} w_p$ is a monotone increasing function of w_p one would naturally require that any reasonable lower bound also be monotone increasing in w_p.

The monotonicity properties of $\widehat{y}_{p,L1}(\gamma, \boldsymbol{u})$ are equivalent to the monotonicity properties of

$$H_\gamma(w_p) = w_p - z_\gamma \frac{1}{\sqrt{\widehat{W}\rho}} \sqrt{\rho \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + (w_p - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b})^2} \ .$$

Note that

$$H'_\gamma(w_p) = \frac{\partial H_\gamma(w_p)}{\partial w_p} = 1 - \psi_\gamma \frac{w_p - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b}}{\sqrt{\rho \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + (w_p - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b})^2}}$$

with $\psi_\gamma = z_\gamma / \sqrt{\widehat{W}\rho}$.

When $|\psi_\gamma| \leq 1$ we clearly have $H'_\gamma(w_p) > 0$ for all w_p and in that case the lower confidence bounds $\widehat{y}_{p,L1}(\gamma, \boldsymbol{u})$ are monotone increasing in w_p. Since $H'_\gamma(w_p) \approx w_p(1-\psi_\gamma)$

for $|w_p| \approx \infty$ it follows that in this case $\widehat{y}_{p,L1}(\gamma, \boldsymbol{u})$ increases from $-\infty$ to ∞ as p increases from 0 to 1.

When $\psi_\gamma > 1$ one has

$$H'_\gamma(w_p) > 0 \quad \text{for} \quad w_p < \bar{w} \quad \text{and} \quad H'_\gamma(w_p) < 0 \quad \text{for} \quad w_p > \bar{w}$$

and for $\psi_\gamma < -1$ one has

$$H'_\gamma(w_p) < 0 \quad \text{for} \quad w_p < \bar{w} \quad \text{and} \quad H'_\gamma(w_p) > 0 \quad \text{for} \quad w_p > \bar{w}$$

where

$$\bar{w} = \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b} - \text{sign}(\psi_\gamma)\sqrt{\frac{\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\psi_\gamma^2 - 1}}.$$

The monotonicity of the lower bound $\widehat{y}_{p,L1}(\gamma, \boldsymbol{u})$ depends on the magnitude of $|\psi_\gamma|$ and that is governed by the proportionality factor z_γ but also by the quantity $\widehat{W}\rho$. From (5) we have that $\widehat{W}\rho \geq r$ so that $|\psi_\gamma| \ll |z_\gamma|/\sqrt{r}$ which shows the impact of the number of uncensored observations on this issue.

Aside from the monotonicity issue there is also the issue of approximation quality and hence the coverage of the resulting bounds. The quality of the invoked normal approximation

$$\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma}w_p \sim \mathcal{N}(\boldsymbol{u}'\boldsymbol{\beta} + \sigma w_p, \tau^2(\boldsymbol{u}, w_p))$$

will clearly depend on the factor w_p. For example, when $w_p = 0$, i.e., when $p = 1 - \exp(-1) \approx .63212$, the effect of the possibly poor normal approximation for $\widehat{\sigma}$ would not be felt, while for p close to 0 or 1, i.e., for extreme quantiles, the term $\widehat{\sigma}w_p$ would tend to dominate in the above approximation and then the normal approximation would be poor in small samples or when there are few failures. The same issue arises for all p when $\boldsymbol{u} = \boldsymbol{0}$ or approximately so when $\boldsymbol{u} \approx \boldsymbol{0}$. Both issues (the possible lack of monotonicity and the poor normal approximation of $\widehat{\sigma}$) are addressed by the new proposal in the next section.

3.2 New Quantile Lower Confidence Bound

Another construction of a lower confidence bound is

$$\widehat{y}_{p,L2}(\gamma, p, \boldsymbol{u}) = \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma}w_p - k\widehat{\sigma}$$

where one finds the factor $k = k(w_p)$ such that $P\left(\widehat{y}_{p,L2}(\gamma, p, \boldsymbol{u}) \le y_p(\boldsymbol{u})\right) \approx \gamma$, namely

$$\begin{aligned}
\gamma &\approx P\left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - k)\widehat{\sigma} \le \boldsymbol{u}'\boldsymbol{\beta} + w_p\sigma\right) \\
&= P\left(\frac{\boldsymbol{u}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\widehat{\sigma}} + w_p(1 - \sigma/\widehat{\sigma}) \le k\right) \\
&= P\left(V_1 + w_p[1 - \exp(-V_2)] \le k\right) = P\left(V_1 \le k - w_p + w_p \exp(-V_2)\right) \;,
\end{aligned} \tag{14}$$

where for

$$V_1 = \frac{\boldsymbol{u}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\widehat{\sigma}} \quad \text{and} \quad V_2 = \widehat{\lambda} - \lambda$$

we invoke the bivariate normal distribution given in (9). Using this approximation we can write the defining equation for $k = k(w_p)$ as

$$\begin{aligned}
\gamma &= P\left(V_1 \le k - w_p + w_p \exp(-V_2)\right) \\
&= \int_{-\infty}^{\infty} \Phi\left(\frac{k - w_p + w_p \exp(-v_2) - v_2\sigma_{12}/\sigma_2^2}{\sigma_{1|2}}\right) \frac{1}{\sigma_2}\varphi(v_2/\sigma_2)\, dv_2 \;,
\end{aligned} \tag{15}$$

where $\Phi(z)$ and $\varphi(z)$ are the standard normal distribution function and density, $\sigma_2^2 = 1/(\widehat{W}\rho)$ is the variance of V_2, and $v_2\sigma_{12}/\sigma_2^2$ and $\sigma_{1|2}^2 = \sigma_{11}\left(1 - \sigma_{12}^2/(\sigma_{11}\sigma_{22})\right)$ are the mean and variance of the conditional distribution of V_1 given V_2 with $\sigma_{22} = \sigma_2^2$, $\sigma_{12} = c_{12}/(\widehat{W}\rho)$, and $\sigma_{11} = c_{11}/(\widehat{W}\rho)$. It is easily seen that $\sigma_{1|2}^2 = \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}/\widehat{W}$.

Solving the equation (15) for $k = k(w_p)$ involves combining a numerical integration with a root solving algorithm. One easily sees that there is a unique solution $k = k(w_p)$ for each $w_p \in R$, hence for each $p \in (0, 1)$. Also, it is easily shown that $h(w_p) = w_p - k(w_p)$ is strictly increasing from $-\infty$ to ∞ as w_p increases from $-\infty$ to ∞. This avoids the monotonicity problems that can occur for the traditional confidence bounds when dealing with small samples, few failures, or extreme confidence levels (high values of $|z_\gamma|$).

The above invoked strictly increasing behavior of $h(w)$, although quite evident from equation (15), will now be stated and proved in a more general setting for more general joint distributions of (V_1, V_2) than the bivariate normal distribution.

Lemma 1: For any joint distribution of (V_1, V_2) the function $h(w)$, defined as the largest value satisfying $\gamma \le P[h(w) \le w \exp(-V_2) - V_1]$, is nondecreasing in w.

If in addition Lebesgue measure λ_2 on R^2 and the joint distribution of (V_1, V_2) are absolutely continuous with respect to each other, i.e., with A denoting a Borel set in R^2

$$\lambda_2(A) = 0 \quad \Longleftrightarrow \quad P[(V_1, V_2) \in A] = 0 \;,$$

then $h(w)$ is strictly increasing in w and is defined uniquely by

$$\gamma = P[h(w) \leq w \exp(-V_2) - V_1] \; .$$

Proof: Let $w_1 < w_2$ and assume that $h(w_1) > h(w_2)$. By definition of $h(w_2)$ we have

$$\gamma \leq P[h(w_2) \leq w_2 \exp(-V_2) - V_1]$$

and since $h(w_2)$ is the largest such value we must have

$$\gamma > P[h(w_1) \leq w_2 \exp(-V_2) - V_1]$$

and from $w_1 < w_2$ we then get

$$\gamma > P[h(w_1) \leq w_2 \exp(-V_2) - V_1] \geq P[h(w_1) \leq w_1 \exp(-V_2) - V_1] \geq \gamma \; ,$$

where the last inequality comes from the definition of $h(w_1)$. Thus we have a contradiction and we must have $h(w_1) \leq h(w_2)$.
The absolute continuity of the (V_1, V_2) distribution with respect to λ_2 implies that $w \exp(-V_2) - V_1$ has a continuous distribution and thus the largest $h(w)$ with $\gamma \leq P[h(w) \leq w \exp(-V_2) - V_1]$ actually uniquely satisfies $\gamma = P[h(w) \leq w \exp(-V_2) - V_1]$. The uniqueness follows similarly as the argument given below.
Now let $w_1 < w_2$ and assume that $h(w_1) = h(w_2) = h_0$. Then we have

$$\gamma = P[h_0 \leq w_1 \exp(-V_2) - V_1] \quad \text{and} \quad \gamma = P[h_0 \leq w_2 \exp(-V_2) - V_1]$$

or

$$P[(h_0 + V_1) \exp(V_2) \leq w_1] = \gamma \quad \text{and} \quad P[(h_0 + V_1) \exp(V_2) \leq w_2] = \gamma$$

which implies that

$$P[w_1 < (h_0 + V_1) \exp(V_2) \leq w_2] = 0 \; . \tag{16}$$

Considering the 1-1 mapping $(V_1, V_2) \longrightarrow (U_1, U_2)$ with $U_1 = (h_0 + V_1) \exp(V_2)$ and $U_2 = V_2$ we have that the distributions of (V_1, V_2) and (U_1, U_2) dominate each other and equation 16 implies that $P(a_1 < U_1 < b_1, a_2 < U_2 < b_2) = 0$ for some nondegenerate rectangle $A = (a_1, b_1) \times (a_2, b_2)$ and since $\lambda_2(A) = (b_1 - a_1)(b_2 - a_2) > 0$ we have arrived at a contradiction. Thus $w_1 < w_2 \Longrightarrow h(w_1) < h(w_2)$ q.e.d.

