Inference for the Weibull Distribution

Stat 498B Industrial Statistics

Fritz Scholz

May 22, 2008

1 The Weibull Distribution

The 2-parameter Weibull distribution function is defined as

$$F_{\alpha,\beta}(x) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \qquad \text{for} \quad x \geq 0 \quad \text{and} \quad F_{\alpha,\beta}(x) = 0 \quad \text{for } t < 0.$$

We also write $X \sim \mathcal{W}(\alpha, \beta)$ when X has this distribution function, i.e., $P(X \leq x) = F_{\alpha,\beta}(x)$. The parameters $\alpha > 0$ and $\beta > 0$ are referred to as scale and shape parameter, respectively. The Weibull density has the following form

$$f_{\alpha,\beta}(x) = F'_{\alpha,\beta}(x) = \frac{d}{dx} F_{\alpha,\beta}(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] .$$

For $\beta = 1$ the Weibull distribution coincides with the exponential distribution with mean α. In general, α represents the .632-quantile of the Weibull distribution regardless of the value of β since $F_{\alpha,\beta}(\alpha) = 1 - \exp(-1) \approx .632$ for all $\beta > 0$. Figure 1 shows a representative collection of Weibull densities. Note that the spread of the Weibull distributions around α gets smaller as β increases. The reason for this will become clearer later when we discuss the log-transform of Weibull random variables.

The m^{th} moment of the Weibull distribution is

$$\mathrm{E}(X^m) = \alpha^m \Gamma(1 + m/\beta)$$

and thus the mean and variance are given by

$$\mu = E(X) = \alpha\Gamma(1 + 1/\beta) \qquad \text{and} \qquad \sigma^2 = \alpha^2\left[\Gamma(1 + 2/\beta) - \{\Gamma(1 + 1/\beta)\}^2\right] .$$

Its p-quantile, defined by $P(X \leq x_p) = p$, is

$$x_p = \alpha(-\log(1 - p))^{1/\beta} .$$

For $p = 1 - \exp(-1) \approx .632$ (i.e., $-\log(1 - p) = 1$) we have $x_p = \alpha$ regardless of β, as pointed out previously. For that reason one also calls α the *characteristic life* of the Weibull distribution. The term *life* comes from the common use of the Weibull distribution in modeling lifetime data. More on this later.

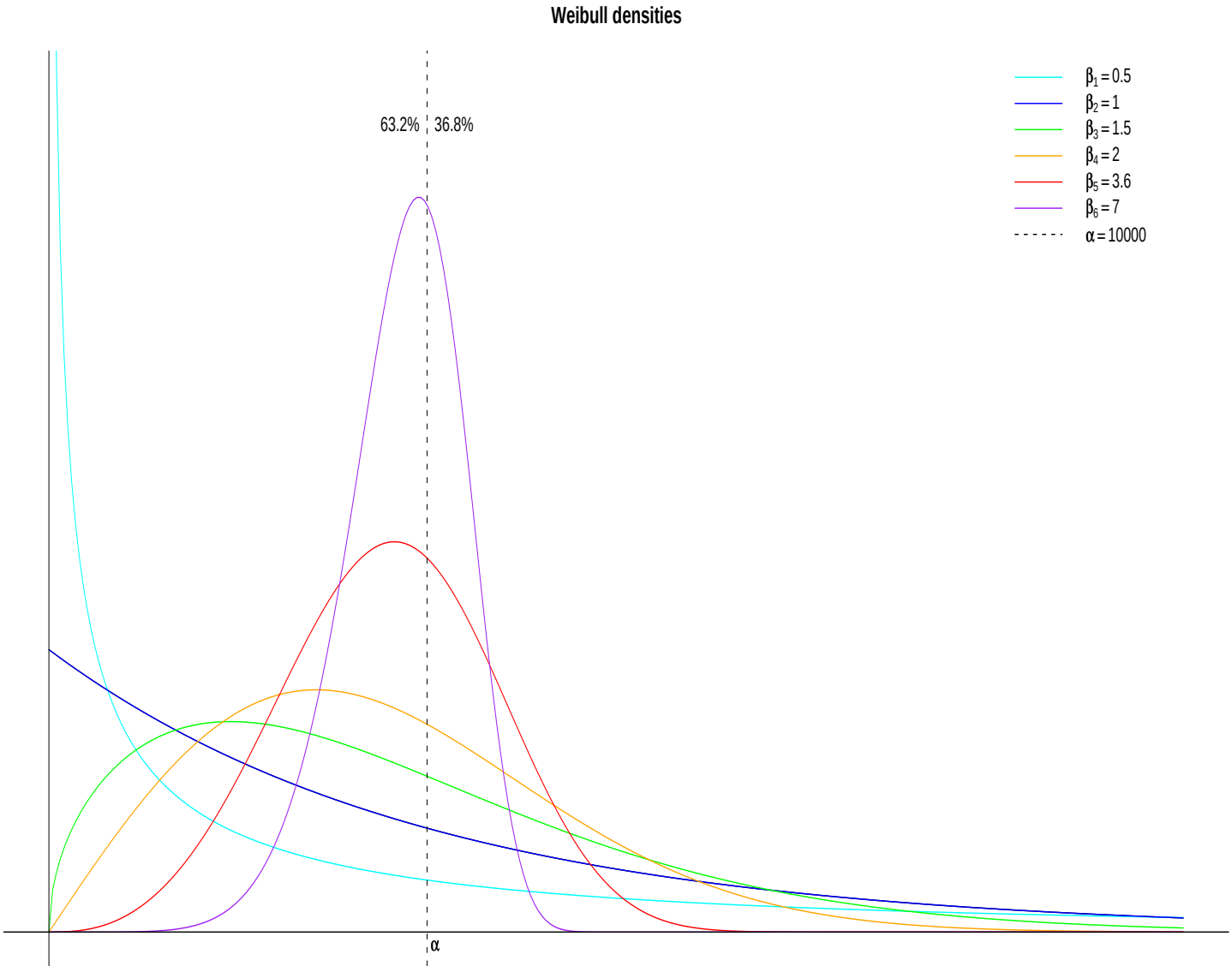


Figure 1: A Collection of Weibull Densities with $\alpha = 10000$ and Various Shapes

2 Minimum Closure and Weakest Link Property

The Weibull distribution has the following minimum closure property: If $X_1, \ldots, X_n$ are independent with $X_i \sim \mathcal{W}(\alpha_i, \beta),\ i = 1, \ldots, n$, then

$$\begin{aligned} P(\min(X_1, \ldots, X_n) > t) &= P(X_1 > t, \ldots, X_n > t) = \prod_{i=1}^{n} P(X_i > t) \\ &= \prod_{i=1}^{n} \exp\left[-\left(\frac{t}{\alpha_i}\right)^{\beta}\right] = \exp\left[-t^{\beta} \sum_{i=1}^{n} \frac{1}{\alpha_i^{\beta}}\right] \\ &= \exp\left[-\left(\frac{t}{\alpha^{\star}}\right)^{\beta}\right] \quad \text{with} \quad \alpha^{\star} = \left(\sum_{i=1}^{n} \frac{1}{\alpha_i^{\beta}}\right)^{-1/\beta}, \end{aligned}$$

i.e., $\min(X_1, \ldots, X_n) \sim \mathcal{W}(\alpha^{\star}, \beta)$. This is reminiscent of the closure property for the normal distribution under summation, i.e., if $X_1, \ldots, X_n$ are independent with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ then

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

This summation closure property plays an essential role in proving the central limit theorem: Sums of independent random variables (not necessarily normally distributed) have an approximate normal distribution, subject to some mild conditions concerning the distribution of such random variables. There is a similar result from Extreme Value Theory that says: The minimum of independent, identically distributed random variables (not necessarily Weibull distributed) has an approximate Weibull distribution, subject to some mild conditions concerning the distribution of such random variables. This is also referred to as the "weakest link" motivation for the Weibull distribution.

The Weibull distribution is appropriate when trying to characterize the random strength of materials or the random lifetime of some system. This is related to the weakest link property as follows. A piece of material can be viewed as a concatenation of many smaller material cells, each of which has its random breaking strength X_i when subjected to stress. Thus the strength of the concatenated total piece is the strength of its weakest link, namely $\min(X_1, \ldots, X_n)$, i.e., approximately Weibull. Similarly, a system can be viewed as a collection of many parts or subsystems, each of which has a random lifetime X_i. If the system is defined to be in a failed state whenever any one of its parts or subsystems fails, then the system lifetime is $\min(X_1, \ldots, X_n)$, i.e., approximately Weibull.

Figure 2 gives a sense of usage of the Weibull distribution and Figure 3 shows the "real thing." Googling "Weibull distribution" produced 185,000 hits while "normal distribution" had 2,420,000 hits.

Societies and worked to the last day of his remarkable life. He died on October 12, 1979, in Annecy, France. * 6.18.1887

Perhaps the greatest tribute to anyone's work is the extent to which it is used and cited by subsequent researchers. The table and graph below, supplied by Mr. Göran Weibull, show the rapidly increasing use of the Weibull distribution and the wide range of fields in which it has been applied. Although the tabulation indicates the variety of applications, it underestimates the total number of references. Another 600 have appeared since the tabulations were prepared. 1984

As some of you know, I am of Hungarian origin. In the Hungarian language, "Waloddi", spelled "Valodi", means "the Real Thing". Waloddi Weibull was the real thing.

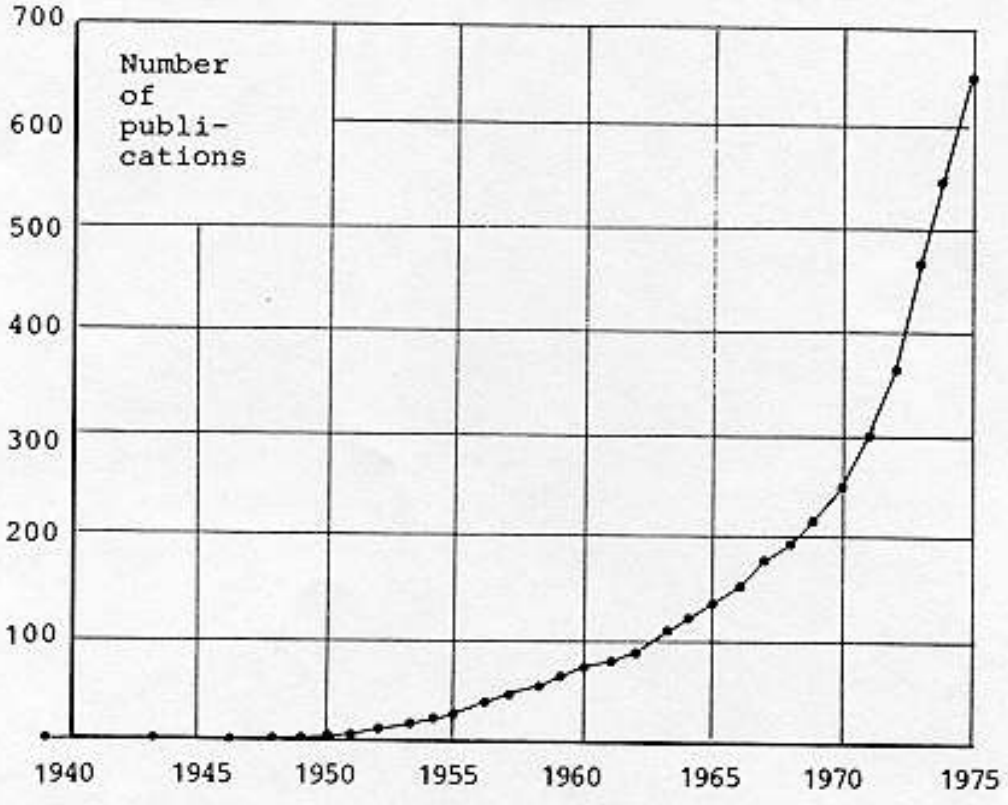


Figure 2: Publications on the Weibull Distribution

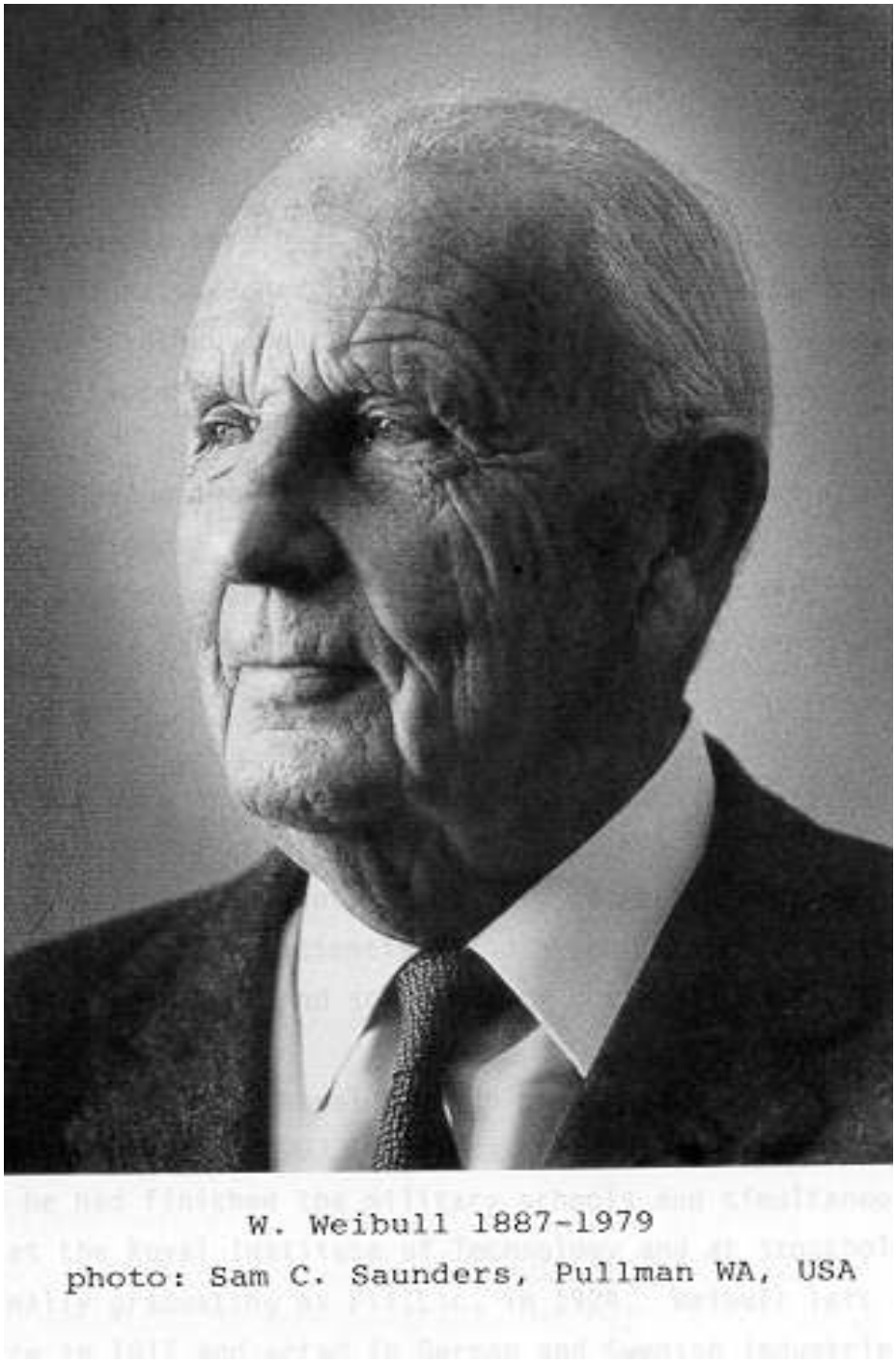


Figure 3: Waloddi Weibull

The Weibull distribution is very popular among engineers. One reason for this is that the Weibull cdf has a closed form which is not the case for the normal cdf $\Phi(x)$. However, in today's computing environment one could argue that point since typically the computation of even $\exp(x)$ requires computing. That this can be accomplished on most calculators is also moot since many calculators also give you $\Phi(x)$. Another reason for the popularity of the Weibull distribution among engineers may be that Weibull's most famous paper, originally submitted to a statistics journal and rejected, was eventually published in an engineering journal: Waloddi Weibull (1951) "A statistical distribution function of wide applicability." *Journal of Applied Mechanics*, **18**, 293-297.

"... he tried to publish an article in a well-known British journal. At this time, the distribution function proposed by Gauss was dominating and was distinguishingly called the normal distribution. By some statisticians it was even believed to be the only possible one. The article was refused with the comment that it was interesting but of no practical importance. That was just the same article as the highly cited one published in 1951." (Göran W. Weibull, 1981, `http://www.garfield.library.upenn.edu/classics1981/A1981LD32400001.pdf`)

Sam Saunders (1975): 'Professor Wallodi (sic) Weibull recounted to me that the now famous paper of his "A Statistical Distribution of Wide Applicability", in which was first advocated the "Weibull" distribution with its failure rate a power of time, was rejected by the Journal of the American Statistical Association as being of no interrest. Thus one of the most influential papers in statistics of that decade was published in the Journal of Applied Mechanics. See [35]. (Maybe that is the reason it was so influential!)'

3 The Hazard Function

The hazard function for any nonnegative random variable with cdf $F(x)$ and density $f(x)$ is defined as $h(x) = f(x)/(1-F(x))$. It is usually employed for distributions that model random lifetimes and it relates to the probability that a lifetime comes to an end within the next small time increment of length d given that the lifetime has exceeded x so far, namely

$$P(x < X \le x + d | X > x) = \frac{P(x < X \le x + d)}{P(X > x)} = \frac{F(x+d) - F(x)}{1 - F(x)} \approx \frac{d \times f(x)}{1 - F(x)} = d \times h(x) \; .$$

In the case of the Weibull distribution we have

$$h(x) = \frac{f_{\alpha,\beta}(x)}{1 - F_{\alpha,\beta}(x)} = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \; .$$

Various other terms are used equivalently for the hazard function, such as hazard rate, failure rate (function), or force of mortality. In the case of the Weibull hazard rate function we observe that it

is increasing in x when $\beta > 1$, decreasing in x when $\beta < 1$ and constant when $\beta = 1$ (exponential distribution with memoryless property).

When $\beta > 1$ the part or system, for which the lifetime is modeled by a Weibull distribution, is subject to *aging* in the sense that an older system has a higher chance of failing during the next small time increment d than a younger system.

For $\beta < 1$ (less common) the system has a better chance of surviving the next small time increment d as it gets older, possibly due to hardening, maturing, or curing. Often one refers to this situation as one of infant mortality, i.e., after initial early failures the survival gets better with age. However, one has to keep in mind that we may be modeling parts or systems that consist of a mixture of defective or weak parts and of parts that practically can live forever. A Weibull distribution with $\beta < 1$ may not do full justice to such a mixture distribution.

For $\beta = 1$ there is no aging, i.e., the system is as good as new given that it has survived beyond x, since for $\beta = 1$ we have

$$P(X > x + h | X > x) = \frac{P(X > x + h)}{P(X > x)} = \frac{\exp(-(x + h)/\alpha)}{\exp(-x/\alpha)} = \exp(-h/\alpha) = P(X > h) \ ,$$

i.e., it is again exponential with same mean α. One also refers to this as a random failure model in the sense that failures are due to external shocks that follow a Poisson process with rate $\lambda = 1/\alpha$. The random times between shocks are exponentially distributed with mean α. Given that there are k such shock events in an interval $[0, T]$ one can view the k occurrence times as being uniformly distributed over the interval $[0, T]$, hence the allusion to random failures.

4 Location-Scale Property of $\log(X)$

Another useful property, of which we will make strong use, is the following location-scale property of the log-transformed Weibull distribution. By that we mean that: $X \sim \mathcal{W}(\alpha, \beta) \Longrightarrow \log(X) = Y$ has a location-scale distribution, namely its cumulative distribution function (cdf) is

$$\begin{aligned} P(Y \leq y) &= P(\log(X) \leq y) = P(X \leq \exp(y)) = 1 - \exp\left[- \left(\frac{\exp(y)}{\alpha} \right)^{\beta} \right] \\ &= 1 - \exp\left[- \exp\left\{ (y - \log(\alpha)) \times \beta \right\} \right] = 1 - \exp\left[- \exp\left(\frac{y - \log(\alpha)}{1/\beta} \right) \right] \\ &= 1 - \exp\left[- \exp\left(\frac{y - u}{b} \right) \right] \end{aligned}$$

with location parameter $u = \log(\alpha)$ and scale parameter $b = 1/\beta$. The reason for referring to such parameters this way is the following. If $Z \sim G(z)$ then $Y = \mu + \sigma Z \sim G((y - \mu)/\sigma)$ since

$$H(y) = P(Y \leq y) = P(\mu + \sigma Z \leq y) = P(Z \leq (y - \mu)/\sigma) = G((y - \mu)/\sigma) \ .$$

The form $Y = \mu + \sigma X$ should make clear the notion of location scale parameter, since Z has been scaled by the factor σ and is then shifted by μ. Two prominent location-scale families are

1. $Y = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$, where $Z \sim \mathcal{N}(0, 1)$ is standard normal with cdf $G(z) = \Phi(z)$ and thus Y has cdf $H(y) = \Phi((y - \mu)/\sigma)$,

2. $Y = u + bZ$ where Z has the standard extreme value distribution with cdf $G(z) = 1 - \exp(-\exp(z))$ for $z \in R$, as in our log-transformed Weibull example above.

In any such a location-scale model there is a simple relationship between the p-quantiles of Y and Z, namely $y_p = \mu + \sigma z_p$ in the normal model and $y_p = u + bw_p$ in the extreme value model (using the location and scale parameters u and b resulting from log-transformed Weibull data). We just illustrate this in the extreme value location-scale model.

$$p = P(Z \leq w_p) = P(u + bZ \leq u + bw_p) = P(Y \leq u + bw_p) \qquad \Longrightarrow \qquad y_p = u + bw_p$$

with $w_p = \log(-\log(1 - p))$. Thus y_p is a linear function of $w_p = \log(-\log(1 - p))$, the p-quantile of G. While w_p is known and easily computable from p, the same cannot be said about y_p, since it involves the typically unknown parameters u and b. However, for appropriate $p_i = (i - .5)/n$ one can view the i^{th} ordered sample value $Y_{(i)}$ $(Y_{(1)} \leq \ldots \leq Y_{(n)})$ as a good approximation for y_{p_i}. Thus the plot of $Y_{(i)}$ against w_{p_i} should look approximately linear. This is the basis for Weibull probability plotting (and the case of plotting $Y_{(i)}$ against z_{p_i} for normal probability plotting), a very appealing graphical procedure which gives a visual impression of how well the data fit the assumed model (normal or Weibull) and which also allows for a crude estimation of the unknown location and scale parameters, since they relate to the slope and intercept of the line that may be fitted to the perceived linear point pattern. For more in relation to Weibull probability plotting we refer to Scholz (2008).

5 Maximum Likelihood Estimation

There are many ways to estimate the parameters $\theta = (\alpha, \beta)$ based on a random sample $X_1, \ldots, X_n \sim \mathcal{W}(\alpha, \beta)$. Maximum likelihood estimation (MLE) is generally the most versatile and popular method. Although MLE in the Weibull case requires numerical methods and a computer, that is no longer an issue in today's computing environment. Previously, estimates that could be computed by hand had been investigated, but they are usually less efficient than mle's (estimates derived by MLE). By efficient estimates we loosely refer to estimates that have the smallest sampling variance. MLE tends to be efficient, at least in large samples. Furthermore, under regularity conditions MLE produces estimates that have an approximate normal distribution in large samples.

When $X_1, \ldots, X_n \sim F_\theta(x)$ with density $f_\theta(x)$ then the maximum likelihood estimate of θ is that value $\theta = \hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ which maximizes the likelihood

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f_\theta(x_i)$$

over θ, i.e., which gives highest local probability to the observed sample $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$

$$L(x_1, \ldots, x_n, \hat{\theta}) = \sup_\theta \left\{ \prod_{i=1}^{n} f_\theta(x_i) \right\} .$$

Often such maximizing values $\hat{\theta}$ are unique and one can obtain them by solving, i.e.,

$$\frac{\partial}{\partial \theta_j} \prod_{i=1}^{n} f_\theta(x_i) = 0 \qquad j = 1, \ldots, k ,$$

where k is the number of parameters involved in $\theta = (\theta_1, \ldots, \theta_k)$. These above equations reflect the fact that a smooth function has a horizontal tangent plane at its maximum (minimum or saddle point). Thus solving such equations is necessary but not sufficient, since it still needs to be shown that it is the location of a maximum.

Since taking derivatives of a product is tedious (product rule) one usually resorts to maximizing the log of the likelihood, i.e.,

$$\ell(x_1, \ldots, x_n, \theta) = \log\left(L(x_1, \ldots, x_n, \theta)\right) = \sum_{i=1}^{n} \log\left(f_\theta(x_i)\right)$$

since the value of θ that maximizes $L(x_1, \ldots, x_n, \theta)$ is the same as the value that maximizes $\ell(x_1, \ldots, x_n, \theta)$, i.e.,

$$\ell(x_1, \ldots, x_n, \hat{\theta}) = \sup_\theta \left\{ \sum_{i=1}^{n} \log\left(f_\theta(x_i)\right) \right\} .$$

It is a lot simpler to deal with the likelihood equations

$$\frac{\partial}{\partial \theta_j} \ell(x_1, \ldots, x_n, \hat{\theta}) = \frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} \log(f_\theta(x_i)) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log(f_\theta(x_i)) = 0 \qquad j = 1, \ldots, k$$

when solving for $\theta = \hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$.

