

University of Washington



STATISTICS

STAT 498 B

Industrial Statistics

Weibull Analysis

Fritz Scholz

Spring Quarter 2008

The Weibull Distribution

The 2-parameter Weibull distribution function is defined as

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

Write $X \sim \mathcal{W}(\alpha, \beta)$ when X has this distribution function, i.e., $P(X \leq x) = F_{\alpha,\beta}(x)$.

$\alpha > 0$ and $\beta > 0$ are referred to as **scale** and **shape** parameter, respectively.

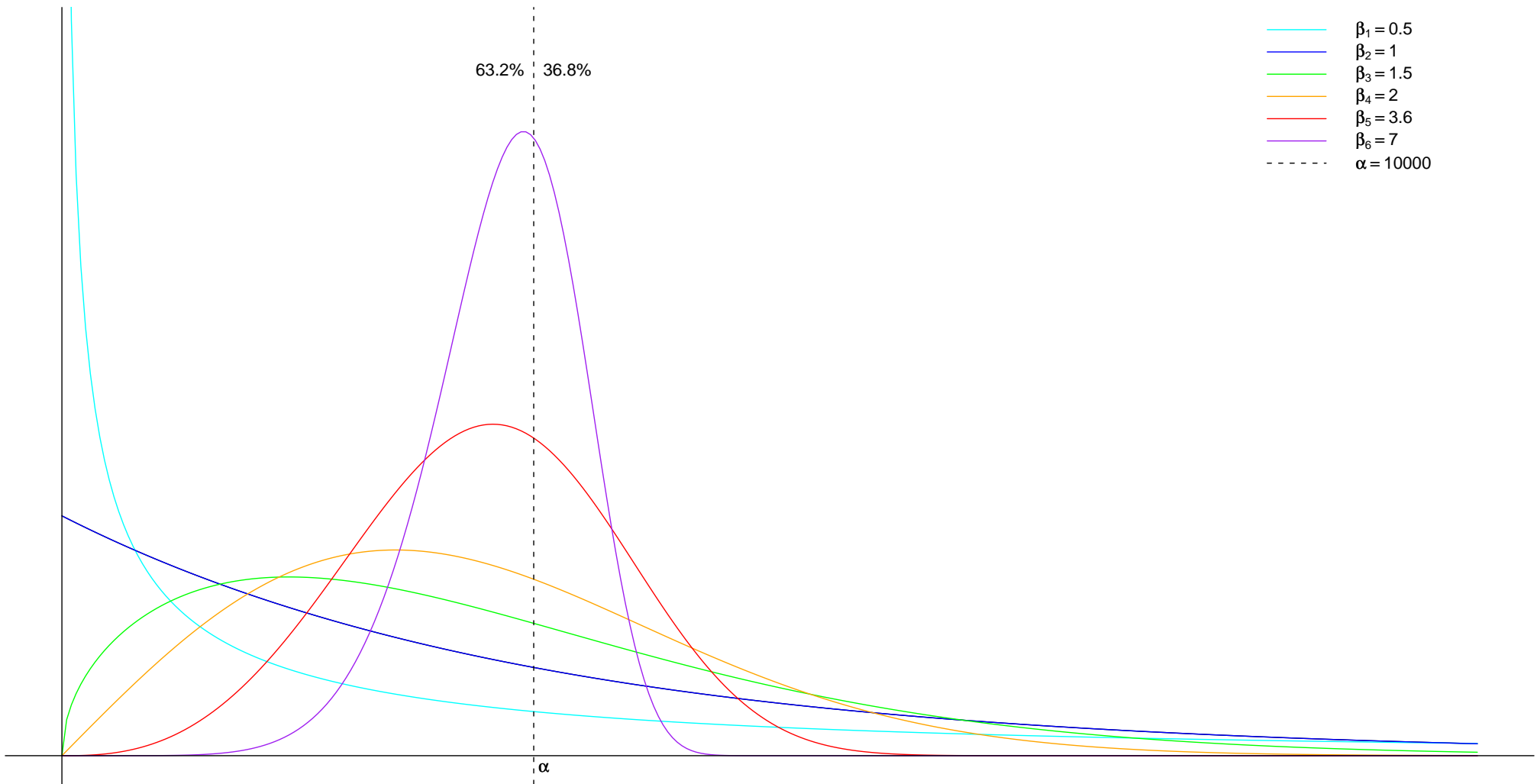
The Weibull density has the following form

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

For $\beta = 1$ the Weibull distribution = 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 $\beta > 0$.

Weibull densities



Note that the Weibull distribution spread around α \searrow as β \nearrow .

The reason becomes clearer later when we discuss the log-transform $Y = \log(X)$.

Moments and Quantiles

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

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

and thus the mean and variance are given by

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

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 $x_p = \alpha$ for all $\beta > 0$

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.

Minimum Closure Property

If X_1, \dots, X_n are independent with $X_i \sim \mathcal{W}(\alpha_i, \beta)$, $i = 1, \dots, n$, then

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

i.e., $\min(X_1, \dots, X_n) \sim \mathcal{W}(\alpha^*, \beta)$.

Similar to the closure property for the normal distribution under summation, i.e.,

if X_1, \dots, X_n are independent with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ then

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

Limit Theorems

This summation closure property is essential 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.

Weakest Link Motivation for Weibull Modeling

The Weibull distribution is appropriate when trying to characterize the random strength of materials or the random lifetime of some system.

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 $\min(X_1, \dots, X_n)$, 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 \implies system lifetime is $\min(X_1, \dots, X_n)$, i.e., approximately Weibull.

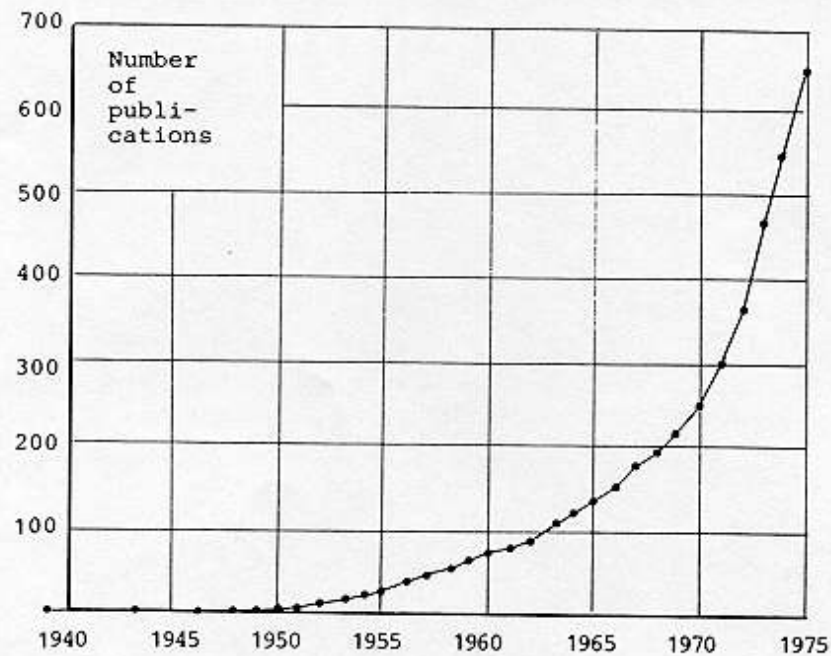
Societies and worked to the last day of his remarkable life. He died on October 12, 1979, in Annecy, France.

* 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. Goran 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.

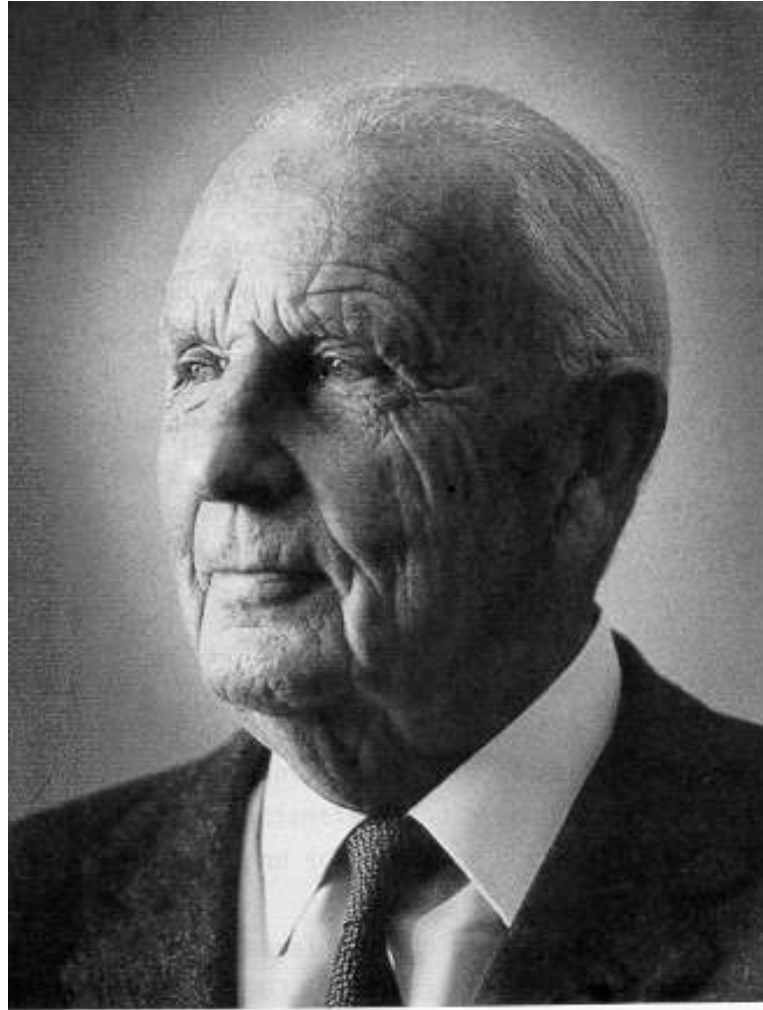
1984

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.



Publications concerning the Weibull distribution, theoretical properties and practical applications

Waloddi Weibull



W. Weibull 1887-1979

photo: Sam C. Saunders, Pullman WA, USA

Weibull Distribution Popularity

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 $\Phi(x)$.

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.

Göran W. Weibull (1981):

“... 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.”

<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 interest. Thus one of the most influential papers in statistics of that decade was published in the Journal of Applied Mechanics. . .

(Maybe that is the reason it was so influential!)

Novel ideas are often misunderstood.

The Hazard Function

The **hazard function** for a nonnegative random variable $X \sim F(x)$ with 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

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

Various other terms are used equivalently for the hazard function, such as **hazard rate**, **failure rate (function)**, or **force of mortality**.

The Weibull Hazard Function

In the case of the Weibull distribution we have

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

The Weibull hazard rate function is

↗ in x when $\beta > 1$,

↘ in x when $\beta < 1$

and constant when $\beta = 1$ (exponential distribution with memoryless property)

Aging & Infant Mortality

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.

Constant Failure Rate

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.

Location-Scale Property of $Y = \log(X)$

A useful property, of which we will make strong use, is the following location-scale property of the log-transformed Weibull distribution.

If $X \sim \mathcal{W}(\alpha, \beta) \implies \log(X) = Y$ has a location-scale distribution, namely its cumulative distribution function (cdf) is

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

with location parameter $u = \log(\alpha)$ and scale parameter $b = 1/\beta$.

Location-Scale Families

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 $z \in \mathbb{R}$, as in our $Y = \log(X)$ Weibull example above.

Quantiles in Location-Scale Families

In 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.

We just illustrate this in the extreme value location-scale model.

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

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 .

Probability Plotting

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 $Y_{(i)}$ ($Y_{(1)} \leq \dots \leq Y_{(n)}$) 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 linear point pattern.

For more on Weibull probability plotting we refer to

<http://www.stat.washington.edu/fritz/DATAFILES498B2008/WeibullPaper.pdf>

Maximum Likelihood Estimation

There are many ways to estimate the parameters $\theta = (\alpha, \beta)$ based on a random sample $X_1, \dots, X_n \sim \mathcal{W}(\alpha, \beta)$.

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, many estimates, computable by hand, had been investigated. They are usually less efficient than **mle's (maximum likelihood estimates)**.

By **efficient estimates** we loosely refer to estimates that have the smallest sampling variance. mle's tend to be efficient, at least in large samples.

Furthermore, under regularity conditions mle's have an approximate normal distribution in large samples.

Maximum Likelihood Estimates

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

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

over $\theta = (\theta_1, \dots, \theta_k)$, i.e., which gives highest local probability to the observed sample $(X_1, \dots, X_n) = (x_1, \dots, x_n)$

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

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

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

These above equations reflect the fact that a smooth function has a horizontal tangent plane at its maximum (minimum or saddle point).

These equations are a necessary but not sufficient condition for a maximum.

The Log-Likelihood

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

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

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

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

It is a lot simpler to deal with the **likelihood equations**

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

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

MLE's in Normal Case

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

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

These are good and intuitively appealing estimates, however $\hat{\sigma}$ is biased.

Although there are many data model situations with explicit mle's, there are even more where mle's are not explicit, but need to be found numerically by computer.

In today's world that is no longer a problem.

Likelihood Equations in the Weibull Case

In the case of a Weibull sample we take the further simplifying step of dealing with the log-transformed sample $(y_1, \dots, y_n) = (\log(x_1), \dots, \log(x_n))$.

Recall that $Y_i = \log(X_i)$ has cdf $F(y) = 1 - \exp(-\exp((y-u)/b)) = G((y-u)/b)$ with $G(z) = 1 - \exp(-\exp(z))$ with $g(z) = G'(z) = \exp(z - \exp(z))$.

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

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

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

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

and thus as **likelihood equations**

$$0 = -\frac{n}{b} + \frac{1}{b} \sum_{i=1}^n \exp\left(\frac{y_i - u}{b}\right) \quad \& \quad 0 = -\frac{n}{b} - \frac{1}{b} \sum_{i=1}^n \frac{y_i - u}{b} + \frac{1}{b} \sum_{i=1}^n \frac{y_i - u}{b} \exp\left(\frac{y_i - u}{b}\right)$$

Simplified Likelihood Equations

These equations can be simplified to a single equation in b and an expression for u in terms of b . We give the latter first and then use it to simplify the other equation.

$$\sum_{i=1}^n \exp\left(\frac{y_i - u}{b}\right) = n \quad \text{or} \quad \exp(u) = \left[\frac{1}{n} \sum_{i=1}^n \exp\left(\frac{y_i}{b}\right) \right]^b$$

Using both of these expressions in the second equation we get a single equation in b

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

$$\text{with} \quad w_i(b) = \frac{\exp(y_i/b)}{\sum_{j=1}^n \exp(y_j/b)} \quad \text{and} \quad \sum_{i=1}^n w_i(b) = 1.$$

$\sum_{i=1}^n w_i(b) y_i - b - \bar{y}$ decreases strictly from $M - \bar{y} > 0$ to $-\infty$ as $0 \nearrow^b \infty$, provided $M = \max(y_1, \dots, y_n) > \bar{y}$. Thus the above equation has a unique solution in b , if not all the y_i coincide. $y_1 = \dots = y_n$ is a degenerate case: $\hat{b} = 0$ & $\hat{u} = y_1$.