3.3 Bracketing the Root $k = k(w_p)$

Solving $P(V_1 - w_p \exp(-V_2) + w_p \le k) = \gamma$ for k usually requires that the root be bracketed. Although there are generic algorithms for this, they typically require a few function evaluations. Here we give some simple brackets. Denote by $v_i(\alpha)$ the α-quantile of V_i, i.e.,

$$v_{1,\alpha} = z_\alpha \sqrt{\frac{\rho \boldsymbol{u}' \boldsymbol{A}^{-1} \boldsymbol{u} + (\boldsymbol{u}' \boldsymbol{A}^{-1} \boldsymbol{b})^2}{\widehat{W} \rho}} \quad \text{and} \quad v_{2,\alpha} = \frac{z_\alpha}{\sqrt{\widehat{W} \rho}} , \tag{17}$$

where z_α is the α-quantile of the standard normal distribution. We distinguish two cases, $w_p \ge 0$ and $w_p < 0$.

Case $w_p \ge 0$: Let $\alpha = (1 - \gamma)/2$ and invoke the Bonferroni inequality

$$P(V_1 < v_{1,1-\alpha}, V_2 < v_{2,1-\alpha}) \ge 1 - 2\alpha = \gamma .$$

Then

$$\begin{aligned} \gamma &= P(V_1 - w_p \exp(-V_2) + w_p \le k) \le P(V_1 < v_{1,1-\alpha}, V_2 < v_{2,1-\alpha}) \\ &\le P(V_1 - w_p \exp(-V_2) + w_p < v_{1,1-\alpha} - w_p \exp(-v_{2,1-\alpha}) + w_p) \end{aligned}$$

and this implies $k \le v_{1,1-\alpha} - w_p \exp(-v_{2,1-\alpha}) + w_p$.

Next let $\beta = \gamma/2$ and by Boole's inequality we have

$$P(V_1 < v_{1,\beta} \cup V_2 < v_{2,\beta}) \le 2\beta = \gamma .$$

Then

$$\begin{aligned} &P(V_1 - w_p \exp(-V_2) + w_p < v_{1,\beta} - w_p \exp(-v_{2,\beta}) + w_p) \\ &\qquad \le P(V_1 < v_{1,\beta} \cup V_2 < v_{2,\beta}) \le \gamma = P(V_1 - w_p \exp(-V_2) + w_p \le k) \end{aligned}$$

and this implies $v_{1,\beta} - w_p \exp(-v_{2,\beta}) + w_p \le k$. We thus can bracket the solution k as follows

$$v_{1,\beta} - w_p \exp(-v_{2,\beta}) + w_p \le k \le v_{1,1-\alpha} - w_p \exp(-v_{2,1-\alpha}) + w_p .$$

Case $w_p < 0$: With $\alpha = (1 - \gamma)/2$ we get

$$P(V_1 < v_{1,1-\alpha}, V_2 > v_{2,\alpha}) \ge 1 - 2\alpha = \gamma .$$

Then

$$\begin{aligned} \gamma &= P(V_1 - w_p \exp(-V_2) + w_p \le k) \le P(V_1 < v_{1,1-\alpha}, V_2 > v_{2,\alpha}) \\ &\le P(V_1 - w_p \exp(-V_2) + w_p < v_{1,1-\alpha} - w_p \exp(-v_{2,\alpha}) + w_p) \end{aligned}$$

and this implies $k \leq v_{1,1-\alpha} - w_p \exp(-v_{2,\alpha}) + w_p$.
Next let $\beta = \gamma/2$ and by Boole's inequality we have

$$P(V_1 < v_{1,\beta} \cup V_2 > v_{2,1-\beta}) \leq 2\beta = \gamma \ .$$

Then

$$\begin{aligned}\gamma &= P(V_1 - w_p \exp(-V_2) + w_p \leq k) \geq P(V_1 < v_{1,\beta} \cup V_2 > v_{2,1-\beta}) \\ &\geq P(V_1 - w_p \exp(-V_2) + w_p < v_{1,\beta} - w_p \exp(-v_{2,1-\beta}) + w_p)\end{aligned}$$

and this implies $v_{1,\beta} - w_p \exp(-v_{2,1-\beta}) + w_p \leq k$. This leads to the following bracketing of k:

$$v_{1,\beta} - w_p \exp(-v_{2,1-\beta}) + w_p \leq k \leq v_{1,1-\alpha} - w_p \exp(-v_{2,\alpha}) + w_p \ .$$

4 Two Upper Confidence Bounds for the CDF

In this section we present two methods for constructing upper confidence bounds for the cumulative distribution function $G((y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma)$ for given threshold y_0 and covariate vector $\boldsymbol{u}$. We focus on upper bounds here, since they are dual to quantile lower bounds and since a $100\gamma\%$ upper bound is also a $100(1-\gamma)\%$ lower bound.

The first upper bound uses a traditional method (and there are others) and the second is based on the inversion of the new quantile lower bounds and is thus in line with classical theory when exact methods are possible, i.e., no normal approximations are invoked.

Of course one could also invert the traditional lower bounds for quantiles, but that may not always be possible because of monotonicity issues. Similarly one could invert the traditional upper bounds for the CDF to get lower bounds for quantiles, but again that runs into possible monotonicity problems, as will be shown. Thus we will not deal with inversions of traditional approximate bounds on quantiles or the CDF.

4.1 Traditional Upper Confidence Bounds for the CDF

Using the normal approximation (7) one gets the following $100\gamma\%$ upper confidence bounds for $(y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma$

$$\frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} + z_\gamma \frac{1}{\sqrt{\rho \widehat{W}}} \sqrt{\rho \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left(\frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b}\right)^2}$$

and thus as $100\gamma\%$ upper confidence bounds for $G((y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma)$ we have

$$\widehat{p}_{y,U1}(\gamma,\boldsymbol{u}) = G\left(\frac{y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} + z_\gamma \frac{1}{\sqrt{\rho\widehat{W}}}\sqrt{\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left(\frac{y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b}\right)^2}\right) .$$

This upper bound is guaranteed to give values within the interval $(0,1)$. This would not necessarily be the case when invoking an approximate normal distribution (again obtained by delta method) for $G((y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma})$ directly.

Since $G((y-\boldsymbol{u}'\boldsymbol{\beta})/\sigma)$ is monotone increasing in y one would expect that $\widehat{p}_{y,U1}(\gamma,\boldsymbol{u})$ also be monotone increasing in y. That is not always the case. The monotonicity properties of $\widehat{p}_{y,U1}(\gamma,\boldsymbol{u})$ are equivalent to the monotonicity properties of

$$\tilde{H}_\gamma(y) = \frac{y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} + z_\gamma \frac{1}{\sqrt{\rho\widehat{W}}}\sqrt{\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + \left(\frac{y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} - \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b}\right)^2} ,$$

which essentially already appeared previously in the form of $H_\gamma(w_p)$, if one identifies w_p with $(y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$ and changes z_γ to $-z_\gamma$. The latter sign change just means switching the confidence level from γ to $1-\gamma$, i.e.,

$$\tilde{H}_\gamma(y) = H_{1-\gamma}\left(\frac{y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right) \quad \text{and thus} \quad \tilde{H}'_\gamma(y) = H'_{1-\gamma}\left(\frac{y-\boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)\frac{1}{\widehat{\sigma}} .$$

Thus we can restate the monotonicity results as follows, using the notation

$$\psi_{1-\gamma} = z_{1-\gamma}/\sqrt{\widehat{W}\rho} = -z_\gamma/\sqrt{\widehat{W}\rho} = -\tilde{\psi}_\gamma .$$

When $|\tilde{\psi}_\gamma| = |-\psi_{1-\gamma}| \leq 1$, then the upper bound $\widehat{p}_{y,U1}(\gamma,\boldsymbol{u})$ is strictly increasing in y and increases from 0 to 1 as y increases from $-\infty$ to ∞.