In the case of a normal random sample we have $\theta = (\mu, \sigma)$ with $k = 2$ and the unique solution of the likelihood equations results in the explicit expressions

$$\hat{\mu} = \bar{x} = \sum_{i=1}^{n} x_i/n \qquad \text{and} \qquad \hat{\sigma} = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2/n} \qquad \text{and thus} \qquad \hat{\theta} = (\hat{\mu}, \hat{\sigma}) .$$

In the case of a Weibull sample we take the further simplifying step of dealing with the log-transformed sample $(y_1, \ldots, y_n) = (\log(x_1), \ldots, \log(x_n))$. Recall that $Y_i = \log(X_i)$ has cdf $F(y) = 1 - \exp(-\exp((x-u)/b)) = G((y-u)/b)$ with $G(z) = 1 - \exp(-\exp(z))$ with $g(z) = G'(z) = \exp(z - \exp(z))$. Thus

$$f(y) = F'(y) = \frac{d}{dy}F(y) = \frac{1}{b}g((y-u)/b))$$

with

$$\log(f(y)) = -\log(b) + \frac{y-u}{b} - \exp\left(\frac{y-u}{b}\right) .$$

As partial derivatives of $\log(f(y))$ with respect to u and b we get

$$\begin{aligned}
\frac{\partial}{\partial u}\log(f(y)) &= -\frac{1}{b} + \frac{1}{b}\exp\left(\frac{y-u}{b}\right) \\
\frac{\partial}{\partial b}\log(f(y)) &= -\frac{1}{b} - \frac{1}{b}\frac{y-u}{b} + \frac{1}{b}\frac{y-u}{b}\exp\left(\frac{y-u}{b}\right)
\end{aligned}$$

and thus as likelihood equations

$$\begin{aligned}
0 &= -\frac{n}{b} + \frac{1}{b}\sum_{i=1}^{n}\exp\left(\frac{y_i-u}{b}\right) \quad \text{or} \quad \sum_{i=1}^{n}\exp\left(\frac{y_i-u}{b}\right) = n \quad \text{or} \quad \exp(u) = \left[\frac{1}{n}\sum_{i=1}^{n}\exp\left(\frac{y_i}{b}\right)\right]^b , \\
0 &= -\frac{n}{b} - \frac{1}{b}\sum_{i=1}^{n}\frac{y_i-u}{b} + \frac{1}{b}\sum_{i=1}^{n}\frac{y_i-u}{b}\exp\left(\frac{y_i-u}{b}\right) .
\end{aligned}$$

i.e., we have a solution $u = \hat{u}$ once we have a solution $b = \hat{b}$. Substituting this expression for $\exp(u)$ into the second likelihood equation we get (after some cancelation and manipulation)

$$0 = \frac{\sum_{i=1}^{n} y_i \exp(y_i/b)}{\sum_{i=1}^{n}\exp(y_i/b)} - b - \frac{1}{n}\sum_{i=1}^{n} y_i .$$

Analyzing the solvability of this equation is more convenient in terms of $\beta = 1/b$ and we thus write

$$0 = \sum_{i=1}^{n} y_i w_i(\beta) - \frac{1}{\beta} - \bar{y} \qquad \text{where} \qquad w_i(\beta) = \frac{\exp(y_i\beta)}{\sum_{j=1}^{n}\exp(y_j\beta)} \qquad \text{with} \qquad \sum_{i=1}^{n} w_i(\beta) = 1 .$$

Note that the derivative of these weights with respect to β take the following form

$$w_i'(\beta) = \frac{d}{d\beta}w_i(\beta) = y_i w_i(\beta) - w_i(\beta)\sum_{j=1}^{n} y_j w_j(\beta) .$$

Hence

$$\frac{d}{d\beta}\left\{\sum_{i=1}^{n} y_i w_i(\beta) - \frac{1}{\beta} - \bar{y}\right\} = \sum_{i=1}^{n} y_i w_i'(\beta) + \frac{1}{\beta^2} = \sum_{i=1}^{n} y_i^2 w_i(\beta) - \left(\sum_{j=1}^{n} y_j w_j(\beta)\right)^2 + \frac{1}{\beta^2} > 0$$

since

$$\mathrm{var}_w(y) = \sum_{i=1}^{n} y_i^2 w_i(\beta) - \left(\sum_{j=1}^{n} y_j w_j(\beta)\right)^2 = E_w(y^2) - [E_w(y)]^2 \geq 0$$

can be interpreted as a variance of the n values of $y = (y_1, \ldots, y_n)$ with weights or probabilities given by $w = (w_1(\beta), \ldots, w_n(\beta))$. Thus the reduced second likelihood equation $\sum y_i w_i(\beta) - 1/\beta - \bar{y} = 0$ has a unique solution (if it has a solution at all) since the the equation's left side is strictly increasing. Note that $w_i(\beta) \to 1/n$ as $\beta \to 0$. Thus $\sum y_i w_i(\beta) - 1/\beta - \bar{y} \approx -1/\beta \to -\infty$ as $\beta \to 0$.
Furthermore, with $M = \max(y_1, \ldots, y_n)$ and $\beta \to \infty$ we have

$$w_i(\beta) = \exp(\beta(y_i - M))/\sum_{j=1}^{n} \exp(\beta(y_j - M)) \to 0 \quad \text{when } y_i < M \text{ and} \quad w_i(\beta) \to 1/r \quad \text{when } y_i = M,$$

where $r \geq 1$ is the number of y_i coinciding with M. Thus

$$\sum y_i w_i(\beta) - 1/\beta - \bar{y} \approx M - 1/\beta - \bar{y} \to M - \bar{y} > 0 \quad \text{as } \beta \to \infty$$

where $M - \bar{y} > 0$ assumes that not all y_i coincide (a degenerate case with probability 0). That this unique solution corresponds to a maximum and thus a unique global maximum takes some extra effort and we refer to Scholz (1996) for an even more general treatment that covers Weibull analysis with censored data and covariates.
However, a somewhat loose argument can be given as follows. If we consider the likelihood of the log-transformed Weibull data we have

$$L(y_1, \ldots, y_n, u, b) = \frac{1}{b^n} \prod_{i=1}^{n} g\left(\frac{y_i - u}{b}\right) .$$

Contemplate this likelihood for fixed $y = (y_1, \ldots, y_n)$ and for parameters u with $|u| \to \infty$ (the location moves away from all observed data values $y_1, \ldots, y_n$) and b with $b \to 0$ (the spread becomes very concentrated on some point and cannot simultaneously do so at all values $y_1, \ldots, y_n$, unless they are all the same, excluded as a zero probability degeneracy) and $b \to \infty$ (in which case all probability is diffused thinly over the whole half plane $\{(u, b) : u \in R, b > 0\}$), it is then easily seen that this likelihood approaches zero in all cases. Since this likelihood is positive everywhere (but approaching zero near the fringes of the parameter space, the above half plane) it follows that it

must have a maximum somewhere with zero partial derivatives. We showed there is only one such point (uniqueness of the solution to the likelihood equations) and thus there can only be one unique (global) maximum, which then is also the unique maximum likelihood estimate $\hat{\theta} = (\hat{u}, \hat{b})$.

In solving $0 = \sum y_i \exp(y_i/b)/\sum \exp(y_i/b) - b - \bar{y}$ it is numerically advantageous to solve the equivalent equation $0 = \sum y_i \exp((y_i - M)/b)/\sum \exp((y_i - M)/b) - b - \bar{y}$ where $M = \max(y_1, \ldots, y_n)$. This avoids overflow or accuracy loss in the exponentials when the y_i tend to be large.

The above derivations go through with very little change when instead of observing a full sample $Y_1, \ldots, Y_n$ we only observe the $r \geq 2$ smallest sample values $Y_{(1)} < \ldots < Y_{(r)}$. Such data is referred to as type II censored data. This situation typically arises in a laboratory setting when several units are put on test (subjected to failure exposure) simultaneously and the test is terminated (or evaluated) when the first r units have failed. In that case we know the first r failure times $X_{(1)} < \ldots < X_{(r)}$ and thus $Y_{(i)} = \log(X_{(i)}), i = 1, \ldots, r$, and we know that the lifetimes of the remaining units exceed $X_{(r)}$ or that $Y_{(i)} > Y_{(r)}$ for $i > r$. The advantage of such data collection is that we do not have to wait until all n units have failed. Furthermore, if we put a lot of units on test (high n) we increase our chance of seeing our first r failures before a fixed time y. This is a simple consequence of the following binomial probability statement:

$$P(Y(r) \leq y) = P(\text{at least } r \text{ failures} \leq y \text{ in } n \text{ trials}) = \sum_{i=r}^{n} \binom{n}{i} P(Y \leq y)^i (1 - P(Y \leq y))^{n-i}$$

which is strictly increasing in n for any fixed y and $r \geq 1$ (exercise).

The joint density of $Y_{(1)}, \ldots, Y_{(n)}$ at $(y_1, \ldots, y_n)$ with $y_1 < \ldots < y_n$ is

$$f(y_1, \ldots, y_n) = n! \prod_{i=1}^{n} \frac{1}{b} g\left(\frac{y_i - u}{b}\right) = n! \prod_{i=1}^{n} f(y_i)$$

where the multiplier $n!$ just accounts for the fact that all $n!$ permutations of $y_1, \ldots, y_n$ could have been the order in which these values were observed and all of these orders have the same density (probability). Integrating out $y_n > y_{n-1} > \ldots > y_{r+1} (> y_r)$ and using $\bar{F}(y) = 1 - F(y)$ we get after $n - r$ successive integration steps the joint density of the first r failure times $y_1 < \ldots < y_r$ as

$$\begin{aligned}
f(y_1, \ldots, y_{n-1}) &= n! \prod_{i=1}^{n-1} f(y_i) \times \int_{y_{n-1}}^{\infty} f(y_n) dy_n = n! \prod_{i=1}^{n-1} f(y_i) \bar{F}(y_{n-1}) \\
f(y_1, \ldots, y_{n-2}) &= n! \prod_{i=1}^{n-2} f(y_i) \times \int_{y_{n-2}}^{\infty} f(y_{n-1}) \bar{F}(y_{n-1}) dy_{n-1} = n! \prod_{i=1}^{n-2} f(y_i) \times \frac{1}{2} \bar{F}^2(y_{n-2}) \\
f(y_1, \ldots, y_{n-3}) &= n! \prod_{i=1}^{n-3} f(y_i) \times \int_{y_{n-3}}^{\infty} f(y_{n-2}) \bar{F}^2(y_{n-2})/2 dy_{n-2} = n! \prod_{i=1}^{n-3} f(y_i) \times \frac{1}{3!} \bar{F}^3(y_{n-3})
\end{aligned}$$

$$\ldots$$

$$f(y_1,\ldots,y_r) = n!\prod_{i=1}^{r} f(y_i) \times \frac{1}{(n-r)!}\bar{F}^{n-r}(y_r) = \left[\frac{n!}{(n-r)!}\prod_{i=1}^{r} f(y_i)\right] \times [1-F(y_r)]^{n-r}$$

$$= r!\prod_{i=1}^{r}\frac{1}{b}g\left(\frac{y_i-u}{b}\right) \times \binom{n}{r}\left[1-G\left(\frac{y_r-u}{b}\right)\right]^{n-r}$$

with log-likelihood

$$\ell(y_1,\ldots,y_r,u,b) = \log\left(\frac{n!}{(n-r)!}\right) - r\log(b) + \sum_{i=1}^{r}\frac{y_i-u}{b} - \sum_{i=1}^{r\star}\exp\left(\frac{y_i-u}{b}\right)$$

where we use the notation

$$\sum_{i=1}^{r\star} x_i = \sum_{i=1}^{r} x_i + (n-r)x_r \ .$$

The likelihood equations are

$$0 = \frac{\partial}{\partial u}\ell(y_1,\ldots,y_r,u,b) = -\frac{r}{b} + \frac{1}{b}\sum_{i=1}^{r\star}\exp\left(\frac{y_i-u}{b}\right) \quad \text{or} \quad \exp(u) = \left[\frac{1}{r}\sum_{i=1}^{r\star}\exp\left(\frac{y_i}{b}\right)\right]^b$$

$$0 = \frac{\partial}{\partial b}\ell(y_1,\ldots,y_r,u,b) = -\frac{r}{b} - \frac{1}{b}\sum_{i=1}^{r}\frac{y_i-u}{b} + \frac{1}{b}\sum_{i=1}^{r\star}\frac{y_i-u}{b}\exp\left(\frac{y_i-u}{b}\right)$$

where again the transformed first equation gives us a solution $\hat{u}$ once we have a solution $\hat{b}$ for b. Using this in the second equation it transforms to a single equation in b alone, namely

$$\sum_{i=1}^{r\star} y_i\exp(y_i/b)\Big/\sum_{i=1}^{r\star}\exp(y_i/b) - b - \frac{1}{r}\sum_{i=1}^{r} y_i = 0 \ .$$

Again it is advisable to use the equivalent but computationally more stable form

$$\sum_{i=1}^{r\star} y_i\exp((y_i-y_r)/b)\Big/\sum_{i=1}^{r\star}\exp((y_i-y_r)/b) - b - \frac{1}{r}\sum_{i=1}^{r} y_i = 0 \ .$$

As in the complete sample case one sees that this equation has a unique solution $\hat{b}$ and that $(\hat{u},\hat{b})$ gives the location of the (unique) global maximimum of the likelihood function, i.e., $(\hat{u},\hat{b})$ are the mle's.

6 Computation of Maximum Likelihood Estimates in R

The computation of the mle's of the Weibull parameters α and β is facilitated by the function `survreg` which is part of the R package `survival`. Here `survreg` is used in its most basic form in the context of Weibull data (full sample or type II censored Weibull data). `survreg` does a whole lot more than compute the mle's but we will not deal with these aspects here, at least for now. The following is an R function, called `Weibull.mle`, that uses `survreg` to compute these estimates. Note that it tests for the existence of `survreg` before calling it. This function is part of the R work space that is posted on the class web site.

```
Weibull.mle <- function (x=NULL,n=NULL){
# This function computes the maximum likelihood estimates of alpha and beta
# for complete or type II censored samples assumed to come from a 2-parameter
# Weibull distribution. Here x is the sample, either the full sample or the first
# r observations of a type II censored sample. In the latter case one must specify
# the full sample size n, otherwise x is treated as a full sample.
# If x is not given then a default full sample of size n=10, namely
# c(7,12.1,22.8,23.1,25.7,26.7,29.0,29.9,39.5,41.9) is analyzed and the returned
# results should be
# $mles
# alpha.hat  beta.hat
# 28.914017  2.799793
#
# In the type II censored usage
#  Weibull.mle(c(7,12.1,22.8,23.1,25.7),10)
# $mles
# alpha.hat  beta.hat
# 30.725992  2.432647
if(is.null(x))x <- c(7,12.1,22.8,23.1,25.7,26.7,29.0,29.9,39.5,41.9)
r <- length(x)
if(is.null(n)){n<-r}else{if(r>n||r<2){
return("x must have length r with: 2 <= r <= n")}}
xs <- sort(x)
if(!exists("survreg"))library(survival)
#tests whether survival package is loaded, if not, then it loads survival
if(r<n){
statusx <- c(rep(1,r),rep(0,n-r))
dat.weibull <- data.frame(c(xs,rep(xs[r],n-r)),statusx)
```

```
}else{statusx <- rep(1,n)
dat.weibull <- data.frame(xs,statusx)}
names(dat.weibull)<-c("time","status")
out.weibull <- survreg(Surv(time,status)~1,dist="weibull",data=dat.weibull)
alpha.hat <- exp(out.weibull$coef)
beta.hat <- 1/out.weibull$scale
parms <- c(alpha.hat,beta.hat)
names(parms)<-c("alpha.hat","beta.hat")
list(mles=parms)}
```

Note that `survreg` analyzes objects of class `Surv`. Here such an object is created by the function `Surv` and it basically adjoins the failure times with a status vector of same length. The status is 1 when a time corresponds to an actual failure time. It is 0 when the corresponding time is a censoring time, i.e., we only know that the unobserved actual failure time exceeds the reported censoring time. In the case of type II censored data these censoring times equal the r^{th} largest failure time.

To get a sense of the calculation speed of this function we ran `Weibull.mle` a 1000 times, which tells us that the time to compute the mle's in a sample of size $n = 10$ is roughly $5.91/1000 = .00591$. This fact plays a significant role later on in the various inference procedures which we will discuss.

```
system.time(for(i in 1:1000){Weibull.mle(rweibull(10,1))})
   user  system elapsed
   5.79    0.00    5.91
```

For $n = 100, 500, 1000$ the elapsed times came to $8.07, 15.91$ and 25.87, respectively. The relationship of computing time to n appears to be quite linear, but with slow growth, as Figure 4 shows.

7 Location and Scale Equivariance of Maximum Likelihood Estimates

The maximum likelihood estimates $\hat{u}$ and $\hat{b}$ of the location and scale parameters u and b have the following equivariance properties which will play a strong role in the later pivot construction and resulting confidence intervals.

Based on data $\mathbf{z} = (z_1, \ldots, z_n)$ we denote the estimates of u and b more explicitly by $\hat{u}(z_1, \ldots, z_n) = \hat{u}(\mathbf{z})$ and $\hat{b}(z_1, \ldots, z_n) = \hat{b}(\mathbf{z})$. If we transform $\mathbf{z}$ to $\mathbf{r} = (r_1, \ldots, r_n)$ with $r_i = A + Bz_i$, where $A \in R$ and $B > 0$ are arbitrary constant, then

$$\hat{u}(r_1, \ldots, r_n) = A + B\hat{u}(z_1, \ldots, z_n) \qquad \text{or} \qquad \hat{u}(\mathbf{r}) = \hat{u}(A + B\mathbf{z}) = A + B\hat{u}(\mathbf{z})$$

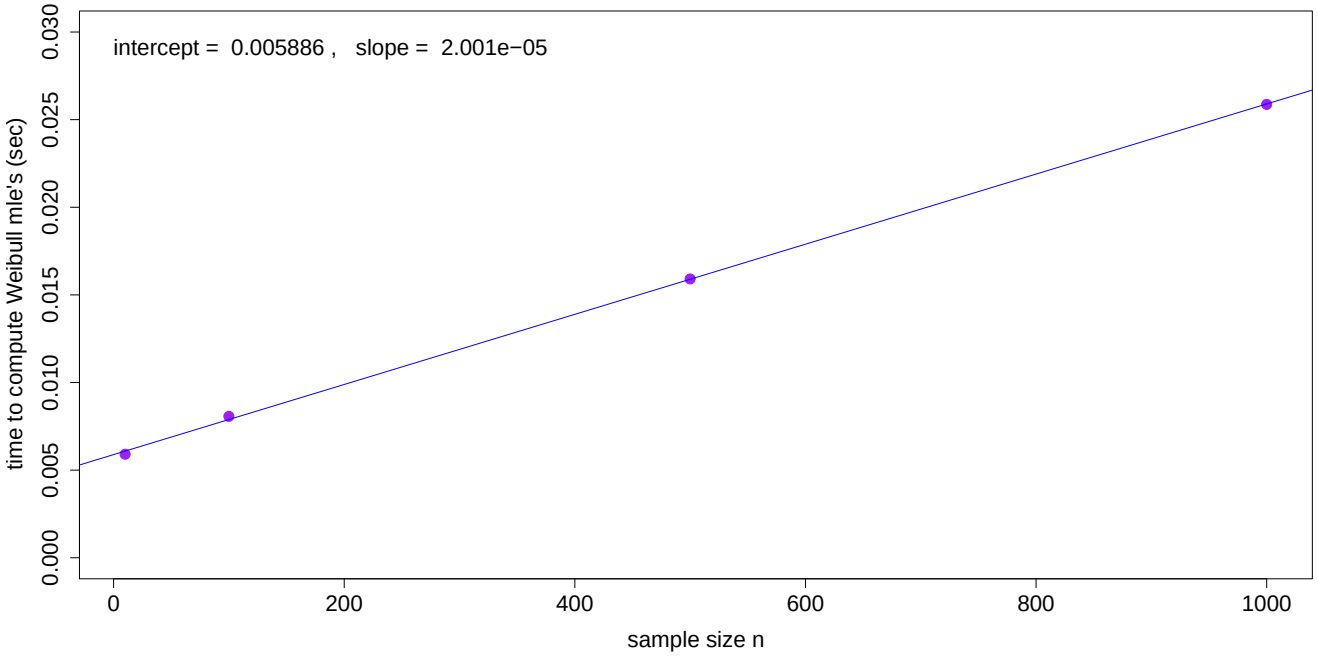


Figure 4: Weibull Parameter MLE Computation Time in Relation to Sample Size n

and

$$\hat{b}(r_1,\ldots,r_n) = B\hat{b}(z_1,\ldots,z_n) \qquad \text{or} \qquad \hat{b}(\mathbf{r}) = \hat{b}(A+B\mathbf{z}) = B\hat{b}(\mathbf{z})\ .$$

These properties are naturally desirable for any location and scale estimates and for mle's they are indeed true.

Proof: Observe the following defining properties of the mle's in terms of $\mathbf{z} = (z_1,\ldots,z_n)$ and $\mathbf{r} = (r_1,\ldots,r_n)$

$$\begin{aligned}
\sup_{u,b}\left\{\frac{1}{b^n}\prod_{i=1}^n g((z_i-u)/b)\right\} &= \frac{1}{\hat{b}^n(\mathbf{z})}\prod_{i=1}^n g((z_i-\hat{u}(\mathbf{z}))/\hat{b}(\mathbf{z})) \\
\sup_{u,b}\left\{\frac{1}{b^n}\prod_{i=1}^n g((r_i-u)/b)\right\} &= \frac{1}{\hat{b}^n(\mathbf{r})}\prod_{i=1}^n g((r_i-\hat{u}(\mathbf{r}))/\hat{b}(\mathbf{r})) \\
&= \frac{1}{B^n}\frac{1}{(\hat{b}(\mathbf{r})/B)^n}\prod_{i=1}^n g((z_i-(\hat{u}(\mathbf{r})-A)/B)/(\hat{b}(\mathbf{r})/B))
\end{aligned}$$

but also

$$\begin{aligned}
\sup_{u,b}\left\{\frac{1}{b^n}\prod_{i=1}^{n} g((r_i-u)/b)\right\} &= \sup_{u,b}\left\{\frac{1}{b^n}\prod_{i=1}^{n} g((A+Bz_i-u)/b)\right\} \\
&= \sup_{u,b}\left\{\frac{1}{B^n}\frac{1}{(b/B)^n}\prod_{i=1}^{n} g((z_i-(u-A)/B)/(b/B))\right\} \\
\begin{array}{c}\tilde{u}=(u-A)/B \\ \tilde{b}=b/B\end{array} \Rightarrow \quad &= \sup_{\tilde{u},\tilde{b}}\left\{\frac{1}{B^n}\frac{1}{\tilde{b}^n}\prod_{i=1}^{n} g((z_i-\tilde{u})/\tilde{b})\right\} = \frac{1}{B^n}\frac{1}{\hat{b}^n(\mathbf{z})}\prod_{i=1}^{n} g((z_i-\hat{u}(\mathbf{z}))/\hat{b}(\mathbf{z}))
\end{aligned}$$

Thus by the uniqueness of the mle's we have

$$\hat{u}(\mathbf{z}) = (\hat{u}(\mathbf{r})-A)/B \qquad \text{and} \qquad \hat{b}(\mathbf{z}) = \hat{b}(\mathbf{r})/B$$

or

$$\hat{u}(\mathbf{r}) = \hat{u}(A+B\mathbf{z}) = A+B\hat{u}(\mathbf{z}) \qquad \text{and} \qquad \hat{b}(\mathbf{r}) = \hat{b}(A+B\mathbf{z}) = B\hat{b}(\mathbf{z}) \qquad q.e.d.$$

The same equivariance properties hold for the mle's in the context of type II censored samples, as is easily verified.

8 Tests of Fit Based on the Empirical Distribution Function

Relying on subjective assessment of linearity in Weibull probability plots in order to judge whether a sample comes from a 2-parameter Weibull population takes a fair amount of experience. It is simpler and more objective to employ a formal test of fit which compares the empirical distribution function $\hat{F}_n(x)$ of a sample with the fitted Weibull distribution function $\hat{F}(x) = F_{\hat{\alpha},\hat{\beta}}(x)$ using one of several common discrepancy metrics.

The empirical distribution function (EDF) of a sample $X_1, \ldots, X_n$ is defined as

$$\hat{F}_n(x) = \frac{\#\text{ of observations} \le x}{n} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}$$

where $I_A = 1$ when A is true, and $I_A = 0$ when A is false. The fitted Weibull distribution function (using mle's $\hat{\alpha}$ and $\hat{\beta}$) is

$$\hat{F}(x) = F_{\hat{\alpha},\hat{\beta}}(x) = 1 - \exp\left(-\left(\frac{x}{\hat{\alpha}}\right)^{\hat{\beta}}\right) .$$

From the law of large numbers (LLN) we see that for any x we have that $\hat{F}_n(x) \longrightarrow F_{\alpha,\beta}(x)$ as $n \longrightarrow \infty$. Just view $\hat{F}_n(x)$ as a binomial proportion or as an average of Bernoulli random variables.

From MLE theory we also know that $\hat{F}(x) = F_{\hat{\alpha},\hat{\beta}}(x) \longrightarrow F_{\alpha,\beta}(x)$ as $n \longrightarrow \infty$ (also derived from the LLN).
Since the limiting cdf $F_{\alpha,\beta}(x)$ is continuous in x one can argue that these convergence statements can be made uniformly in x, i.e.,

$$\sup_x |\hat{F}_n(x) - F_{\alpha,\beta}(x)| \longrightarrow 0 \quad \text{and} \quad \sup_x |F_{\hat{\alpha},\hat{\beta}}(x) - F_{\alpha,\beta}(x)| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

and thus $$\sup_x |\hat{F}_n(x) - F_{\hat{\alpha},\hat{\beta}}(x)| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty \qquad \text{for all } \alpha > 0 \text{ and } \beta > 0.$$

The distance $$D_{\mathrm{KS}}(F, G) = \sup_x |F(x) - G(x)|$$

is known as the Kolmogorov-Smirnov distance between two cdf's F and G.
Figures 5 and 6 give illustrations of this Kolmogorov-Smirnov distance between EDF and fitted Weibull distribution and show the relationship between sampled true Weibull distribution, fitted Weibull distribution, and empirical distribution function.
Some comments:

1. It can be noted that the closeness between $\hat{F}_n(x)$ and $F_{\hat{\alpha},\hat{\beta}}(x)$ is usually more pronounced than their respective closeness to $F_{\alpha,\beta}(x)$, in spite of the sequence of the above convergence statements.

2. This can be understood from the fact that both $\hat{F}_n(x)$ and $F_{\hat{\alpha},\hat{\beta}}(x)$ fit the data, i.e., try to give a good representation of the data. The fit of the true distribution, although being the origin of the data, is not always good due to sampling variation.