Do We Get MLE's?

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.

Consider
$$L(y_1, \dots, y_n, u, b) = \frac{1}{b^n} \prod_{i=1}^n g\left(\frac{y_i - u}{b}\right) \quad \text{for fixed } (y_1, \dots, y_n).$$

Let $|u| \rightarrow \infty$ (the location moves away from all observed data values y_1, \dots, y_n)

and b with $b \rightarrow 0$ (the density is very concentrated near u) and $b \rightarrow \infty$

(all probability is diffused thinly over the half plane $\mathcal{H} = \{(u, b) : u \in \mathcal{R}, b > 0\}$).

It is then easily seen that this likelihood approaches zero in all cases.

Since $L > 0$ but $L \rightarrow 0$ near the fringes of the parameter space \mathcal{H} , it follows that L must have a maximum somewhere with zero partial derivatives. We showed there is only one such point \implies unique maximum likelihood estimate $\hat{\theta} = (\hat{u}, \hat{b})$.

Numerical Stability Considerations

In solving

$$0 = \frac{\sum y_i \exp(y_i/b)}{\sum \exp(y_i/b)} - b - \bar{y}$$

it is numerically advantageous to solve the equivalent equation

$$0 = \frac{\sum y_i \exp((y_i - M)/b)}{\sum \exp((y_i - M)/b)} - b - \bar{y} \quad \text{where } M = \max(y_1, \dots, y_n) .$$

This avoids overflow or accuracy loss in the exponentials for large y_i .
Of course, one could have expressed the y_i in terms of higher units,
say in terms of 1000's, but that is essentially what we are doing.

Type II Censored Data

The above derivations go through with very little change when instead of observing a full sample Y_1, \dots, Y_n we only observe the $r \geq 2$ smallest sample values $Y_{(1)} < \dots < Y_{(r)}$. 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 $X_{(1)} < \dots < X_{(r)}$ and thus $Y_{(i)} = \log(X_{(i)})$, $i = 1, \dots, r$, 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.

Strategy with Type II Censored Data

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 $r \geq 1$.

If $B_n \sim \text{Binomial}(n, p)$ then

$$\begin{aligned} P(B_{n+1} \geq r) &= P(B_n + V_{n+1} \geq r) = p \times P(B_n \geq r - 1) + (1 - p) \times P(B_n \geq r) \\ &= P(B_n \geq r) + p \times P(B_n = r - 1) > P(B_n \geq r) \end{aligned}$$

where V_{n+1} is the Bernoulli random variable for the $(n + 1)^{\text{st}}$ independent trial.

Joint Density of $Y_{(1)}, \dots, Y_{(r)}$

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

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

where the multiplier $n!$ just accounts for the fact that all $n!$ permutations of y_1, \dots, 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 $y_n > y_{n-1} > \dots > y_{r+1} (> y_r)$ and using $\bar{F}(y) = 1 - F(y)$ we get the joint density of the first r failure times $y_1 < \dots < y_r$ as

$$\begin{aligned} f(y_1, \dots, y_r) &= n! \prod_{i=1}^r f(y_i) \times \frac{1}{(n-r)!} \bar{F}^{n-r}(y_r) \\ &= r! \prod_{i=1}^r \frac{1}{b} g\left(\frac{y_i - u}{b}\right) \times \binom{n}{n-r} \left[1 - G\left(\frac{y_r - u}{b}\right)\right]^{n-r} \end{aligned}$$

Likelihood Equations for $Y_{(1)}, \dots, Y_{(r)}$

Log-likelihood

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

where we use the notation $\sum_{i=1}^{r*} x_i = \sum_{i=1}^r x_i + (n-r)x_r$. The likelihood equations are

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

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

$$\frac{\sum_{i=1}^{r*} y_i \exp(y_i/b)}{\sum_{i=1}^{r*} \exp(y_i/b)} - b - \frac{1}{r} \sum_{i=1}^r y_i = 0 \quad \text{again with unique solution for } b \implies \text{mle's } (\hat{u}, \hat{b})$$

For computation again use

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

Computation of MLE's in R

The computation of the mle's $\hat{\alpha}$ and $\hat{\beta}$ 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

```
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  
#
```

Weibull.mle

```
# 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
# the package 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)
```

Weibull.mle

```
}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)}
```

Computation Time for Weibull.mle

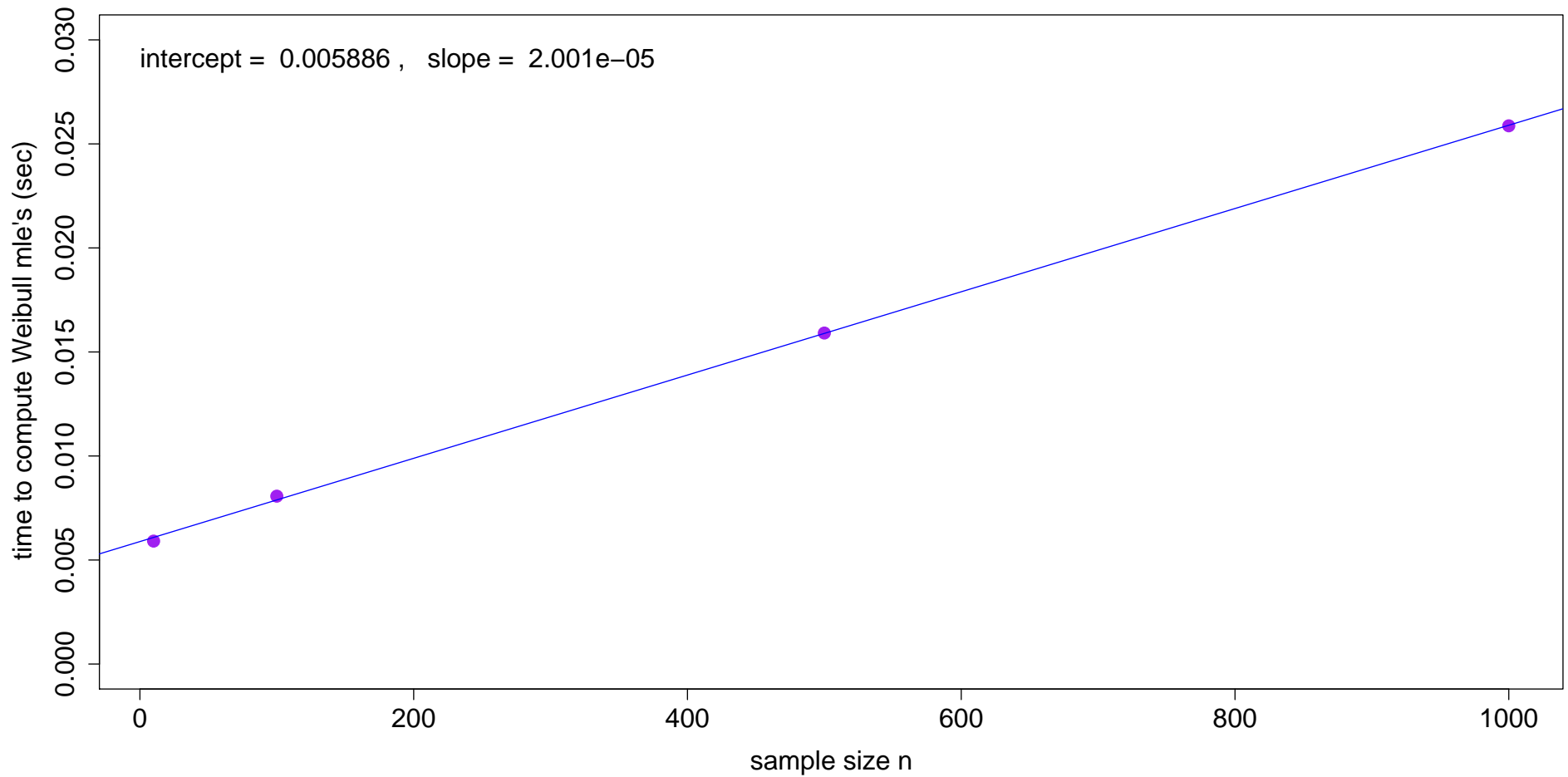
```
system.time(for(i in 1:1000){Weibull.mle(rweibull(10,1))})  
  user  system elapsed  
 5.79   0.00   5.91
```

This 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 procedure which we will discuss.

For $n = 100, 500, 1000$ the elapsed times came to 8.07, 15.91 and 25.87.

The relationship of computing time to n appears to be quite linear, but with slow growth, as the next slide shows.

Computation Time Graph for Weibull.mle



Location and Scale Equivariance of MLE's

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

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

and

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

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

Proof of Equivariance of MLE's

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

Proof of Equivariance of MLE's (contd)

Thus by the uniqueness of the mle's we have

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

or

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

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

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, \dots, X_n is defined as

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

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

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

Large Sample Considerations

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 $n \rightarrow \infty$. 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 $n \rightarrow \infty$ (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.,

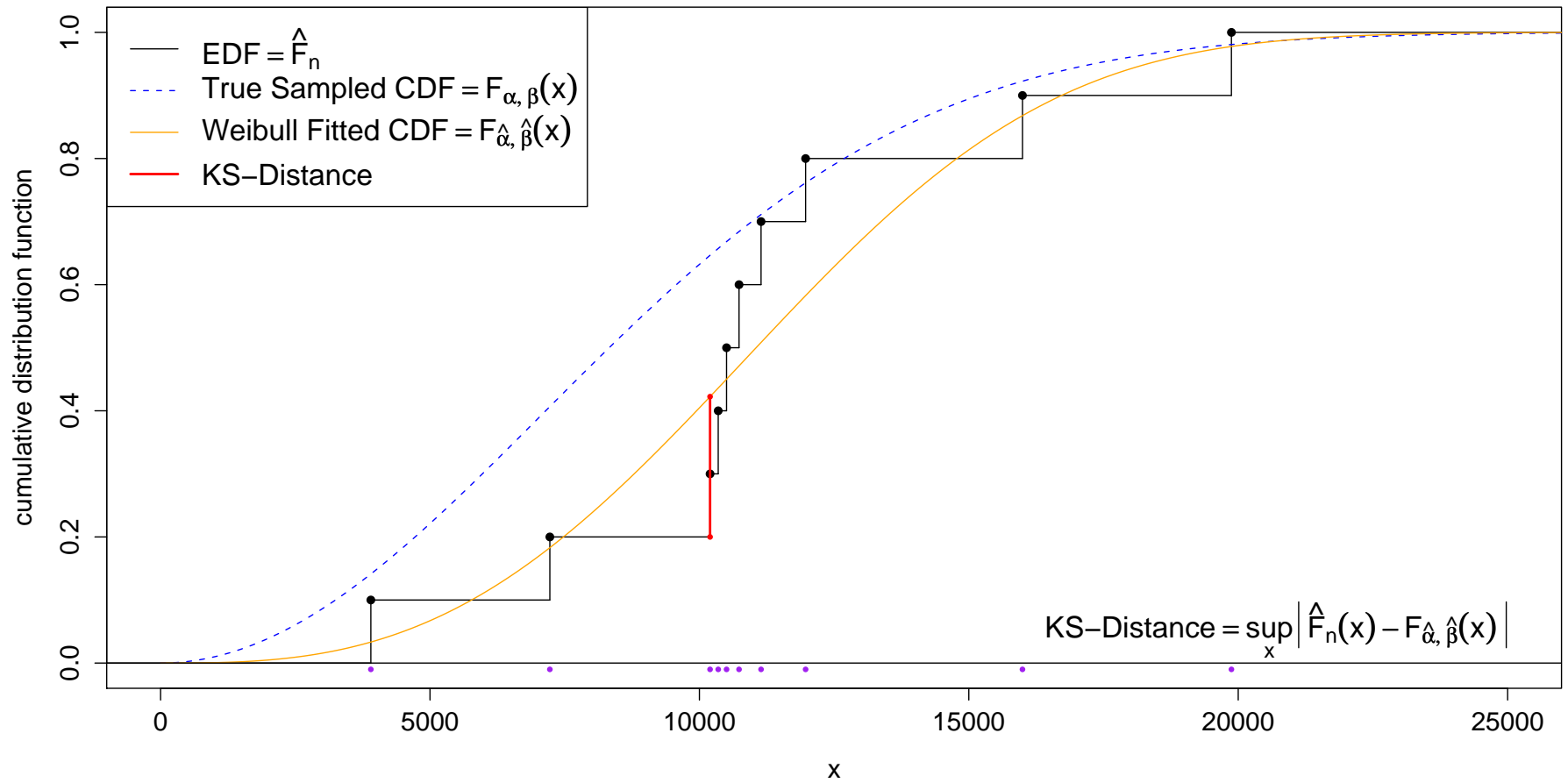
$$\sup_x |\hat{F}_n(x) - F_{\alpha,\beta}(x)| \longrightarrow 0 \quad \text{and} \quad \sup_x |F_{\hat{\alpha},\hat{\beta}}(x) - F_{\alpha,\beta}(x)| \longrightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

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

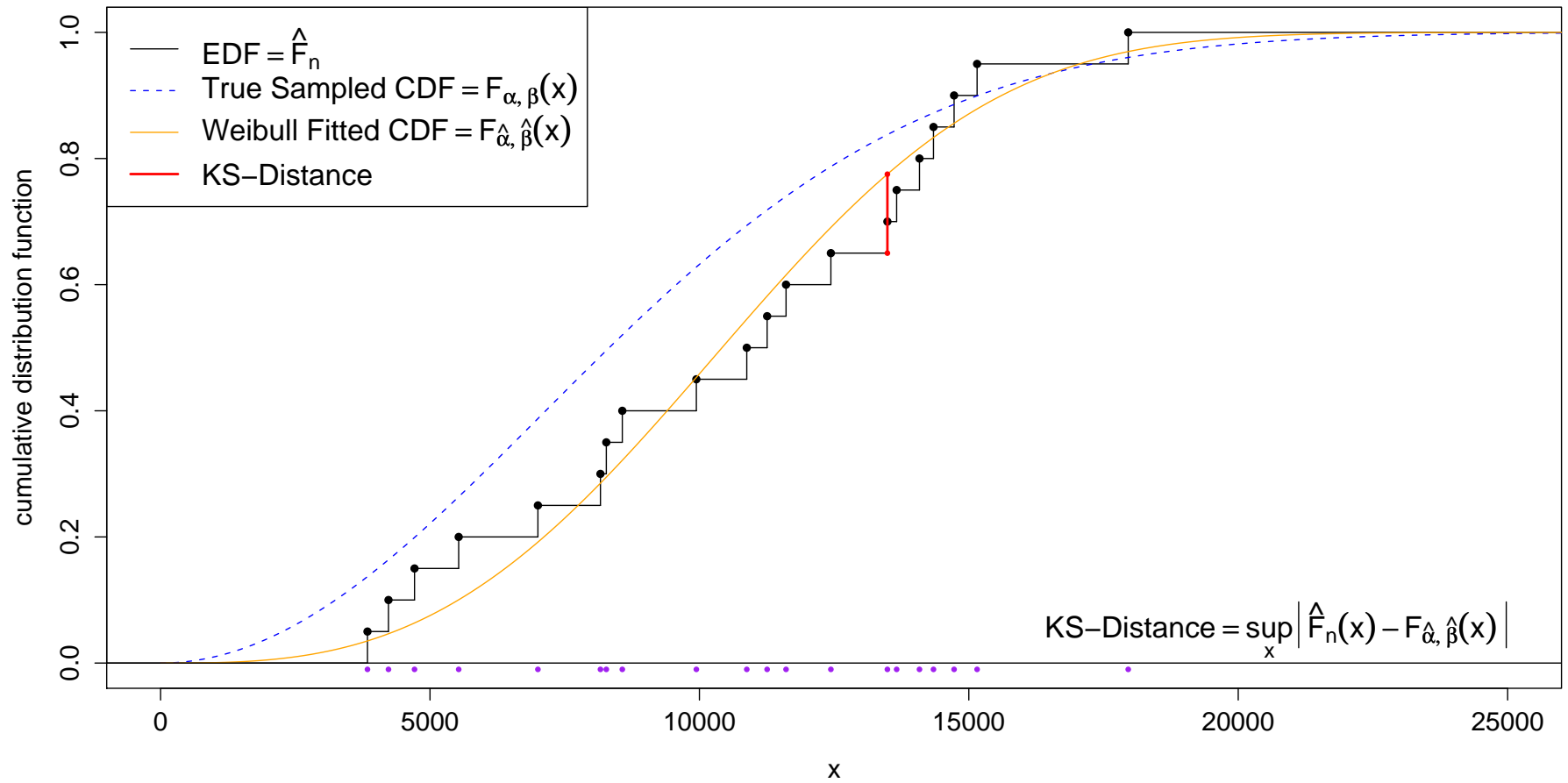
The last discrepancy metric can be evaluated since we know $\hat{F}_n(x)$ and $F_{\hat{\alpha},\hat{\beta}}(x)$.