When $\tilde{\psi}_\gamma > 1$, i.e., $\psi_{1-\gamma} < -1$, one has

$$\tilde{H}'_\gamma(y) < 0 \quad \text{for} \quad y < \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma}\bar{w} \quad \text{and} \quad \tilde{H}'_\gamma(y) > 0 \quad \text{for} \quad y > \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma}\bar{w}$$

and for $\tilde{\psi}_\gamma < -1$, i.e., $\psi_{1-\gamma} > 1$, one has

$$\tilde{H}'_\gamma(y) > 0 \quad \text{for} \quad y < \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma}\bar{w} \quad \text{and} \quad \tilde{H}'_\gamma(y) < 0 \quad \text{for} \quad y > \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + \widehat{\sigma}\bar{w} ,$$

where

$$\bar{w} = \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b} - \text{sign}(\psi_{1-\gamma})\sqrt{\frac{\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\psi_{1-\gamma}^2 - 1}} = \boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b} + \text{sign}(\tilde{\psi}_\gamma)\sqrt{\frac{\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u}}{\tilde{\psi}_\gamma^2 - 1}} .$$

4.2 New Upper Confidence Bounds for the CDF

This method is based on an inversion of the new method for quantile lower bounds. Because of the clean monotonicity properties of h, which was introduced in the construction of the new quantile lower bounds, one can invert the quantile lower confidence bounds to get upper confidence bounds for the cumulative distribution function of Y for any given covariate vector $\boldsymbol{u}$ and threshold value y, namely for

$$p_0 = p_0(y;\boldsymbol{u}) = P(Y \le y) = G\left(\frac{y - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\right) . \tag{18}$$

This is done as follows: For fixed $\boldsymbol{\beta}, \sigma$, and y let p_0 be as defined in (18) and hence $w_{p_0} = (y - \boldsymbol{u}'\boldsymbol{\beta})/\sigma$ or $y = \boldsymbol{u}'\boldsymbol{\beta} + \sigma w_{p_0}$ and thus

$$\begin{aligned}
\gamma &= P\left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + h(w_{p_0})\widehat{\sigma} \le \boldsymbol{u}'\boldsymbol{\beta} + \sigma w_{p_0}\right) = P\left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + h(w_{p_0})\widehat{\sigma} \le y\right) \\
&= P\left(h(w_{p_0}) \le \frac{y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right) = P\left(w_{p_0} \le h^{-1}\left(\frac{y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)\right) \\
&= P\left(p_0 \le G\left(h^{-1}\left(\frac{y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)\right)\right) .
\end{aligned}$$

Hence we can treat

$$\widehat{p}_{y,U2}(\gamma, \boldsymbol{u}) = G\left(h^{-1}\left(\frac{y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)\right)$$

as $100\gamma\%$ upper bound for p_0.

Since the evaluation of h^{-1} would involve another root solving step it would be useful if one could avoid this root solver on top of a root solver, since that increases the number of function evaluations quadratically. This can indeed be done directly. We need to find $w(y) = h^{-1}((y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma})$, i.e., solve $h(w(y)) = (y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$ for $w(y)$. Recall that when finding the quantile lower confidence bound we had to solve

$$\gamma = \int_{-\infty}^{\infty} \Phi\left(\frac{w_p \exp(-v_2) - h(w_p) - v_2\sigma_{12}/\sigma_2^2}{\sigma_{1|2}}\right) \frac{1}{\sigma_2}\varphi(v_2/\sigma_2)\, dv_2$$

for $h(w_p)$ for a given value of w_p. Since there is a 1-1 relationship between $h(w_p)$ and w_p we can use this defining equation also in reverse, namely solve

$$\gamma = \int_{-\infty}^{\infty} \Phi\left(\frac{w(y) \exp(-v_2) - h_y - v_2\sigma_{12}/\sigma_2^2}{\sigma_{1|2}}\right) \frac{1}{\sigma_2}\varphi(v_2/\sigma_2)\, dv_2$$

for $w(y)$ for a given value of $h_y = (y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$. We then take $G(w(y))$ as the desired upper bound for $P(Y \le y)$.

4.3 Bracketing the Root $w = w(y)$

Here we want to solve $P(V_1 - w \exp(-V_2) \le -h_y) = P((V_1 + h_y)\exp(V_2) \le w) = \gamma$ for $w = w(y)$ for given h_y. To facilitate the root solving we again provide brackets for w.
With $\gamma = 1 - 2\alpha$ and $v_{1,1-\alpha} + h_y \ge 0$ we invoke the Bonferroni inequality

$$P(V_1 + h_y \le v_{1,1-\alpha} + h_y, \exp(V_2) \le \exp(v_{2,1-\alpha})) \ge 1 - 2\alpha = \gamma \; .$$

Then

$$P((V_1 + h_y)\exp(V_2) \le w) = \gamma \le P((V_1 + h_y)\exp(V_2) \le (v_{1,1-\alpha} + h_y)\exp(v_{2,1-\alpha}))$$

which implies $w \le (v_{1,1-\alpha} + h_y)\exp(v_{2,1-\alpha})$.
For $v_{1,1-\alpha} + h_y < 0$ we invoke the Bonferroni inequality

$$P(V_1 + h_y \le v_{1,1-\alpha} + h_y, \exp(V_2) \ge \exp(v_{2,\alpha})) \ge 1 - 2\alpha = \gamma \; .$$

Then

$$P((V_1 + h_y)\exp(V_2) \le w) = \gamma \le P((V_1 + h_y)\exp(V_2) \le (v_{1,1-\alpha} + h_y)\exp(v_{2,\alpha}))$$

which implies $w \le (v_{1,1-\alpha} + h_y)\exp(v_{2,\alpha})$.
With $\beta = \gamma/2$ and $v_{1,\beta} + h_y \ge 0$ we invoke the Bonferroni inequality

$$\begin{aligned} 1 - \gamma &\le P(V_1 + h_y \ge v_{1,\beta} + h_y, \exp(V_2) \ge \exp(v_{2,\beta})) \\ &\le P((V_1 + h_y)\exp(V_2) \ge (v_{1,\beta} + h_y)\exp(v_{2,\beta})) \end{aligned}$$

and thus

$$P((V_1 + h_y)\exp(V_2) < (v_{1,\beta} + h_y)\exp(v_{2,\beta})) \le \gamma = P((V_1 + h_y)\exp(V_2) \le w)$$

which implies $(v_{1,\beta} + h_y)\exp(v_{2,\beta}) \le w$.
With $\beta = \gamma/2$ and $v_{1,\beta} + h_y < 0$ we invoke the Bonferroni inequality

$$\begin{aligned} 1 - \gamma &\le P(V_1 + h_y \ge v_{1,\beta} + h_y, \exp(V_2) \le \exp(v_{2,1-\beta})) \\ &\le P((V_1 + h_y)\exp(V_2) \ge (v_{1,\beta} + h_y)\exp(v_{2,1-\beta})) \end{aligned}$$

and

$$P((V_1 + h_y)\exp(V_2) < (v_{1,\beta} + h_y)\exp(v_{2,1-\beta})) \le \gamma = P((V_1 + h_y)\exp(V_2) \le w)$$

which implies $(v_{1,\beta} + h_y)\exp(v_{2,1-\beta}) \le w$.
Depending on which of the four combinations of inequalities $v_{1,1-\alpha} + h_y \ge (<) 0$ and $v_{1,\beta} + h_y \ge (<) 0$ applies we obtain four different bracketing intervals.

5 Confidence Bounds for Parameters

Confidence bounds for the parameters $\beta_1, \ldots, \beta_p$ and σ can be obtained by using traditional quantile bounds with specific choices for $\boldsymbol{u}$ and w_p.

For example, if we denote by $\boldsymbol{u}_i$ a p-vector with a 1 in the i^{th} and with 0 in all other positions then $\boldsymbol{u}_i'\boldsymbol{\beta} = \beta_i$ and if we choose $w_p = 0$, i.e., $p = 1 - \exp(-1) \approx .63212$ then the p-quantile $y_p(\boldsymbol{u}_i)$ is the same as β_i. In that case the traditional lower confidence bounds and the new bounds for this quantile coincide. This comes about because of $w_p = 0$.

Similarly one can get confidence bounds for σ by using quantile bounds for $y_p(\boldsymbol{u})$ with $\boldsymbol{u} = \boldsymbol{0}$ and $p = 1 - \exp(-\exp(1)) \approx .93401$ because then $y_p(\boldsymbol{u}) = \sigma$. However here the traditional and new bounds do not coincide. While the "traditional" bound for σ uses the approximate normality of $\widehat{\sigma}$ (previously pointed out as a poor approximation), the new method uses the approximate normality of $\log(\widehat{\sigma})$ which usually is employed in favor of the "traditional" bound. The traditional method has the inherent flaw that it treats $\widehat{\sigma}$ as approximately normal. Depending on the choice of $\boldsymbol{u}$ and w_p this will have a negative impact when dealing with small to moderate sized samples or with intensive censoring. The new method does not suffer from this and its approximation quality is completely determined by the quality of the joint normal approximation for $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}$ and $\log(\widehat{\sigma})$, because the remaining step to converting this to quantile or CDF confidence bounds is an exact process. However, this does not preclude approximation errors for (V_1, V_2) being magnified by w_p when dealing with the distribution of $V_1 + w_p[1 - \exp(-V_2)]$, for example.

6 Bootstrap Confidence Bounds

As pointed out at the end of the previous section the coverage quality of the new type of quantile or CDF confidence bounds hinges entirely on the quality of the invoked joint normal approximation for $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}$ and $\log(\widehat{\sigma})$. For small samples or small number of observed failures this approximation may still be lacking. One simple way out of this is to bootstrap the distribution of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}$ and $\log(\widehat{\sigma}/\sigma)$. Here the bootstrap approach will depend on how much is known about the censoring times $c_1, \ldots, c_n$.

There are situations where the observed and the potential censoring times are all known. For example, all n items were put into service at various times and at some fixed time t_0 stock is taken concerning the failure times of those items that failed and concerning the times that the other items have been exposed without failure. For the latter these times are the censoring times and for the former one deduces the potential censoring times by subtracting the times at which these failed items were placed into service from t_0. With the knowledge of all censoring times (observed or potential) it is a simple matter to generate new right censored log-Weibull samples using a random sample of size n from a log-Weibull distribution with parameters $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ (using the same covariates as before in each case) and comparing these simulated observations with the corresponding censoring times to produce a right censored sample. For each such censored sample one then again computes the estimates $\widehat{\boldsymbol{\beta}}^\star$ and $\widehat{\sigma}^\star$. Repeating this resampling/right censoring and estimation a large number of B times one can get an empirically generated distribution of $(\widehat{\boldsymbol{\beta}}^\star - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^\star$ and $\log(\widehat{\sigma}^\star) - \log(\widehat{\sigma})$ which may serve as a better representation of the true joint distribution (other than the $(p+1)$-variate normal distribution) of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}$ and $\log(\widehat{\sigma}) - \log(\sigma)$. Because we are no longer concerned with getting a close to normal distribution we may equivalently generate the bootstrap distribution of $(\widehat{\boldsymbol{\beta}}^\star - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^\star$ and $\widehat{\sigma}^\star/\widehat{\sigma}$ as approximation to the joint distribution of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}$ and $\widehat{\sigma}/\sigma$.