3. The closeness between all three distributions improves as n gets larger.

Several other distances between cdf's F and G have been proposed and investigated in the literature. We will only discuss two of them, the Cramér-von Mises distance D_{CvM} and the Anderson-Darling distance D_{AD}. They are defined respectively as follows

$$D_{\mathrm{CvM}}(F, G) = \int_{-\infty}^{\infty} (F(x) - G(x))^2 \, dG(x) = \int_{-\infty}^{\infty} (F(x) - G(x))^2 \, g(x) \, dx$$

and

$$D_{\mathrm{AD}}(F, G) = \int_{-\infty}^{\infty} \frac{(F(x) - G(x))^2}{G(x)(1 - G(x))} \, dG(x) = \int_{-\infty}^{\infty} \frac{(F(x) - G(x))^2}{G(x)(1 - G(x))} \, g(x) \, dx \, .$$

Rather than focussing on the very local phenomenon of a maximum discrepancy at some point x as in D_{KS}, these alternate distances or discrepancy metrics integrate these distances in squared form over all x, weighted by $g(x)$ in the case of $D_{\mathrm{CvM}}(F, G)$ and by $g(x)/[G(x)(1 - G(x))]$ in the case

$D_{\rm AD}(F,G)$. In the latter case, the denominator increases the weight in the tails of the G distribution, i.e., compensates to some extent for the tapering off in the density $g(x)$. Thus $D_{\rm AD}(F,G)$ is favored in situations where judging tail behavior is important, e.g., in risk situations. Because of the integration nature of these last two metrics they have more global character. There is no easy graphical representation of these metrics, except to suggest that when viewing the previous figures illustrating $D_{\rm KS}$ one should look at all vertical distances (large and small) between $\hat{F}_n(x)$ and $\hat{F}(x)$, square them and accumulate these squares in the appropriately weighted fashion. For example, when one cdf is shifted relative to the other by a small amount (no large vertical discrepancy), these small vertical discrepancies (squared) will add up and indicate a moderately large difference between the two compared cdf's.

We point out the asymmetric nature of these last two metrics, i.e., we typically have

$$D_{\rm CvM}(F,G) \neq D_{\rm CvM}(G,F) \qquad \text{and} \qquad D_{\rm AD}(F,G) \neq D_{\rm AD}(G,F)\ .$$

When using these metrics for tests of fit one usually takes the cdf with a density (the model distribution to be tested) as the one with respect to which the integration takes place, while the other cdf is taken to be the EDF.

As complicated as these metrics may look at first glance, their computation is quite simple. We will give the following computational expressions (without proof):

$$D_{\rm KS}(\hat{F}_n(x),\hat{F}(x)) = D = \max\left[\max\left\{i/n - V_{(i)}\right\},\ \max\left\{V_{(i)} - (i-1)/n\right\}\right]$$

where $V_{(1)} \leq \ldots \leq V_{(n)}$ are the ordered values of $V_i = \hat{F}(X_i), i = 1,\ldots,n$.

For the other two test of fit criteria we have

$$D_{\rm CvM}(\hat{F}_n(x),\hat{F}(x)) = W^2 = \sum_{i=1}^{n}\left\{V_{(i)} - \frac{2i-1}{2n}\right\}^2 + \frac{1}{12n}$$

and

$$D_{\rm AD}(\hat{F}_n(x),\hat{F}(x)) = A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log(V_{(i)}) + \log(1-V_{(n-i+1)})\right]\ .$$

In order to carry out these tests of fit we need to know the null distributions of D, W^2 and A^2. Quite naturally we would reject the hypothesis of a sampled Weibull distribution whenever D or W^2 or A^2 are too large. The null distribution of D, W^2 and A^2 does not depend on the unknown parameters α and β, being estimated by $\hat{\alpha}$ and $\hat{\beta}$ in $V_i = \hat{F}(X_i) = F_{\hat{\alpha},\hat{\beta}}(X_i)$. The reason for this is that the V_i have a distribution that is independent of the unknown parameters α and β. This is seen as follows. Using our prior notation we write $\log(X_i) = Y_i = u + bZ_i$ and since

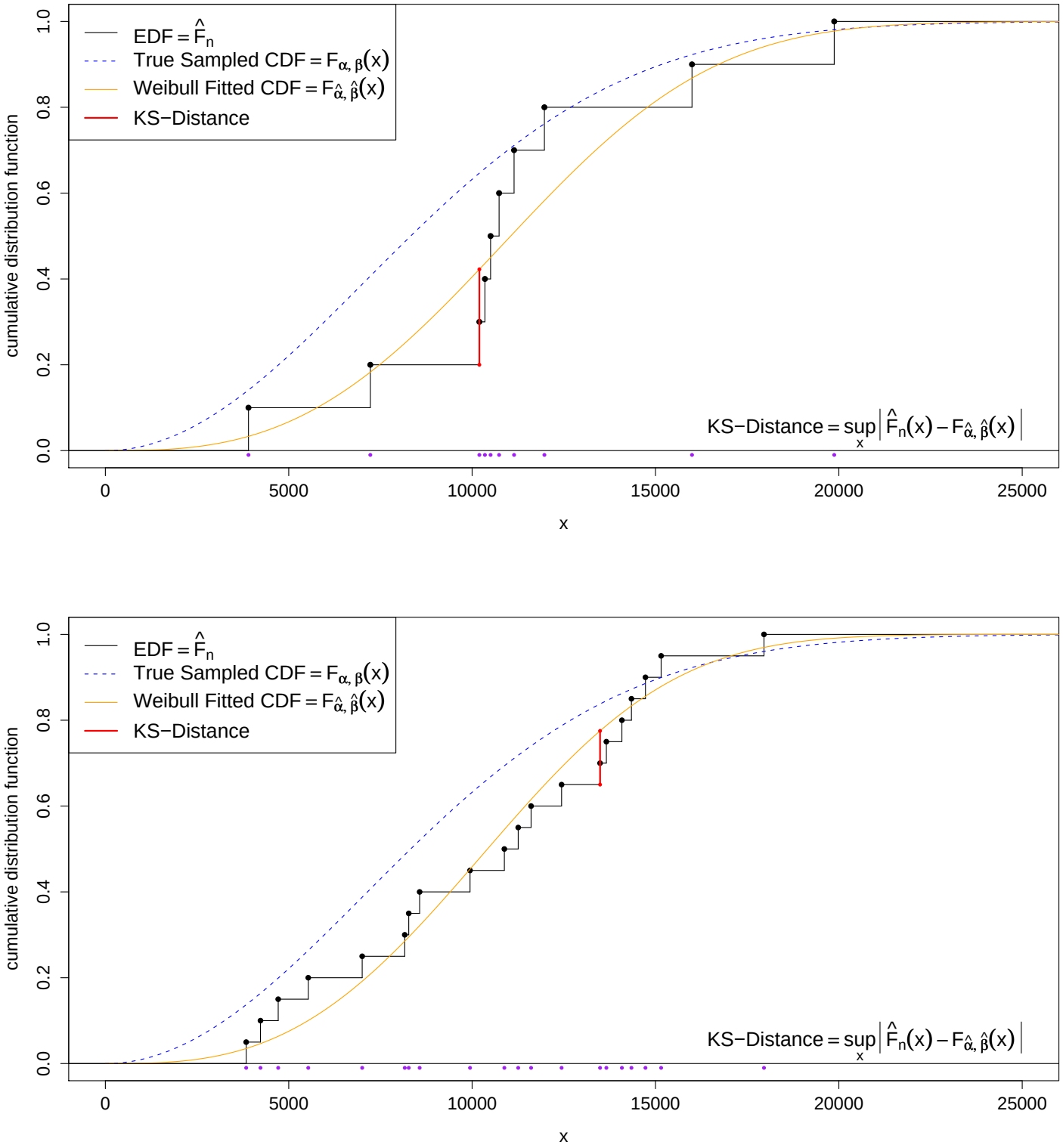


Figure 5: Illustration of Kolmogorov-Smirnov Distance for $n = 10$ and $n = 20$

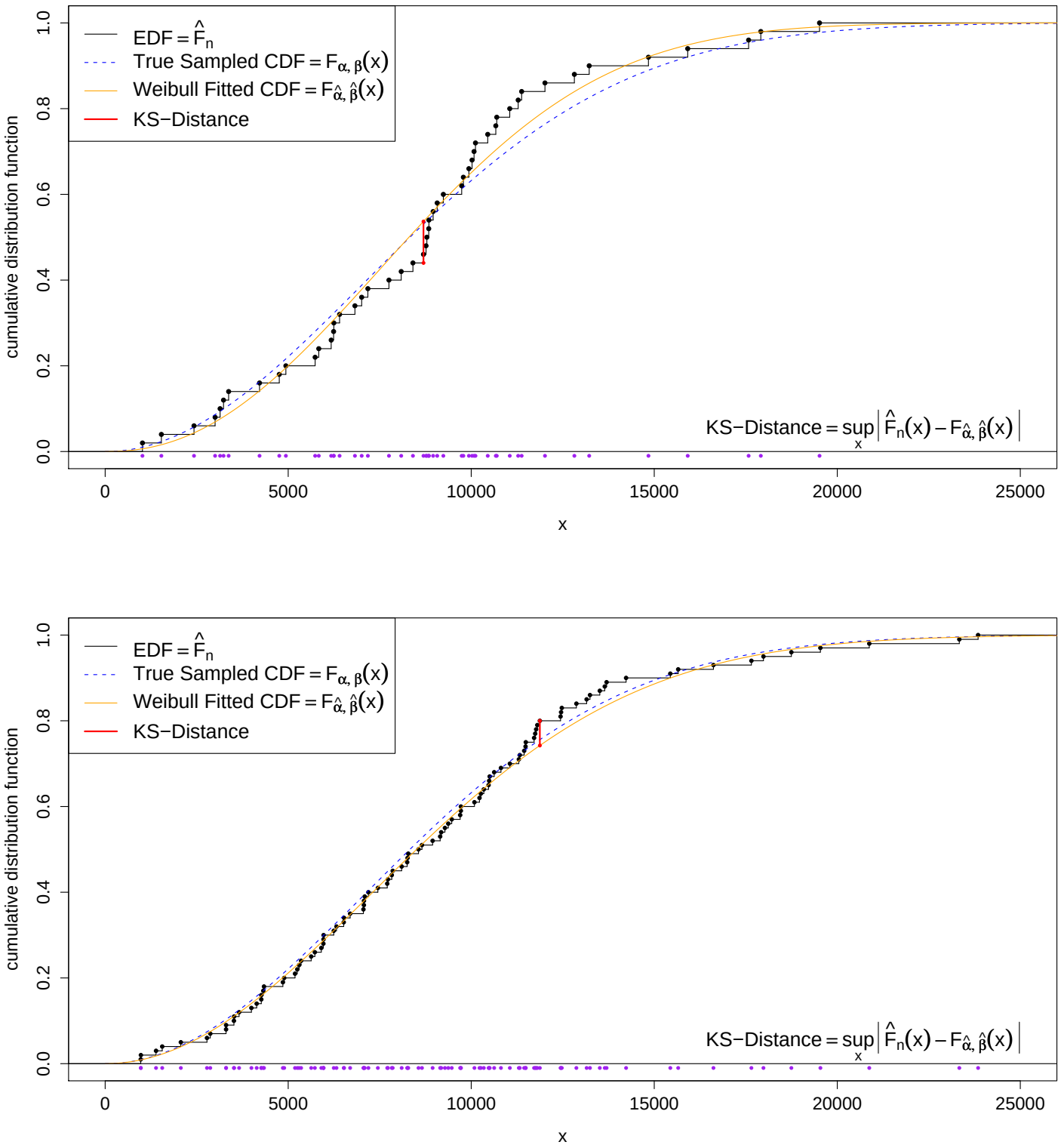


Figure 6: Illustration of Kolmogorov-Smirnov Distance for $n = 50$ and $n = 100$

$$F(x) = P(X \le x) = P(\log(X) \le \log(x)) = P(Y \le y) = 1 - \exp(-\exp((y-u)/b))$$

and thus

$$\begin{aligned} V_i = \hat{F}(X_i) &= 1 - \exp(-\exp((Y_i - \hat{u}(\mathbf{Y}))/\hat{b}(\mathbf{Y}))) \\ &= 1 - \exp(-\exp((u + bZ_i - \hat{u}(u + b\mathbf{Z}))/\hat{b}(u + b\mathbf{Z}))) \\ &= 1 - \exp(-\exp((u + bZ_i - u - b\hat{u}(\mathbf{Z}))/[b\,\hat{b}(\mathbf{Z}])) \\ &= 1 - \exp(-\exp((Z_i - \hat{u}(\mathbf{Z}))/\hat{b}(\mathbf{Z}))) \end{aligned}$$

and all dependence on the unknown parameters $u = \log(\alpha)$ and $b = 1/\beta$ has canceled out.

This opens up the possibility of using simulation to find good approximations to these null distributions for any n, especially in view of the previously reported timing results for computing the mle's $\hat{\alpha}$ and $\hat{\beta}$ of α and β. Just generate samples $\mathbf{X}^\star = (X_1^\star, \ldots, X_n^\star)$ from $\mathcal{W}(\alpha = 1, \beta = 1)$ (standard exponential distribution), compute the corresponding $\hat{\alpha}^\star = \hat{\alpha}(\mathbf{X}^\star)$ and $\hat{\beta}^\star = \hat{\beta}(\mathbf{X}^\star)$, then $V_i^\star = \hat{F}(X_i^\star) = F_{\hat{\alpha}^\star,\hat{\beta}^\star}(X_i^\star)$ (where $F_{\alpha,\beta}(x)$ is the cdf of $\mathcal{W}(\alpha,\beta)$) and from that the values $D^\star = D(\mathbf{X}^\star)$, $W^{2\star} = W^2(\mathbf{X}^\star)$ and $A^{2\star} = A^2(\mathbf{X}^\star)$. Calculating all three test of fit criteria makes sense since the main calculation effort is in getting the mle's $\hat{\alpha}^\star$ and $\hat{\beta}^\star$. Repeating this a large number of times, say $N_{\text{sim}} = 10000$, should give us a reasonably good approximation to the desired null distribution and from it one can determine appropriate p-values for any sample $X_1, \ldots, X_n$ for which one wishes to assess whether the Weibull distribution hypothesis is tenable or not. If $C(\mathbf{X})$ denotes the used test of fit criterion then the estimated p-value of this sample is simply the proportion of $C(\mathbf{X}^\star)$ that are $\ge C(\mathbf{X})$.

Prior to the ease of current computing Stephens (1986) provided tables for the $(1-\alpha)$-quantiles $q_{1-\alpha}$ of these null distributions. For the n-adjusted versions $A^2(1 + .2/\sqrt{n})$ and $W^2(1 + .2/\sqrt{n})$ these null distributions appear to be independent of n and $(1-\alpha)$-quantiles were given for $\alpha = .25, .10, .05, .025, .01$. Plotting $\log(\alpha/(1-\alpha))$ against $q_{1-\alpha}$ shows a mildly quadratic pattern which can be used to interpolate or extrapolate the appropriate p-value (observed significance level α) for any observed n-adjusted value $A^2(1 + .2/\sqrt{n})$ and $W^2(1 + .2/\sqrt{n})$, as is illustrated in Figure 7.

For $\sqrt{n}D$ the null distribution still depends on n (in spite of the normalizing factor $\sqrt{n}$) and $(1-\alpha)$-quantiles for $\alpha = .10, .05, .025, .01$ were tabulated for $n = 10, 20, 50, \infty$ by Stephens (1986). Here a double inter- and extrapolation scheme is needed, first by plotting these quantiles against $1/\sqrt{n}$, fitting quadratics in $1/\sqrt{n}$ and reading off the four interpolated quantile values for the needed n_0 (the sample size at issue) and as a second step perform the interpolation or extrapolation scheme as it was done previously, but using a cubic this time. This is illustrated in Figure 8.

Functions for computing these p-values (via interpolation from Stephens' tabled values) are given in the Weibull R work space provided at the class web site. They are `GOF.KS.test`, `GOF.CvM.test`, and `GOF.AD.test` for computing p-values for n-adjusted test criteria $\sqrt{n}D$, $W^2(1+.2/\sqrt{n})$, and $A^2(1+.2/\sqrt{n})$, respectively. These functions have an optional argument `graphic` where `graphic = T` causes the interpolation graphs shown in Figures 7 and 8 to be produced, otherwise only the p-values are given. The function `Weibull.GOF.test` does a Weibull goodness of fit test on any given sample, returning p-values for all three test criteria. You also find there the function `Weibull.mle` that was listed earlier, and several other functions not yet documented here.

One could easily reproduce and extend the tables given by Stephens (1986) so that extrapolations becomes less of an issue. For $n = 100$ it should take less than 1.5 minutes to simulate the null distributions based on $N_{\text{sim}} = 10,000$ and the previously given timing of 8.07 sec for $N_{\text{sim}} = 1000$.

9 Pivots

Based on the previous equivariance properties of $\hat{u}(\mathbf{Y})$ and $\hat{b}(\mathbf{Y})$ we have the following pivots, namely functions $W = \psi(\hat{u}(\mathbf{Y}), \hat{b}(\mathbf{Y}), \vartheta)$ of the estimates and an unknown parameter ϑ of interest such that W has a fixed and known distribution and the function ψ is strictly monotone in the unknown parameter ϑ, so that it is invertible with respect to ϑ.

Recall that for a Weibull random sample $\mathbf{X} = (X_1, \ldots, X_n)$ we have $Y_i = \log(X_i) \sim G((y-u)/b)$ with $b = 1/\beta$ and $u = \log(\alpha)$. Then $Z_i = (Y_i - u)/b \sim G(z) = 1 - \exp(-\exp(z))$, which is a known distribution (does not depend on unknown parameters). This is seen as follows:

$$P(Z_i \le z) = P((Y_i - u)/b \le z) = P(Y_i \le u + bz) = G(([u + bz] - u)/b) = G(z) \ .$$

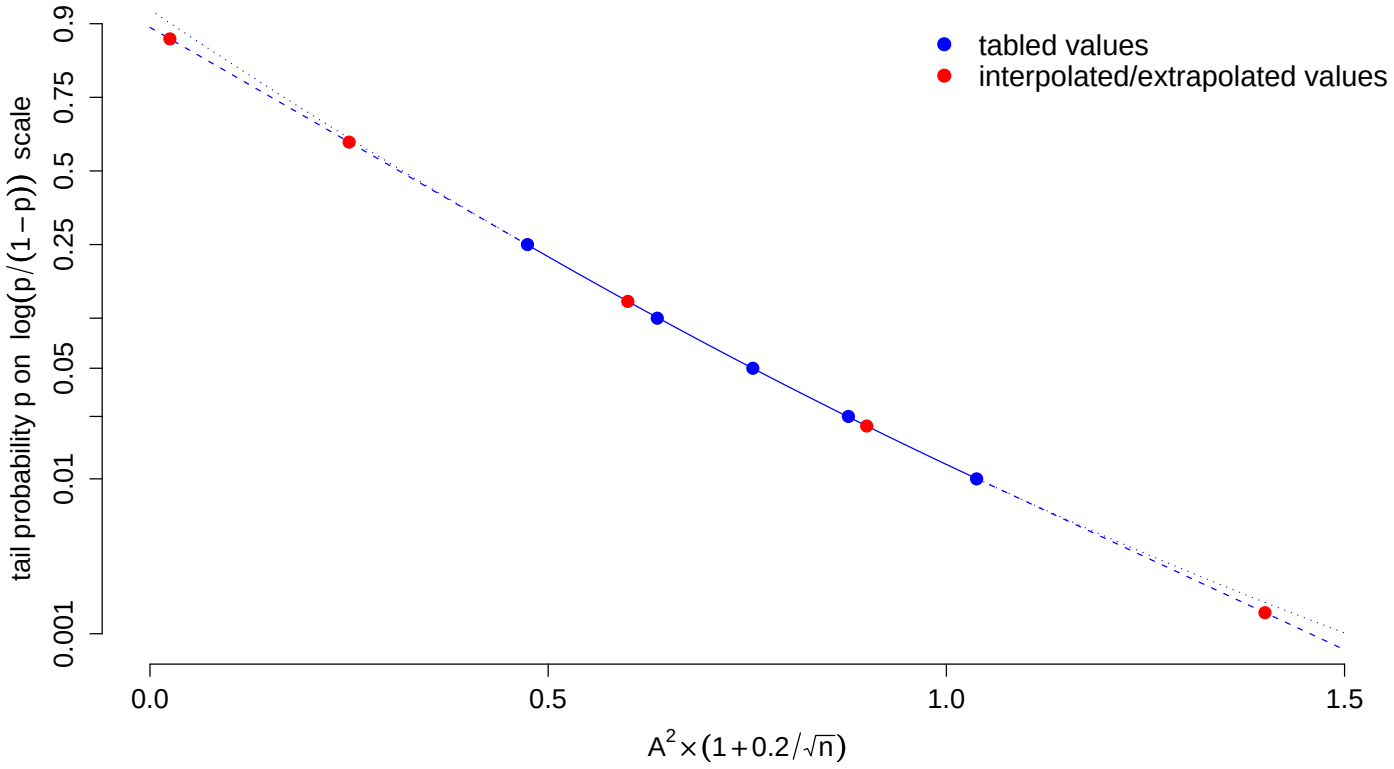
It is this known distribution of $\mathbf{Z} = (Z_1, \ldots, Z_n)$ that is instrumental in knowing the distribution of the four pivots that we discuss below. There we utilize the representation $Y_i = u + bZ_i$ or $\mathbf{Y} = u + b\mathbf{Z}$ in vector form.

9.1 Pivot for the Scale Parameter b

As natural pivot for the scale parameter $\vartheta = b$ we take

$$W_1 = \frac{\hat{b}(\mathbf{Y})}{b} = \frac{\hat{b}(u + b\mathbf{Z})}{b} = \frac{b\hat{b}(\mathbf{Z})}{b} = \hat{b}(\mathbf{Z}) \ .$$

The right side, being a function of $\mathbf{Z}$ alone, has a distribution that does not involve unknown parameters and $W_1 = \hat{b}(\mathbf{Y})/b$ is strictly monotone in b.



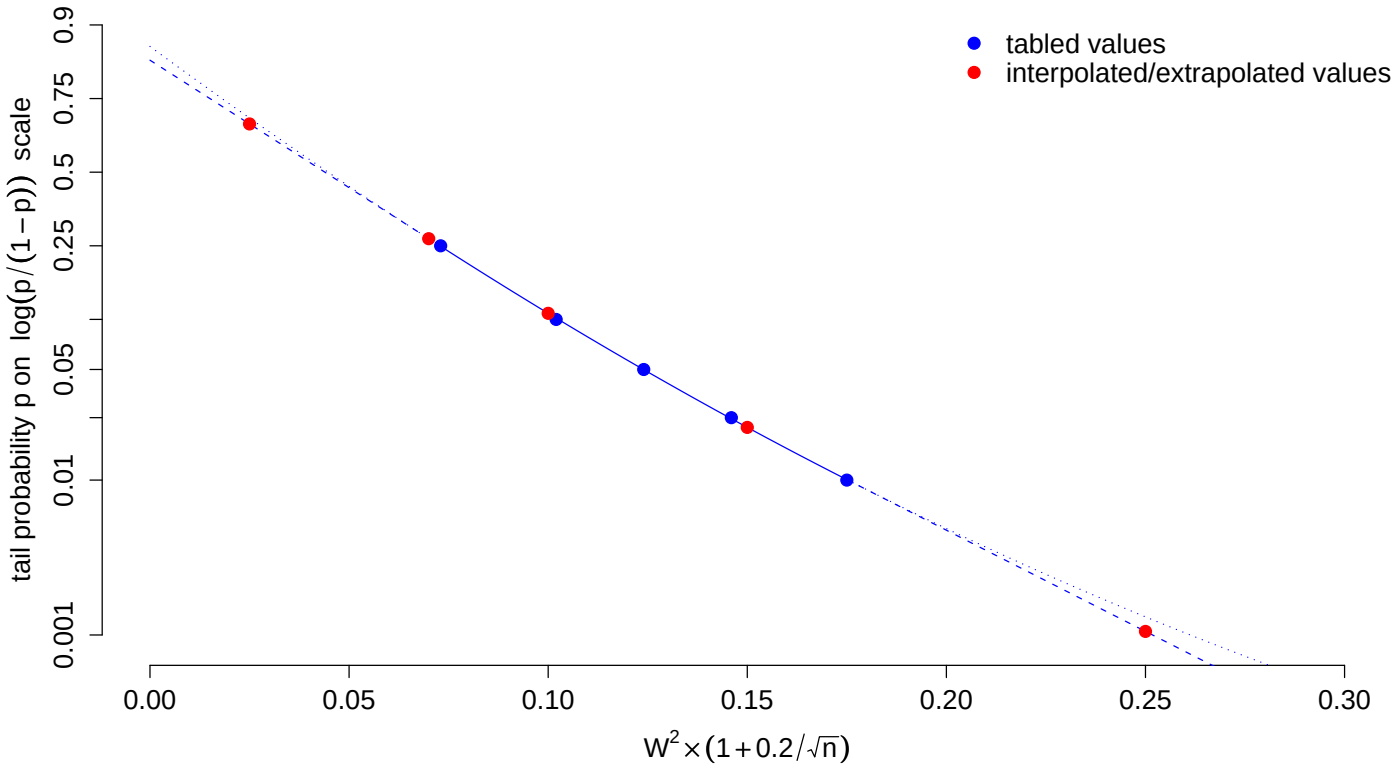
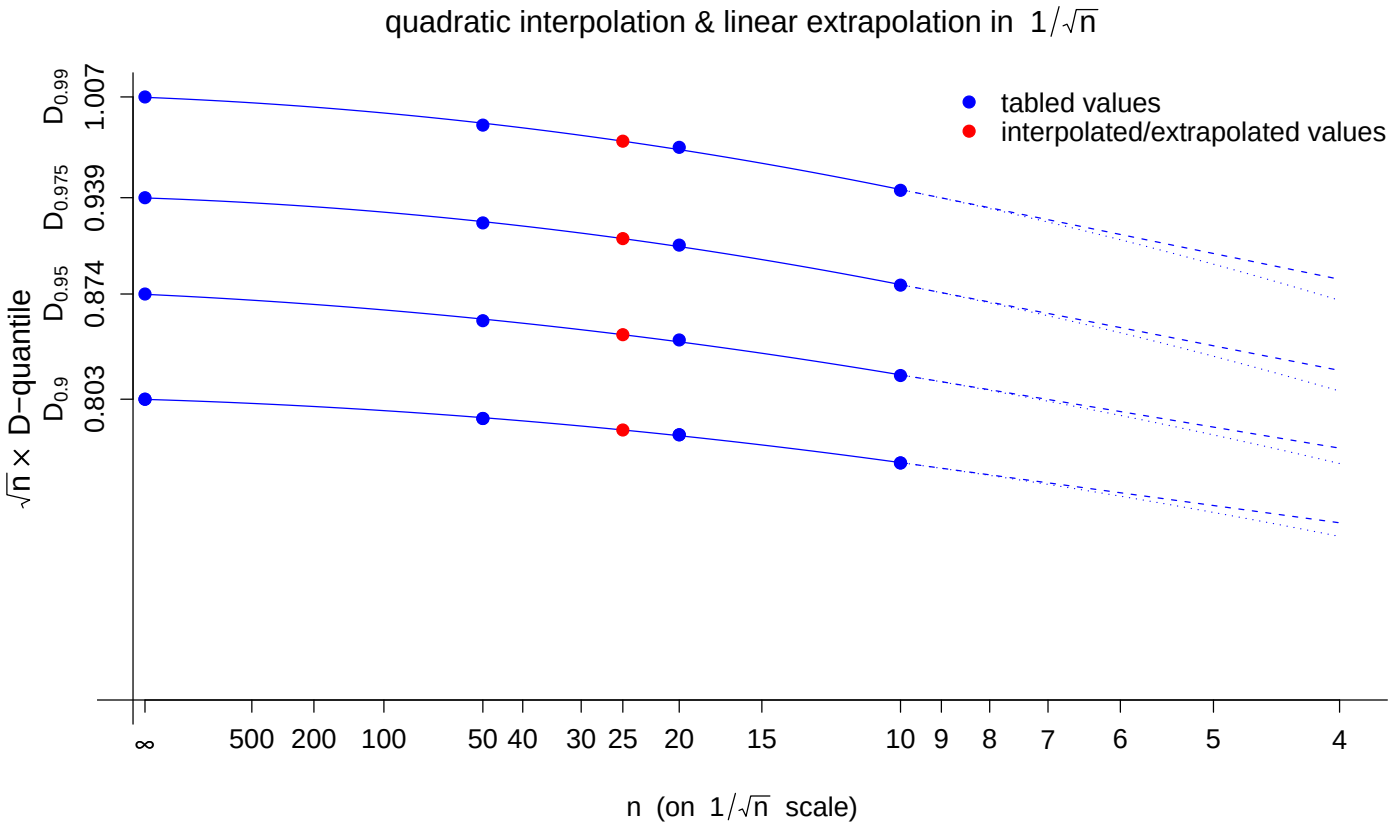