It is known as the **Kolmogorov-Smirnov (KS) distance** $D(\hat{F}_n, F_{\hat{\alpha},\hat{\beta}})$.

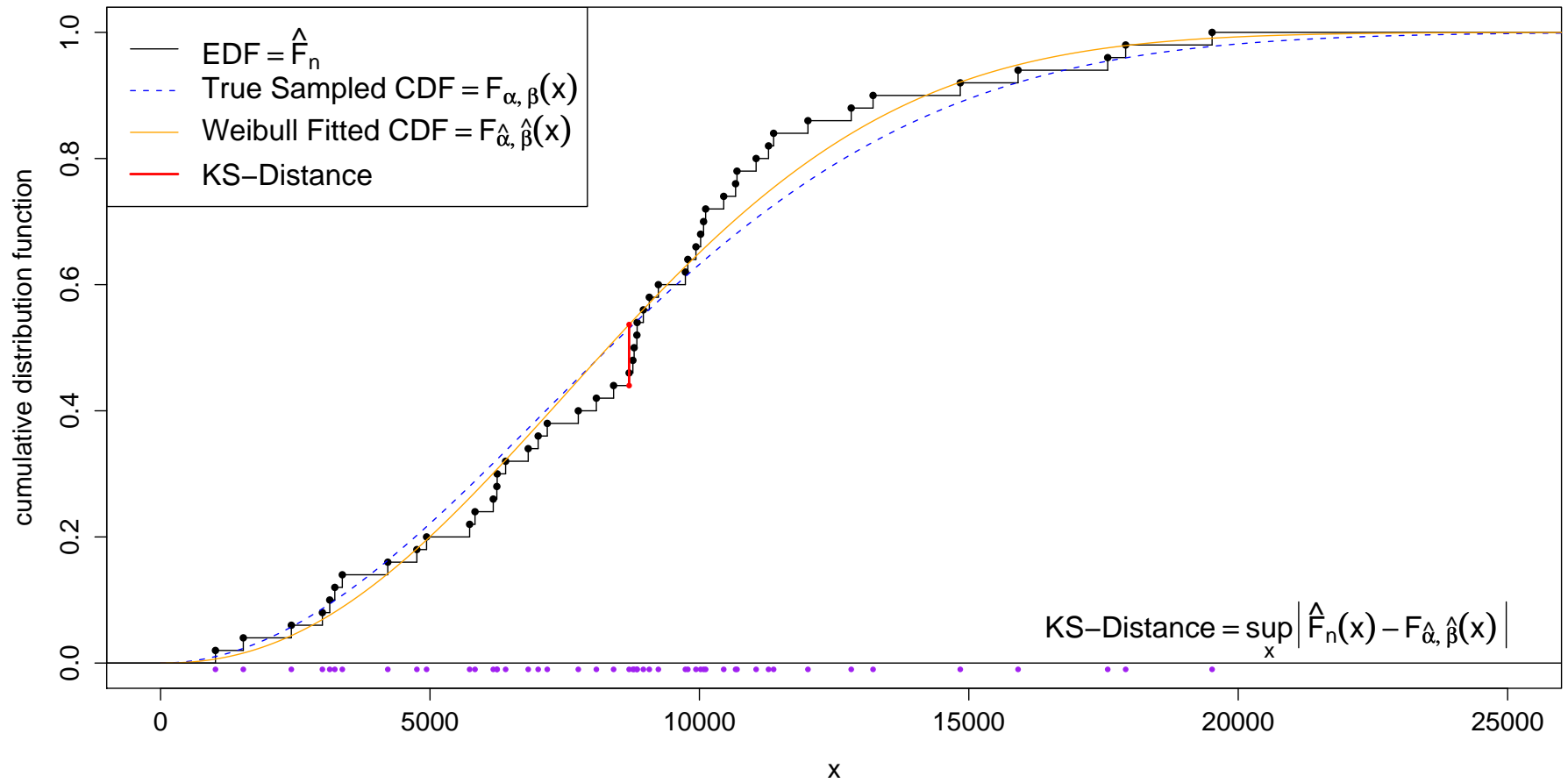
$n = 10$



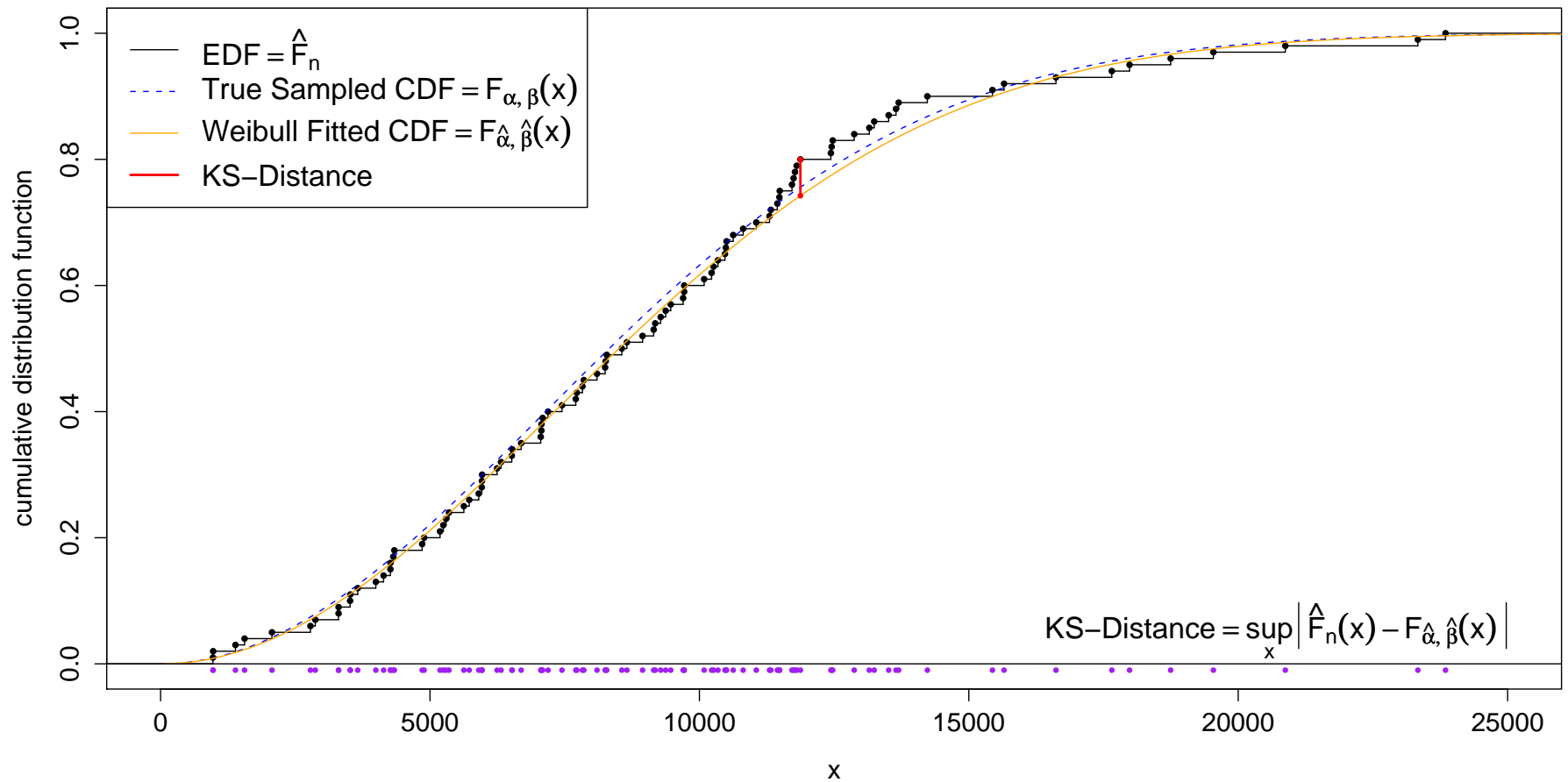
$n = 20$



$n = 50$



$n = 100$



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.

Other Discrepancy Metrics

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} .

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

and

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

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 $D_{\text{CvM}}(F, G)$ and by $g(x)/[G(x)(1 - G(x))]$ in the case $D_{\text{AD}}(F, G)$.

In $D_{\text{AD}}(F, G)$ 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 $g(x)$.

Some Comments

Thus $D_{AD}(F, G)$ is favored in situations where judging distribution tail behavior is important, e.g., in risk situations.

The integration nature gives these last two metrics a more global character.

There is no easy graphical representation of these metrics, except to suggest that when viewing the previous figures illustrating D_{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.

The KS distance won't react so readily to such shifts.

Distance?

We point out the asymmetric nature of $D_{\text{CvM}}(F, G)$ and $D_{\text{AD}}(F, G)$.

We typically have

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

since we integrate w.r.t. to the density of the second argument.

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.

Computational Formulas

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

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

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

For the other two test of fit criteria we have

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

and

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

Null Distributions

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 .

The Distribution of V_i (Ancillarity)

This is seen as follows. Using our prior notation we write $\log(X_i) = Y_i = u + bZ_i$ and since

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

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.

Simulated Null Distributions

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 $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$ from $\mathcal{W}(\alpha = 1, \beta = 1)$ compute the corresponding $\hat{\alpha}^* = \hat{\alpha}(\mathbf{X}^*)$ and $\hat{\beta}^* = \hat{\beta}(\mathbf{X}^*)$, then $V_i^* = \hat{F}(X_i^*) = F_{\hat{\alpha}^*, \hat{\beta}^*}(X_i^*)$ (where $F_{\alpha, \beta}(x)$ is the cdf of $\mathcal{W}(\alpha, \beta)$) and from that the values $D^* = D(\mathbf{X}^*)$, $W^{2*} = W^2(\mathbf{X}^*)$ and $A^{2*} = A^2(\mathbf{X}^*)$.

Calculating all three test of fit criteria makes sense since the main calculation effort is in getting the mle's $\hat{\alpha}^*$ and $\hat{\beta}^*$. Repeating this a large number of times, say $N_{\text{sim}} = 10000$, should give us a reasonably good approximation to the desired three null or reference distributions.

P-Value from the Simulated Null Distributions

From these null distributions one can determine appropriate p -values for any sample X_1, \dots, 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 (discrepancy metric) then the estimated p -value of this sample is simply the proportion of $C(\mathbf{X}_i^*), i = 1, \dots, N_{\text{sim}}$ that are $\geq C(\mathbf{X})$

$$\text{P-value} = \frac{\#C(\mathbf{X}_i^*) \geq C(\mathbf{X})}{N_{\text{sim}}} .$$

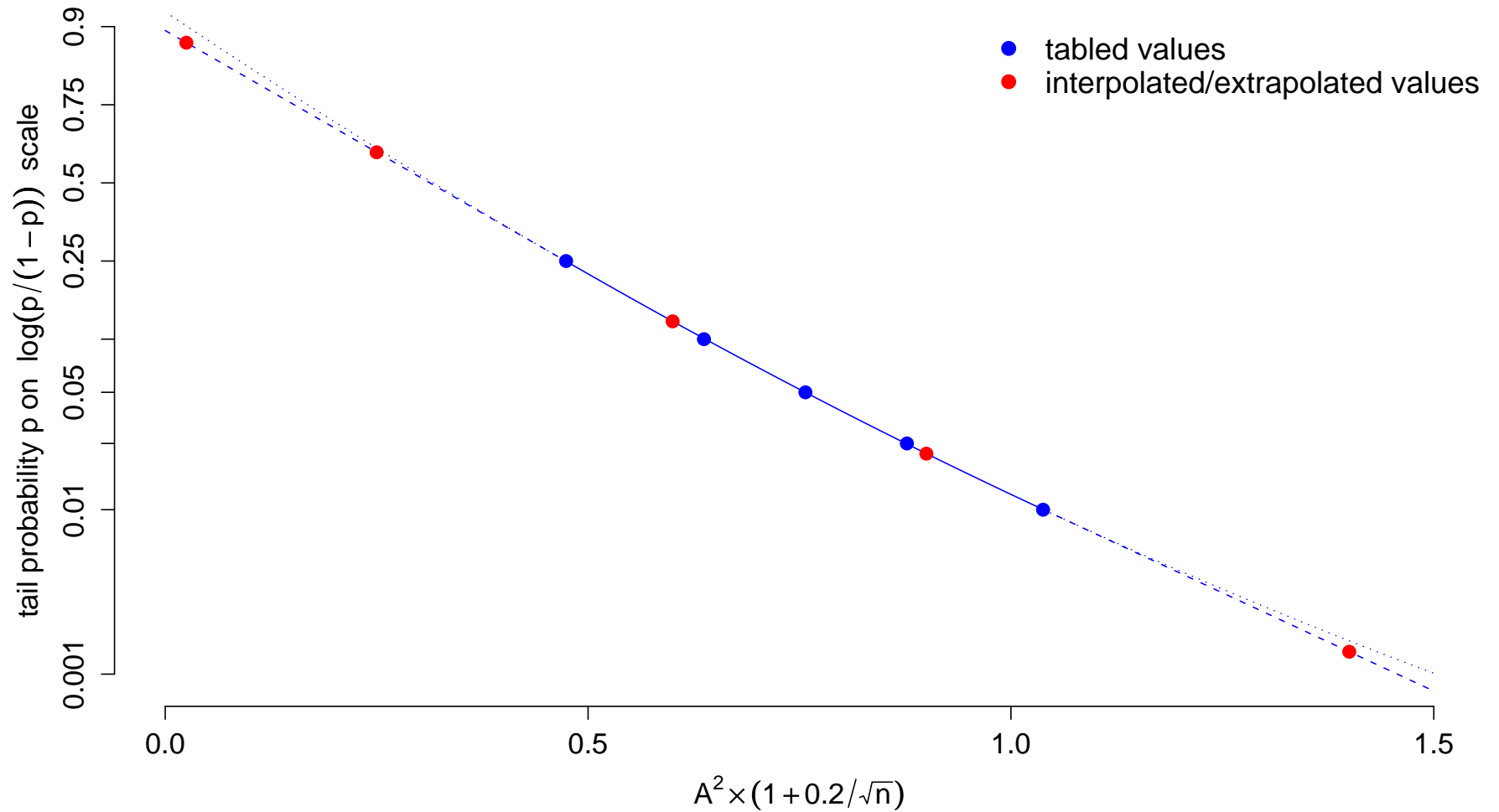
P-Value Tables for A^2 and W^2

Prior to the ease of current computing Stephens (1986) provided tables for the $(1 - \alpha)$ -quantiles $q_{1-\alpha}$ of these null distributions.