However, there are also other situations in which the potential censoring times are not known. If in that case one can view the censoring times as a random sample from some unknown distribution H one could estimate H by using the Kaplan-Meier estimate $\widehat{H}$, i.e., treating the observed censoring times as "failures" and the observed failures as "censoring times" on the unobserved censoring times. In generating new bootstrap samples it makes most sense to treat the observed censoring times as given for those cases for which they occurred (i.e., condition on the known ancillaries), while any unknown potential censoring time for an observed failure y could be randomly generated from the conditional distribution of $\widehat{H}$ given that the censoring time exceeds y. This is the scheme proposed originally by Robinson [8] and then again by Hjort [1] and Kim [3].

The calculation of $100\gamma\%$ lower confidence bounds for p-quantiles by the bootstrap method is straightforward and follows directly from equation 14. For given $\boldsymbol{u}$ and p we sort the bootstrapped values of

$$\frac{\boldsymbol{u}'(\widehat{\boldsymbol{\beta}}_i^\star - \widehat{\boldsymbol{\beta}})}{\widehat{\sigma}_i^\star} + w_p\left(1 - \frac{\widehat{\sigma}}{\widehat{\sigma}_i^\star}\right) , \qquad i = 1, \ldots, B$$

and pick off (by linear interpolation) the γ-quantile of these values. That will give us the bootstrap approximation $k^\star = k^\star(w_p)$ to the desired quantity $k = k(w_p)$ and the resulting lower bound is then

$$\widehat{y}^\star_{p,L2} = \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - k^\star)\widehat{\sigma} \ .$$

To be able to get this γ-quantile one needs to choose B large enough. Of course, the larger B the better the approximation to the value of $k^\star$ when $B = \infty$.

We won't go into examining whether the absolute continuity conditions of Lemma 1 are satisfied to claim the strict monotonicity of $h^\star(w_p) = w_p - k^\star(w_p)$ when $B = \infty$. We take that as given. The discrete nature of the bootstrap procedure for finite B is handled through the suggested linear interpolation step.

For sparse failure data it is occasionally impossible to get the bootstrapped values of $(\widehat{\boldsymbol{\beta}}^{\star} - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^{\star}$ and $\widehat{\sigma}^{\star}/\widehat{\sigma}$, because the m.l.e.'s don't exist. Our position is to leave out these corresponding bootstrap sample cases, i.e., we work with a reduced bootstrap sample size $\tilde{B} < B$, or we could also get new bootstrap samples until we have the desired number B of "valid" bootstrap samples. The rationale is that we would not even attempt to calculate confidence bounds if the original estimates had not been possible in the first place. Thus we only factor in the uncertainty of the estimates conditional on the assumption that they exist.

As pointed out before, confidence bounds for p-quantiles will also give us confidence bounds for the parameters σ and β_j. It remains to describe the bootstrap process for getting confidence upper bounds for the CDF, $p_0(y;\boldsymbol{u}) = G((y - \boldsymbol{u}'\boldsymbol{\beta})/\sigma)$, that is more direct than trying to invert the quantile bounds, i.e., for given y solving $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + h(w_p)\widehat{\sigma} = y$ or $(y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma} = h(w_p)$ for w_p and thus for p. To do this more directly, we rewrite equation 14 as follows

$$\gamma \approx P\left(\left[\frac{\boldsymbol{u}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\widehat{\sigma}} + h(w_p)\right]\frac{\widehat{\sigma}}{\sigma} \leq w_p\right)$$

and exploit the 1-1 relationship between w and $h(w)$. Thus we let $h_y = (y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$ and sort the bootstrap values

$$\left(\frac{\boldsymbol{u}'(\widehat{\boldsymbol{\beta}}_i^{\star} - \widehat{\boldsymbol{\beta}})}{\widehat{\sigma}_i^{\star}} + h_y\right)\frac{\widehat{\sigma}_i^{\star}}{\widehat{\sigma}}\ , \qquad i = 1, \ldots, B$$

and pick off (by linear interpolation) the γ-quantile of these values. That will give us the bootstrap approximation $w_{p_0}^{\star}$ of $h^{-1}(h_y)$ and then $\hat{p}_{y,U2}^{\star}(\gamma, \boldsymbol{u}) = G(w_{p_0}^{\star})$ is the bootstrap $100\gamma\%$ upper confidence bound for $p_0(y;\boldsymbol{u})$.

7 Examining Bivariate Normality

The new method, based on the approximate bivariate normality of (V_1, V_2), assumes that $((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}, \widehat{\lambda} - \lambda) = ((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}, \log(\widehat{\sigma}/\sigma))$ has an approximate $(p+1)$-variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\widehat{\boldsymbol{B}}^{-1}$. To examine this assumption we took a particular data set, namely the shock absorber data as given in Meeker and Escobar [7] with $n = 38$ and $p = 1$. We treated its estimated Weibull model as the truth and generated bootstrap samples from it with the same sample size. Whenever the censoring times were not known from the sample we generated censoring times by Robinson's [8] conditional method as oulined above. These censoring times were then used to generate randomly right censored Weibull samples with the same sample size of 38. This then leads to new maximum likelihood estimates $(\widehat{\boldsymbol{\beta}}^{\star}_i, \widehat{\sigma}^{\star}_i)$, $i = 1, \ldots, N_b$, where N_b is the number of bootstrap replications of this sampling process.

Note that the covariance matrix $\widehat{\boldsymbol{B}}^{-1}$ of the assumed $(p+1)$-variate normal distribution of $((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\sigma}, \widehat{\lambda} - \lambda)$ depends not only on the estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ but also more generally on the data, namely on the $\hat{z}_i = (x_i - \boldsymbol{u}_i'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$, the covariates $\boldsymbol{u}_i$, and on r, the number of failures in the sample. Thus when judging the multivariate normality of the standardized bootstrap estimates $((\widehat{\boldsymbol{\beta}}^{\star}_i - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^{\star}_i, \log(\widehat{\sigma}^{\star}_i/\widehat{\sigma}))$, $i = 1, \ldots, N_b$, we are faced with a different multivariate normal distribution for each i, namely with mean $\mathbf{0}$ and covariance matrix $\widehat{\boldsymbol{B}}^{\star -1}_i$ which depends on the respective bootstrap sample.

In order to judge the multivariate normality of the standardized bootstrap estimates $((\widehat{\boldsymbol{\beta}}^{\star}_i - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^{\star}_i, \log(\widehat{\sigma}^{\star}_i/\widehat{\sigma}))$, $i = 1, \ldots, N_b$, on a common scale we further standardize each $((\widehat{\boldsymbol{\beta}}^{\star}_i - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^{\star}_i, \log(\widehat{\sigma}^{\star}_i/\widehat{\sigma}))$ by premultiplying it by the Cholesky factor Q_i of the respective $\widehat{\boldsymbol{B}}^{\star}_i$ (i.e., $\widehat{\boldsymbol{B}}^{\star}_i = Q_i'Q_i$ with Cholesky factor Q_i, an upper triangular matrix with positive diagonal elements).