Figure 7: Interpolation & Extrapolation for $A^2(1 + .2/\sqrt{n})$ and $W^2(1 + .2/\sqrt{n})$



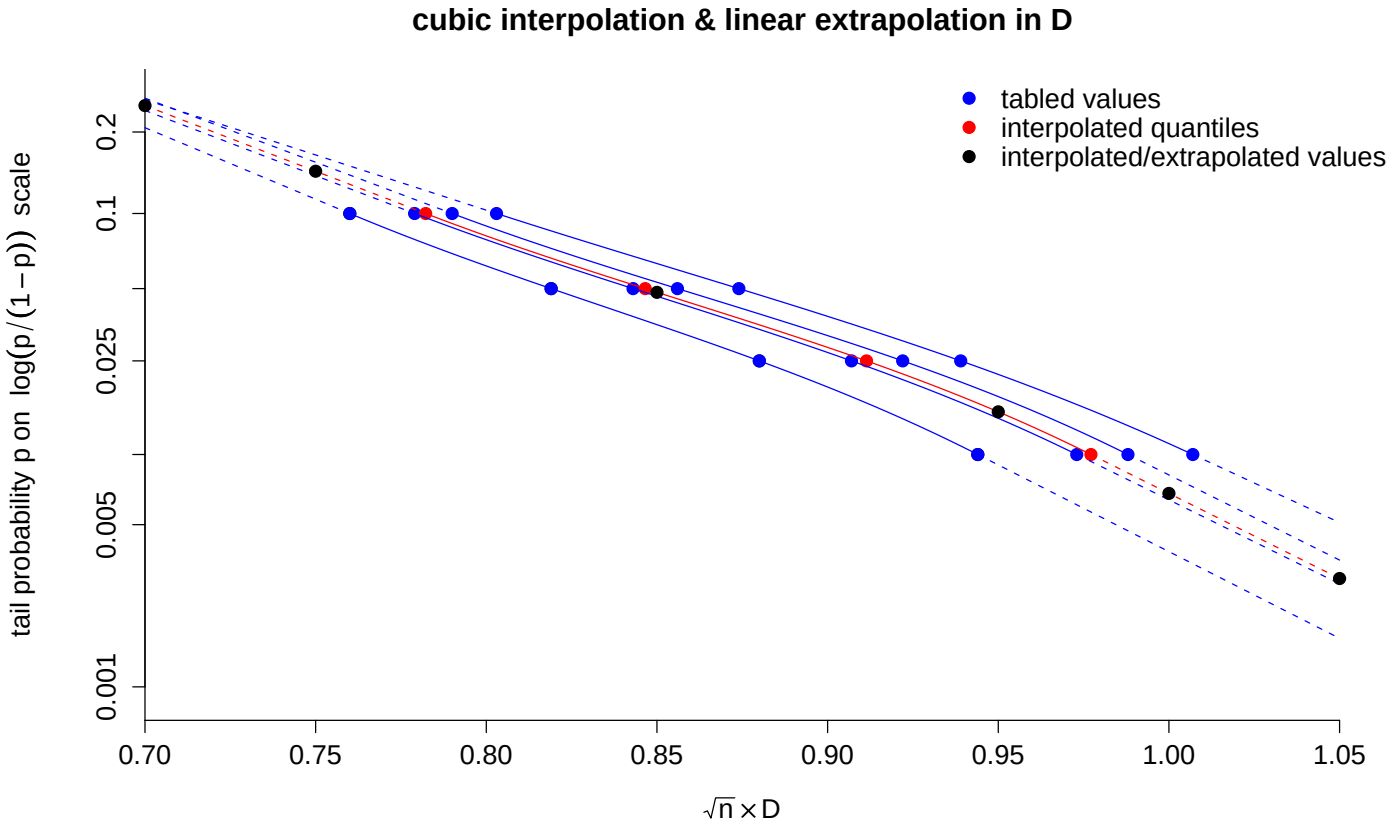


Figure 8: Interpolation & Extrapolation for $\sqrt{n} \times D$

How do we obtain the distribution of $\hat{b}(\mathbf{Z})$? An analytical approach does not seem possible. The approach followed here is that presented in Bain (1978), Bain and Engelhardt (1991) and originally in Thoman et al. (1969, 1970), who provided tables for this distribution (and for those of the other pivots discussed here) based on N_{sim} simulated values of $\hat{b}(\mathbf{Z})$ (and $\hat{u}(\mathbf{Z})$), where $N_{\text{sim}} = 20000$ for $n = 5$, $N_{\text{sim}} = 10000$ for $n = 6, 8, 10, 15, 20, 30, 40, 50, 75$, and $N_{\text{sim}} = 6000$ for $n = 100$.

In these simulations one simply generates samples $\mathbf{Z}^\star = (Z_1, \ldots, Z_n) \sim G(z)$ and finds $\hat{b}(\mathbf{Z}^\star)$ (and $\hat{u}(\mathbf{Z}^\star)$ for the other pivots discussed later) for each such sample $\mathbf{Z}^\star$. By simulating this process $N_{\text{sim}} = 10000$ times we obtain $\hat{b}(\mathbf{Z}_1^\star), \ldots, \hat{b}(\mathbf{Z}_{N_{\text{sim}}}^\star)$. The empirical distribution function of these simulated estimates $\hat{b}(\mathbf{Z}_i^\star)$, denoted by $\hat{H}_1(w)$, provides a fairly reasonable estimate of the sampling distribution $H_1(w)$ of $\hat{b}(\mathbf{Z})$ and thus also of the pivot distribution of $W_1 = \hat{b}(\mathbf{Y})/b$. From this simulated distribution we can estimate any γ-quantile of $H_1(w)$ to any practical accuracy, provided N_{sim} is sufficiently large. Values of γ closer to 0 or 1 require higher N_{sim}. For $.005 \leq \gamma \leq .995$ a simulation level of $N_{\text{sim}} = 10000$ should be quite adequate.

If we denote the γ-quantile of $H_1(w)$ by $\eta_1(\gamma)$, i.e.,

$$\gamma = H_1(\eta_1(\gamma)) = P(\hat{b}(\mathbf{Y})/b \leq \eta_1(\gamma)) = P(\hat{b}(\mathbf{Y})/\eta_1(\gamma) \leq b)$$

we see that $\hat{b}(\mathbf{Y})/\eta_1(\gamma)$ can be viewed as a $100\gamma\%$ lower bound to the unknown parameter b. We do not know $\eta_1(\gamma)$ but we can estimate it by the corresponding quantile $\hat{\eta}_1(\gamma)$ of the simulated distribution $\hat{H}_1(w)$ which serves as proxy for $H_1(w)$. We then use $\hat{b}(\mathbf{Y})/\hat{\eta}_1(\gamma)$ as an approximate $100\gamma\%$ lower bound to the unknown parameter b. For large N_{sim}, say $N_{\text{sim}} = 10000$, this approximation is practically quite adequate.

We note here that a $100\gamma\%$ lower bound can be viewed as a $100(1-\gamma)\%$ upper bound, because $1-\gamma$ is the chance of the lower bound falling on the wrong side of its target, namely above. To get $100\gamma\%$ upper bounds one simply constructs $100(1-\gamma)\%$ lower bounds by the above method. Similar comments apply to the pivots obtained below, where we only give one-sided bounds (lower or upper) in each case.

Based on the relationship $b = 1/\beta$ the respective $100\gamma\%$ approximate lower and upper confidence bounds for the Weibull shape parameter would be

$$\frac{\hat{\eta}_1(1-\gamma)}{\hat{b}(\mathbf{Y})} = \hat{\eta}_1(1-\gamma) \times \hat{\beta}(\mathbf{X}) \qquad \text{and} \qquad \frac{\hat{\eta}_1(\gamma)}{\hat{b}(\mathbf{Y})} = \hat{\eta}_1(\gamma) \times \hat{\beta}(\mathbf{X})$$

and an approximate $100\gamma\%$ confidence interval for β would be

$$\left[\hat{\eta}_1((1-\gamma)/2) \times \hat{\beta}(\mathbf{X}),\ \hat{\eta}_1((1+\gamma)/2) \times \hat{\beta}(\mathbf{X})\right]$$

since $(1+\gamma)/2 = 1 - (1-\gamma)/2$. Here $\mathbf{X} = (X_1, \ldots, X_n)$ is the untransformed Weibull sample.

9.2 Pivot for the Location Parameter u

For the location parameter $\vartheta = u$ we have the following pivot

$$W_2 = \frac{\hat{u}(\mathbf{Y}) - u}{\hat{b}(\mathbf{Y})} = \frac{\hat{u}(u + b\mathbf{Z}) - u}{\hat{b}(u + b\mathbf{Z})} = \frac{u + b\hat{u}(\mathbf{Z}) - u}{b\hat{b}(\mathbf{Z})} = \frac{\hat{u}(\mathbf{Z})}{\hat{b}(\mathbf{Z})} \, .$$

It has a distribution that does not depend on any unknown parameter, since it only depends on the known distribution of $\mathbf{Z}$. Furthermore W_2 is strictly decreasing in u. Thus W_2 is a pivot with respect to u. Denote this pivot distribution of W_2 by $H_2(w)$ and its γ-quantile by $\eta_2(\gamma)$. As before this pivot distribution and its quantiles can be approximated sufficiently well by simulating $\hat{u}(\mathbf{Z}^\star)/\hat{b}(\mathbf{Z}^\star)$ a sufficient number N_{sim} times and using the empirical cdf $\hat{H}_2(w)$ of the $\hat{u}(\mathbf{Z}_i^\star)/\hat{b}(\mathbf{Z}_i^\star)$ as proxy for $H_2(w)$.
As in the previous pivot case we can exploit this pivot distribution as follows

$$\gamma = H_2(\eta_2(\gamma)) = P\left(\frac{\hat{u}(\mathbf{Y}) - u}{\hat{b}(\mathbf{Y})} \leq \eta_2(\gamma)\right) = P(\hat{u}(\mathbf{Y}) - \hat{b}(\mathbf{Y})\eta_2(\gamma) \leq u)$$

and thus we can view $\hat{u}(\mathbf{Y}) - \hat{b}(\mathbf{Y})\eta_2(\gamma)$ as a $100\gamma\%$ lower bound for the unknown parameter u. Using the γ-quantile $\hat{\eta}_2(\gamma)$ obtained from the empirical cdf $\hat{H}_2(w)$ we then treat $\hat{u}(\mathbf{Y}) - \hat{b}(\mathbf{Y})\hat{\eta}_2(\gamma)$ as an approximate $100\gamma\%$ lower bound for the unknown parameter u.
Based on the relation $u = \log(\alpha)$ this translates into an approximate $100\gamma\%$ lower bound

$$\exp(\hat{u}(\mathbf{Y}) - \hat{b}(\mathbf{Y})\hat{\eta}_2(\gamma)) = \exp(\log(\hat{\alpha}(\mathbf{X})) - \hat{\eta}_2(\gamma)/\hat{\beta}(\mathbf{X})) = \hat{\alpha}(\mathbf{X})\exp(-\hat{\eta}_2(\gamma)/\hat{\beta}(\mathbf{X})) \qquad \text{for } \alpha.$$

Upper bounds and intervals are handled as in the previous situation for b or β.

9.3 Pivot for the p-quantile y_p

With respect to the p-quantile $\vartheta = y_p = u + b\log(-\log(1-p)) = u + bw_p$ of the Y distribution the natural pivot is

$$\begin{aligned} W_p = \frac{\hat{y}_p(\mathbf{Y}) - y_p}{\hat{b}(\mathbf{Y})} &= \frac{\hat{u}(\mathbf{Y}) + \hat{b}(\mathbf{Y})w_p - (u + bw_p)}{\hat{b}(\mathbf{Y})} = \frac{\hat{u}(u + b\mathbf{Z}) + \hat{b}(u + b\mathbf{Z})w_p - (u + bw_p)}{\hat{b}(u + b\mathbf{Z})} \\ &= \frac{u + b\hat{u}(\mathbf{Z}) + b\hat{b}(\mathbf{Z})w_p - (u + bw_p)}{b\hat{b}(\mathbf{Z})} = \frac{\hat{u}(\mathbf{Z}) + (\hat{b}(\mathbf{Z}) - 1)w_p}{\hat{b}(\mathbf{Z})} \, . \end{aligned}$$

Again its distribution only depends on the known distribution of $\mathbf{Z}$ and not on the unknown parameters u and b and the pivot W_p is a strictly decreasing function of y_p. Denote this pivot distribution

function by $H_p(w)$ and its γ-quantile by $\eta_p(\gamma)$. As before, this pivot distribution and its quantiles can be approximated sufficiently well by simulating $\left\{\hat{u}(\mathbf{Z}) + (\hat{b}(\mathbf{Z}) - 1)w_p\right\}/\hat{b}(\mathbf{Z})$ a sufficient number N_{sim} times. Denote the empirical cdf of such simulated values by $\hat{H}_p(w)$ and the corresponding γ-quantiles by $\hat{\eta}_p(\gamma)$.
As before we proceed with

$$\gamma = H_p(\eta_p(\gamma)) = P\left(\frac{\hat{y}_p(\mathbf{Y}) - y_p}{\hat{b}(\mathbf{Y})} \leq \eta_p(\gamma)\right) = P\left(\hat{y}_p(\mathbf{Y}) - \eta_p(\gamma)\hat{b}(\mathbf{Y}) \leq y_p\right)$$

and thus we can treat $\hat{y}_p(\mathbf{Y}) - \eta_p(\gamma)\hat{b}(\mathbf{Y})$ as a $100\gamma\%$ lower bound for y_p. Again we can treat $\hat{y}_p(\mathbf{Y}) - \hat{\eta}_p(\gamma)\hat{b}(\mathbf{Y})$ as an approximate $100\gamma\%$ lower bound for y_p.
Since

$$\hat{y}_p(\mathbf{Y}) - \eta_p(\gamma)\hat{b}(\mathbf{Y}) = \hat{u}(\mathbf{Y}) + w_p\hat{b}(\mathbf{Y}) - \eta_p(\gamma)\hat{b}(\mathbf{Y}) = \hat{u}(\mathbf{Y}) - k_p(\gamma)\hat{b}(\mathbf{Y})$$

with $k_p(\gamma) = \eta_p(\gamma) - w_p$, we could have obtained the same lower bound by the following argument that does not use a direct pivot, namely

$$\begin{aligned}
\gamma = P(\hat{u}(\mathbf{Y}) - k_p(\gamma)\hat{b}(\mathbf{Y}) \leq y_p) &= P(\hat{u}(\mathbf{Y}) - k_p(\gamma)\hat{b}(\mathbf{Y}) \leq u + bw_p) \\
&= P(\hat{u}(\mathbf{Y}) - u - k_p(\gamma)\hat{b}(\mathbf{Y}) \leq bw_p) \\
&= P\left(\frac{\hat{u}(\mathbf{Y}) - u}{b} - k_p(\gamma)\frac{\hat{b}(\mathbf{Y})}{b} \leq w_p\right) \\
&= P(\hat{u}(\mathbf{Z}) - k_p(\gamma)\hat{b}(\mathbf{Z}) \leq w_p) = P\left(\frac{\hat{u}(\mathbf{Z}) - w_p}{\hat{b}(\mathbf{Z})} \leq k_p(\gamma)\right)
\end{aligned}$$

and we see that $k_p(\gamma)$ can be taken as the γ-quantile of the distribution of $(\hat{u}(\mathbf{Z}) - w_p)/\hat{b}(Z)$.
This distribution can be estimated by the empirical cdf of N_{sim} simulated values $(\hat{u}(\mathbf{Z}_i^\star) - w_p)/\hat{b}(\mathbf{Z}_i^\star)$, $i = 1, \ldots, N_{\text{sim}}$ and its γ-quantile $\hat{k}_p(\gamma)$ serves as a good approximation to $k_p(\gamma)$.
It is easily seen that this produces the same quantile lower bound as before. However, in this approach one sees one further detail, namely that $h(p) = -k_p(\gamma)$ is strictly increasing in p[1], since w_p is strictly increasing in p.

[1]Suppose $p_1 < p_2$ and $h(p_1) \geq h(p_2)$ with $\gamma = P(\hat{u}(\mathbf{Z}) + h(p_1)\hat{b}(\mathbf{Z}) \leq w_{p_1})$ and $\gamma = P(\hat{u}(\mathbf{Z}) + h(p_2)\hat{b}(\mathbf{Z}) \leq w_{p_2}) = P(\hat{u}(\mathbf{Z}) + h(p_1)\hat{b}(\mathbf{Z}) \leq w_{p_1} + (w_{p_2} - w_{p_1}) + (h(p_1) - h(p_2))\hat{b}(\mathbf{Z})) \geq P(\hat{u}(\mathbf{Z}) + h(p_1)\hat{b}(\mathbf{Z}) \leq w_{p_1} + (w_{p_2} - w_{p_1})) > \gamma$ (i.e., $\gamma > \gamma$, a contradiction) since $P(w_{p_1} < \hat{u}(\mathbf{Z}) + h(p_1)\hat{b}(\mathbf{Z}) \leq w_{p_1} + (w_{p_2} - w_{p_1})) > 0$. A thorough argument would show that $\hat{b}(\mathbf{z})$ and thus $\hat{u}(\mathbf{z})$ are continuous functions of $\mathbf{z} = (z_1, \ldots, z_n)$ and since there is positive probability in any neighborhood of any $\mathbf{z} \in R$ there is positive probability in any neighborhood of $(\hat{u}(\mathbf{z}), \hat{b}(\mathbf{z}))$.

Of course it makes intuitive sense that quantile lower bounds should be increasing in p since its target p-quantiles are increasing in p. This strictly increasing property allows us to immediately construct upper confidence bounds for left tail probabilities as is shown in the next section.
Since $x_p = \exp(y_p)$ is the corresponding p-quantile of the Weibull distribution we can view

$$\exp\left(\hat{y}_p(\mathbf{Y}) - \hat{\eta}_p(\gamma)\hat{b}(\mathbf{Y})\right) = \hat{\alpha}(\mathbf{X})\exp\left((w_p - \hat{\eta}_p(\gamma))/\hat{\beta}(\mathbf{X})\right) = \hat{\alpha}(\mathbf{X})\exp\left(-\hat{k}_p(\gamma)/\hat{\beta}(\mathbf{X})\right)$$

as an approximate $100\gamma\%$ lower bound for $x_p = \exp(u + bw_p) = \alpha(-\log(1-p))^{1/\beta}$.
Since α is the $(1 - \exp(-1))$-quantile of the Weibull distribution, lower bounds for it can be seen as a special case of quantile lower bounds. Indeed, this particular quantile lower bound coincides with the one given previously.

9.4 Upper Confidence Bounds for the Tail Probability $p(y) = P(Y \leq y)$

As far as an appropriate pivot for $p(y) = P(Y \leq y)$ is concerned, the situation here is not as straightforward as in the previous three cases. Clearly

$$\hat{p}(y) = G\left(\frac{y - \hat{u}(\mathbf{Y})}{\hat{b}(\mathbf{Y})}\right) \quad \text{is the natural estimate (mle) of} \quad p(y) = P(Y \leq y) = G\left(\frac{y-u}{b}\right)$$

and one easily sees that the distribution function H of this estimate depends on u and b only through $p(y)$, namely

$$\hat{p}(y) = G\left(\frac{y - \hat{u}(\mathbf{Y})}{\hat{b}(\mathbf{Y})}\right) = G\left(\frac{(y-u)/b - (\hat{u}(\mathbf{Y}) - u)/b}{\hat{b}(\mathbf{Y})/b}\right) = G\left(\frac{G^{-1}(p(y)) - \hat{u}(\mathbf{Z})}{\hat{b}(\mathbf{Z})}\right) \sim H_{p(y)}\ .$$

Thus by the probability integral transform it follows that

$$W_{p(y)} = H_{p(y)}\left(\hat{p}(y)\right) \sim U(0,1)$$

i.e., $W_{p(y)}$ is a true pivot, contrary to what is stated in Bain (1978) and Bain and Engelhardt (1991). Rather than using this pivot we will go a more direct route as was indicated by the strictly increasing property of $h(p) = h_\gamma(p)$ in the previous section. Denote by $h^{-1}(\cdot)$ the inverse function to $h(\cdot)$. We then have

$$\gamma = P(\hat{u}(\mathbf{Y}) + h(p)\hat{b}(\mathbf{Y}) \leq y_p) = P(h(p) \leq (y_p - \hat{u}(\mathbf{Y}))/\hat{b}(\mathbf{Y})) = P\left(p \leq h^{-1}\left((y_p - \hat{u}(\mathbf{Y}))/\hat{b}(\mathbf{Y})\right)\right)\ ,$$

for any $p \in (0,1)$. If we parameterize such p via $p(y) = P(Y \leq y) = G((y-u)/b)$ we have $y_{p(y)} = y$ and thus also

$$\gamma = P\left(p(y) \leq h^{-1}\left((y - \hat{u}(\mathbf{Y}))/\hat{b}(\mathbf{Y})\right)\right)$$

for any $y \in R$ and $u \in R$ and $b > 0$. Hence $\hat{p}_U(y) = h^{-1}\left((y - \hat{u}(\mathbf{Y}))/\hat{b}(\mathbf{Y})\right)$ can be viewed as $100\gamma\%$ upper confidence bound for $p(y)$ for any given threshold y.

The only remaining issue is the computation of such bounds. Does it require the inversion of h and the concomitant calculations of many $h(p) = -k(p)$ for the iterative convergence of such an inversion? It turns out that there is a direct path just as we had it in the previous three confidence bound situations.

Note that $h^{-1}(x)$ solves $-k_p = x$ for p. We claim that $h^{-1}(x)$ is the γ-quantile of the $G(\hat{u}(\mathbf{Z})+x\hat{b}(\mathbf{Z}))$ distribution which we can simulate by calculating as before $\hat{u}(\mathbf{Z})$ and $\hat{b}(\mathbf{Z})$ a large number N_{sim} times. The above claim concerning $h^{-1}(x)$ is seen as follows. If for any $x = h(p)$ we have

$$\begin{aligned} P(G(\hat{u}(\mathbf{Z}) + x\hat{b}(\mathbf{Z})) \le h^{-1}(x)) &= P(G(\hat{u}(\mathbf{Z}) + h(p)\hat{b}(\mathbf{Z})) \le p) \\ &= P(\hat{u}(\mathbf{Z}) + h(p)\hat{b}(\mathbf{Z}) \le w_p) \\ &= P(\hat{u}(\mathbf{Z}) - k_\gamma(p)\hat{b}(\mathbf{Z}) \le w_p) = \gamma \ , \end{aligned}$$

as seen in the previous section. Thus $h^{-1}(x)$ is the γ-quantile of the $G(\hat{u}(\mathbf{Z}) + x\hat{b}(\mathbf{Z}))$ distribution. If we observe $\mathbf{Y} = \mathbf{y}$ and obtain $\hat{u}(\mathbf{y})$ and $\hat{b}(\mathbf{y})$ as our maximum likelihood estimates for u and b we get our $100\gamma\%$ upper bound for $p(y) = G((y-u)/b)$ as follows: For the fixed value of $x = (y - \hat{u}(\mathbf{y}))/\hat{b}(\mathbf{y}) = G^{-1}(\hat{p}(y))$ simulate the $G(\hat{u}(\mathbf{Z}) + x\hat{b}(\mathbf{Z}))$ distribution (with sufficiently high N_{sim}) and calculate the γ-quantile of this distribution as the desired approximate $100\gamma\%$ upper bound for $p(y) = P(Y \le y) = G((y-u)/b)$.

10 Tabulation of Confidence Quantiles $\eta(\gamma)$

For the pivots for b, u and y_p it is possible to carry out simulations once and for all for a desired set of confidence levels γ, sample sizes n and choices of p, and tabulate the required confidence quantiles $\hat{\eta}_1(\gamma)$, $\hat{\eta}_2(\gamma)$, and $\hat{\eta}_p(\gamma)$. This has essentially been done (with $\sqrt{n}$ scaling modifications) and such tables are given in Bain (1978), Bain and Engelhardt (1991) and Thoman et al. (1969,1970). Similar tables for bounds on $p(y)$ are not quite possible since the appropriate bounds depend on the observed value of $\hat{p}(y)$, which varies from sample to sample. Instead Bain (1978), Bain and Engelhardt (1991) and Thoman et al. (1970) tabulate confidence bounds for $p(y)$ for a reasonably fine grid of values for $\hat{p}(y)$, which can then serve for interpolation purposes with the actually observed value of $\hat{p}(y)$.