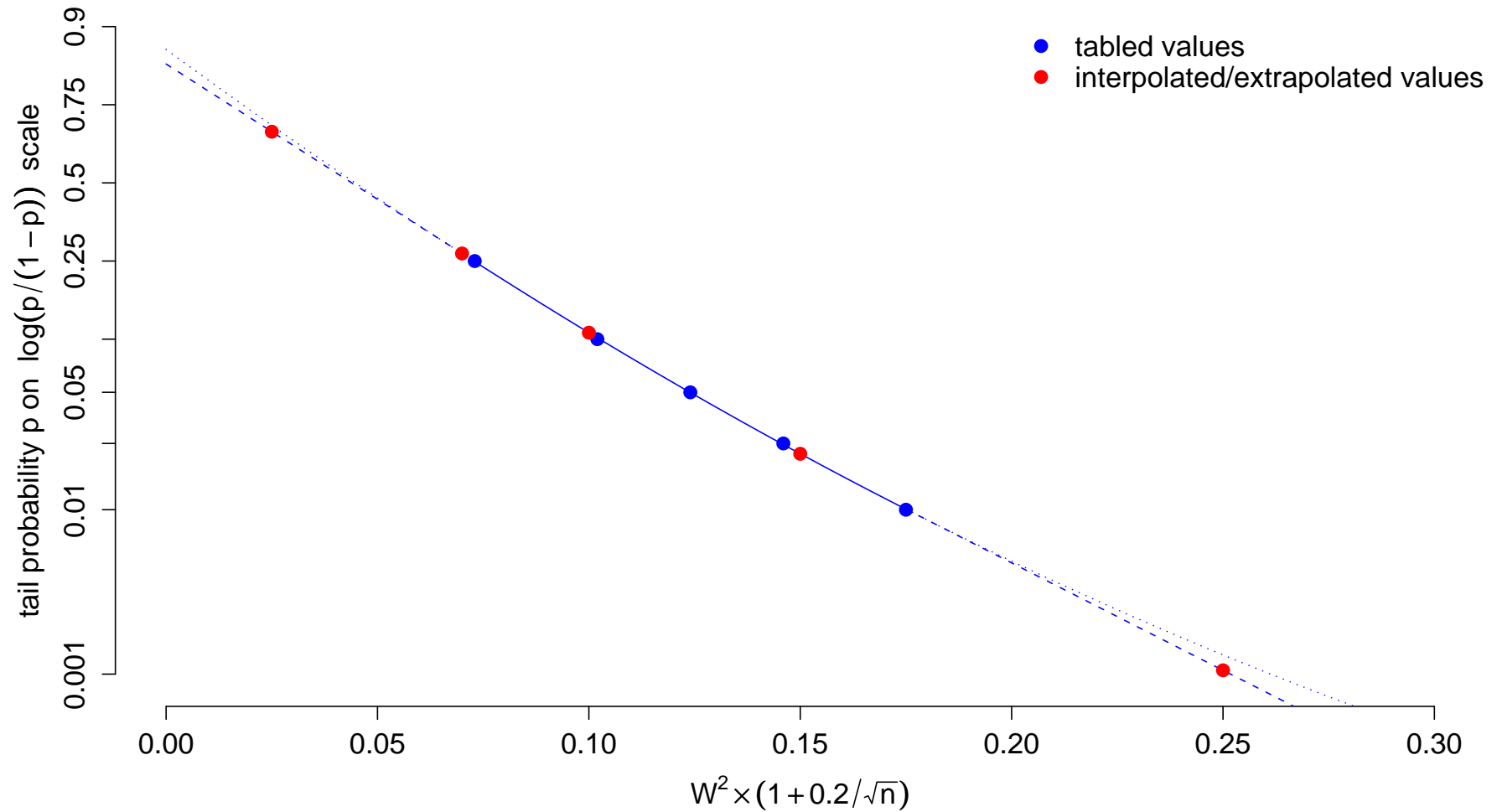
For the n -adjusted versions $A^2(1 + .2/\sqrt{n})$ and $W^2(1 + .2/\sqrt{n})$ these null distributions appear to be independent of n and $(1 - \alpha)$ -quantiles were given by Stephens for $\alpha = .25, .10, .05, .025, .01$.

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 on the next 2 slides.

P-Value Interpolation for A^2



P-Value Interpolation for W^2

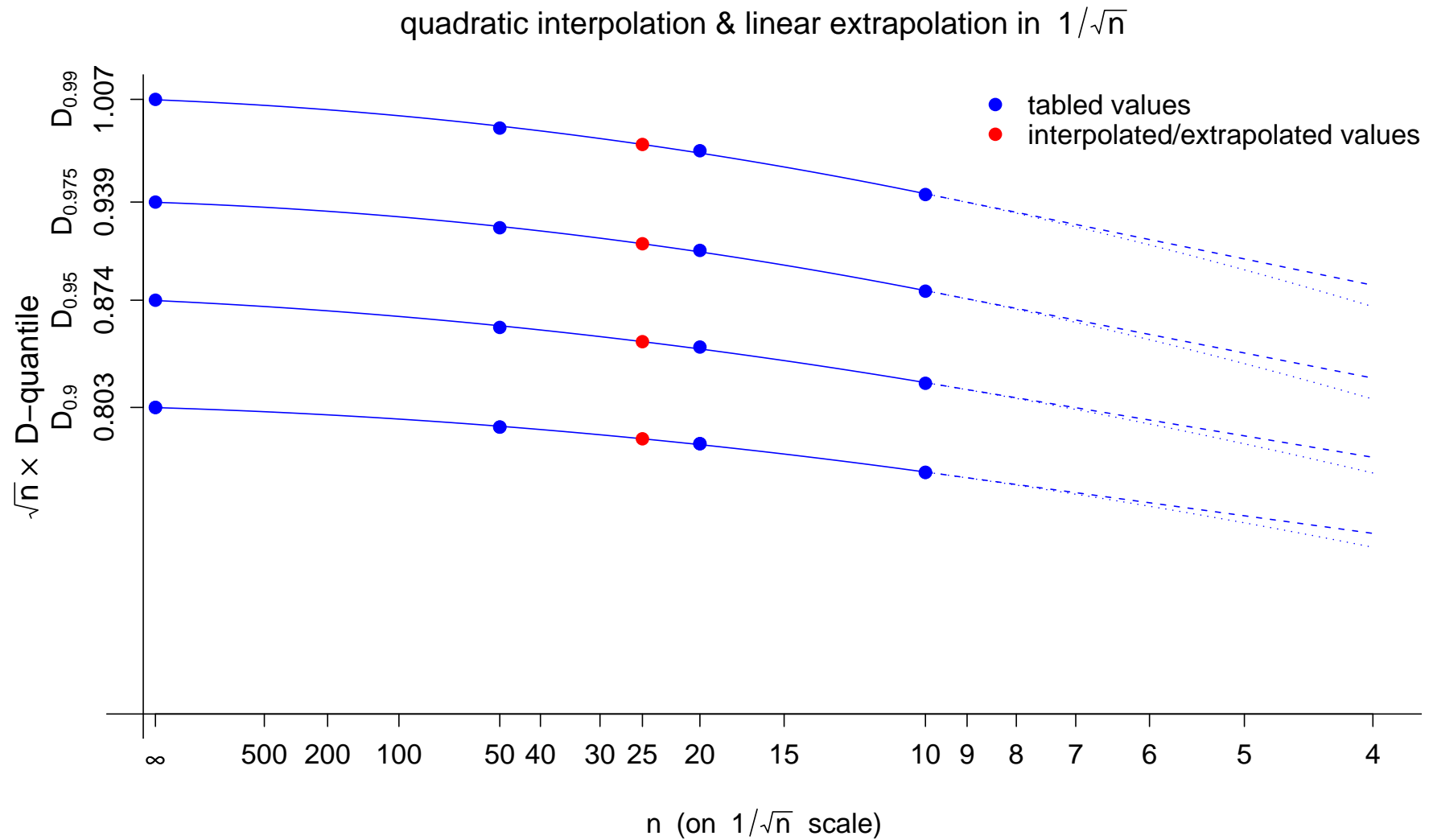


P-Value Tables for D

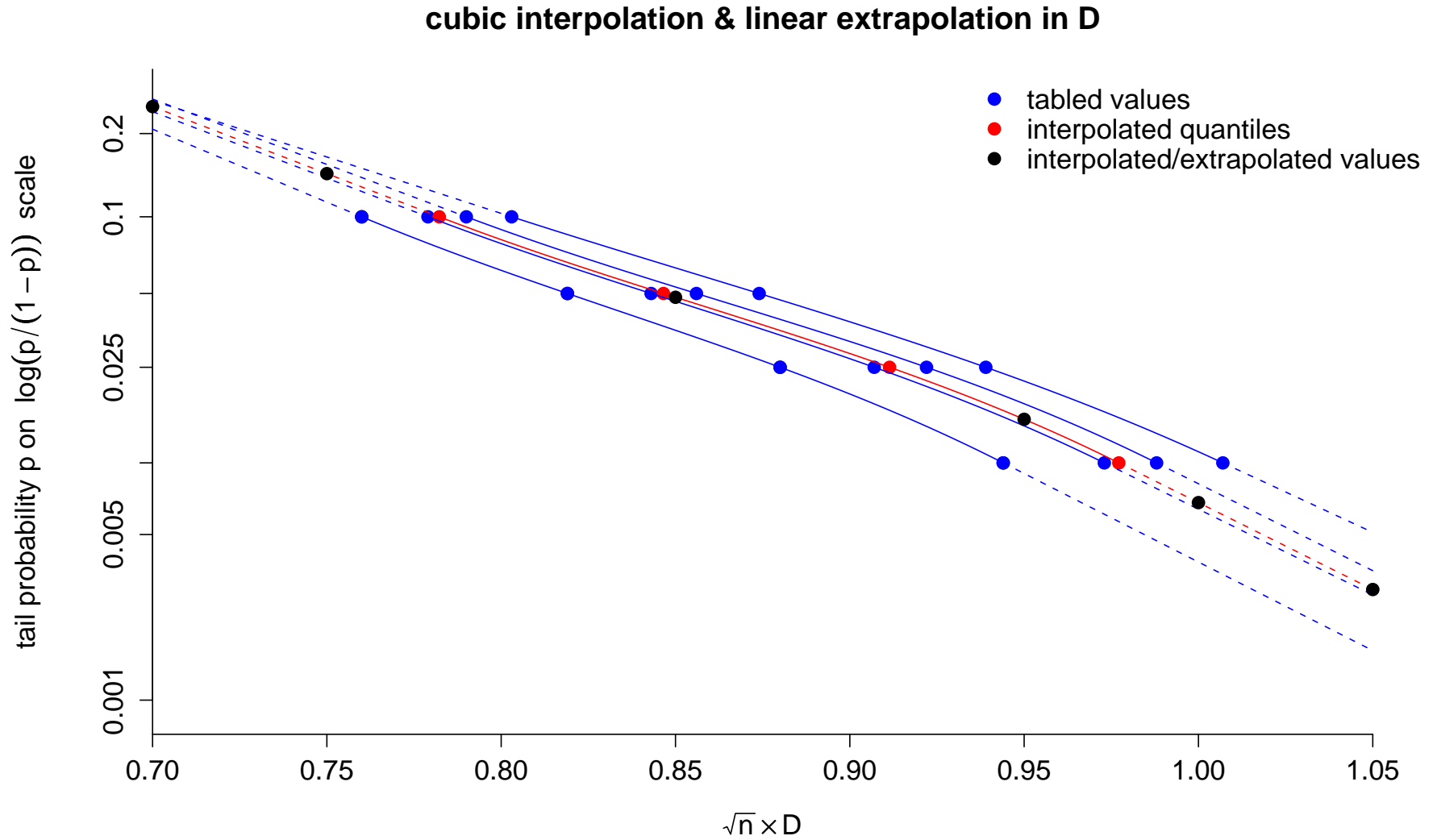
For $\sqrt{n}D$ the null distribution still depends on n (in spite of the normalizing factor \sqrt{n}) and $(1 - \alpha)$ -quantiles for $\alpha = .10, .05, .025, .01$ were tabulated for $n = 10, 20, 50, \infty$ by Stephens (1986).

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 on the next 2 slides.

Quantile Interpolation for D



Quantile Interpolation for D



R Functions for P-Values of D , A^2 and W^2

Functions for computing these p-values (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` generates the interpolation graphs shown in the previous slides, 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, based on the interpolation scheme of the tables.

Pivots in General

A **pivot** is a function $W = \psi(\mathbf{Y}, \vartheta)$ of the data 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 .

Let $\psi(\mathbf{Y}, \vartheta)$ be strictly increasing in ϑ and let $\psi^{-1}(\cdot, \mathbf{Y})$ denote its inverse w.r.t. ϑ .

By η_γ denote the γ -quantile of the known W distribution, then

$$\gamma = P(W \leq \eta_\gamma) = P(\psi(\mathbf{Y}, \vartheta) \leq \eta_\gamma) = P(\vartheta \leq \psi^{-1}(\eta_\gamma, \mathbf{Y}))$$

i.e., we can view $\hat{\vartheta}_{U,\gamma} = \psi^{-1}(\eta_\gamma, \mathbf{Y})$ as a 100 $\gamma\%$ upper bound for ϑ .

Similarly, when $\psi(\mathbf{Y}, \vartheta)$ is strictly decreasing in ϑ , then

$\hat{\vartheta}_{L,\gamma} = \psi^{-1}(\eta_\gamma, \mathbf{Y})$ is a 100 $\gamma\%$ lower bound for ϑ , or

$\hat{\vartheta}_{U,\gamma} = \psi^{-1}(\eta_{1-\gamma}, \mathbf{Y})$ is a 100 $\gamma\%$ upper bound for ϑ .