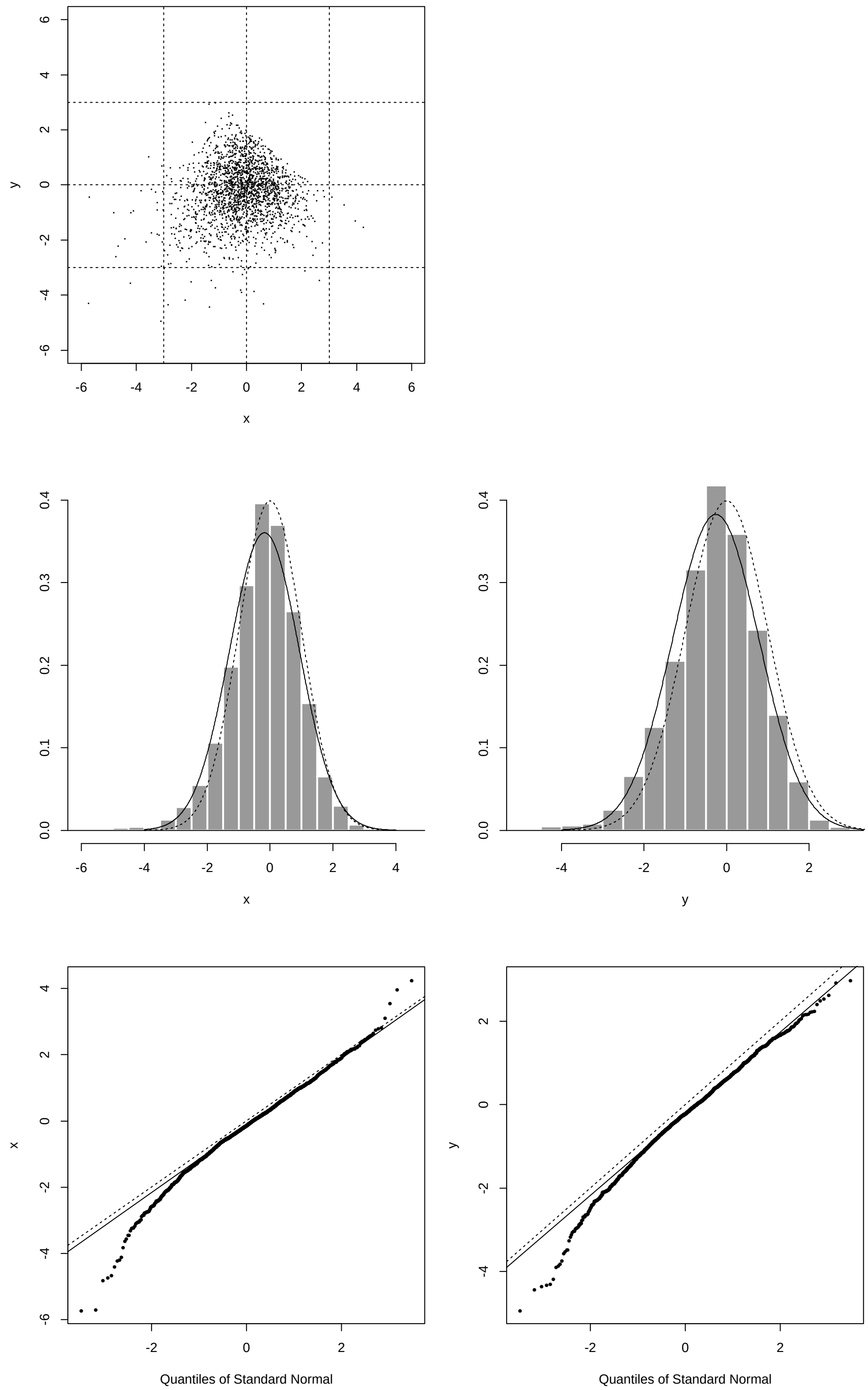
The effect of this transformation is that the transformed values of $Q_i((\widehat{\boldsymbol{\beta}}^{\star}_i - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^{\star}_i$, $\log(\widehat{\sigma}^{\star}_i/\widehat{\sigma}))^t$ then have a $(p+1)$-variate normal distribution with mean $\mathbf{0}$ and $(p+1) \times (p+1)$ identity matrix $\boldsymbol{I}_{p+1}$ as common covariance matrix. All of this holds in an approximate sense since we start with an approximation. However, if normality holds after transformation it is equally valid before transformation and vice versa, i.e., the transformation does not make the normality better or worse. It only serves to make its quality visible in the aggregate over all simulated bootstrap cases.

Figure 1 shows the results of such a bootstrap simulation with $N_b = 2000$. The scatter of the Cholesky transformed values of $((\widehat{\boldsymbol{\beta}}^{\star}_i - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^{\star}_i$, $\log(\widehat{\sigma}^{\star}_i/\widehat{\sigma}))$ is shown in the upper left plot and it roughly looks like the expected circular bivariate normal scatter centered on zero and with variance one in each direction. On closer examination one sees some deformation from this pattern (a flattening effect) in the upper right-hand corner and some broader scatter toward the lower left corner.

The next two plots in this figure show the histograms of the respective x and y dimensions of the scatter with superimposed fitted normal (solid line) and expected standard normal (dotted line) density curves.

The last two plots in Figure 1 show the corresponding normal QQ-plots. These should show a roughly linear pattern if the data were indeed normal. The solid line goes through the first and third quartile of the data while the dotted line represents the

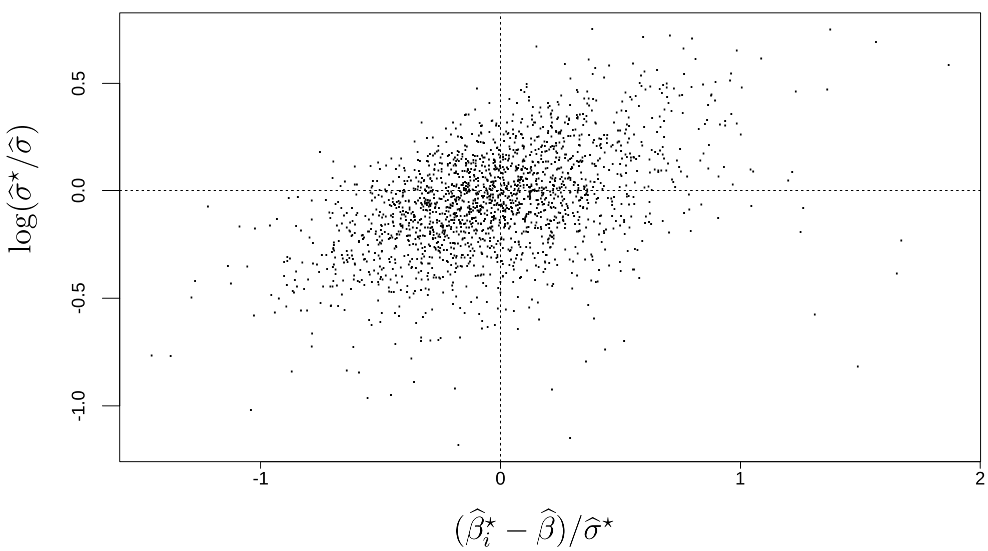
Figure 1: Examining Bivariate Normality



expected standard normal distribution. Again one can see some deviation from the expected normal behavior at the low end of the x and y values. It is not clear for which values of $(\widehat{\boldsymbol{\beta}}_i^\star, \widehat{\sigma}_i^\star)$ this would be the case. To get some sense of that we also plotted the standardized bootstrap estimates $\log(\widehat{\sigma}_i^\star/\widehat{\sigma})$ versus $(\widehat{\boldsymbol{\beta}}_i^\star - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}_i^\star$ in Figure 2.

Again the plot shows the expected ellipsoidal bivariate normal scatter with some flattening in the upper left corner. This seems to suggest that large values of $\widehat{\sigma}_i^\star$ will limit the effect of $\widehat{\boldsymbol{\beta}}_i^\star$ exceeding $\widehat{\boldsymbol{\beta}}$ in the ratio $(\widehat{\boldsymbol{\beta}}_i^\star - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}_i^\star$.

Figure 2: Bivariate Normality of Bootstrap Estimates



8 Confidence Bounds for the Failure Rate Function

As noted previously, the failure rate functions in the Weibull and Gumbel case differ by the factor $1/t_0$, aside from the natural transformations resulting from the log-transform of the first model into the second. This should not affect the methodolgy of constructing confidence bounds, except to implement this factor when appropriate.

Our approach based on the bivariate normality or the bootstrapped distribution of (V_1, V_2) does not transfer quite as cleanly to the failure rate function. We indicate the source of the difficulty and offer a slight modification that neglects a second order term. We start by giving the details of constructing confidence bounds by the traditional method, then discuss the difficulty with the new approach and, based on the probability integral transform, give a solution using a slight modification. In the process the confidence bounds for the CDF are rederived based on the probability integral transform. Finally the corresponding bootstrap approach is presented.

8.1 Traditional Failure Rate Lower Confidence Bounds

For the case of the Gumbel distribution we use the normal approximation given in (12) and get approximate $100\gamma\%$ lower confidence bounds for $r_G(y_0, \boldsymbol{u})$ as follows:

$$\widehat{r}_G(y_0, \boldsymbol{u}) \times \exp\left(-z_\gamma \widehat{\tau}_G(y_0, \boldsymbol{u})\right)$$

and we note again that a $100(1-\gamma)\%$ lower bound also serves as a $100\gamma\%$ upper confidence bound for $r_G(y_0, \boldsymbol{u})$.

The failure rate function $r_G(y_0, \boldsymbol{u})$ is monotone increasing in y_0, in fact $\log(r_G(y_0, \boldsymbol{u}))$ is linear in y_0 with positive slope $1/\sigma$. However, the confidence bounds do not necessarily share the monotonicity property. This is similar to what was observed in the context of quantile bounds and the examination for monotonicity is parallel to that previous case and is not repeated here. Again the violation of monotonicity can occur when the number of failed cases is low or the confidence level is extremely high.

For the case of the Weibull distribution we use the normal approximation given in (13) and get approximate $100\gamma\%$ lower confidence bounds for $r_W(t_0)$ as follows:

$$\widehat{r}_W(t_0, \boldsymbol{u}) \times \exp\left(-z_\gamma \widehat{\tau}_W(t_0, \boldsymbol{u})\right) \ .$$

Note that $\widehat{\tau}_W(t_0, \boldsymbol{u}) = \widehat{\tau}_G(\log(t_0), \boldsymbol{u})$ and $\widehat{r}_W(t_0, \boldsymbol{u}) = \widehat{r}_G(\log(t_0), \boldsymbol{u})/t_0$.

Although $r_G(y_0, \boldsymbol{u})$ is always monotone increasing in y_0 the same is not true for $r_W(t_0, \boldsymbol{u})$, i.e., it is not always monotone increasing in t_0. In fact, $\log(r_W(t_0, \boldsymbol{u}))$ is a linear function in $\log(t_0)$ with slope $1/\sigma - 1 = \eta - 1$. Thus we have monotone increasing behavior in t_0 when $\eta > 1$, monotone decreasing behavior in t_0 when $\eta < 1$, and a constant failure rate function when $\eta = 1$. What can be said about the corresponding confidence bounds? The estimated failure rate function $\widehat{r}_W(t_0, \boldsymbol{u})$ is monotone increasing when $\widehat{\eta} > 1$, monotone decreasing when $\widehat{\eta} < 1$ and constant when $\widehat{\eta} = 1$. However, even when the estimated failure rate function is monotone in one direction it does not necessarily follow that the bounds will be monotone. Again criteria for monotone behavior can be worked out based on the approach given in the quantile case, but here the case for insisting on similar monotone behavior is not so compelling, since the direction of monotonicity of the true failure rate function is not always the same. Thus we will not go into a detailed discussion of this issue.