It should be quite clear that all this requires extensive tabulation. The use of these tables is not easy. Table 4 in Bain (1978) does not have a consistent format and using these tables would require delving deeply into the text for each new use, unless one does this kind of calculation all the time. In fact, in the second edition, Bain and Engelhardt (1991), Table 4 has been greatly reduced to just cover the confidence factors dealing with the location parameter u, and it now

leaves out the confidence factors for general p-quantiles. For the p-quantiles one is referred to the same interpolation scheme that is needed when getting confidence bounds for $p(y)$, using Table 7 in Bain and Engelhardt (1991). The example that they present (page 248) would have benefitted by showing some intermediate steps in the interpolation process. They point out that the resulting confidence bound for x_p is slightly different (14.03) from that obtained using the confidence quantiles of the original Table 4, namely 13.92. They attribute the difference to round-off errors or other discrepancies. Among the latter one may consider that possibly different simulations were involved.

Further, note that some entries in the tables given in Bain (1978) seem to have typos. Presumably they were transcribed by hand from computer output, just as the book (and its second edition) itself is typed and not typeset. We give just give a few examples. In Bain (1978) Table 4A, p.235, bottom row, the second entry from the right should be 3.625 instead of 3.262. This discrepancy shows up clearly when plotting the row values against $\log(p/(1-p))$, see a similar plot for a later example. In Table 3A, p.222, row 3 column 5 shows a double minus sign (still present in the 1991 second edition). In comparing the values of these tables with our own simulation of pivot distribution quantiles, just to validate our simulation for $n = 40$, we encountered an apparent error in Table 4A, p. 235 with last column entry of 4.826. Plotting $\log(p/(1-p))$ against the corresponding row value (γ-quantiles) one clearly sees a change in pattern, see the top plot in Figure 9. We suspect that the whole last column was calculated for $p = .96$ instead of the indicated $p = .98$. The bottom plot shows our simulated values for these quantiles as solid dots with the previous points (circles) superimposed.

The agreement is good for the first 8 points. Our simulated γ-quantile was 5.725 (corresponding to the 4.826 above) and it fits quite smoothly into the pattern of the previous 8 points. Given that this was the only case chosen for comparison it leaves some concern in fully trusting these tables. However, this example also shows that the great majority of tabled values are valid.

11 The R Function `WeibullPivots`

Rather than using these tables we will resort to direct simulations ourselves since computing speed has advanced sufficiently over what was common prior to 1978. Furthermore, computing availability has changed dramatically since then. It may be possible to further increase computing speed by putting the loop over N_{sim} calculations of mle's into compiled form rather than looping within R for each simulation iteration. For example, using `qbeta` in vectorized form reduced the computing time to almost 1/3 of the time compared to looping within R itself over the elements in the argument vector of `qbeta`.

However, such an increase in speed would require writing C-code (or Fortran code) and linking that in compiled form to R. Such extensions of R are possible, see chapter 5 System and foreign language interfaces in the Writing R Extensions manual available under the toolbar Help in R.

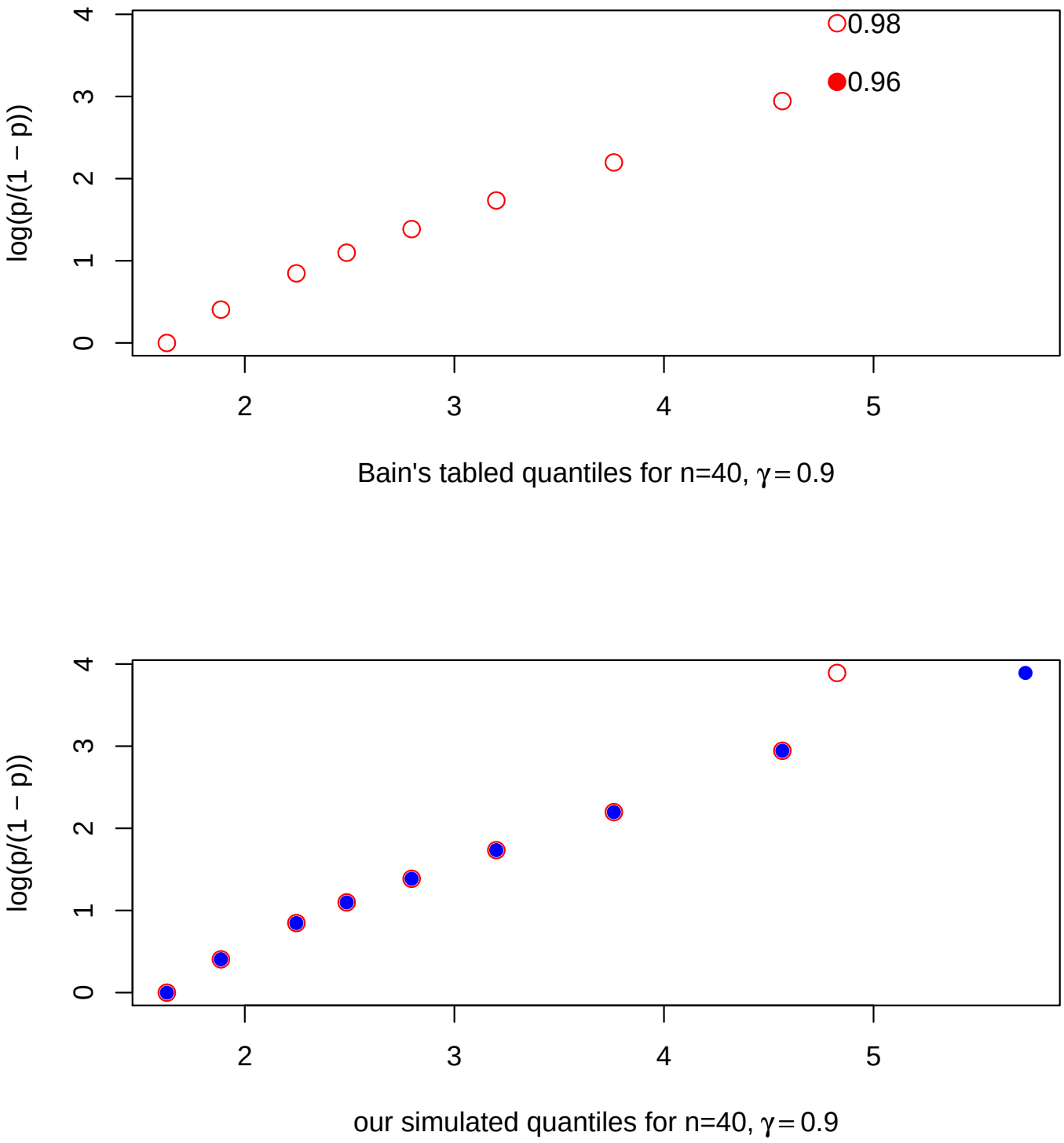


Figure 9: Abnormal Behavior of Tabulated Confidence Quantiles

For the R function `WeibullPivots` (available within the R work space for Weibull Distribution Applications on the class web site) the call

$$\texttt{system.time}(\texttt{WeibullPivots}(\texttt{Nsim} = 10000, \texttt{n} = 10, \texttt{r} = 10, \texttt{graphics} = \texttt{F}))$$

gave an elapsed time of 59.76 seconds. Here the default sample size $n = 10$ was used and $r = 10$ (also default) indicates that the 10 lowest sample values are given and used, i.e., in this case the full sample. Also, an internally generated Weibull data set was used, since the default in the call to `WeibullPivots` is `weib.sample=NULL`. For sample sizes $n = 100$ with $r = 100$ and $n = 1000$ with $r = 1000$ the corresponding calls resulted in elapsed times of 78.22 and 269.32 seconds, respectively. These three computing times suggest strong linear behavior in n as is illustrated in Figure 10. The intercept 57.35 and slope of .2119 given here are fairly consistent with the intercept .005886 and slope of 2.001×10^{-5} given in Figure 4. The latter give the calculation time of a single set of mle's while in the former case we calculate $N_{\text{sim}} = 10000$ such mle's, i.e., the previous slope and intercept for a single mle calculation need to be scaled up by the factor 10000.

For all the previously discussed confidence bounds, be they upper or lower bounds for their respective targets, all that is needed is the set of $(\hat{u}(\mathbf{z}_i), \hat{b}(\mathbf{z}_i))$ for $i = 1, \ldots, N_{\text{sim}}$. Thus we can construct confidence bounds and intervals for u and b, for y_p for any collection of p values, and for $p(y)$ and $1 - p(y)$ for any collection of threshold values y and we can do this for any set of confidence levels that make sense for the simulated distributions, i.e., we don't have to run the simulations over and over for each target parameter, confidence level, p or y, unless one wants independent simulations for some reason.

Proper use of this function only requires understanding the calling arguments, purpose, and output of this function, and the time to run the simulations. The time for running the simulation should easily beat the time spent in dealing with tabulated confidence quantiles in order to get desired confidence bounds, especially since `WeibullPivots` does such calculations all at once for a broad spectrum of y_p and $p(y)$ and several confidence levels without greatly impacting the computing time. Furthermore, `WeibullPivots` does all this not only for full samples but also for type II censored samples, for which appropriate confidence factors are available only sparsely in tables.

We will now explain the calling sequence of `WeibullPivots` and its output. The calling sequence with all arguments given with their default values is as follows:

```
WeibullPivots(weib.sample=NULL,alpha=10000,beta=1.5,n=10,r=10,
                                                    Nsim=1000,threshold=NULL,graphics=T)
```

Here `Nsim` $= N_{\text{sim}}$ has default value 1000 which is appropriate when trying to get a feel for the function for any particular data set. The sample size is input as `n` $= n$ and `r` $= r$ indicates the number of smallest sample values available for analysis. When $r < n$ we are dealing with a type II censored data set where observation stops as soon as the smallest r lifetimes have been observed. We need $r > 1$ and at least two distinct observations among $X_{(1)}, \ldots, X_{(r)}$ in order to estimate

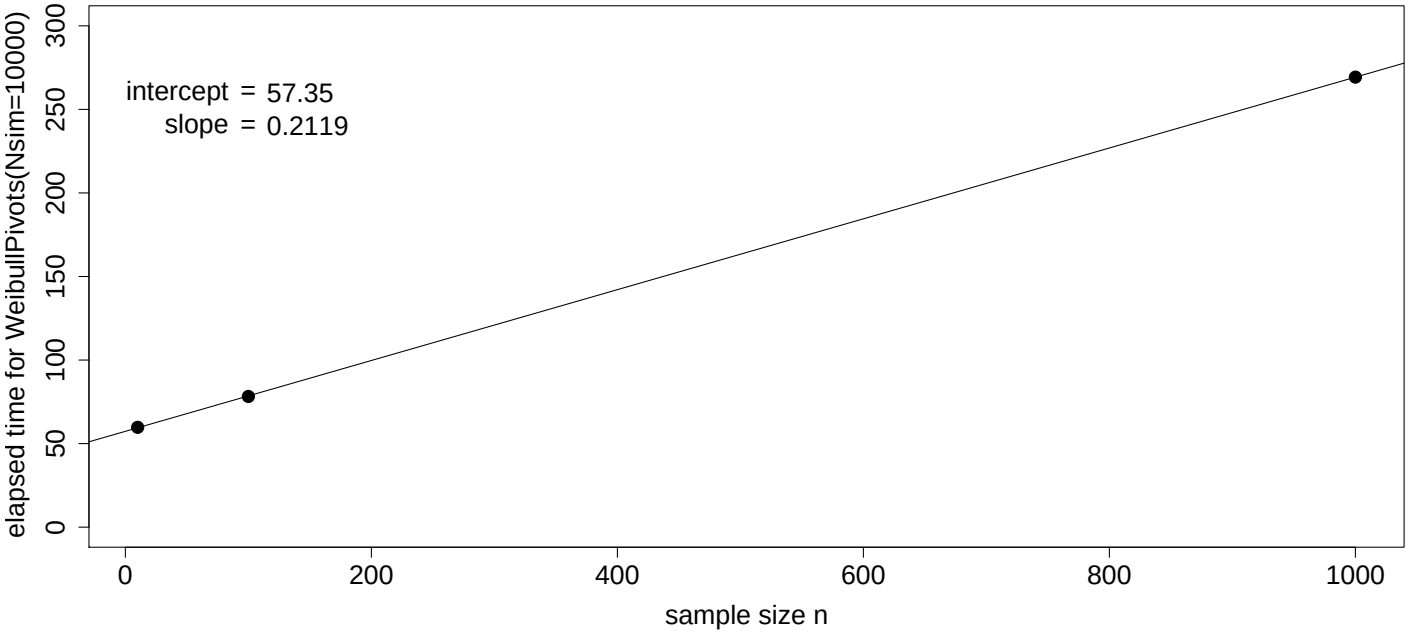


Figure 10: Timings for `WeibullPivots` for Various n

any spread in the data. The available sample values $X_1, \ldots, X_r$ (not necessarily ordered) are given as vector input to `weib.sample`. When `weib.sample=NULL` (the default), an internal data set is generated as input sample from from $\mathcal{W}(\alpha, \beta)$ with $\alpha =$ `alpha` $= 10000$ (default) and $\beta =$ `beta` $= 1.5$ (default), either by using the full sample $X_1, \ldots, X_n$ or a type II censored sample $X_1, \ldots, X_r$ when $r < n$ is specified. The input `thresh` (= `NULL` by default) is a vector of thresholds y for which we desire upper confidence bounds for $p(y)$. The input `graphics` (default `T`) indicates whether graphical output is desired.

Confidence levels γ are set internally as .005, .01, .025, .05, .10, .02, .8, .9, .95, .975, .99, .995 and these levels indicate the coverage probability for the individual one-sided bounds. A .025 lower bound is reported as a .975 upper bound, and a pair of .975 lower and upper bounds constitute a 95% confidence interval. The values of p for which confidence bounds or intervals for x_p are provided are also set internally as .001, .005, .01, .025, .05, .1, (.1), .9, .95, .975, .99, .995, .999.

The output from `WeibullPivots` is a list with components:

`$alpha.hat`, `$alpha.hat`, `$alpha.beta.bounds`,`$p.quantile.estimates`, `$p.quantile.bounds`, `$Tail.Probability.Estimates`, and `$Tail.Probability.Bounds`. The structure and meaning of these components will become clear from the example output given below.

```
$alpha.hat
(Intercept)
     8976.2

$beta.hat
[1] 1.95

$alpha.beta.bounds
      alpha.L alpha.U beta.L beta.U
99.5%  5094.6   16705  0.777   3.22
99%    5453.9   15228  0.855   3.05
97.5%  5948.6   13676  0.956   2.82
95%    6443.9   12608  1.070   2.64
90%    7024.6   11600  1.210   2.42
80%    7711.2   10606  1.390   2.18


$p.quantile.estimates
0.001-quantile 0.005-quantile  0.01-quantile 0.025-quantile  0.05-quantile
         259.9          593.8          848.3         1362.5         1957.0
  0.1-quantile   0.2-quantile   0.3-quantile   0.4-quantile   0.5-quantile
        2830.8         4159.4         5290.4         6360.5         7438.1
  0.6-quantile   0.7-quantile   0.8-quantile   0.9-quantile  0.95-quantile
        8582.7         9872.7        11457.2        13767.2        15756.3
0.975-quantile  0.99-quantile 0.995-quantile 0.999-quantile
       17531.0        19643.5        21107.9        24183.6


$p.quantile.bounds
                     99.5%     99%   97.5%     95%     90%     80%
0.001-quantile.L       1.1     2.6     6.0    12.9    28.2    60.1
0.001-quantile.U    1245.7  1094.9   886.7   729.4   561.4   403.1
0.005-quantile.L       8.6    16.9    31.9    57.4   106.7   190.8
0.005-quantile.U    2066.9  1854.9  1575.1  1359.2  1100.6   845.5
0.01-quantile.L       20.1    36.7    65.4   110.1   186.9   315.3
0.01-quantile.U     2579.8  2361.5  2021.5  1773.9  1478.4  1165.8
0.025-quantile.L      62.8   103.5   169.7   259.3   398.1   611.0
0.025-quantile.U    3498.8  3206.6  2827.2  2532.5  2176.9  1783.5
```

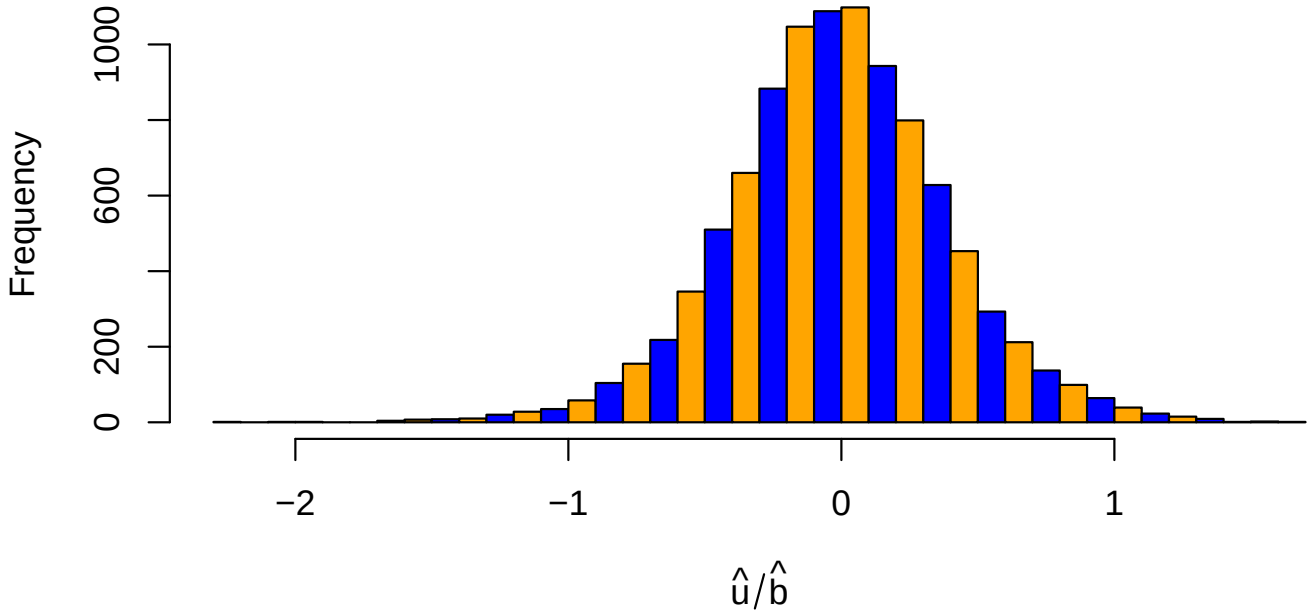
```
0.05-quantile.L     159.2   229.2   352.6   497.5   700.0  1011.9
0.05-quantile.U    4415.7  4081.3  3673.7  3329.5  2930.0  2477.2
0.1-quantile.L      398.3   506.3   717.4   962.2  1249.5  1679.1
0.1-quantile.U     5584.6  5261.9  4811.6  4435.7  3990.8  3474.1
0.2-quantile.L     1012.6  1160.2  1518.8  1882.9  2287.1  2833.2
0.2-quantile.U     7417.1  6978.2  6492.8  6031.2  5543.2  4946.9
0.3-quantile.L     1725.4  1945.2  2383.9  2820.0  3305.0  3929.0
0.3-quantile.U     8919.8  8460.0  7939.8  7384.1  6865.0  6211.4
0.4-quantile.L     2548.0  2848.2  3345.2  3806.6  4353.9  5008.2
0.4-quantile.U    10616.3 10130.4  9380.3  8778.2  8139.3  7421.2
0.5-quantile.L     3502.4  3881.1  4415.1  4873.3  5443.0  6107.3
0.5-quantile.U    12809.0 11919.1 10992.9 10226.8  9485.1  8703.4
0.6-quantile.L     4694.0  5022.6  5573.8  6052.8  6624.4  7300.4
0.6-quantile.U    15626.1 14350.6 12941.3 11974.8 11041.1 10106.2
0.7-quantile.L     6017.1  6399.0  6876.6  7345.8  7938.2  8628.0
0.7-quantile.U    19271.6 17679.9 15545.8 14181.1 12958.0 11784.2
0.8-quantile.L     7601.3  7971.0  8465.4  8933.5  9504.0 10244.2
0.8-quantile.U    24765.2 22445.0 19286.0 17236.0 15605.6 13952.2
0.9-quantile.L     9674.7 10033.7 10538.6 11031.1 11653.0 12460.3
0.9-quantile.U    35233.4 31065.3 26037.4 22670.5 19835.3 17417.5
0.95-quantile.L   11203.6 11584.6 12145.2 12660.2 13365.5 14311.2
0.95-quantile.U   46832.9 40053.3 32863.1 27904.7 23903.9 20703.0
0.975-quantile.L  12434.7 12833.5 13449.7 14030.5 14781.8 15909.1
0.975-quantile.U  59783.1 49209.9 39397.8 33118.7 27938.4 23773.7
0.99-quantile.L   13732.6 14207.7 14876.0 15530.0 16431.7 17783.1
0.99-quantile.U   76425.0 61385.4 48625.4 40067.3 33233.8 27729.8
0.995-quantile.L  14580.4 15115.4 15810.0 16530.6 17551.8 19081.0
0.995-quantile.U  89690.9 71480.4 55033.4 45187.1 36918.7 30505.0
0.999-quantile.L  16377.7 16885.9 17642.5 18557.1 19792.4 21744.7
0.999-quantile.U 121177.7 95515.7 71256.5 56445.5 45328.1 36739.2


$Tail.Probability.Estimates
 p(6000)  p(7000)  p(8000)  p(9000) p(10000) p(11000) p(12000) p(13000)
 0.36612  0.45977  0.55018  0.63402  0.70900  0.77385  0.82821  0.87242
p(14000) p(15000)
 0.90737  0.93424
```

```
$Tail.Probability.Bounds
               99.5%     99%   97.5%     95%     90%     80%
p(6000).L    0.12173 0.13911 0.16954 0.19782 0.23300 0.28311
p(6000).U    0.69856 0.67056 0.63572 0.59592 0.54776 0.49023
p(7000).L    0.17411 0.20130 0.23647 0.26985 0.31017 0.36523
p(7000).U    0.76280 0.73981 0.70837 0.67426 0.62988 0.57670
p(8000).L    0.23898 0.26838 0.30397 0.34488 0.38942 0.44487
p(8000).U    0.82187 0.80141 0.77310 0.74260 0.70414 0.65435
p(9000).L    0.30561 0.33149 0.37276 0.41748 0.46448 0.52203
p(9000).U    0.87042 0.85462 0.82993 0.80361 0.77045 0.72545
p(10000).L   0.36871 0.39257 0.44219 0.48549 0.53589 0.59276
p(10000).U   0.91227 0.89889 0.87805 0.85624 0.82667 0.78641
p(11000).L   0.41612 0.45097 0.50030 0.54631 0.59749 0.65671
p(11000).U   0.94491 0.93318 0.91728 0.89891 0.87425 0.83973
p(12000).L   0.46351 0.50388 0.55531 0.60133 0.65374 0.71215
p(12000).U   0.96669 0.95936 0.94650 0.93210 0.91231 0.88377
p(13000).L   0.50876 0.54776 0.60262 0.65055 0.70218 0.76148
p(13000).U   0.98278 0.97742 0.96794 0.95756 0.94149 0.91745
p(14000).L   0.54668 0.58696 0.64619 0.69359 0.74451 0.80178
p(14000).U   0.99201 0.98837 0.98205 0.97459 0.96267 0.94321
p(15000).L   0.58089 0.62534 0.68389 0.73194 0.78068 0.83590
p(15000).U   0.99653 0.99449 0.99092 0.98596 0.97764 0.96268
```

The above output was produced with

$$\texttt{WeibullPivots}(\texttt{threshold} = \texttt{seq}(6000, 15000, 1000), \texttt{Nsim} = 10000, \texttt{graphics} = \texttt{T}).$$

Since we entered `graphics=T` as argument we also got two pieces of graphical output. The first gives the two intrinsic pivot distributions of $\hat{u}/\hat{b}$ and $\hat{b}$ in Figure 11. The second gives a Weibull plot of the generated sample with a variety of information and with several types of confidence bounds, see Figure 12. The legend in the upper left gives the mle's of α, β (agreeing with the output above), and the mean $\mu = \alpha\Gamma(1 + 1/\beta)$ together with 95% confidence intervals, based on respective normal approximation theory for the mle's. The legend in the lower right explains the red fitted line (representing the mle fit) and the various point-wise confidence bound curves, giving 95% confidence intervals (blue dashed curves) for p-quantiles x_p for any p on the ordinate and 95% confidence intervals (green dot-dashed line) for $p(y)$ for any y on the abscissa. Both of these interval types use normal approximations from large sample mle theory. Unfortunately these two types of bounds are not dual to each other, i.e., don't coincide or to say it differently, one is not the inverse to the other.