Pivots Based on $\hat{u}(\mathbf{Y})$ and $\hat{b}(\mathbf{Y})$

The equivariance properties of $\hat{u}(\mathbf{Y})$ and $\hat{b}(\mathbf{Y})$ allow pivots of the following form $W = \psi(\hat{u}(\mathbf{Y}), \hat{b}(\mathbf{Y}), \vartheta)$, i.e., they depend on the data through the mle's.

For a Weibull sample $\mathbf{X} = (X_1, \dots, X_n)$ 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).

It is this distribution of the Z_i that drives the distribution of our pivots, i.e., we will show

$$W = \psi(\hat{u}(\mathbf{Y}), \hat{b}(\mathbf{Y}), \vartheta) \stackrel{\text{in distribution}}{=} \zeta(\mathbf{Z})$$

for some function $\zeta(\cdot)$ that does not depend on unknown parameters.

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 .

Note that $\hat{b}(\mathbf{Z}) = \zeta(\mathbf{Z})$ is not an explicit function of \mathbf{Z} , but it is nevertheless a well defined function. For each \mathbf{Z} the likelihood equations yield a unique solution $\hat{b}(\mathbf{Z})$.

The Pivot Distribution of $\hat{b}(\mathbf{Z})$

How do we obtain the distribution of $\hat{b}(\mathbf{Z})$?

An analytical approach does not seem possible.

We follow Thoman et al. (1969, 1970), Bain (1978), Bain and Engelhardt (1991).

They provided tables for this distribution (and other pivot distributions) 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 $n = 100$.

Simulation Details

In these simulations one simply generates samples $\mathbf{Z} = (Z_1, \dots, Z_n) \sim G(z)$ and finds $\hat{b}(\mathbf{Z})$ (and $\hat{u}(\mathbf{Z})$ for the other pivots discussed later) for each such sample \mathbf{Z} .

By simulating this process $N_{\text{sim}} = 10000$ times we obtain $\hat{b}(\mathbf{Z}_1), \dots, \hat{b}(\mathbf{Z}_{N_{\text{sim}}})$.

The empirical distribution function of these simulated estimates $\hat{b}(\mathbf{Z}_i)$, 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.

Lower Confidence Bound for b

Let $\eta_1(\gamma)$ denote the γ -quantile of $H_1(w)$, 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 for 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} ($N_{\text{sim}} = 10000$) this approximation is practically quite adequate.

Upper Confidence Bound and Interval for b

A $100\gamma\%$ lower bound can be viewed as a $100(1 - \gamma)\%$ upper bound, since $1 - \gamma$ 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.

Similar comments apply to later pivots.

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

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

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 $\mathbf{X} = (X_1, \dots, X_n)$ is the untransformed Weibull sample.

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 $\hat{u}(\mathbf{Z})/\hat{b}(\mathbf{Z})$ a sufficient number N_{sim} times and using the empirical cdf $\hat{H}_2(w)$ of the $\hat{u}(\mathbf{Z}_i)/\hat{b}(\mathbf{Z}_i)$ as proxy for $H_2(w)$.

Lower Confidence Bound for u

As in the previous pivot case we can exploit this pivot distribution as follows

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

Thus we can view $\hat{u}(\mathbf{Y}) - \hat{b}(\mathbf{Y})\eta_2(\gamma)$ as a $100\gamma\%$ lower bound for the unknown 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 u .

Based on $u = \log(\alpha)$ this translates into an approximate $100\gamma\%$ lower bound

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

for α .

Upper bounds and intervals are handled as in the case of b or β .

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

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

Its distribution only depends on the known distribution of \mathbf{Z} and not on the unknown parameters u and b .

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 $\eta_p(\gamma)$.

The Lower Bounds for y_p

This pivot distribution and its quantiles can be approximated sufficiently well by simulating $\{\hat{u}(\mathbf{Z}) + (\hat{b}(\mathbf{Z}) - 1)w_p\} / \hat{b}(\mathbf{Z})$ 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

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

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 treat $\hat{y}_p(\mathbf{Y}) - \hat{\eta}_p(\gamma)\hat{b}(\mathbf{Y})$ as an approximate $100\gamma\%$ lower bound for y_p .

An Alternate Approach

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

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

$k_p(\gamma)$ can be taken as the γ -quantile of the distribution of $(\hat{u}(\mathbf{Z}) - w_p)/\hat{b}(\mathbf{Z})$.

Monotonicity of y_p Bounds in p

The distribution of $(\hat{u}(\mathbf{Z}) - w_p)/\hat{b}(\mathbf{Z})$ can again be approximated by the empirical cdf of N_{sim} simulated values $(\hat{u}(\mathbf{Z}_i) - w_p)/\hat{b}(\mathbf{Z}_i)$, $i = 1, \dots, N_{\text{sim}}$.

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 , since w_p is strictly increasing in p .

The Monotonicity Argument

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

$$\begin{aligned}\gamma &= P(\hat{u}(\mathbf{Z}) + h(p_2)\hat{b}(\mathbf{Z}) \leq w_{p_2}) \\ &= P(\hat{u}(\mathbf{Z}) + h(p_1)\hat{b}(\mathbf{Z}) \leq w_{p_1} + (w_{p_2} - w_{p_1}) + (h(p_1) - h(p_2))\hat{b}(\mathbf{Z})) \\ &\geq P(\hat{u}(\mathbf{Z}) + h(p_1)\hat{b}(\mathbf{Z}) \leq w_{p_1} + (w_{p_2} - w_{p_1})) > \gamma\end{aligned}$$

(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 $\mathbf{z} = (z_1, \dots, z_n)$ and since there is positive probability in any neighborhood of any $\mathbf{z} \in R$ there is positive probability in any neighborhood of $(\hat{u}(\mathbf{z}), \hat{b}(\mathbf{z}))$.

Monotonicity is Intuitive

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 p -quantile of the Weibull distribution we can take

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

as approximate 100 γ % 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.

Upper Confidence Bounds for $p(y) = P(Y \leq y)$

A pivot for $p(y) = P(Y \leq y)$ is not as straightforward as in the previous three cases.

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

The cdf H of this estimate depends on u and b only through $p(y)$, namely

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

Thus by the probability integral transform it follows that

$$W_{p(y)} = H_{p(y)}(\hat{p}(y)) \sim U(0, 1) \quad H_{p(y)}(w) \nearrow \text{strictly in } p(y),$$

i.e., $W_{p(y)}$ is a true pivot.

This is contrary to what is stated in Bain (1978) and Bain and Engelhardt (1991).

Upper Confidence Bounds for $p(y) = P(Y \leq y)$

Rather than using this pivot $W_{p(y)}$ 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

$$\begin{aligned}\gamma &= P(\hat{u}(\mathbf{Y}) + h(p)\hat{b}(\mathbf{Y}) \leq y_p) \\ &= P(h(p) \leq (y_p - \hat{u}(\mathbf{Y})) / \hat{b}(\mathbf{Y})) \\ &= P\left(p \leq h^{-1}\left((y_p - \hat{u}(\mathbf{Y})) / \hat{b}(\mathbf{Y})\right)\right), \quad \text{for any } p \in (0, 1).\end{aligned}$$

For $p = p(y) = P(Y \leq y) = G((y - u)/b)$ we have $y_{p(y)} = y$ and thus also

$$\gamma = P\left(p(y) \leq h^{-1}\left((y - \hat{u}(\mathbf{Y})) / \hat{b}(\mathbf{Y})\right)\right) \quad \text{for any } y \in R \text{ and } u \in R \text{ and } b > 0.$$

Hence $\hat{p}_U(y) = h^{-1}\left((y - \hat{u}(\mathbf{Y})) / \hat{b}(\mathbf{Y})\right)$ is a $100\gamma\%$ upper confidence bound for $p(y)$ for any given threshold y .

Computation of $\hat{p}_U(y)$

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 (and in the corresponding noncentral t application).

Note that $h^{-1}(x)$ solves $-k_p = x$ for p .

$h^{-1}(x)$ 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.

For x we use $x = (y - \hat{u}(\mathbf{Y})) / \hat{b}(\mathbf{Y})$ which is hard to tabulate upfront.

$h^{-1}(x)$ is the γ -quantile of $G(\hat{u}(\mathbf{Z}) + x\hat{b}(\mathbf{Z}))$

For any $x = h(p)$ we have

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

as seen in the derivations of bounds for y_p .

Thus $h^{-1}(x)$ is the γ -quantile of the $G(\hat{u}(\mathbf{Z}) + x\hat{b}(\mathbf{Z}))$ distribution.

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 $\hat{p}(y)$ (sample dependent).

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 $\hat{p}(y)$.

Usage of the Tables

It should be quite clear that all this requires extensive tabulation.

The use of these tables is not easy and often still requires interpolation.

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.

Usage of the Tables (continued)

For the p -quantiles they referred to the 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. Possibly different simulations were involved.

Typos in the Tables

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.

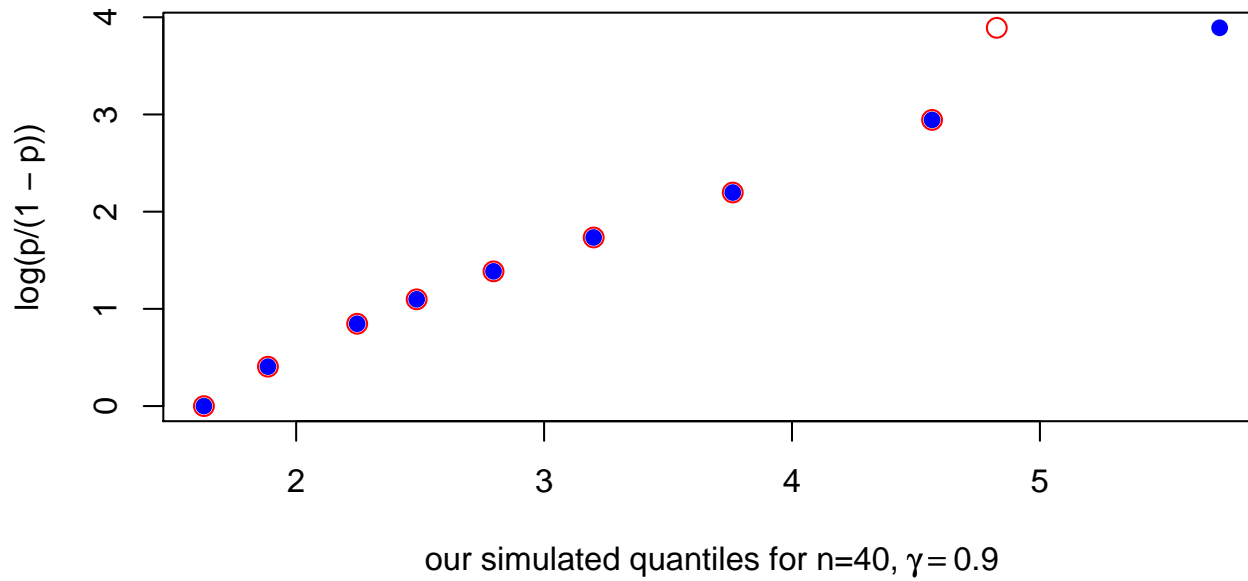
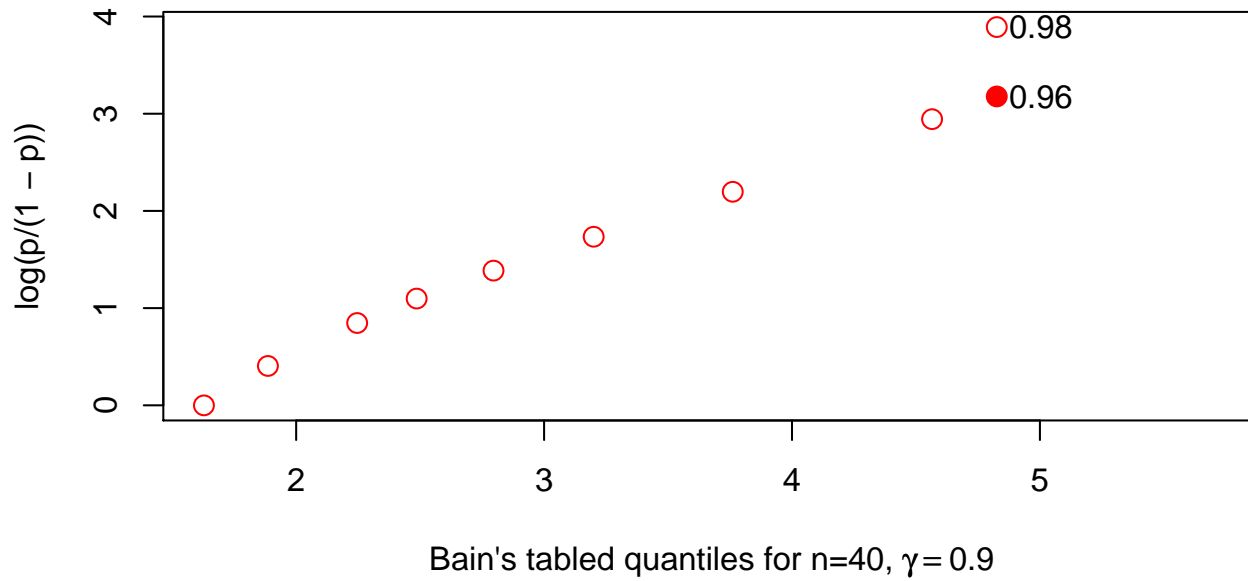
In Table 3A, p.222, row 3 column 5 shows a double minus sign (still present in the 1991 second edition).

Typos in the Tables (continued)

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 next slide. 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.



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](#).