8.2 Modified New Approach

We discuss mainly the Gumbel case and reduce the Weibull case to it at the end. Note that the estimate

$$\log\left(\widehat{r}_G(y_0,\boldsymbol{u})\right) = -\log(\widehat{\sigma}) + \frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} \quad \text{of} \quad \log\left(r_G(y_0,\boldsymbol{u})\right) = -\log(\sigma) + \frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}$$

can be rewritten as follows

$$\begin{aligned}
\log\left(\widehat{r}_G(y_0,\boldsymbol{u})\right) &= -\log(\widehat{\sigma}/\sigma) - \log(\sigma) + \frac{\boldsymbol{u}'\boldsymbol{\beta} - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} + \frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma} + \frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\left(\frac{\sigma}{\widehat{\sigma}} - 1\right) \\
&= \log\left(r_G(y_0,\boldsymbol{u})\right) - \log(\widehat{\sigma}/\sigma) + \frac{\boldsymbol{u}'\boldsymbol{\beta} - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} + \frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\left(\frac{\sigma}{\widehat{\sigma}} - 1\right) \\
&= \log\left(r_G(y_0,\boldsymbol{u})\right) - V_2 - V_1 + \frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma}\left(\exp(-V_2) - 1\right) \\
&= \log\left(r_G(y_0,\boldsymbol{u})\right)\exp(-V_2) - V_1 - V_2 + \log(\sigma)\left(\exp(-V_2) - 1\right) \qquad (19)
\end{aligned}$$

and one sees that the distribution of $\widehat{r}_G(y_0,\boldsymbol{u})$ depends on the unknown parameters not only through the target quantity $r_G(y_0,\boldsymbol{u})$ but also through σ alone.

Recall, that in the case of confidence bounds for the CDF we faced the following situation

$$\frac{y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} = \exp(-V_2)\frac{y_0 - \boldsymbol{u}'\boldsymbol{\beta}}{\sigma} - V_1 \; ,$$

and here the distribution of the estimate $(y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$ is seen to depend only on the target quantity $(y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma$. It is understood that (V_1, V_2) has a known albeit estimated distribution. Note that here, in contrast to the previous situation, there is no dependence on additional parameters.

We now give another derivation of the upper confidence bound for the target parameter $\theta = \theta(y_0,\boldsymbol{u}) = (y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma$, which through composition with G leads to an upper confidence bound for the CDF, i.e., for $G\left((y_0 - \boldsymbol{u}'\boldsymbol{\beta})/\sigma\right)$. This derivation is more direct than our previous inversion of quantile lower bounds but the result is the same. It also points the way on how to deal with confidence bounds for the failure rate.

Let $\widehat{\theta} = \widehat{\theta}(y_0,\boldsymbol{u}) = (y_0 - \boldsymbol{u}'\widehat{\boldsymbol{\beta}})/\widehat{\sigma}$ and denote by $H_\theta(y) = P_\theta\left(\widehat{\theta} \le y\right)$ its distribution function. From the previous representation of $\widehat{\theta}$ in terms of (V_1, V_2) we have

$$H_\theta(y) = P\left(\theta\exp(-V_2) - V_1 \le y\right)$$

and assuming a known bivariate normal distribution for (V_1, V_2) it is clear that $H_\theta(y)$ is strictly decreasing in θ for fixed y. Denote this strictly decreasing function of θ by $\psi_y(\theta)$ and for $u \in (0,1)$ denote its inverse (in θ) by $\psi_y^{-1}(u)$. By the bivariate normality of (V_1, V_2) we can treat $\widehat{\theta}$ as a continuous random variable and thus can invoke the probability integral transform result, namely the random variable $H_\theta(\widehat{\theta})$ has a uniform distribution on the interval $(0,1)$, i.e., $\psi_{\widehat{\theta}}(\theta) = H_\theta(\widehat{\theta})$ is a pivot. Thus

$$P\left(\psi_{\widehat{\theta}}(\theta) \leq 1-\gamma\right) = 1-\gamma \ .$$

From this one gets

$$P\left(\theta \geq \psi_{\widehat{\theta}}^{-1}(1-\gamma)\right) = 1-\gamma \quad \text{or} \quad P\left(\theta \leq \psi_{\widehat{\theta}}^{-1}(1-\gamma)\right) = \gamma$$

where $x = \psi_{\widehat{\theta}}^{-1}(1-\gamma)$ solves $\psi_{\widehat{\theta}}(x) = 1-\gamma$ or $H_x(\widehat{\theta}) = P(x\exp(-V_2) - V_1 \leq \widehat{\theta}) = 1-\gamma$. Here $\widehat{\theta}$ is the observed value and should not be treated as random within the last probability statement. This leads to solving

$$\gamma = P(V_1 \leq x\exp(-V_2) - \widehat{\theta}) = \int_{-\infty}^{\infty} \Phi\left(\frac{x\exp(-v_2) - \widehat{\theta} - v_2\sigma_{12}/\sigma_2^2}{\sigma_{1|2}}\right)\frac{1}{\sigma_2}\phi(v_2/\sigma_2)\, dv_2$$

for x, which is the same task that was faced previously when inverting quantile lower bounds.

Returning our focus to the failure rate we now let $\widehat{\theta} = \widehat{\theta}(y_0, \boldsymbol{u}) = \log(\widehat{r}_G(y_0, \boldsymbol{u}))$ represent the estimate of $\theta = \log(r_G(y_0, \boldsymbol{u}))$. From (19) we know that the distribution of $\widehat{\theta}$ depends both on θ and $\lambda = \log(\sigma)$, i.e.,

$$H_{\theta,\lambda}(y) = P(\widehat{\theta} \leq y) = P(\theta\exp(-V_2) - V_1 - V_2 + \lambda(\exp(-V_2) - 1) \leq y) \ . \tag{20}$$

As pointed out before, the difficulty is the term $\lambda(\exp(-V_2) - 1)$ with the unknown parameter λ. Our modification in approach is to replace the unknown λ by its maximum likelihood estimate $\widehat{\lambda} = \log(\widehat{\sigma})$. This amounts to neglecting the second order term $-V_2(\exp(-V_2) - 1)$ in

$$\begin{aligned}\lambda(\exp(-V_2) - 1) &= \left(\lambda - \widehat{\lambda}\right)(\exp(-V_2) - 1) + \widehat{\lambda}(\exp(-V_2) - 1) \\ &= -V_2(\exp(-V_2) - 1) + \widehat{\lambda}(\exp(-V_2) - 1) \ .\end{aligned}$$

The notion "second order term" derives from the fact that both V_2 and $\exp(-V_2) - 1$ converge to zero in large samples. Thus we propose to work with the approximate pivot $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ when developing the confidence bound for the parameter θ.

Recall that the probability integral transform result gives us that $H_{\theta,\lambda}(\widehat{\theta})$ is a pivot with uniform distribution on $(0,1)$. We will use the same pivot distribution for $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ banking on the negligibility of the above second order term.

Of course it is possible to do a more refined analysis of the distribution of $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ (using only the known and estimated distribution of (V_1, V_2)) but this involves a second level of integration. Furthermore, the unknown parameter θ then enters in two different places, once as subscript to H and once through the distribution of $\widehat{\theta}$, and the monotonicity in θ is no longer clear. However, this is not an issue when trying to obtain quantiles of the $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ distribution. On top of the double integration there is still the extra layer of simulation to validate this refined method. Thus it seems reasonable and more practical to stay with the uniform distribution on $(0,1)$ as the approximate pivot distribution for $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$.

For $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ we have the following representation

$$H_{\theta,\widehat{\lambda}}(\widehat{\theta}) = \int_{-\infty}^{\infty} \Phi\left(\frac{\widehat{\theta} + v_2 - \widehat{\lambda}(\exp(-v_2) - 1) - \theta \exp(-v_2) + v_2\sigma_{12}/\sigma_2^2}{\sigma_{1|2}}\right) \frac{1}{\sigma_2}\phi(v_2/\sigma_2)dv_2.$$

From

$$P(H_{\theta,\widehat{\lambda}}(\widehat{\theta}) \leq 1-\gamma) = 1-\gamma$$

and the fact that $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ is strictly decreasing in θ we obtain a $100(1-\gamma)\%$ lower confidence bound $\widehat{\theta}_L(1-\gamma, y_0, \boldsymbol{u})$ for θ by solving $H_{\theta,\widehat{\lambda}}(\widehat{\theta}) = 1-\gamma$ for $\theta = \widehat{\theta}_L(1-\gamma, y_0, \boldsymbol{u})$. Thus we also have a $100\gamma\%$ upper confidence bound $\widehat{\theta}_U(\gamma, y_0, \boldsymbol{u}) = \widehat{\theta}_L(1-\gamma, y_0, \boldsymbol{u})$ for θ. With that one gets in $\exp(\widehat{\theta}_U(\gamma, y_0, \boldsymbol{u}))$ a $100\gamma\%$ upper confidence bound for the failure rate function $r_G(y_0, \boldsymbol{u})$.

From the construction of $\widehat{\theta}_L(1-\gamma, y_0, \boldsymbol{u})$ one easily sees that it is an increasing function of $\widehat{\theta} = \widehat{\theta}(y_0, \boldsymbol{u})$ and thus also an increasing function of y_0. Hence these confidence bounds have the same monotonicity property as the target failure rate function.