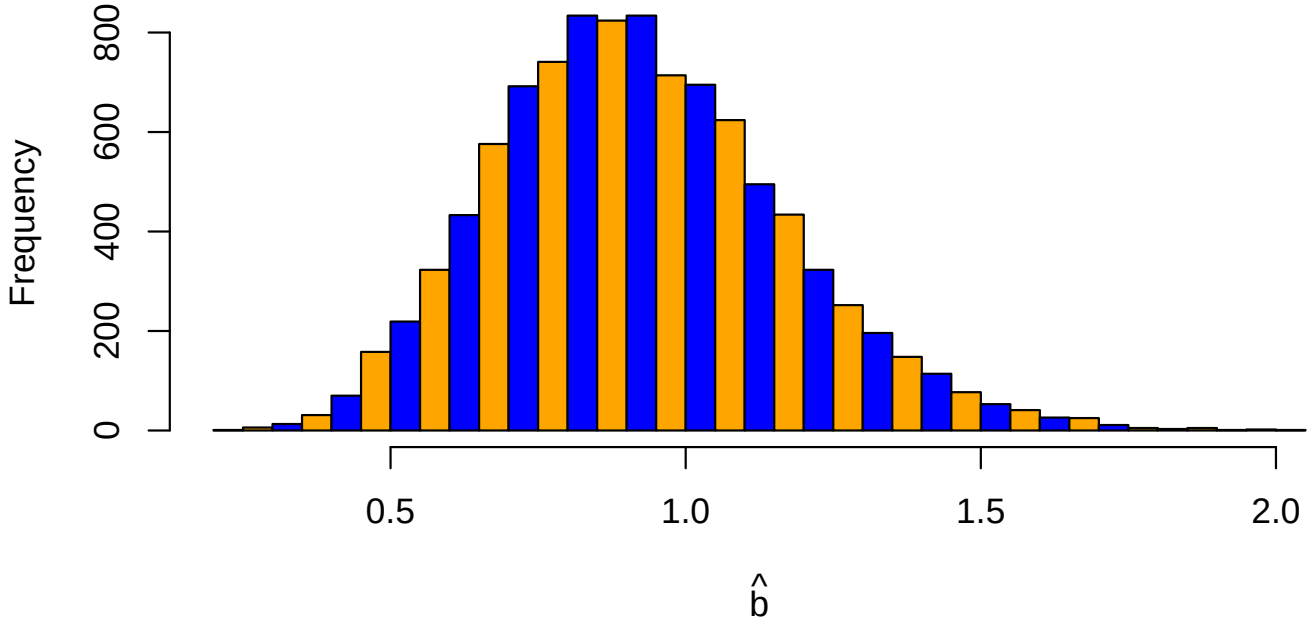


Figure 11: Pivot Distributions of $\hat{u}/\hat{b}$ and $\hat{b}$

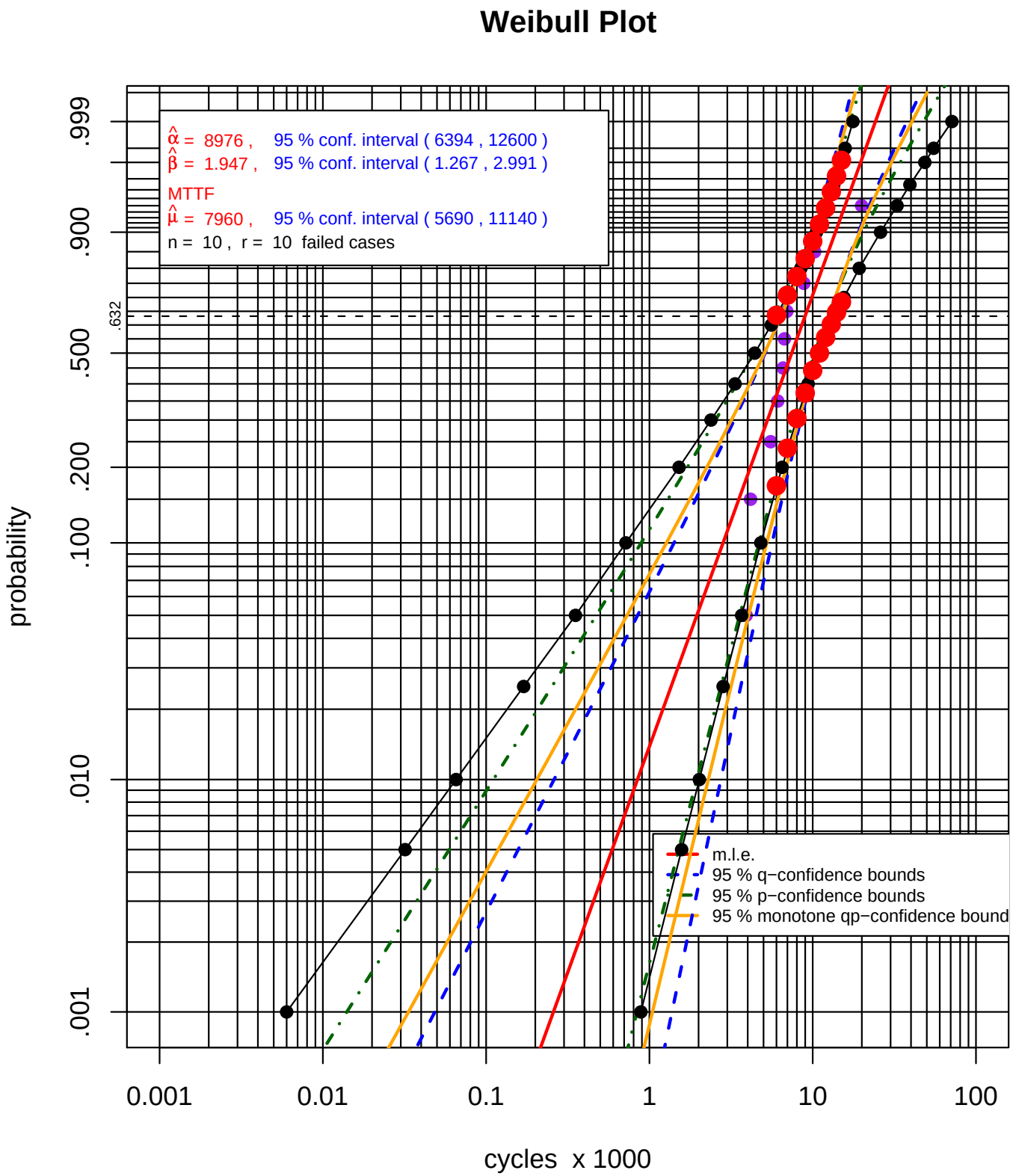


Figure 12: Weibull Plot Corresponding to Previous Output

A third type of bound is presented in the orange curve which simultaneously provides 95% confidence intervals for x_p and $p(x)$, depending on the direction in which the curves are used. We either read sideways from p and down from the curve (at that p level) to get upper and lower bounds for x_p, or we read vertically up from an abscissa value x to read off upper and lower bounds for $p(x)$ on the ordinate axis as we go from the respective curves at that x value to the left. These latter bounds are also based on normal mle approximation theory and the approximation will naturally suffer for small sample sizes. However, the principle behind these bounds is a unifying one in that the same curve is used for quantile and tail probability bounds. If instead of using the approximating normal distribution one uses the parametric bootstrap approach (simulating samples from an estimated Weibull distribution) the unifying principle reduces to the pivot simulation approach, i.e., is basically exact except for the simulation aspect $N_{\text{sim}} < \infty$.

The curves representing the latter (pivots with simulated distributions) are the solid black lines connecting the solid black dots which represent the x_p 95% confidence intervals (using the 97.5% lower and upper bounds to x_p given in our output example above. Also seen on these curves are solid red dots that correspond to the abscissa values $x = 6000, (1000), 15000$ and viewed vertically they represent 95% confidence intervals for $p(x)$. This illustrates that the same curves are used.

Figure 13 represents an extreme case where we have a sample of size $n = 2$ and here another aspect becomes apparent. Both of the first two types of bounds (blue and green) are no longer monotone in p or x respectively. This is in the nature of an imperfect normal approximation for these two approaches. Thus we could not (at least not generally) have taken either to take the role of serving both purposes, i.e., as providing bounds for x_p and $p(x)$ simultaneously. However, the orange curve is still monotone and still serves that dual purpose, although its coverage probability properties are bound to be affected badly by the small sample size $n = 2$. The pivot based curves are also strictly monotone and they have exact coverage probability, subject to the $N_{\text{sim}} < \infty$ limitation.

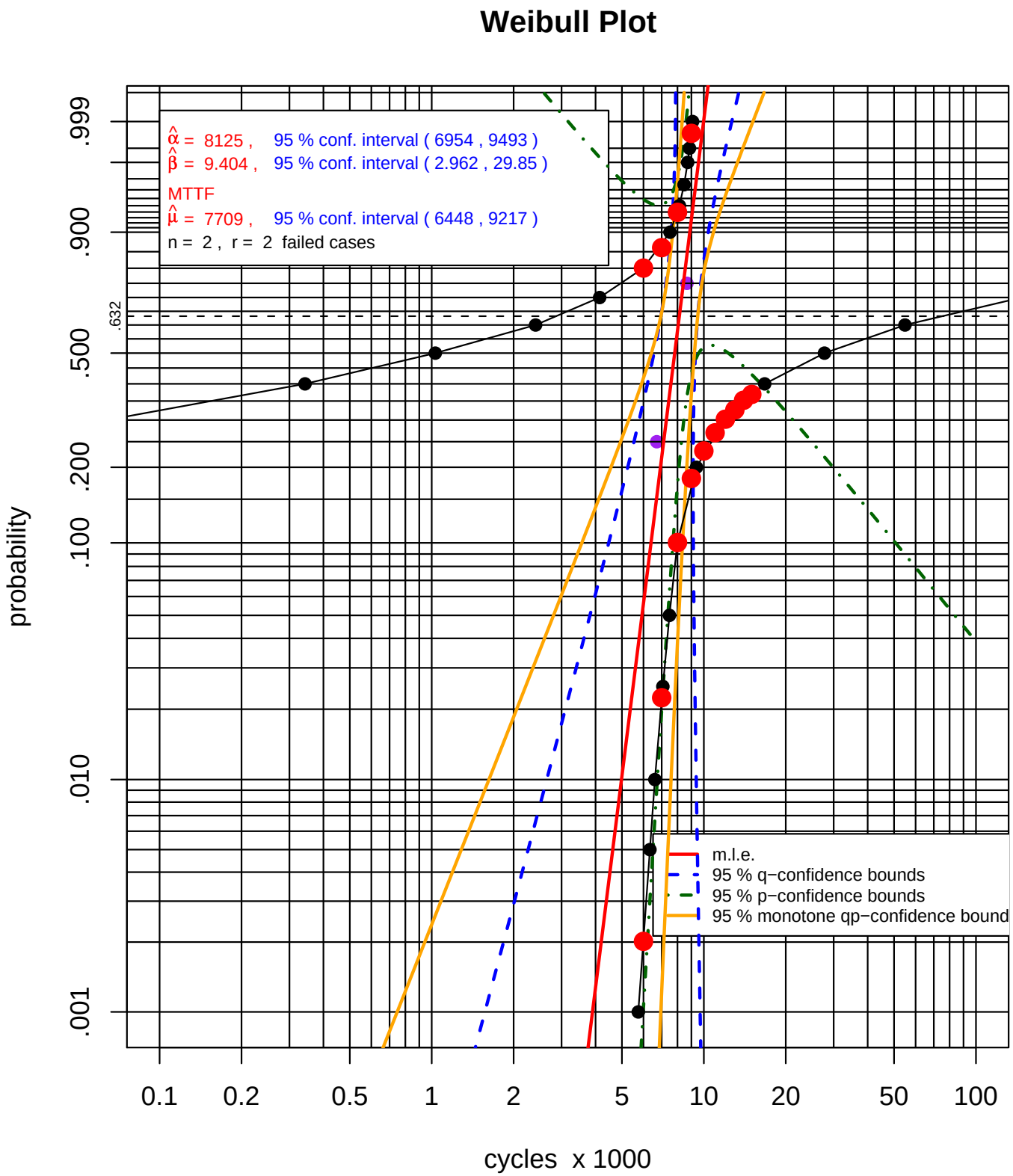


Figure 13: Weibull Plot for Weibull Sample of Size $n = 2$

12 Regression

Here we extend the results of the previous location/scale model for log-transformed Weibull samples to the more general regression model where the location parameter u for $Y_i = \log(X_i)$ can vary with i, specifically it varies as a linear combination of known covariates $c_{i,1}, \ldots, c_{i,k}$ for the i^{th} observation as follows:

$$u_i = \zeta_1 c_{i,1} + \ldots + \zeta_k c_{i,k} \ , \quad i = 1, \ldots, n \ ,$$

while the scale parameter b stays constant. Thus we have the following model for $Y_1, \ldots, Y_n$

$$Y_i = u_i + bE_i = \zeta_1 c_{i,1} + \ldots + \zeta_k c_{i,k} + bZ_i \ , \quad i = 1, \ldots, n \ ,$$

with independent $Z_i \sim G(z) = 1 - \exp(-\exp(z))$, $i = 1, \ldots, n$, and unknown parameters $b > 0$ and $\zeta_1, \ldots, \zeta_k \in R$.

Two concrete examples of this general linear model will be discussed in detail later on. The first is the simple linear regression model and the other is the k-sample model, which exemplifies ANOVA situations.

It can be shown (Scholz, 1996) that the mle's $\hat{\zeta}' = (\hat{\zeta}_1, \ldots, \hat{\zeta}_k)$ and $\hat{b}$ of ζ and b exist and are unique provided the covariate matrix C, consisting of the rows $c_i' = (c_{i,1}, \ldots, c_{i,k})$, $i = 1, \ldots, n$, has full rank k and $n > k$. It is customary that the first column of C is a vector of n 1's. Alternatively, one can also only specify the remaining $k - 1$ columns and implicitly invoke the default option in `survreg` that augments those columns with such a 1-vector. These two usages are illustrated in the function `WeibullReg` which is given on the next page.

It is very instructive to run this function as part of the following call:

{\tt system.time(for(i in 1:1000)WeibullReg())} ,

i.e., we execute the function `WeibullReg` a thousand times in close succession.

The rapidly varying plots give a good visual image of the sampling uncertainty and the resulting sampling variation of the fitted lines. The fixed line represents the true line with respect to which the Weibull data are generated by simulation. Of course, the log-Weibull data are plotted because of its more transparent relationship to the true line. It is instructive to see the variability of the data clouds around the true line, but also the basic stability of the overall cloud pattern as a whole.

On my laptop the elapsed time for this call is about 15 seconds, and this includes the plotting time. When the plotting commands are commented out the elapsed time reduces to about 9 seconds. This promises reasonable behavior with respect to the computing times that can be anticipated for the confidence bounds to be discussed below.

```
WeibullReg <- function (n=50,x=NULL,alpha=10000,beta=1.5,slope=.05)
{
# We can either input our own covariate vector x of length n
# or such a vector is generated for us (default).
#
if(is.null(x)) x <- (1:n-(n+1)/2)
uvec <- log(alpha)+slope*x
b <- 1/beta
# Create the Weibull data
time <- exp(uvec+b*log(-log(1-runif(n))))
# Creating good vertical plotting limits
m <- min(uvec)+b*log(-log(1-1/(3*n+1)))
M <- max(uvec)+b*log(-log(1/(3*n+1)))

plot(x,log(time),ylim=c(m,M))
dat <- data.frame(time,x)
out <- survreg(Surv(time)~x,data=dat,dist="weibull")
# The last two lines would give the same result as the next three lines
# after removing the # signs.
# x0 <- rep(1,n)
# dat <- data.frame(time,x0,x)
# survreg(formula = Surv(time) ~ x0 + x - 1, data = dat, dist = "weibull")
# Here we created the vector x0 of ones explicitly and removed the implicit
# vector of ones by the -1 in ~ x0+x-1.
# Note also, that we did not use a status vector (of ones) in the creation
# of dat, since survreg will use status = 1 for each observation, i.e,
# treat the given time as a failure time as default.
abline(log(alpha),slope) #true line
# estimated line
abline(out$coef[1],out$coef[2],col="blue",lty=2)
# Here out has several components, of which only
# out$coef and out$scale of of interest to us.
# The estimate out$scale is the mle of b=1/beta
# and out$coef is a vector that gives the mle's
# of intercept and the various regression coefficients.
out
}
```

12.1 Equivariance Properties

From the existence and uniqueness of the mle's we can again deduce the following equivariance properties for the mle's, namely for $\mathbf{r} = C\mathbf{a} + \sigma\mathbf{z}$ we have

$$\hat{\zeta}(\mathbf{r}) = \mathbf{a} + \sigma\hat{\zeta}(\mathbf{z}) \quad \text{and} \quad \hat{b}(\mathbf{r}) = \sigma\hat{b}(\mathbf{z}) \; .$$

The proof follows the familiar line used in the location/scale case. With $r_i = c_i'\mathbf{a} + \sigma z_i$ we have

$$\begin{aligned} \sup_{b,\zeta}\left\{\prod_{i=1}^{n}\frac{1}{b}\,g\left(\frac{r_i - c_i'\zeta}{b}\right)\right\} &= \frac{1}{\sigma^n}\sup_{b,\zeta}\left\{\prod_{i=1}^{n}\frac{1}{b/\sigma}\,g\left(\frac{z_i - c_i'(\zeta-\mathbf{a})/\sigma}{b/\sigma}\right)\right\} \\ \text{using} \quad \tilde{\zeta} = (\zeta-\mathbf{a})/\sigma \quad \text{and} \quad \tilde{b} = b/\sigma \qquad &= \frac{1}{\sigma^n}\sup_{\tilde{b},\tilde{\zeta}}\left\{\prod_{i=1}^{n}\frac{1}{\tilde{b}}\,g\left(\frac{z_i - c_i'\tilde{\zeta}}{\tilde{b}}\right)\right\} \\ &= \frac{1}{\sigma^n}\prod_{i=1}^{n}\frac{1}{\hat{b}(\mathbf{z})}\,g\left(\frac{z_i - c_i'\hat{\zeta}(\mathbf{z})}{\hat{b}(\mathbf{z})}\right) \end{aligned}$$

On the other hand

$$\begin{aligned} \sup_{b,\zeta}\left\{\prod_{i=1}^{n}\frac{1}{b}\,g\left(\frac{r_i - c_i'\zeta}{b}\right)\right\} &= \prod_{i=1}^{n}\frac{1}{\hat{b}(\mathbf{r})}\,g\left(\frac{r_i - c_i'\hat{\zeta}(\mathbf{r})}{\hat{b}(\mathbf{r})}\right) \\ &= \frac{1}{\sigma^n}\prod_{i=1}^{n}\frac{1}{\hat{b}(\mathbf{r})/\sigma}\,g\left(\frac{z_i - c_i'(\hat{\zeta}(\mathbf{r})-\mathbf{a})/\sigma}{\hat{b}(\mathbf{r})/\sigma}\right) \end{aligned}$$

and by the uniqueness of the mle's the equivariance claim is an immediate consequence.

12.2 Pivots and Confidence Bounds

From these equivariance properties it follows that $(\hat{\zeta}-\zeta)/\hat{b}$ and $\hat{b}/b$ have distributions that do not depend on any unknown parameters, i.e., b and ζ. The log-transformed Weibull data have the following regression structure $\boldsymbol{Y} = C\zeta + b\boldsymbol{Z}$, where $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ consists of independent and identically distributed components with known cdf $G(z) = 1 - \exp(-\exp(z))$. From the equivariance property we have that

$$\hat{\zeta}(\boldsymbol{Y}) = \zeta + b\,\hat{\zeta}(\boldsymbol{Z}) \qquad \text{and} \qquad \hat{b}(\boldsymbol{Y}) = b\,\hat{b}(\boldsymbol{Z}) \; .$$

Thus

$$\frac{\hat{\zeta}(\boldsymbol{Y})-\zeta}{\hat{b}(\boldsymbol{Y})} = \frac{b\,\hat{\zeta}(\boldsymbol{Z})}{b\,\hat{b}(\boldsymbol{Z})} = \frac{\hat{\zeta}(\boldsymbol{Z})}{\hat{b}(\boldsymbol{Z})} \quad \text{and} \quad \frac{\hat{b}(\boldsymbol{Y})}{b} = \frac{b\,\hat{b}(\boldsymbol{Z})}{b} = \hat{b}(\boldsymbol{Z}) \; ,$$

which have a distribution free of any unknown parameters. This distribution can be approximated to any desired degree via simulation, just as in the location scale case, except that we will need to incorporate the known covariate matrix C in the call to `survreg` in order to get the N_{sim} simulated parameter vectors $(\hat{\zeta}(\boldsymbol{Z}_1^\star), \hat{b}_1(\boldsymbol{Z}_1^\star)), \ldots, (\hat{\zeta}(\boldsymbol{Z}_{N_{\text{sim}}}^\star), \hat{b}(\boldsymbol{Z}_{N_{\text{sim}}}^\star))$ and thus the empirical distribution of $(\hat{\zeta}(\boldsymbol{Z}_1^\star)/\hat{b}_1(\boldsymbol{Z}_1^\star), \hat{b}_1(\boldsymbol{Z}_1^\star)), \ldots, (\hat{\zeta}(\boldsymbol{Z}_{N_{\text{sim}}}^\star)/\hat{b}(\boldsymbol{Z}_{N_{\text{sim}}}^\star), \hat{b}(\boldsymbol{Z}_{N_{\text{sim}}}^\star))$.

For any target covariate vector $\boldsymbol{c}_0' = (c_{0,1}, \ldots, c_{0,k})$ the distribution of $(\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Y}) - \boldsymbol{c}_0'\zeta)/\hat{b}(\boldsymbol{Y})$ is free of unknown parameters since

$$\frac{\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Y}) - \boldsymbol{c}_0'\zeta}{\hat{b}(\boldsymbol{Y})} = \frac{\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Z})}{\hat{b}(\boldsymbol{Z})}$$

and we can use the simulated values $(\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Z}_i^\star))/\hat{b}(\boldsymbol{Z}_i^\star)$, $i = 1, \ldots, N_{\text{sim}}$, to approximate this parameter free distribution. If $\hat{\eta}_2(\gamma, \boldsymbol{c}_0)$ denotes the γ-quantile of this simulated distribution then we can view $\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Y}) - \hat{\eta}_2(\gamma, \boldsymbol{c}_0)\hat{b}(\boldsymbol{Y})$ as an approximate $100\gamma\%$ lower bound for $\boldsymbol{c}_0'\zeta$, the log of the characteristic life at the covariate vector $\boldsymbol{c}_0$. This can be demonstrated as in the location/scale case for the location parameter u.

Similarly, if $\hat{\eta}_1(\gamma)$ is the γ-quantile of the simulated $\hat{b}(\boldsymbol{Z}_i^\star)$, $i = 1, \ldots, N_{\text{sim}}$, then we can view $\hat{b}(\boldsymbol{Y})/\hat{\eta}_1(\gamma)$ as approximate $100\gamma\%$ lower bound for b. We note here that these quantiles $\hat{\eta}_1(\gamma)$ and $\hat{\eta}_2(\gamma, \boldsymbol{c}_0)$ depend on the original covariance matrix C, i.e., they differ from those used in the location/scale case. The same comment applies to the other confidence bound procedures following below.

For a given covariate vector $\boldsymbol{c}_0$ we can target the p-quantile $y_p(\boldsymbol{c}_0) = \boldsymbol{c}_0'\zeta + bw_p$ of the Y distribution with covariate dependent location parameter $u(\boldsymbol{c}_0) = \boldsymbol{c}_0'\zeta$ and scale parameter b. We can calculate $\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Y}) - \hat{k}_p(\gamma)\hat{b}(\boldsymbol{Y})$ as an approximate $100\gamma\%$ lower bound for $y_p(\boldsymbol{c}_0)$, where $\hat{k}_p(\gamma)$ is the γ-quantile of the simulated $(\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Z}_i^\star) - w_p)/\hat{b}(\boldsymbol{Z}_i^\star)$, $i = 1, \ldots, N_{\text{sim}}$.

For the tail probability $p(y_0) = G((y_0 - \boldsymbol{c}_0'\zeta)/b)$ with given threshold y_0 and covariate vector $\boldsymbol{c}_0$ we obtain an approximate $100\gamma\%$ upper bound by using the γ-quantile of the simulated values

$$G(\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Z}_i^\star) - x\hat{b}(\boldsymbol{Z}_i^\star)), \quad i = 1, \ldots, N_{\text{sim}} \ ,$$

where $x = (y_0 - \boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{y})/\hat{b}(\boldsymbol{y})$ and $\boldsymbol{y}$ is the originally observed sample vector, obtained under the covariate conditions specified through V.

We note here that the above confidence bounds for the log(characteristic life) or regression location, p-quantiles and tail probabilities depend on the covariate vector $\boldsymbol{c}_0$ that is specified. Not only does this dependence arise through the use of $\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Y})$ in each case but also through the simulated distributions which incorporate $\boldsymbol{c}_0$ in each of these three situations. The only exception is the confidence bound for b, which makes sense since we assumed a constant scale for all covariate situations.

12.3 The Simple Linear Regression Model

Here we assume the following simple linear regression model for the $Y_i = \log(X_i)$

$$Y_i = \zeta_1 + \zeta_2 c_i \ , \quad i = 1, \ldots, n \ .$$

In matrix notation this becomes

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & c_1 \\ \vdots & \vdots \\ 1 & c_n \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} + b \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = C\zeta + b\, \boldsymbol{Z} \ .$$

Here ζ_1 and ζ_2 represent the intercept and slope parameters in the straight line regression model for the location parameter and b represents the degree of scatter (scale) around that line. In the context of the general regression model we have $k = 2$ here and $c_{1,i} = 1$ and $c_{2,i} = c_i$ for $i = 1, \ldots, n$. The conditions for existence and uniqueness of the mle's are satisfied when the covariate values $c_1, \ldots, c_n$ are not all the same.

The R function call `system.time(WeibullRegSim(n=20,Nsim=10000))` (done twice and recording an elapsed time of about 76 seconds each) produced each of the plots in Figure 14. Each call generates its own data set of 20 points using 5 different levels of covariate values. The data are generated from a true Weibull distribution with a known true regression line relationship for $\log(\alpha)$ in relation to the covariates, as shown in the plots. Also shown in these plots is the .10-quantile line. Estimated lines are indicated by the corresponding color coded dashed lines.

In contrast, the quantile lower confidence bounds based on $N_{\text{sim}} = 10000$ simulations are represented by a curve. This results from the fact that the factor $\hat{k}_p(\gamma)$ used in the construction of the lower bound, $\hat{\zeta}_1(\boldsymbol{Y}) + \hat{\zeta}_2(\boldsymbol{Y})c - \hat{k}_p(\gamma)\hat{b}(\boldsymbol{Y})$, is the γ-quantile of the simulated values $(\boldsymbol{c}_0'\hat{\zeta}(\boldsymbol{Z}_i^\star) - w_p)/\hat{b}(\boldsymbol{Z}_i^\star)$, $i = 1, \ldots, N_{\text{sim}}$, and these values change depending on which $\boldsymbol{c}_0' = (1, c)$ is involved. This curvature adjusts to some extent to the sampling variation swivel action in the fitted line.

We repeated the above with a sample of size $n = 50$ (taking about 85 seconds for each plot) and the corresponding two plots are shown in Figure 15. We point out two features. In this second set of plots the lower confidence bound curve is generally closer to the fitted quantile line than in the first set of plots. This illustrates the sample size effect, i.e., we are getting better or less conservative in our bounds. The second feature shows up in the bottom plot where the confidence curve crosses the true percentile line, i.e., it gets on the wrong side of it. Such things happen, because we have only 95% confidence in the bound. Note that these bounds should be interpreted point-wise for each covariate value and should not be viewed as simultaneous confidence bands.