Computation Time for WeibullPivots

For the R function `Weibull.Pivots` (available within the [R work space for Weibull Distribution Applications](#) on the class web site) the call

```
system.time(WeibullPivots(Nsim = 10000, n = 10, r = 10, graphics = F))
```

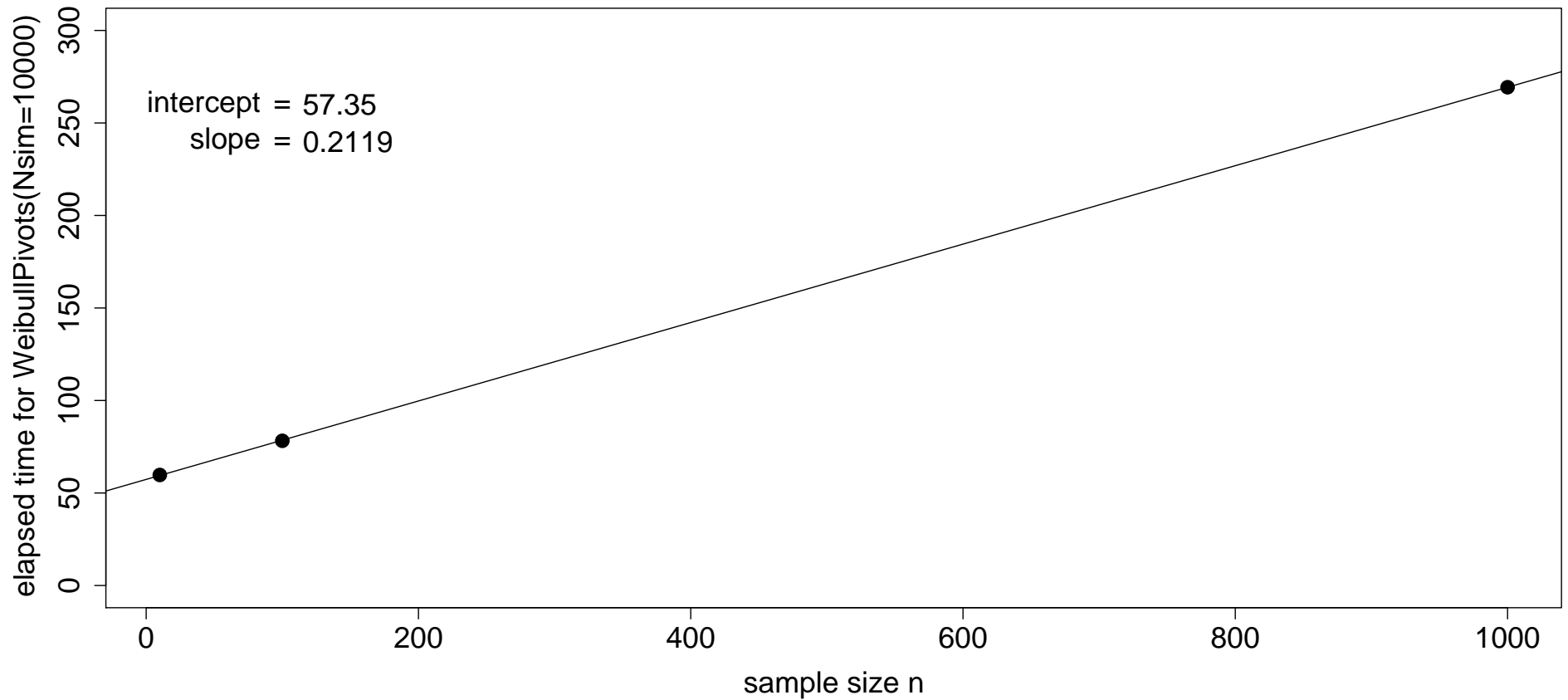
gave an elapsed time of 59.76 seconds.

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 size $n = 100$ with $r = 100$ and $n = 1000$ with $r = 1000$ the corresponding calls resulted in elapsed times of 78.22 and 269.32 seconds.

These three computing times suggest strong linear behavior in n (next slide).



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 for `Weibull.mle`.

What Does WeibullPivots Do?

For all the previously discussed confidence bounds all that is needed is the set of $(\hat{u}(\mathbf{z}_i), \hat{b}(\mathbf{z}_i))$ for $i = 1, \dots, N_{\text{sim}}$.

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 .

Do this for any confidence levels that make sense for the simulated distributions.

No need to run the simulations over and over for each target parameter, confidence level γ , p or y , unless one wants independent simulations for some reason.

Usage of WeibullPivots

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 $\eta(\gamma)$ 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.

Calling Sequence of WeibullPivots

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

Here $N_{\text{sim}} = N_{\text{sim}}$ 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.

We need $r > 1$ and at least two distinct observations among $X_{(1)}, \dots, X_{(r)}$ in order to estimate any spread in the data.

Other Inputs to WeibullPivots

The available sample values X_1, \dots, 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 = 10000` (default) and `beta = 1.5` (default), either by using the full sample X_1, \dots, X_n or a type II censored sample X_1, \dots, 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.

Output of WeibullPivots

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.estimate,  
$p.quantile.bounds, $Tail.Probability.Estimates, $Tail.Probability.Bounds
```

The structure and meaning of these components will become clear from the example output given below.

Sample Output Excerpts of WeibullPivots

```
$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  
...
```

```
$p.quantile.estimates  
0.001-quantile 0.005-quantile 0.01-quantile 0.025-quantile ...  
      259.9          593.8          848.3          1362.5 ...  
.....
```

Sample Output Excerpts of WeibullPivots

```
$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

```
....
```

```
$Tail.Probability.Estimates
```

p(6000)	p(7000)	p(8000)	p(9000)	p(10000)	p(11000)	p(12000)	...
0.36612	0.45977	0.55018	0.63402	0.70900	0.77385	0.82821	...

```
...
```

```
$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

```
....
```

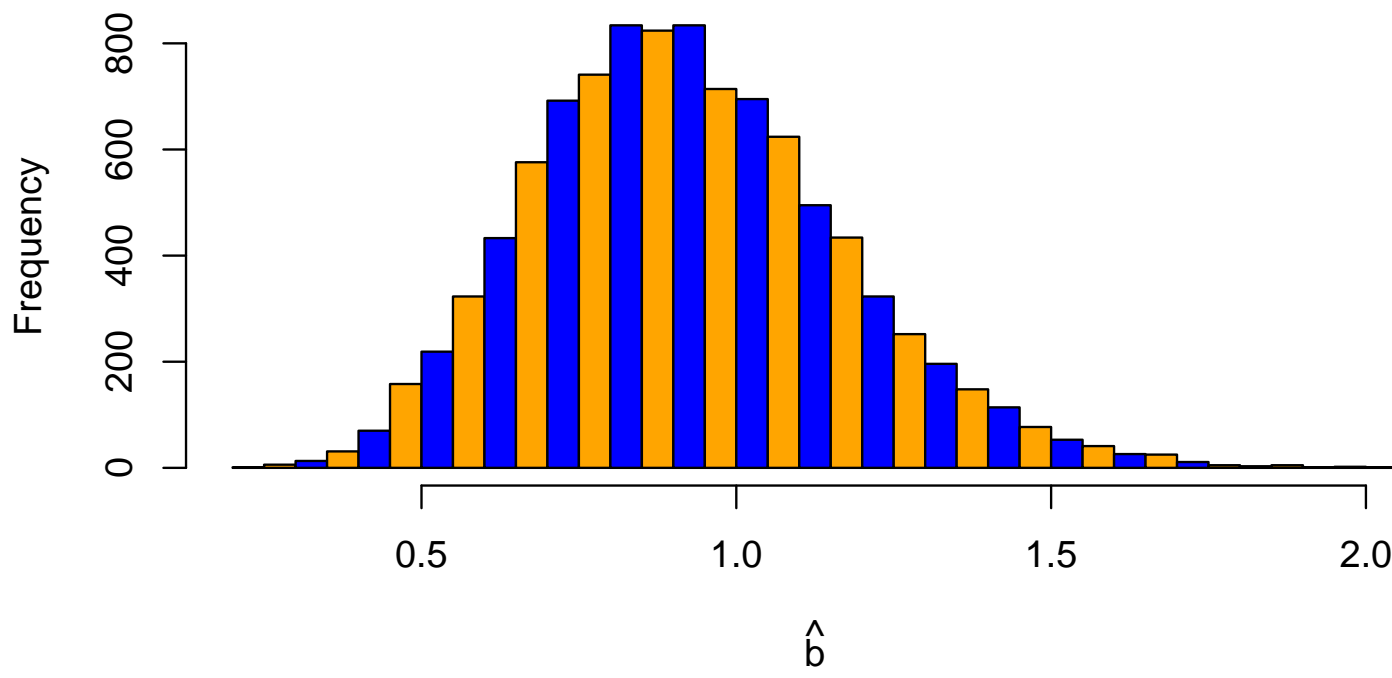
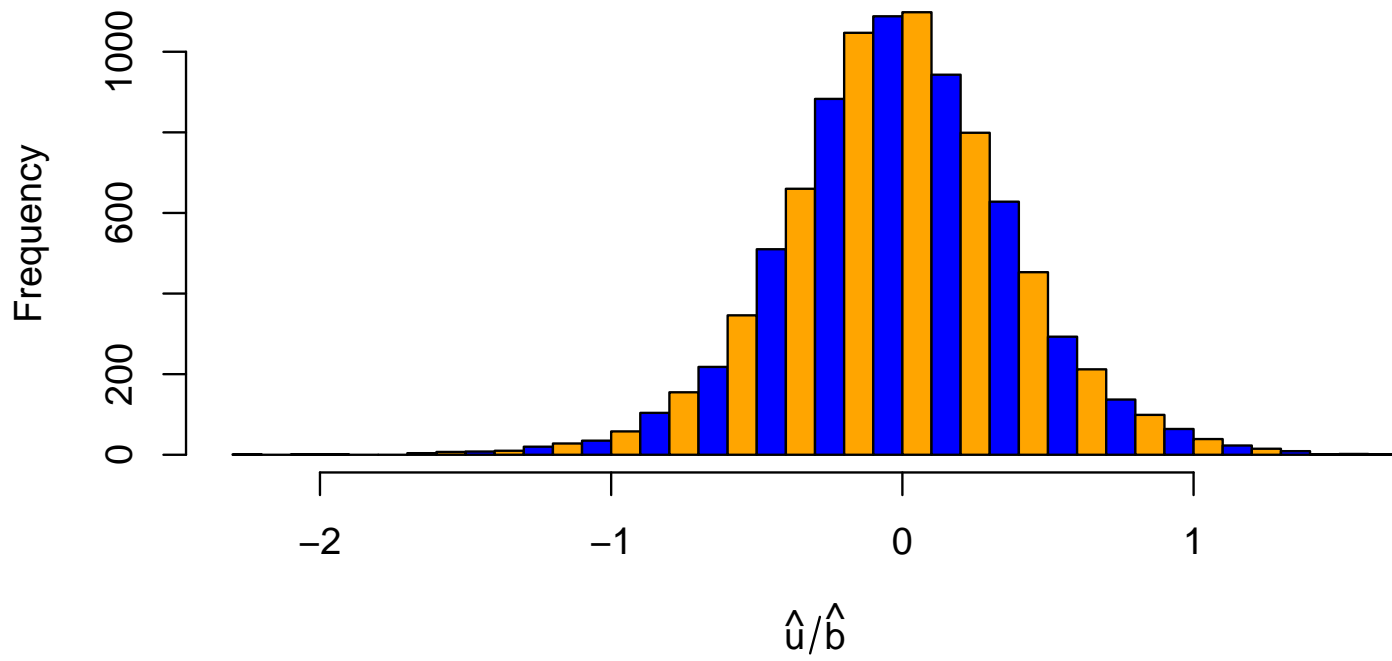
The Call that Produced Previous Output

```
WeibullPivots(threshold = seq(6000, 15000, 1000),  
              Nsim = 10000, graphics = T)
```

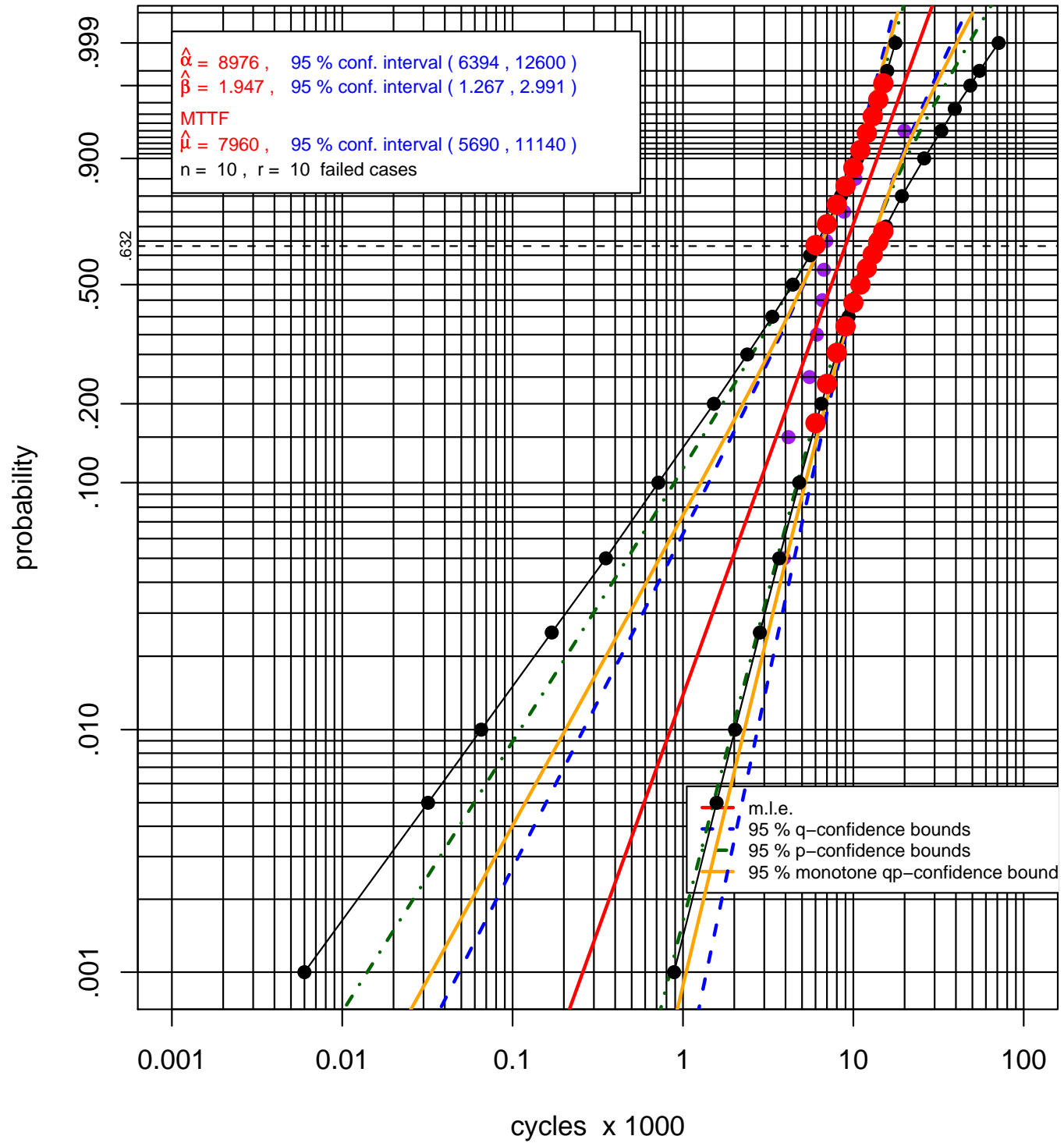
Because of `graphics=T` we also got 2 pieces of graphical output.

The first gives the two intrinsic pivot distributions of \hat{u}/\hat{b} and \hat{b} (next slide).

The second gives a Weibull plot of the generated sample with a variety of information and with several types of confidence bounds (slide after next).



Weibull Plot



Weibull Regression Models

Here the location parameter u_i of $\log(X_i) = Y_i$ can vary as follows in relation to known covariates $c_{i,j}$:

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

The regression coefficients $\zeta' = (\zeta_1, \dots, \zeta_k)$ and the common scale parameter b are unknown.

Thus we have the following model for Y_1, \dots, Y_n

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

with independent $Z_i \sim G(z) = 1 - \exp(-\exp(z))$, $i = 1, \dots, n$.

In terms of the Weibull random variables X_i the characteristic life $\alpha_i = \exp(u_i)$ is modeled in multiplicative dependence in relation to the covariates

$$\alpha_i = \exp(u_i) = \exp(\zeta_1 c_{i,1} + \dots + \zeta_k c_{i,k}) = \exp(\zeta_1 c_{i,1}) \times \dots \times \exp(\zeta_k c_{i,k})$$

MLE's

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

It can be shown that the mle's $\hat{\zeta}' = (\hat{\zeta}_1, \dots, \hat{\zeta}_k)$ and \hat{b} of ζ and b exist and are unique, provided the covariate matrix C , consisting of the rows $c'_i = (c_{i,1}, \dots, c_{i,k})$, $i = 1, \dots, n$, 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 slides.

WeibullReg