To get a corresponding bound in the Weibull case one simply makes use of the relationship $r_W(t_0, \boldsymbol{u}) = r_G(\log(t_0), \boldsymbol{u})/t_0$ after log-transforming the Weibull data into Gumbel form. Here the monotonicity property of the bounds may be reversed (through the division by t_0) because the ratio of two monotone functions may not stay monotone. As noted previously, the case for insisting on the same monotonicity behavior as seen in the estimate is no longer so compelling.

8.3 Bracketing the Root $\widehat{\theta}_L(\gamma, y_0, \boldsymbol{u})$

The lower confidence bound $\widehat{\theta}_L(\gamma, y_0, \boldsymbol{u})$ is defined as the solution in θ of

$$P\left(\theta \geq \exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)\right) = 1 - \gamma \ .$$

In solving for the unique root of this equation it is useful to bracket the solution θ by $L \leq \theta \leq U$.

Using the previous notation of (17) we have

$$P(-v_{i,1-\gamma/4} \leq V_i \leq v_{i,1-\gamma/4}) = 1 - \frac{\gamma}{2}, \quad i = 1, 2$$

and thus by Bonferoni's inequality

$$P(-\tilde{v}_1 \leq V_1 \leq \tilde{v}_1, -\tilde{v}_2 \leq V_2 \leq \tilde{v}_2) \geq 1 - \gamma \ ,$$

where $\tilde{v}_i = v_{i,1-\gamma/4}$. Hence with probability at least $1 - \gamma$ we have for $\widehat{\lambda} \geq 0$

$$\exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)$$
$$\leq \exp(\tilde{v}_2) \max\left(0, \widehat{\theta} + \tilde{v}_1 + \tilde{v}_2 - \widehat{\lambda}(\exp(-\tilde{v}_2) - 1)\right)$$

and for $\widehat{\lambda} < 0$

$$\exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)$$
$$\leq \exp(\tilde{v}_2) \max\left(0, \widehat{\theta} + \tilde{v}_1 + \tilde{v}_2 - \widehat{\lambda}(\exp(\tilde{v}_2) - 1)\right) \ .$$

Combining these two cases and using $\text{sign}(x) = 1$ for $x \geq 0$ and $\text{sign}(x) = -1$ for $x < 0$ we have with probability at least $1 - \gamma$ that

$$\exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)$$
$$\leq \exp(\tilde{v}_2) \max\left(0, \widehat{\theta} + \tilde{v}_1 + \tilde{v}_2 - \widehat{\lambda}(\exp(-\text{sign}(\widehat{\lambda})\tilde{v}_2) - 1)\right) = U \ .$$

This U can thus serve as upper bound to the root θ.

For the lower bound we proceed as follows. Let $\bar{v}_i = v_{i,1-(1-\gamma)/4}$ then

$$P(-\bar{v}_i \leq V_i \leq \bar{v}_i) = \frac{1 + \gamma}{2}$$

and by Bonferoni's inequality

$$P(-\bar{v}_1 \le V_1 \le \bar{v}_1,\ -\bar{v}_2 \le V_2 \le \bar{v}_2) \ge \gamma\ .$$

Hence with probability at least γ we have for $\widehat{\lambda} \ge 0$

$$\exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)$$
$$\ge \exp(\bar{v}_2)\min\left(0, \widehat{\theta} - \bar{v}_1 - \bar{v}_2 - \widehat{\lambda}(\exp(\bar{v}_2) - 1)\right)$$

and for $\widehat{\lambda} < 0$

$$\exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)$$
$$\ge \exp(\bar{v}_2)\min\left(0, \widehat{\theta} - \bar{v}_1 - \bar{v}_2 - \widehat{\lambda}(\exp(-\bar{v}_2) - 1)\right)\ .$$

Combining these two cases we have with probability at least γ that

$$\exp(V_2)\left(\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2) - 1)\right)$$
$$\ge \exp(\bar{v}_2)\min\left(0, \widehat{\theta} - \bar{v}_1 - \bar{v}_2 - \widehat{\lambda}(\exp(\mathrm{sign}(\widehat{\lambda})\bar{v}_2) - 1)\right) = L\ .$$

This L can thus serve as lower bound to the root θ.

8.4 Bootstrap Confidence Bounds for the Failure Rate

Again we let $\widehat{\theta} = \log(\widehat{r}_G(y_0, \boldsymbol{u}))$ represent the estimate of $\theta = \log(r_G(y_0, \boldsymbol{u}))$ and we treat $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$ as an approximate pivot with uniform distribution on $(0, 1)$. Here we no longer treat (V_1, V_2) as distributed according to a bivariate normal distribution but will bootstrap that distribution. This assumption of a uniform pivot distribution is reasonable and practical. The alternative would be to use the double bootstrap approach given in Scholz [9] to get at the distribution and quantiles of $H_{\theta,\widehat{\lambda}}(\widehat{\theta})$. The validation step would put a third level of simulation on top of that. This is an inordinate amount of effort to account for a term that is negligible in all but extreme situations.

Thus we need to find the value θ that solves

$$1 - \gamma = H_{\theta,\widehat{\lambda}}(\widehat{\theta})$$

to get a $100(1-\gamma)\%$ lower bound $\widehat{\theta}^{\star}_L(1-\gamma)$ for θ. Again we can treat $\widehat{\theta}^{\star}_U(\gamma, \boldsymbol{u}) = \widehat{\theta}^{\star}_L(1-\gamma, \boldsymbol{u})$ as $100\gamma\%$ upper bound for θ and $\exp(\widehat{\theta}^{\star}_U(\gamma, \boldsymbol{u}))$ as $100\gamma\%$ upper confidence bound for the failure rate function $r_G(y_0, \boldsymbol{u})$ with the same adaptation as before for Weibull data.

It remains to find $\widehat{\theta}^\star_L(1-\gamma, \boldsymbol{u})$ and that can again be done without iteration via a simple bootstrap step as follows. Using (20) with y and λ replaced by the observed values of $\widehat{\theta}$ and $\widehat{\lambda}$ we have to solve (again treating only (V_1, V_2) as random in the following probability statements)

$$\begin{aligned} 1-\gamma &= P\left(\theta \exp(-V_2) - V_1 - V_2 + \widehat{\lambda}(\exp(-V_2)-1) \le \widehat{\theta}\,\right) \\ &= P\left(\theta \le \left[\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2)-1)\right]\exp(V_2)\right) \end{aligned}$$

or

$$\gamma = P\left(\left[\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2)-1)\right]\exp(V_2) \le \theta\right)$$

for θ. That solution is simply the γ-quantile of the distribution of

$$\left[\widehat{\theta} + V_1 + V_2 - \widehat{\lambda}(\exp(-V_2)-1)\right]\exp(V_2)\ ,$$

while (V_1, V_2) vary according to their assumed bivariate distribution.

This quantile can be obtained by bootstrapping the (V_1, V_2) values as before, namely generate B independent copies $(\widehat{\boldsymbol{\beta}}^\star_i, \widehat{\sigma}^\star_i), i = 1, \ldots B$, (treating $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ as the true population parameters) compute $V^\star_{1,i} = \boldsymbol{u}'(\widehat{\boldsymbol{\beta}}^\star_i - \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^\star_i$ and $V^\star_{2,i} = \log(\widehat{\sigma}^\star_i/\widehat{\sigma})$ and sort the resulting values of

$$\left[\widehat{\theta} + V^\star_{1,i} + V^\star_{2,i} - \widehat{\lambda}(\exp(-V^\star_{2,i})-1)\right]\exp(V^\star_{2,i})\ , \qquad i = 1, \ldots, B\ ,$$

in increasing order and interpolating the γ-quantile of these values. This quantile then serves as the bootstrap solution for $\widehat{\theta}^\star_U(\gamma, \boldsymbol{u})$.

9 Bootstrap-t Solutions

The previously discussed bootstrap solution neither corresponds to a percentile bootstrap nor to a bootstrap-t approach, although the latter comes close since use was made of the bootstrap distribution of the "pivotal" quantities (V_1, V_2). However, the standardization in $V_1 = (\boldsymbol{u}'\widehat{\boldsymbol{\beta}} - \boldsymbol{u}'\boldsymbol{\beta})/\widehat{\sigma}$ uses just $\widehat{\sigma}$ and not the standard error estimate of $\boldsymbol{u}'\widehat{\boldsymbol{\beta}}$, with a corresponding comment applying to V_2. The simulation results in Jeng and Meeker [2] suggest that the bootstrap-t approach works very well as far as coverage properties are concerned. In this approach one takes for any given estimator $\widehat{\theta} = g(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})$ of $\theta = g(\boldsymbol{\beta}, \sigma)$

the quantity $R = (\widehat{\theta} - \theta)/\widehat{se}(\widehat{\theta})$ as the basic "pivot." The standard error estimate $\widehat{se}(\widehat{\theta})$ could be computed by several different methods. We will use the standard deviation of the normal approximation for $\widehat{\theta}$ when using the estimated observed Fisher information. The γ-quantile r_γ of R (if it could be computed) could then be used to form $100\gamma\%$ lower confidence bounds for θ since

$$\gamma = P\left(\frac{\widehat{\theta} - \theta}{\widehat{se}(\widehat{\theta})} \le r_\gamma\right) = P\left(\widehat{\theta} - r_\gamma \widehat{se}(\widehat{\theta}) \le \theta\right) .$$