The function `WeibullRegSim` is part of the R workspace on the class web site. It can easily be modified to handle any simple linear regression Weibull data set. Multiple regression relationships could also be accommodated quite easily. To get a feel for the behavior of the confidence bounds it is useful to exercise this function repeatedly, but using $N_{\text{sim}} = 1000$ for faster response.

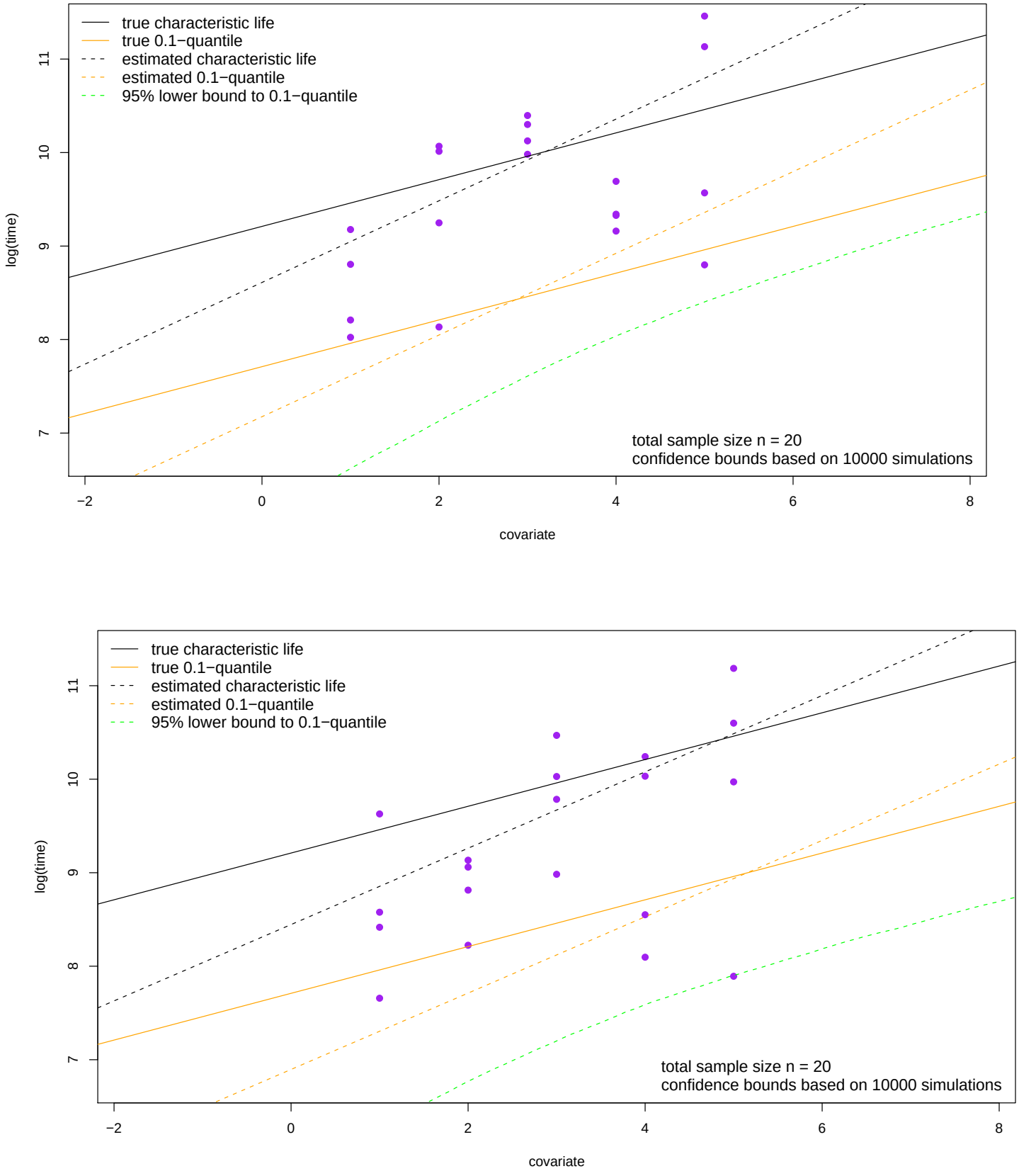


Figure 14: Weibull Regression with Quantile Bounds ($n = 20$)

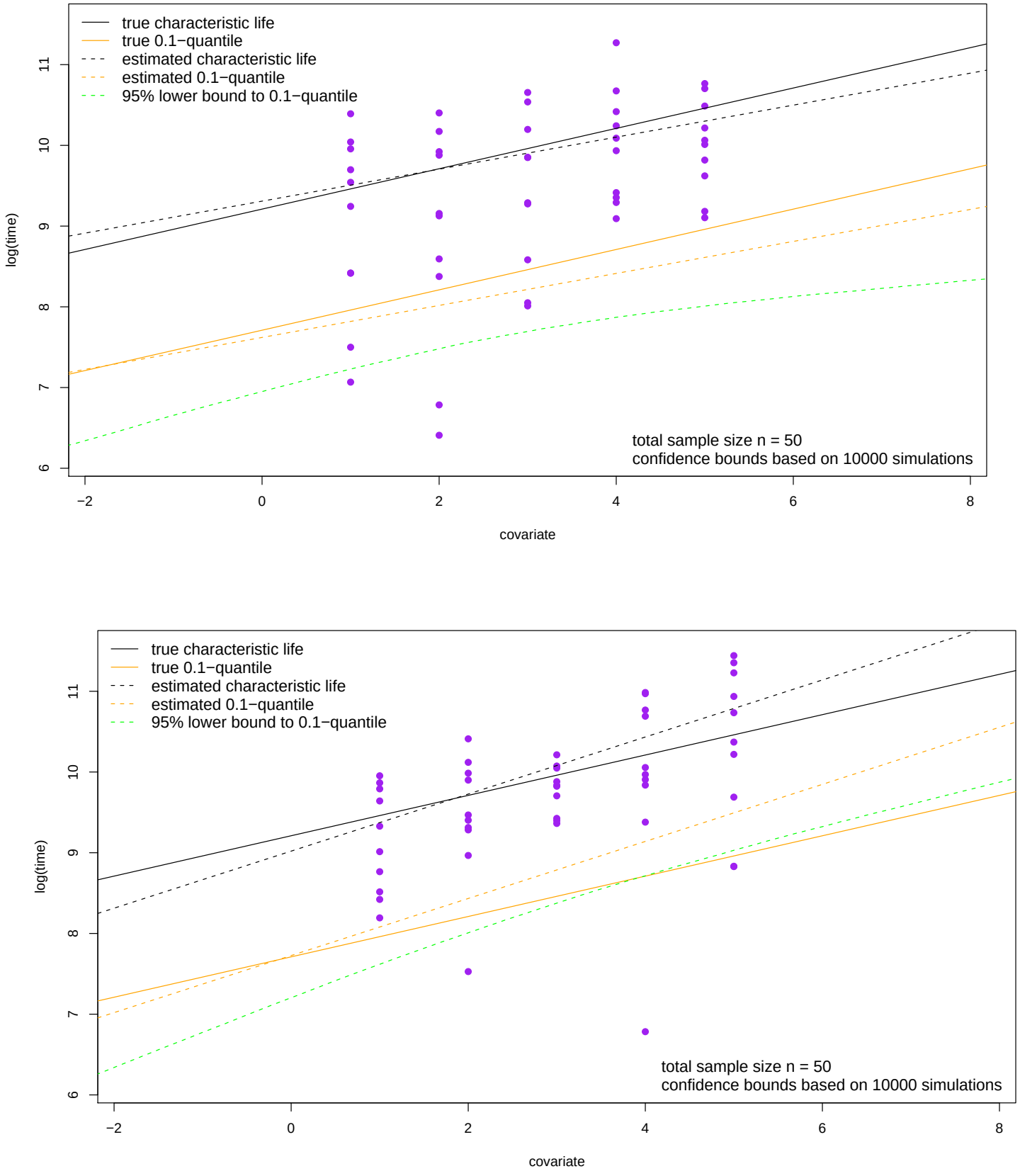


Figure 15: Weibull Regression with Quantile Bounds ($n = 50$)

12.4 The k-Sample Problem

A second illustration example concerns the situation of $k = 3$ samples with same scale but possibly different locations. In terms of the untransformed Weibull data this means that we have possibly different unknown characteristic life parameters $(\alpha_1, \alpha_2, \alpha_3)$ but same unknown shape β for each sample. The modifications for $k \neq 3$ should be obvious. In matrix notation this model is

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ \vdots \\ Y_{n_1+1} \\ \vdots \\ Y_{n_1+n_2} \\ Y_{n_1+n_2+1} \\ \vdots \\ Y_{n_1+n_2+n_3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} + b \begin{pmatrix} Z_1 \\ \vdots \\ Z_{n_1} \\ \vdots \\ Z_{n_1+1} \\ \vdots \\ Z_{n_1+n_2} \\ Z_{n_1+n_2+1} \\ \vdots \\ Z_{n_1+n_2+n_3} \end{pmatrix} = C\zeta + b\, \boldsymbol{Z} \; .$$

Here the Y_i have location $u_1 = \zeta_1$ for the first n_1 observations, they have location $u_2 = \zeta_1 + \zeta_2$ for the next n_2 observations and they have location $u_3 = \zeta_1 + \zeta_3$ for the last n_3 observations. Thus we can consider $u_1 = \zeta_1$ as the baseline location (represented by the first n_1 observations), ζ_2 can be considered as the incremental change from u_1 to u_2 and ζ_3 is the incremental change from u_1 to u_3. If we were interested in the question whether the three samples come from the same location/scale model we would consider testing the hypothesis $H_0 : \zeta_2 = \zeta_3 = 0$ or equivalently $H_0 : u_1 = u_2 = u_3$. Instead of using the likelihood ratio test, which invokes the $\chi^2_{k-1} = \chi^2_2$ distribution as approximate null distribution, we will employ the test statistic suggested in Lawless (1982) (p. 302, equation (6.4.12)) for which the same approximate null distribution is invoked. Our reason for following this choice is its similarity to the standard test statistic used in the corresponding normal distribution model, i.e., when $Z_i \sim \Phi(z)$ instead of $Z_i \sim G(z)$ as in the above regression model. Also, the modification of this test statistic for general $k(\neq 3)$ is obvious. The formal definition of the test statistic proposed by Lawless is as follows:

$$\Lambda_1 = (\hat{\zeta}_2, \hat{\zeta}_3) K_{11}^{-1} (\hat{\zeta}_2, \hat{\zeta}_3)^t \; ,$$

where K_{11} is the asymptotic 2×2 covariance matrix of $(\hat{\zeta}_2, \hat{\zeta}_3)$. Without going into the detailed derivation one can give the following alternate and more transparent expression for Λ_1

$$\Lambda_1 = \frac{\sum_{i=1}^{3} n_i (\hat{u}_i(\boldsymbol{Y}) - \hat{u}(\boldsymbol{Y}))^2}{\hat{b}(\boldsymbol{Y})^2} \; ,$$

where

$$\hat{u}_1(\boldsymbol{Y}) = \hat{\zeta}_1(\boldsymbol{Y}) \ , \ \ \hat{u}_2(\boldsymbol{Y}) = \hat{\zeta}_1(\boldsymbol{Y}) + \hat{\zeta}_2(\boldsymbol{Y}) \ , \ \ \hat{u}_3(\boldsymbol{Y}) = \hat{\zeta}_1(\boldsymbol{Y}) + \hat{\zeta}_3(\boldsymbol{Y}) \ \text{ and } \ \hat{u}(\boldsymbol{Y}) = \sum_{i=1}^{3} (n_i/N) \hat{u}_i(\boldsymbol{Y}) \ ,$$

with $N = n_1 + n_1 + n_3$. In the normal case Λ_1 reduces to the traditional F-test statistic (except for a constant multiplier, namely $(n-k)/((k-1)n) = (n-3)/(2n)$) when writing $\hat{u}_i(\boldsymbol{Y}) = \bar{Y}_{i\bullet}$, $i = 1, 2, 3$ and $\hat{u}(\boldsymbol{Y}) = \bar{Y}_{\bullet\bullet} = (n_1/N)\bar{Y}_{1\bullet} + (n_2/N)\bar{Y}_{2\bullet} + (n_3/N)\bar{Y}_{3\bullet}$ and

$$\hat{b}(\boldsymbol{Y})^2 = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2 \ ,$$

which are the corresponding mle's in the normal case. However, in the normal case one uses the $F_{k-1,N-k}$ distribution as the exact null distribution of the properly scaled Λ_1 and the uncertainty in $\hat{b}(\boldsymbol{Y})^2$ is not ignored by simply referring to the χ^2_{k-1} distribution, using a large sample argument. We don't have to use a large sample approximation either, since the null distribution of Λ_1 (in the log-Weibull case) is free of any unknown parameters and can be simulated to any desired degree of accuracy. This is seen as follows from our equivariance properties. Recall that

$$\frac{\hat{u}_1(\boldsymbol{Y}) - u_1}{\hat{b}(\boldsymbol{Y})} = \frac{\zeta_1(\boldsymbol{Y}) - \zeta_1}{\hat{b}(\boldsymbol{Y})} \ , \ \ \frac{\hat{u}_i(\boldsymbol{Y}) - u_i}{\hat{b}(\boldsymbol{Y})} = \frac{\zeta_1(\boldsymbol{Y}) + \zeta_i(\boldsymbol{Y}) - (\zeta_1 + \zeta_i)}{\hat{b}(\boldsymbol{Y})} \ , \ \ i = 2, 3$$

have distributions free of unknown parameters. Under the hypothesis H_0 when $u_1 = u_2 = u_3 (= u)$ we thus have that

$$\frac{\hat{u}_i(\boldsymbol{Y}) - u}{\hat{b}(\boldsymbol{Y})} \ , \quad \frac{\hat{u}(\boldsymbol{Y}) - u}{\hat{b}(\boldsymbol{Y})} \ , \quad \text{and thus} \quad \frac{\hat{u}_i(\boldsymbol{Y}) - \hat{u}(\boldsymbol{Y})}{\hat{b}(\boldsymbol{Y})} = \frac{\hat{u}_i(\boldsymbol{Y}) - u}{\hat{b}(\boldsymbol{Y})} - \frac{\hat{u}(\boldsymbol{Y}) - u}{\hat{b}(\boldsymbol{Y})}$$

have distributions free of any unknown parameters which in turn implies the above claim about Λ_1. Thus we can estimate the null distribution of Λ_1 by using the N_{sim} simulated values of $\hat{\zeta}_i(\boldsymbol{Z}^\star_j)/\hat{b}(\boldsymbol{Z}^\star_j)$ to create

$$\frac{\hat{u}_1(\boldsymbol{Z}^\star_j)}{\hat{b}(\boldsymbol{Z}^\star_j)} = \frac{\hat{\zeta}_1(\boldsymbol{Z}^\star_j)}{\hat{b}(\boldsymbol{Z}^\star_j)} \ , \ \ \frac{\hat{u}_i(\boldsymbol{Z}^\star_j)}{\hat{b}(\boldsymbol{Z}^\star_j)} = \frac{\hat{\zeta}_1(\boldsymbol{Z}^\star_j) + \hat{\zeta}_i(\boldsymbol{Z}^\star_j)}{\hat{b}(\boldsymbol{Z}^\star_j)} \ , \ \ i = 2, 3 \quad \text{and} \quad \frac{\hat{u}(\boldsymbol{Z}^\star_j)}{\hat{b}(\boldsymbol{Z}^\star_j)} = \frac{\sum_{i=1}^{3} n_i \hat{u}_i(\boldsymbol{Z}^\star_j)/N}{\hat{b}(\boldsymbol{Z}^\star_j)}$$

and thus

$$\Lambda_1(\boldsymbol{Z}^\star_j) = \frac{\sum_{i=1}^{3} n_i (\hat{u}_i(\boldsymbol{Z}^\star_j) - \hat{u}(\boldsymbol{Z}^\star_j))^2}{\hat{b}(\boldsymbol{Z}^\star_j)^2} \qquad j = 1, \ldots, N_{\text{sim}} \ .$$

The distribution of these N_{sim} values $\Lambda_1(\boldsymbol{Z}^\star_j)$ will give a very good approximation for the true null distribution of Λ_1. The accuracy of this approximation is entirely controllable by the choice of N_{sim}. $N_{\text{sim}} = 10000$ should be sufficient for most practical purposes.

The following plots examine the χ^2_2 approximation to the Λ_1 null distribution in the case of 3 samples of respective sizes $n_1 = 5$, $n_2 = 7$ and $n_3 = 9$. This is far from qualifying for a large sample situation. The histogram in Figure 16 is based on $N_{\text{sim}} = 10000$ simulated values of $\Lambda_1(\boldsymbol{Z}^\star)$. Although the superimposed χ^2_2 density is similar in character, there are strong differences. Using the χ^2_2 distribution would result in much smaller p-values than appropriate when these are on the low side.

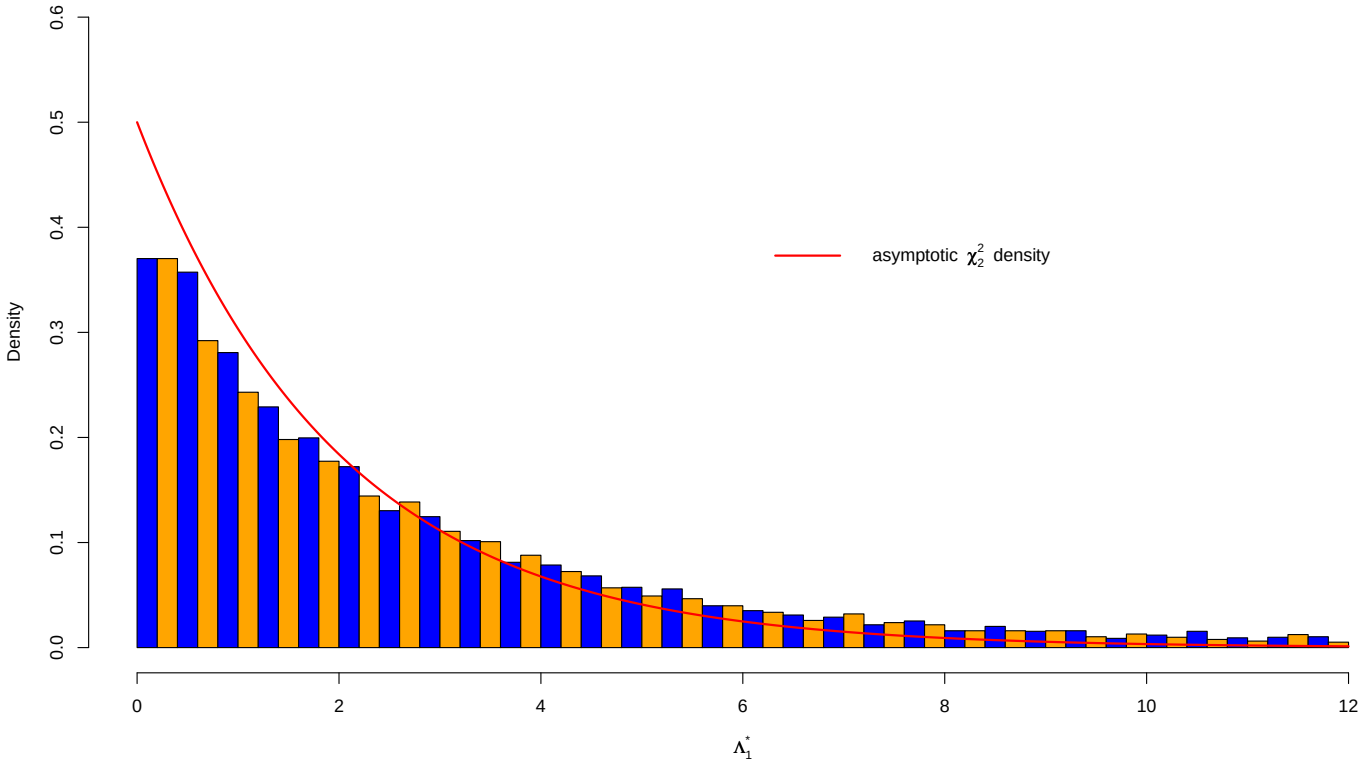


Figure 16: Histogram of Λ_1 Null Distribution with Asymptotic Approximation

Figure 17 shows a QQ-plot of the approximating χ^2_2-quantiles corresponding to the $N_{\text{sim}} = 10000$ simulated and ordered $\Lambda_1(\boldsymbol{Z}^\star_i)$ values. For a good approximation the points should follow the shown main diagonal. The discrepancy is quite strong. Each point on this plot corresponds to a p-quantile where the abscissa value of such a point gives us the approximate p-quantile of the Λ_1 null distribution and the corresponding ordinate gives us the p-quantile of the χ^2_2 distribution which is suggested as asymptotic approximation. The vertical probability scale facilitates the reading off of p for each quantile level on either axis.

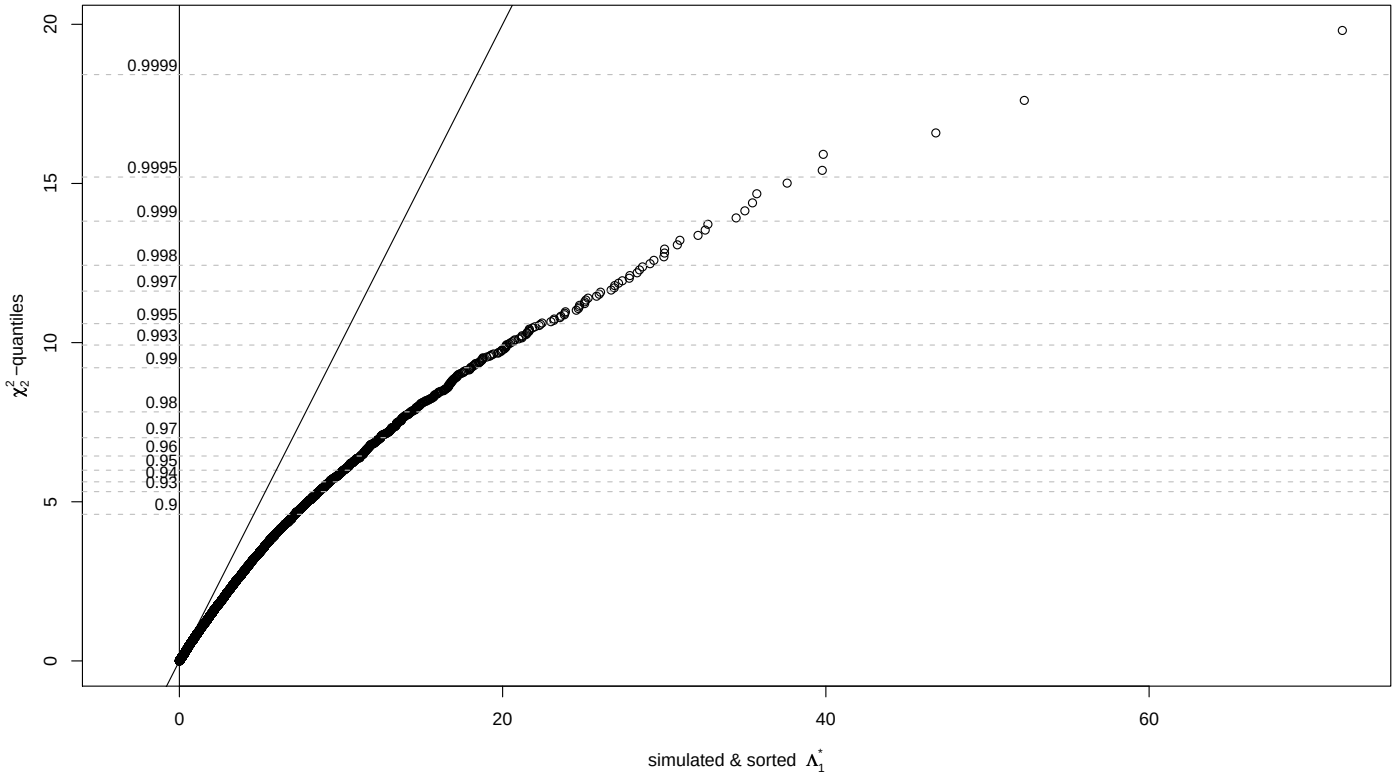


Figure 17: QQ-Plots Comparing $\Lambda_1(\boldsymbol{Z}^\star)$ with χ^2_2

Clearly the p-quantiles of the χ^2_2 distribution are smaller than the corresponding p-quantiles of the the Λ_1 null distribution. If we take the p corresponding to a Λ_1 on the abscissa and look up its p' according to the χ^2_2 scale, we only need to go up to the main diagonal at that abscissa location and look up the p' on the χ^2_2 scale to the left. For example, the .95-quantile for the Λ_1 null distribution would yield a $p' \approx .994$ on the χ^2_2 scale. Thus a true Λ_1 p-value of .05 would translate into a very overstated observed significance level of .006 according to the χ^2_2 approximation.

Figure 18 shows the comparison of the χ^2_2-quantiles with the corresponding quantiles of the $2 \times F_{2,21-3}$ distribution (the factor 2 adjusts for the fact that the numerator of the F statistic is divided by $k - 1 = 2$). This comparison is the counter part to the previous situation if we were to use the asymptotic χ^2_2 distribution as approximation to the exact and true $2 \times F$-ratio distribution had we carried out the corresponding test in a normal data situation, i.e., tested whether three normal random samples with common σ have the same mean. Again, the approximation is not good. The departure from the main diagonal is not as severe as in Figure 17 , but the effect is similar, i.e., the

χ^2_2 distribution would yield much smaller p-values than appropriate when these p-values are on the small side, i.e., when they become critical.