```
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 <- rweibull(n,beta,alpha*exp(slope*x))
# 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)
```


WeibullReg

```
out <- survreg(Surv(time)~x,data=dat,dist="weibull")
# The last two lines would give the same result as the
# next four 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)
```

WeibullReg

```
# Here out has several components, of which only
# out$coef and out$scale are 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  $u=\log(\alpha)$  and the regression slope.
out
}
```

Comments on WeibullReg

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.

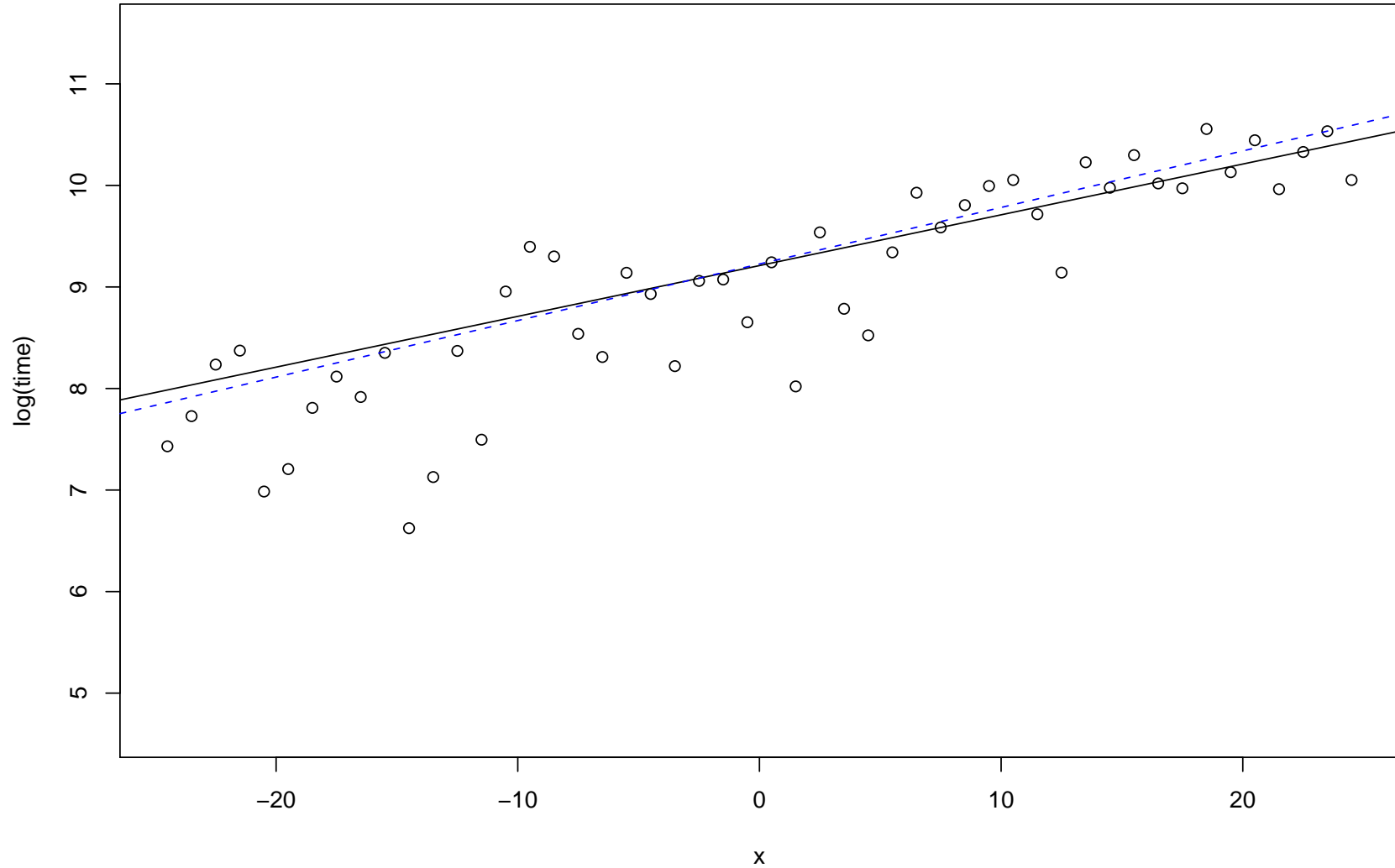
The rapidly varying plots give a good visual image of the sampling uncertainty and the resulting sampling variation of the fitted lines. The fixed line represents the true line with respect to which the Weibull data are generated by simulation.

The plotted log-Weibull data show a more transparent relationship in relation 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 ≈ 15 seconds, ≈ 9 seconds without plotting.

\implies reasonable behavior w.r.t. the anticipated computing times for the confidence bounds to be discussed below.

Example Plot from WeibullReg



Unequal point frequencies below & above line: $1 - e^{-1} = .632$ vs $e^{-1} = .368$

Equivariance Properties of MLE's

Existence and uniqueness of the mle's \implies the following equivariance properties

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

The proof is similar to the location/scale case. With $r_i = c'_i \mathbf{a} + \sigma z_i$ we have

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

The equivariance claim follows from the existence/uniqueness of the mle's.

Pivot Properties of MLE's

Equivariance properties $\implies (\hat{\zeta} - \zeta)/\hat{b}$ 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

$Y = C\zeta + bZ$, where $Z = (Z_1, \dots, Z_n)'$ with $Z_i \sim G(z) = 1 - \exp(-\exp(z))$ i.i.d.

From the equivariance property we have that

$$\hat{\zeta}(Y) = \zeta + b \hat{\zeta}(Z) \quad \text{and} \quad \hat{b}(Y) = b \hat{b}(Z) .$$

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

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 $(\hat{\zeta}(Z_1^*), \hat{b}_1(Z_1^*)), \dots, (\hat{\zeta}(Z_{N_{\text{sim}}}^*), \hat{b}(Z_{N_{\text{sim}}}^*))$ and thus the empirical

distribution of $(\hat{\zeta}(Z_1^*)/\hat{b}_1(Z_1^*), \hat{b}_1(Z_1^*)), \dots, (\hat{\zeta}(Z_{N_{\text{sim}}}^*)/\hat{b}(Z_{N_{\text{sim}}}^*), \hat{b}(Z_{N_{\text{sim}}}^*))$.

Confidence Bounds for $c_0'\zeta$

The distribution of $(c_0'\hat{\zeta}(Y) - c_0'\zeta)/\hat{b}(Y)$ is free of unknown parameters for any target covariate vector $c_0' = (c_{0,1}, \dots, c_{0,k})$, since

$$\frac{c_0'\hat{\zeta}(Y) - c_0'\zeta}{\hat{b}(Y)} = \frac{c_0'\hat{\zeta}(Z)}{\hat{b}(Z)}$$

We use the simulated values $(c_0'\hat{\zeta}(Z_i^*))/\hat{b}(Z_i^*)$, $i = 1, \dots, N_{\text{sim}}$, to approximate this parameter free distribution.

If $\hat{\eta}_2(\gamma, c_0)$ denotes the γ -quantile of this simulated distribution, then we can view $c_0'\hat{\zeta}(Y) - \hat{\eta}_2(\gamma, c_0)\hat{b}(Y)$ as an approximate 100 $\gamma\%$ lower bound for $c_0'\zeta$.

This is demonstrated as in the location/scale case for the location parameter u .

Here $c_0'\zeta$ is the log of the characteristic life at the covariate vector c_0 .

Confidence Bounds for b

Similarly, if $\hat{\eta}_1(\gamma)$ is the γ -quantile of the simulated $\hat{b}(Z_i^*)$, $i = 1, \dots, N_{\text{sim}}$, then we can view $\hat{b}(Y)/\hat{\eta}_1(\gamma)$ as approximate 100 γ % lower bound for b .

We note here that these quantiles $\hat{\eta}_1(\gamma)$ and $\hat{\eta}_2(\gamma, c_0)$ depend on the original covariate 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.

Confidence Bounds for $y_p(c_0)$ and $P(Y(c_0) \leq y_0)$

For a given covariate vector c_0 we can target the p -quantile $y_p(c_0) = c_0' \zeta + b w_p$ of the Y distribution with covariate dependent location parameter $u(c_0) = c_0' \zeta$ and scale parameter b .

Calculate $c_0' \hat{\zeta}(Y) - \hat{k}_p(\gamma) \hat{b}(Y)$ as an approximate 100 $\gamma\%$ lower bound for $y_p(c_0)$, where $\hat{k}_p(\gamma)$ is the γ -quantile of the simulated $(c_0' \hat{\zeta}(Z_i^*) - w_p) / \hat{b}(Z_i^*)$, $i = 1, \dots, N_{\text{sim}}$.

For the tail probability $p(y_0) = G((y_0 - c_0' \zeta) / b)$ with given threshold y_0 and covariate vector c_0 we obtain an approximate 100 $\gamma\%$ upper bound by using the γ -quantile of the simulated values

$$G(c_0' \hat{\zeta}(Z_i^*) - x \hat{b}(Z_i^*)), \quad i = 1, \dots, N_{\text{sim}},$$

where $x = (y_0 - c_0' \hat{\zeta}(y) / \hat{b}(y))$ and y is the originally observed sample vector, obtained under the covariate conditions specified through C .

The Simple Linear Regression Model

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

$$Y_i = \zeta_1 + \zeta_2 c_i + b Z_i, \quad i = 1, \dots, n \quad \text{with } Z_i \sim G(z).$$