Instead of the unknown distribution of R we use its bootstrap distribution, namely that of $R^\star = (\widehat{\theta}^\star - \widehat{\theta})/\widehat{se}^\star(\widehat{\theta}^\star)$ where the estimates with superscript * are obtained through simulation by assuming that the unknown parameters $(\boldsymbol{\beta}, \sigma)$ are replaced by their known maximum likelihood estimates $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})$ when simulating random samples and at the same time emulating the relevant censoring mechanism as discussed previously. From the simulated $R^\star$ distribution ones determines the γ-quantile $r^\star_\gamma$ and then uses $\widehat{\theta} - r^\star_\gamma \widehat{se}(\widehat{\theta})$ as the bootstrap-t lower confidence bound for θ. Although its coverage is good it is not evident whether lower bounds for $\theta = \widehat{y}_p = \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + w_p\sigma$ will be monotone in p since the bootstrap distribution of $R^\star$ involves w_p and may thus present the same difficulties as were encountered with the classical maximum likelihood approach. One could probe this issue by trying this method on some extreme cases.
We denote the bootstrap-t lower bound for $\theta = y_p = \boldsymbol{u}'\boldsymbol{\beta} + w_p\sigma$ by $\widehat{y}^\star_{p,L2,t} = \widehat{y}^\star_{p,L2,t}(\gamma, \boldsymbol{u})$. Correspondingly we denote the bootstrap-t upper bound for $\theta = p_0(y) = G([y - \boldsymbol{u}'\boldsymbol{\beta}]/\sigma)$ by $\widehat{p}^\star_{y,U2,t} = \widehat{p}^\star_{y,U2,t}(\gamma, \boldsymbol{u})$. In both cases the subscript t indicates the bootstrap-t method for obtaining the bounds. As pointed out before, these two bounds are not necessarily inverses of each other.

9.1 Monotone Bootstrap-t Quantile Confidence Bounds

Here we present a hybrid approach that results in monotone quantile confidence bounds while using a hybrid bootstrap-t approach. The method invokes the joint distribution of

$$\tilde{V}_1 = \frac{\boldsymbol{u}'\widehat{\boldsymbol{\beta}} - \boldsymbol{u}'\boldsymbol{\beta}}{\widehat{f}_2\widehat{\sigma}} \quad \text{and} \quad \tilde{V}_2 = \frac{\sigma}{\widehat{f}_2\widehat{\sigma}} ,$$

where

$$\widehat{f}_2\widehat{\sigma} = \widehat{se}(\widehat{\sigma}) = \widehat{\sigma}/\sqrt{\widehat{W}\rho} .$$

We point out that the standardization $\tilde{V}_1$ with $\widehat{f}_2$ in the denominator is not in the spirit of the standard error estimate. To do that one would have to use $\widehat{f}_1$ instead, where

$$\widehat{f}_1\widehat{\sigma} = \widehat{se}(\boldsymbol{u}'\widehat{\boldsymbol{\beta}}\,) = \frac{\widehat{\sigma}\sqrt{\rho\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{u} + (\boldsymbol{u}'\boldsymbol{A}^{-1}\boldsymbol{b})^2}}{\sqrt{\widehat{W}\rho}} \ .$$

We also point out that in the standardization of $\widehat{\sigma}$ we did not use $(\widehat{\sigma}-\sigma)/(\widehat{f}_2\widehat{\sigma})$. Doing it as proposed will preserve the desired monotonicity of the quantile lower bounds to be derived. All this will simply lead to a minor modification of our original bootstrap proposal, namely changing the $\widehat{\sigma}$ in the denominator to $\widehat{f}_2\widehat{\sigma}$. It should be noted that $\widehat{f}_2$ appears as subfactor in $\widehat{f}_1$ and it is this subfactor that is affected most by the number r of failed observations in the sample. This r, when small, can have strong effects during the bootstrap process and a standardization may stabilize that.

We will consider lower bounds of the form $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - k)\widehat{f}_2\widehat{\sigma}$ for $\boldsymbol{u}'\boldsymbol{\beta} + w_p\sigma$, where k has to be determined appropriately. If the joint distribution of $\tilde{V}_1$ and $\tilde{V}_2$ were known one could for given w_p determine the γ-quantile $\tilde{k}_\gamma = \tilde{k}_\gamma(w_p)$ defined by

$$\gamma = P(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - \tilde{k}_\gamma)\widehat{f}_2\widehat{\sigma} \le \boldsymbol{u}'\boldsymbol{\beta} + w_p\sigma) = P\left(\tilde{V}_1 - w_p\tilde{V}_2 \le \tilde{k}_\gamma - w_p\right) \qquad (21)$$

and then we could view $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - \tilde{k}_\gamma)\widehat{f}_2\widehat{\sigma}$ as $100\gamma\%$ lower bound for $y_p = \boldsymbol{u}'\boldsymbol{\beta} + w_p\sigma$. Since the joint distribution of $\tilde{V}_1$ and $\tilde{V}_2$ is not known we will again use its bootstrap distribution and determine the corresponding quantile $\tilde{k}^\star_\gamma = \tilde{k}^\star_\gamma(w_p)$ from the bootstrap distribution of $\tilde{V}^\star_1 - w_p\tilde{V}^\star_2$. As lower confidence bound for $y_p = \boldsymbol{u}'\boldsymbol{\beta} + w_p\sigma$ we then propose to use $\widehat{y}^\star_{p,L2,m} = \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - \tilde{k}^\star_\gamma)\widehat{f}_2\widehat{\sigma}$, where the subscript m on $\widehat{y}^\star_{p,L2,m}$ indicates the monotone bootstrap version that is discussed here.

That $\tilde{h}(w_p) = w_p - \tilde{k}^\star_\gamma(w_p)$ is nondecreasing in w_p and thus in p is shown in exactly the same manner as before.

The corresponding upper bounds for $p_0 = p_0(y; \boldsymbol{u}) = G([y - \boldsymbol{u}'\boldsymbol{\beta}]/\sigma)$ are obtained in analogous fashion as in our first bootstrap approach. For given y we need to solve

$$\widehat{y}^\star_{p,L2,m} = \boldsymbol{u}'\widehat{\boldsymbol{\beta}} + (w_p - \tilde{k}^\star_\gamma)\widehat{f}_2\widehat{\sigma} = y$$

for w_p and thus for p, or solve

$$\tilde{h}_y = \frac{y - \boldsymbol{u}'\widehat{\boldsymbol{\beta}}}{\widehat{f}_2\widehat{\sigma}} = \tilde{h}(w_p)$$

for $w_p = \tilde{h}^{-1}(\tilde{h}_y)$. We rewrite Equation 21 (in bootstrap form) as follows

$$\gamma = P\left(\left[\tilde{V}^\star_1 + \tilde{h}(w_p)\right]/\tilde{V}^\star_2 \le w_p\right) \ .$$

Thus for given $\tilde{h}_y$ sort the bootstrap values

$$\left(\tilde{V}^\star_{1,i} + h_y\right) / \tilde{V}^\star_{2,i} = \left(\frac{\boldsymbol{u}'(\widehat{\boldsymbol{\beta}}^\star_i - \widehat{\boldsymbol{\beta}})}{\widehat{f}^\star_{2,i}\widehat{\sigma}^\star_i} + h_y \right) \frac{\widehat{f}^\star_{2,i}\widehat{\sigma}^\star_i}{\widehat{\sigma}}\ , \qquad i = 1, \ldots, B$$

and pick off (by interpolation) the γ-quantile of these values. That will give us the bootstrap approximation $w^\star_{p_0}$ of $\tilde{h}^{-1}(h_y)$ and then $\widehat{p}^\star_{y,U2,m}(\gamma, \boldsymbol{u}) = G(w^\star_{p_0})$ is the $100\gamma\%$ upper confidence bound for $p_0(y; \boldsymbol{u})$ which inverts the corresponding quantile lower bound. The subscript m on $\widehat{p}^\star_{y,U2,m}(\gamma, \boldsymbol{u})$ indicates again the monotone bootstrap version dealt with here.

The counterpart of this bootstrap hybrid for the case of the failure rate function was unclear and was not pursued further.

References

[1] Hjort, N.L. (1985), "Bootstrapping Cox's regression model," *Technical Report No. 241, November 1985* Department of Statistics, Stanford University.

[2] Jeng, S.-L. and Meeker, W.Q. (2000), "Comparisons of approximate confidence interval procedures for type I censored data," *Technometrics* **42**, 135-148.

[3] Kim, Ji-Hyun (1990). *Conditional Bootstrap Methods for Censored Data*, unpublished Ph.D. dissertation in the Department of Statistics, Florida State University.

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

[5] Le Cam, L. (1991), "Maximum Likelihood; An Introduction," unpublished manuscript, available at `http://www.stat.berkeley.edu/users/rice/LeCam/papers/index.html`.

[6] Meeker, W.Q. and Escobar, L.A. (1995), "Teaching about approximate confidence regions based on maximum likelihood," *The American Statistician* **49**, 48-53.

[7] Meeker, W.Q. and Escobar, L.A. (1998), *Statistical Methods for Reliability Data*, John Wiley & Sons, New York.

[8] Robinson, J.A. (1983), "Bootstrap confidence intervals in location-scale models with progressive censoring," *Technometrics* **25**, 179-187.

[9] Scholz, F.W. (1994). "On exactness of the parametric double bootstrap." *Statistica Sinica* **2**, 477-492.

[10] Scholz, F.W. (1996). "Maximum likelihood estimation for type I censored Weibull data including covariates." *ISSTECH-96-022*, Boeing Information & Support Services, P.O. Box 24346, MS-7L-22, Seattle WA 98124-0346.