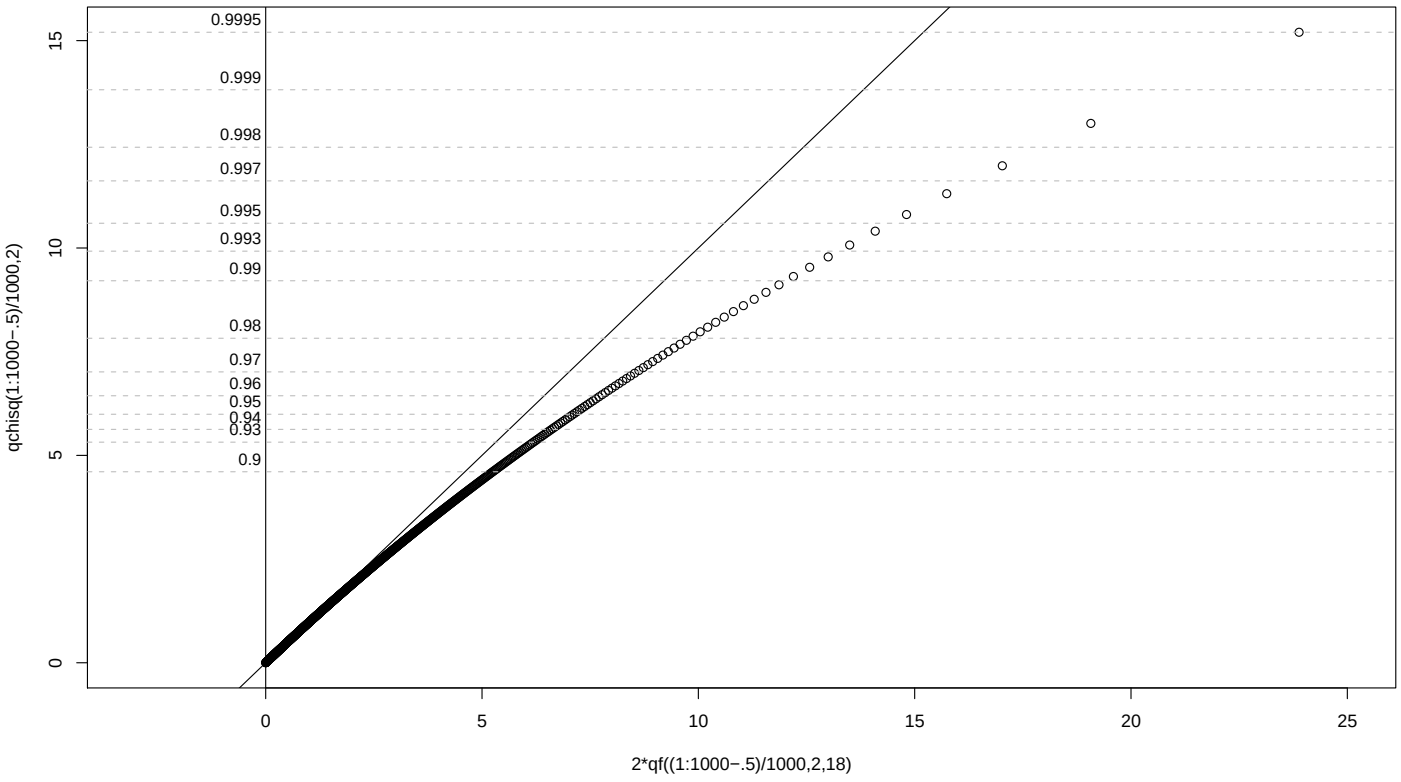


Figure 18: QQ-Plots Comparing $2 \times F_{2,21-3}$ with χ^2_2

Aside from testing the equality of log-Weibull distributions we can also obtain the various types of confidence bounds for the location, scale, p-quantiles and tail probabilities for each sampled population, whether the k locations are the same or not. This is different from doing so for each sample separately since we use all k samples to estimate b, which was assumed to be the same for all k populations. This will result in tighter confidence bounds than what would result from the corresponding confidence bound analysis of individual samples. We omit the specific details since they should be clear from the general situation as applied in the simple linear regression case.

12.5 Goodness of Fit Tests

As in the location/scale case we can exploit the equivariance properties of the mle's in the general regression model to carry out the previously discussed goodness of fit tests by simulation. Using the previous computational formulas for the D, W^2 and A^2 we only need to define the appropriate V_i, namely

$$V_i = G\left(\frac{Y_i - v_i'\hat{\zeta}(\boldsymbol{Y})}{\hat{b}(\boldsymbol{Y})}\right) \ , \quad i = 1, \ldots, n \ .$$

Pierce and Kopecky (1979) showed that the asymptotic null distributions of D, W^2 and A^2, using the sorted values $V_{(1)} \leq \ldots \leq V_{(n)}$ of these modified versions of V_i, are respectively the same as in the location/scale case, i.e., they do not depend on the additional covariates that may be present. This assumes that the covariate matrix C contains a vector of ones. However, for finite sample sizes the effects of these covariates may still be relevant. The effect of using the small sample tables given by Stephens (1986) is not clear.

However, one can easily simulate the null distributions of these statistics since they do not depend on any unknown parameters. Using the data representation $Y_i = \boldsymbol{c}_i'\zeta + bZ_i$ with i.i.d. $Z_i \sim G(z)$, or $\boldsymbol{Y} = C\zeta + b\boldsymbol{Z}$ this is seen from the equivariance properties as follows

$$\frac{Y_i - \boldsymbol{c}_i'\hat{\zeta}(\boldsymbol{Y})}{\hat{b}(\boldsymbol{Y})} \ = \ \frac{\boldsymbol{c}_i'\zeta + bZ_i - \boldsymbol{c}_i'(\zeta + b\hat{\zeta}(\boldsymbol{Z}))}{b\hat{b}(\boldsymbol{Z})} = \frac{Z_i - \boldsymbol{c}_i'\hat{\zeta}(\boldsymbol{Z})}{\hat{b}(\boldsymbol{Z})}$$

and thus

$$V_i \ = \ G\left(\frac{Y_i - \boldsymbol{c}_i'\hat{\zeta}(\boldsymbol{Y})}{\hat{b}(\boldsymbol{Y})}\right) = G\left(\frac{Z_i - \boldsymbol{c}_i'\hat{\zeta}(\boldsymbol{Z})}{\hat{b}(\boldsymbol{Z})}\right) \ .$$

For any covariate matrix C and sample size n the null distributions of D, W^2 and A^2 can be approximated to any desired degree. All we need to do is generate vectors $Z^\star = (Z_1, \ldots, Z_n)'$ i.i.d. $\sim G(z)$, compute the mle's $\hat{\zeta}(\boldsymbol{Z})$, $\hat{b}(\boldsymbol{Z})$, and from that $\boldsymbol{V}^\star$, followed by $D^\star = D(\boldsymbol{V}^\star)$, $W^{2\star} = W^2(\boldsymbol{V}^\star)$ and $A^{2\star} = A^2(\boldsymbol{V}^\star)$. Repeating this a large number of times, say $N_{\text{sim}} = 10000$, would yield values $D_i^\star, W_i^{2\star}, A_i^{2\star}$, $i = 1, \ldots, N_{\text{sim}}$. Their respective empirical distributions would serve as excellent approximations to the desired null distributions of these test of fit criteria.

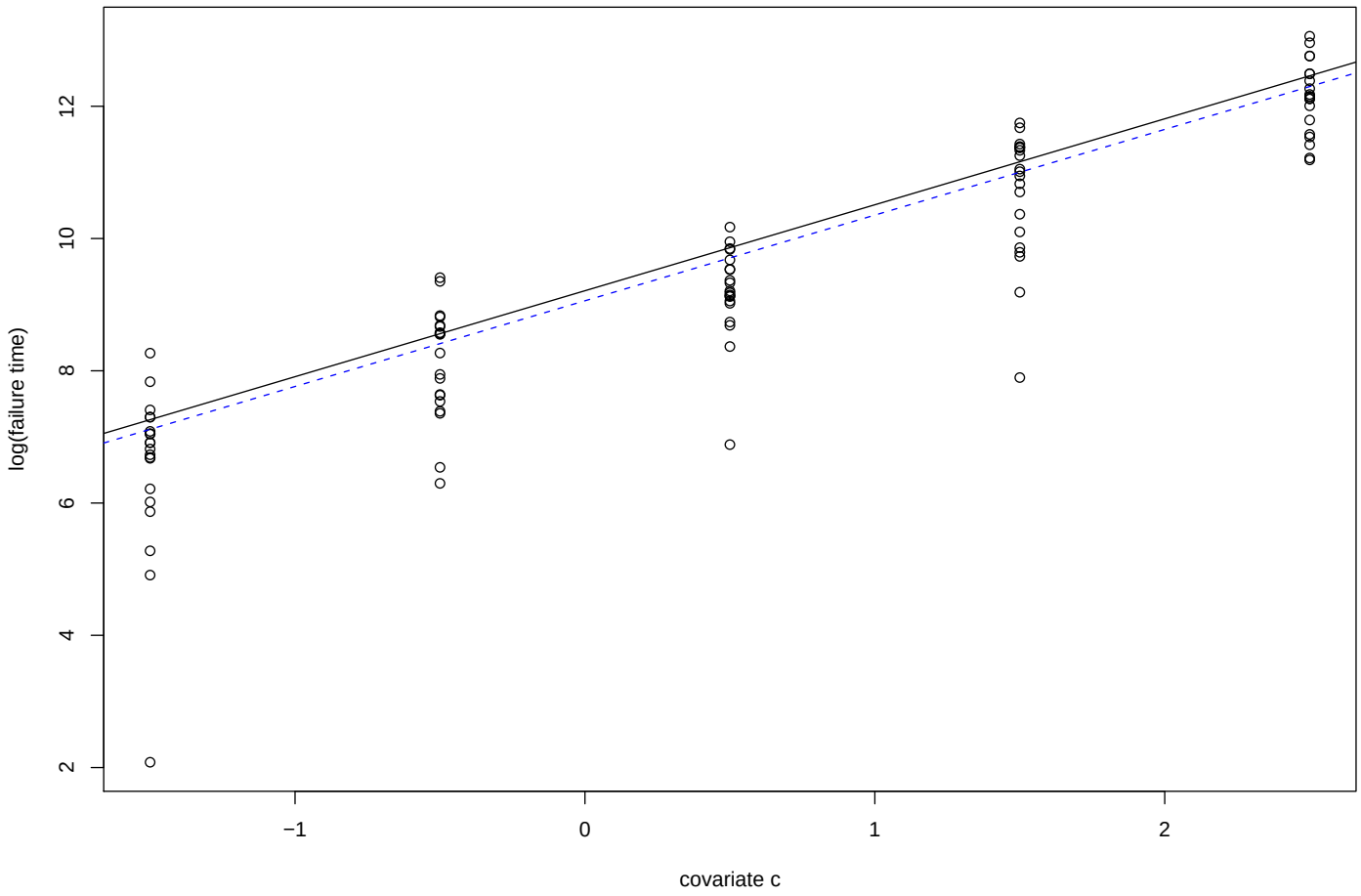


Figure 19: Weibull Regression Example

Figure 19 shows some Weibull regression data which were generated as follows:

$$Y_i = \log(\alpha) + \text{slope} \times c_i + bZ_i \ , \quad i = 1, \ldots, n \qquad \text{with } Z_1, \ldots, Z_n \text{ i.i.d. } \sim G(z) \ ,$$

where $\alpha = 10000, b = 2/3$, slope $= 1.3$ and there were 20 observations each at $c_i = i - 2.5$ for $i = 1, \ldots, 5$.

Here $X_i = \exp(Y_i)$ would be viewed as the Weibull failure time data with common shape parameter $\beta = 1/b = 1.5$ and characteristic life parameters $\alpha_i = \alpha \exp(\text{scale} \times c_i)$ which are affected in a multiplicative manner by the covariate values c_i, $i = 1, \ldots, 5$.

The solid sloped line in Figure 19 indicates the true log(characteristic life) for the Weibull regression data while the dashed line represents its estimate.

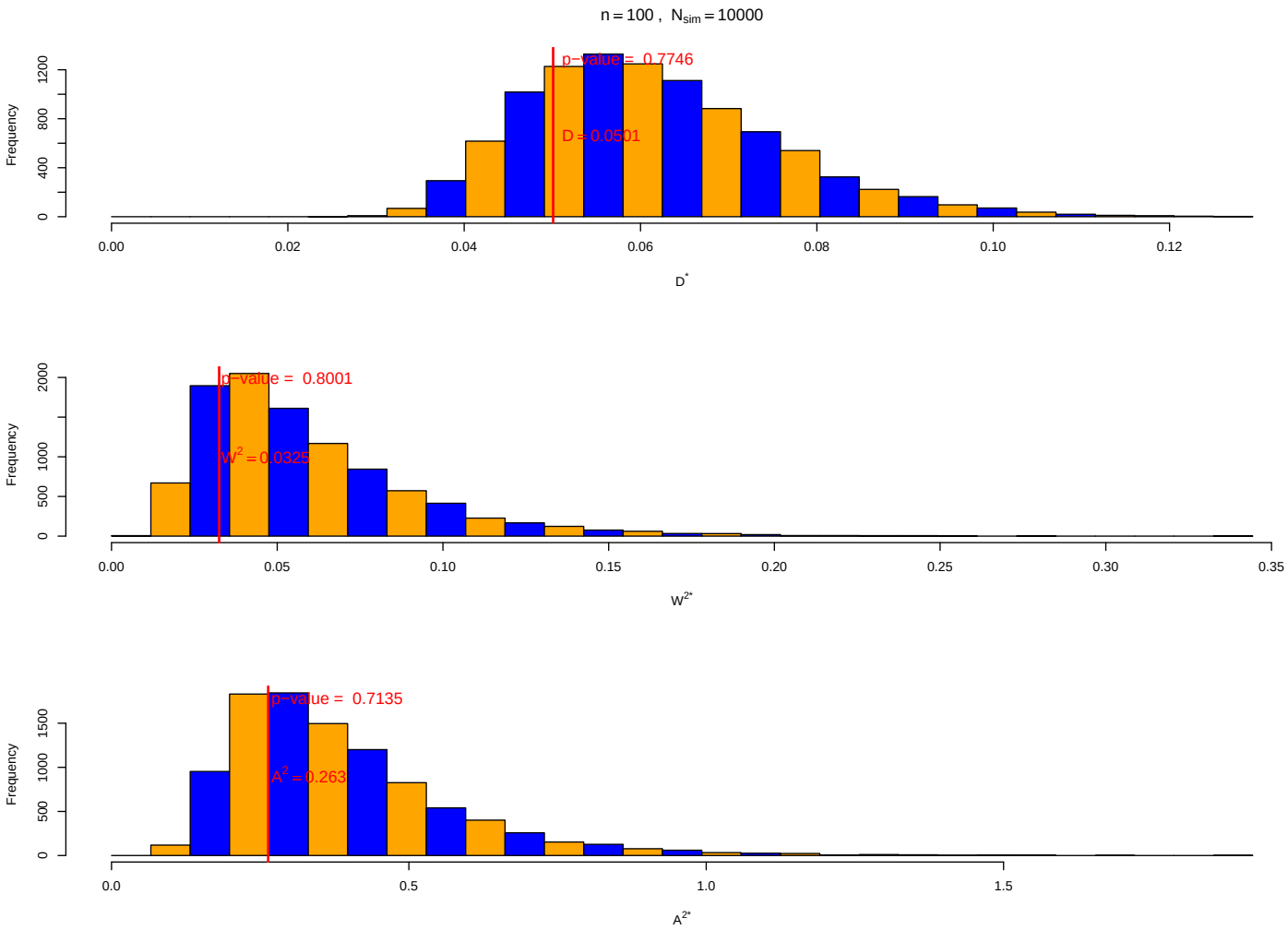


Figure 20: Weibull Goodness of Fit for Weibull Regression Example

For this generated Weibull regression data set Figure 20 shows the results of the Weibull goodness of fit tests in relation to the simulated null distributions for the test criteria D, W^2 and A^2. The hypothesis of a Weibull lifetime distribution cannot be rejected by any of the three test of fit criteria based on the shown p-values.

This example was produced by the R function `WeibullRegGOF` available in the R workspace on the class web site. It took 105 seconds on my laptop. This function performs Weibull goodness of fit tests for any supplied regression data set. When this data set is missing it generates its own Weibull regression data set of size $n = 100$ and uses the indicated covariates and parameters.

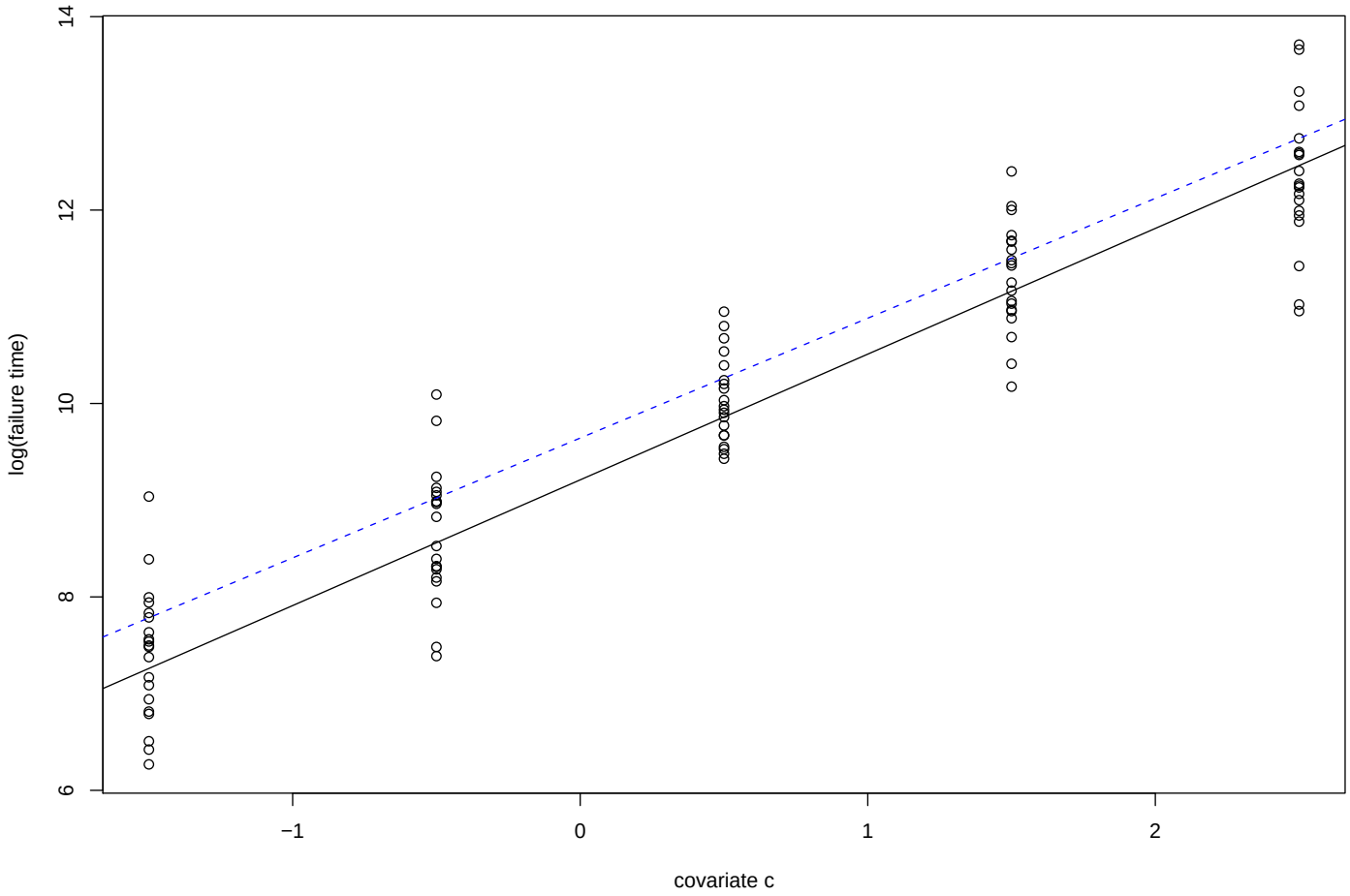


Figure 21: Normal Regression Example

Figure 21 shows some normal regression data which were generated as follows:

$$Y_i = \log(\alpha) + \text{slope} \times c_i + bZ_i\ , \quad i = 1, \ldots, n \qquad \text{with } Z_1, \ldots, Z_n \text{ i.i.d. } \sim \Phi(z)\ ,$$

where $\alpha = 10000, b = 2/3$, slope $= 1.3$ and there were 20 observations each at $c_i = i - 2.5$ for $i = 1, \ldots, 5$. Here $X_i = \exp(Y_i)$ would be viewed as the failure time data. Such data would have a log-normal distribution.

This data set was produced internally within `WeibullRegGOF` by modifying the line that generated the original data sample, so that $Z_i \sim \Phi(z)$, i.e., `Z <- rnorm(n,0,1)` is used. The simulation of the test of fit null distributions remains essentially unchanged except that a different random number starting seed was used.

Here the solid sloped line indicates the mean of the normal regression data while the dashed line represents the estimate according to an assumed Weibull model. Note the much wider discrepancy here compared to the corresponding situation in Figure 19. The reason for this wider gap is that the fitted line aims for the .632-quantile and not the median of that data.

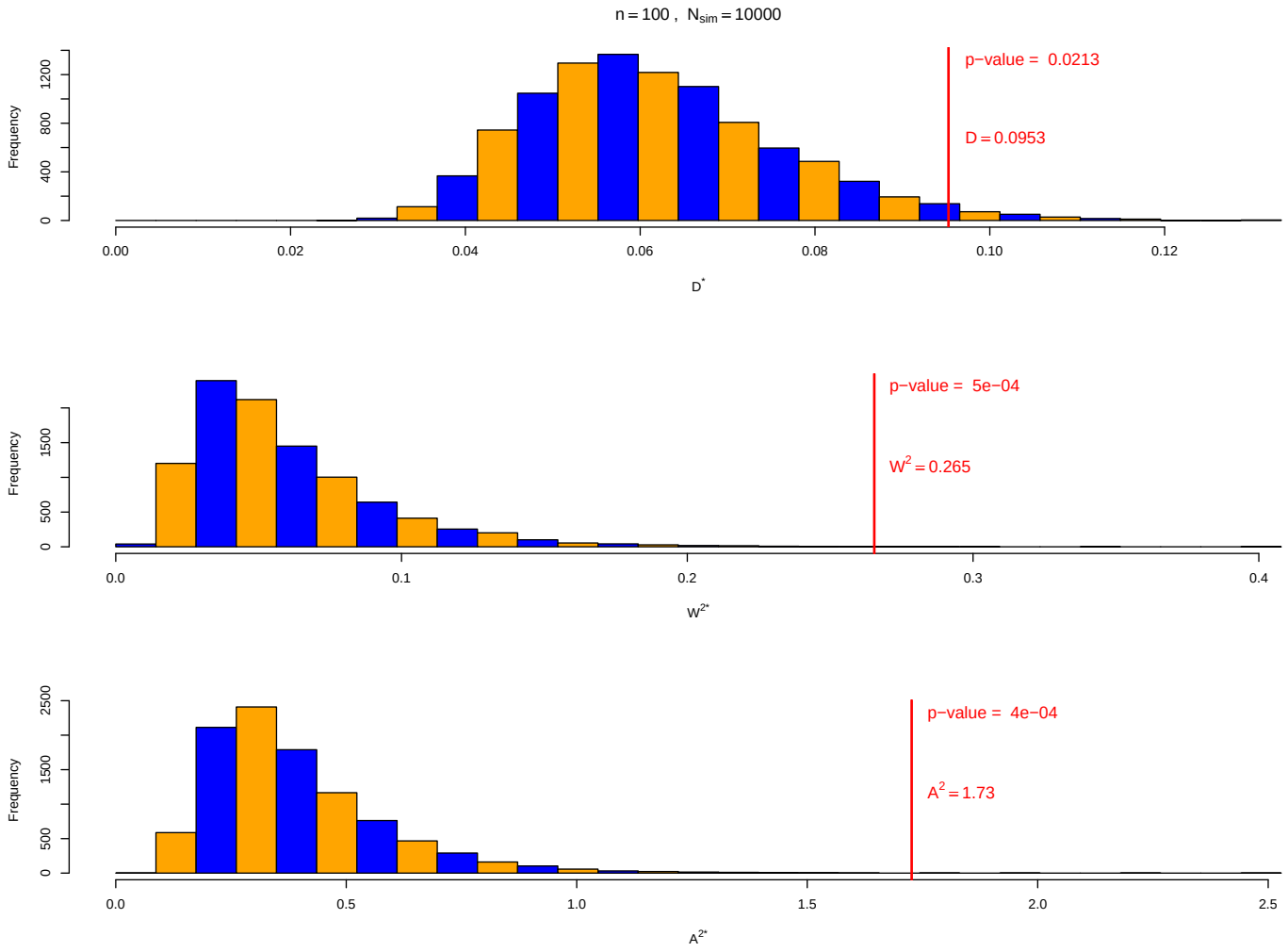


Figure 22: Weibull Goodness of Fit for Normal Regression Example

Here the p-values clearly indicate that the hypothesis of a Weibull distribution should be rejected, although the evidence in the case of D is not very strong. However, for W^2 and A^2 there should be no doubt in the (correct) rejection of the hypothesis.
Any slight differences in the null distributions shown here and in the previous example are due to a different random number seed being used in the two cases.

References

Bain, L.J. (1978) *Statistical Analysis of Reliability and Life-Testing Models*, New York, Dekker.

Bain, L.J. and Engelhardt M.(1991) *Statistical Analysis of Reliability and Life-Testing Models, Theory and Methods, Second Edition*, New York, Dekker.

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

Pierce and Kopecky, K.J. (1979), "Testing Goodness of Fit for the Distribution of Errors in Regression Models," *Biometrika*, Vol. 66, No. 1, 1-5

Saunders, S.C. (1975), "Birnbaum's Contributions to Reliability," *Reliability and Fault Tree Analysis, Theoretical and Applied Aspects of System Reliability and Safety Assessment*,Barlow, R.E, Fussell, J.B., and Singpurwalla, N.D. editors, Society for Industrial and Applied Mathematics, 33 South 17 Street, Philadelphia PA 19103.

Stephens, M.A. (1986), "Tests based on EDF statistics," *Goodness-of-Fit Techniques*, D'Agostino, R. and Stephens, M.A. (editors), Dekker, New York.

Scholz, F.W. (1996) (revised 2001), "Maximum Likelihood Estimation for Type I Censored Weibull Data Including Covariates," *ISSTECH-96-022*, Boeing Information and Support Services.

Scholz, F.W. (2008) "Weibull Probability Paper,"(informal note).

Thoman, D.R., Bain, L.J., and Antle, C.E. (1969), "Inferences on parameters of the Weibull distribution," *Technometrics*, Vol 11, No. 3, 445-460.

Thoman, D.R., Bain, L.J., and Antle, C.E. (1970), "Exact confidence intervals for reliability, and tolerance limits in the Weibull distribution," *Technometrics*, Vol 12, No. 2, 363-371.