In matrix notation this becomes

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

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 $i = 1, \dots, n$. The conditions for existence and uniqueness of the mle's are satisfied when the covariate values c_1, \dots, c_n are not all the same.

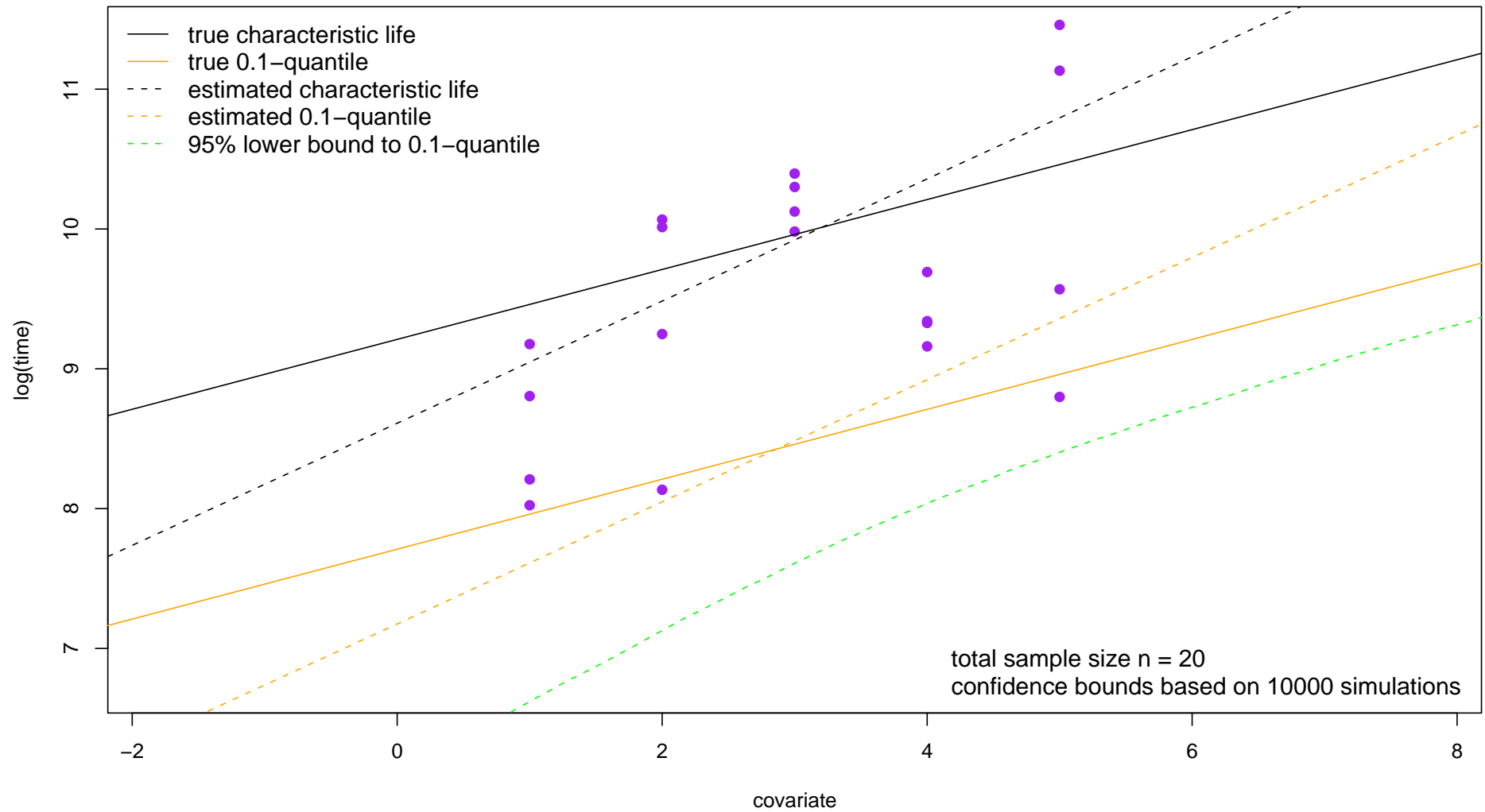
WeibullRegSim

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 following plots.

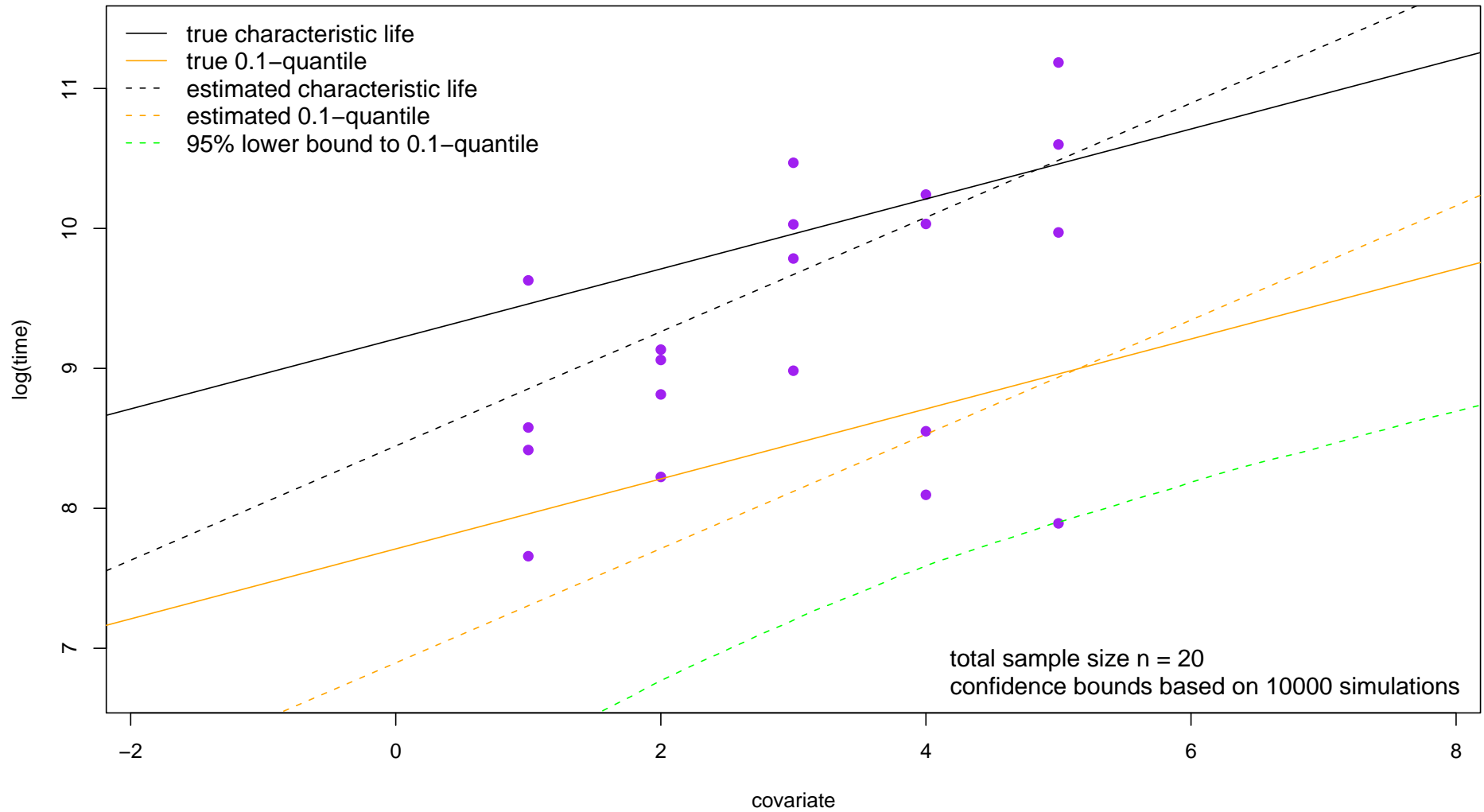
Each call generates its own data set of 20 points using 5 different levels of covariate values. The data are generated from a true Weibull distribution with a known true regression line relationship for $\log(\alpha)$ in relation to the covariates, as shown in the plots. Also shown in these plots is the .10-quantile line. Estimated lines are indicated by the corresponding color coded dashed lines.

We repeated this with $n = 50$ with elapsed time of 85 seconds per plot.

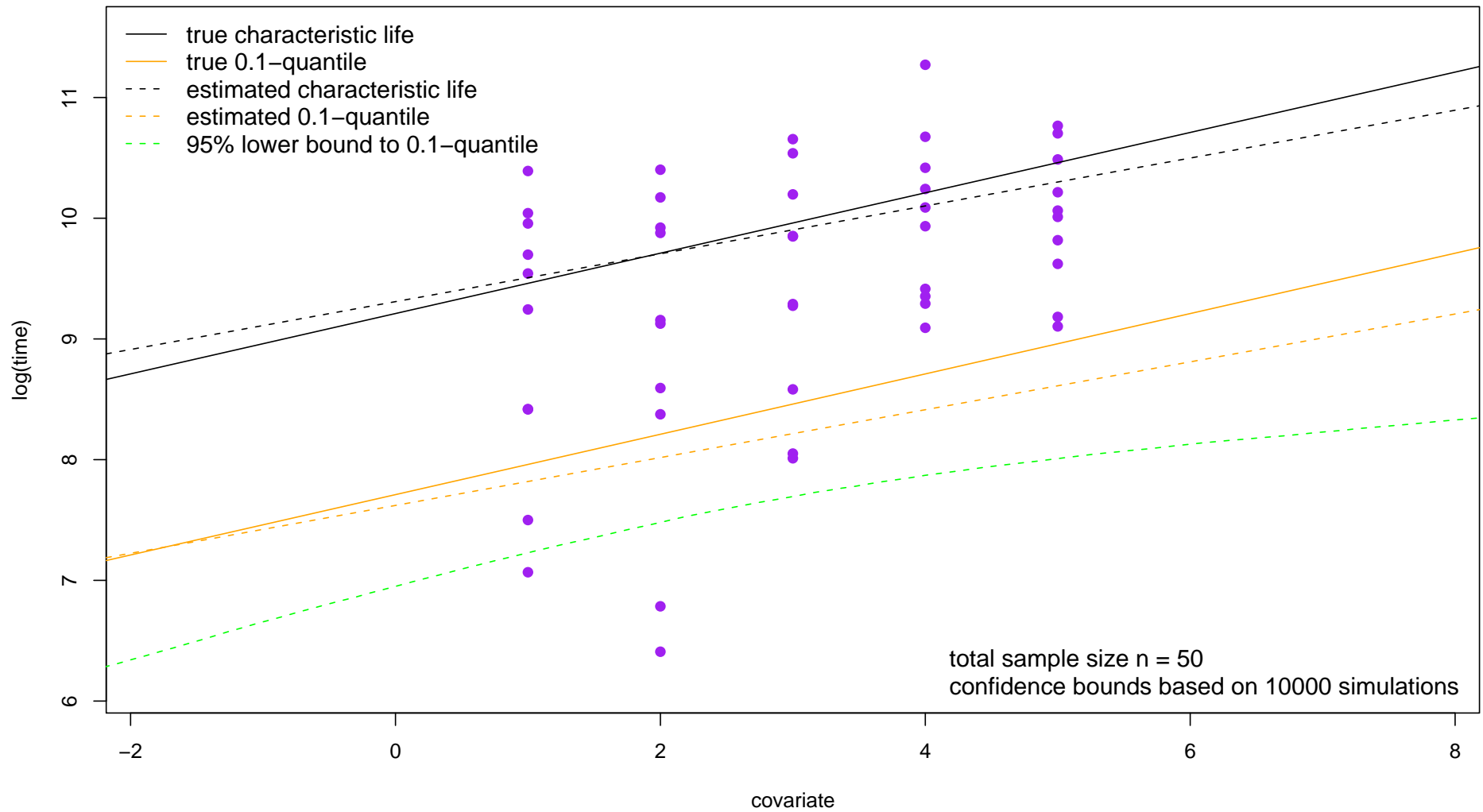
Regression Plot 1 ($n = 20$)



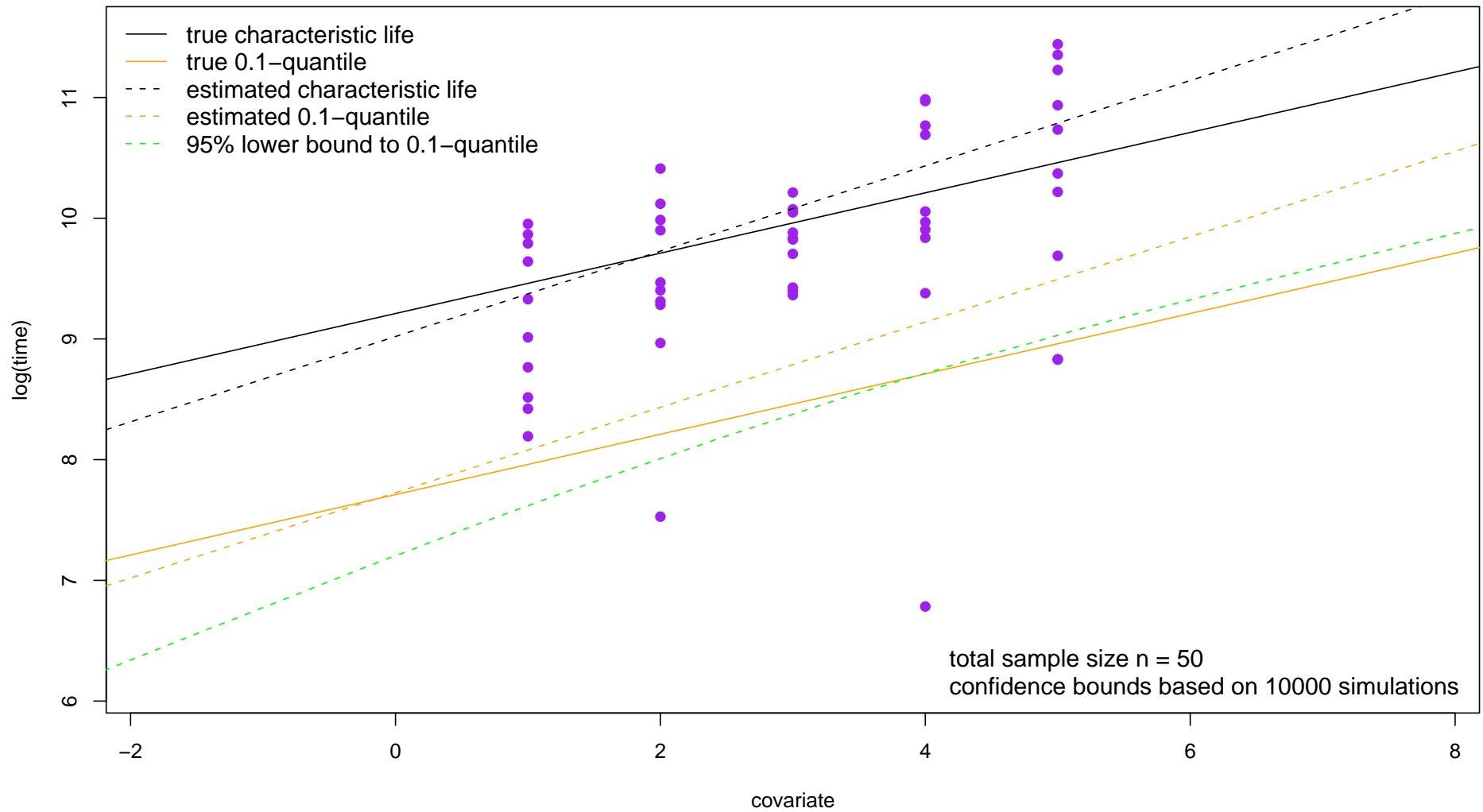
Regression Plot 2 ($n = 20$)



Regression Plot 3 ($n = 50$)



Regression Plot 4 ($n = 50$)



p -Quantile Lower Bound Curve

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, $\hat{\zeta}_1(Y) + \hat{\zeta}_2(Y)c - \hat{k}_p(\gamma)\hat{b}(Y)$, is the γ -quantile of the simulated values $(c'_0\hat{\zeta}(Z_i^*) - w_p)/\hat{b}(Z_i^*)$, $i = 1, \dots, N_{\text{sim}}$, and these values change depending on which $c'_0 = (1, c)$ is involved.

This curvature adjusts to some extent to the sampling variation induced swivel action in the fitted line.

The fitted line will deviate more strongly from the true line the farther away we are from the data scatter used in the mle fitting.

Note the crossing of the true quantile line by the confidence curve in the last plot.

The k -Sample Problem

Here we deal with $k = 3$ samples with same scale but possibly different locations.

The modifications for $k \neq 3$ should be obvious. In matrix notation this model is

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

Here the Y_i have location $u_1 = \zeta_1$ for the first n_1 observations, $u_2 = \zeta_1 + \zeta_2$ for the next n_2 observations and $u_3 = \zeta_1 + \zeta_3$ for the last n_3 observations.

Comments on Parameters

$u_1 = \zeta_1$ is 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 .

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 $H_0 : u_1 = u_2 = u_3$.

Testing $H_0 : \zeta_2 = \zeta_3 = 0$

Instead of using the likelihood ratio test, which invokes the $\chi_{k-1}^2 = \chi_2^2$ 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.

Which test statistic?

The Test Statistic

The formal definition of the test statistic proposed by Lawless is as follows:

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

where C_{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

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

where

$$\hat{u}_1(Y) = \hat{\zeta}_1(Y) , \quad \hat{u}_2(Y) = \hat{\zeta}_1(Y) + \hat{\zeta}_2(Y) , \quad \hat{u}_3(Y) = \hat{\zeta}_1(Y) + \hat{\zeta}_3(Y)$$

$$\text{and} \quad \hat{u}(Y) = \sum_{i=1}^3 \frac{n_i}{N} \hat{u}_i(Y) , \quad \text{with } N = n_1 + n_1 + n_3 .$$

The Normal Case Parallel

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 $\hat{u}_i(Y) = \bar{Y}_{i\cdot}$, $i = 1, 2, 3$ and $\hat{u}(Y) = \bar{Y}_{\cdot\cdot} = (n_1/N)\bar{Y}_{1\cdot} + (n_2/N)\bar{Y}_{2\cdot} + (n_3/N)\bar{Y}_{3\cdot}$ and

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

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 $\hat{b}(Y)^2$ is not ignored by simply referring to the χ_{k-1}^2 distribution, using a large sample argument.

The Null Distribution of Λ_1

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

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

have distributions free of unknown parameters.

Under the hypothesis H_0 when $u_1 = u_2 = u_3 (= u)$ we thus have that

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

have distributions free of any unknown parameters which in turn implies the above claim about Λ_1 .

Simulating the Null Distribution of Λ_1

Thus we can estimate the null distribution of Λ_1 by using the N_{sim} simulated values of $\hat{\zeta}_i(Z_j^*)/\hat{b}(Z_j^*)$ to create

$$\frac{\hat{u}_1(Z_j^*)}{\hat{b}(Z_j^*)} = \frac{\hat{\zeta}_1(Z_j^*)}{\hat{b}(Z_j^*)}, \quad \frac{\hat{u}_i(Z_j^*)}{\hat{b}(Z_j^*)} = \frac{\hat{\zeta}_1(Z_j^*) + \hat{\zeta}_i(Z_j^*)}{\hat{b}(Z_j^*)}, \quad i = 2, 3$$

$$\text{and} \quad \frac{\hat{u}(Z_j^*)}{\hat{b}(Z_j^*)} = \frac{\sum_{i=1}^3 n_i \hat{u}_i(Z_j^*)/N}{\hat{b}(Z_j^*)}$$

and thus

$$\Lambda_1(Z_j^*) = \frac{\sum_{i=1}^3 n_i (\hat{u}_i(Z_j^*) - \hat{u}(Z_j^*))^2}{\hat{b}(Z_j^*)^2} \quad j = 1, \dots, N_{\text{sim}}.$$

The distribution of these N_{sim} values $\Lambda_1(Z_j^*)$ 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 Λ_1 Null Distribution and its χ_2^2 Approximation

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.

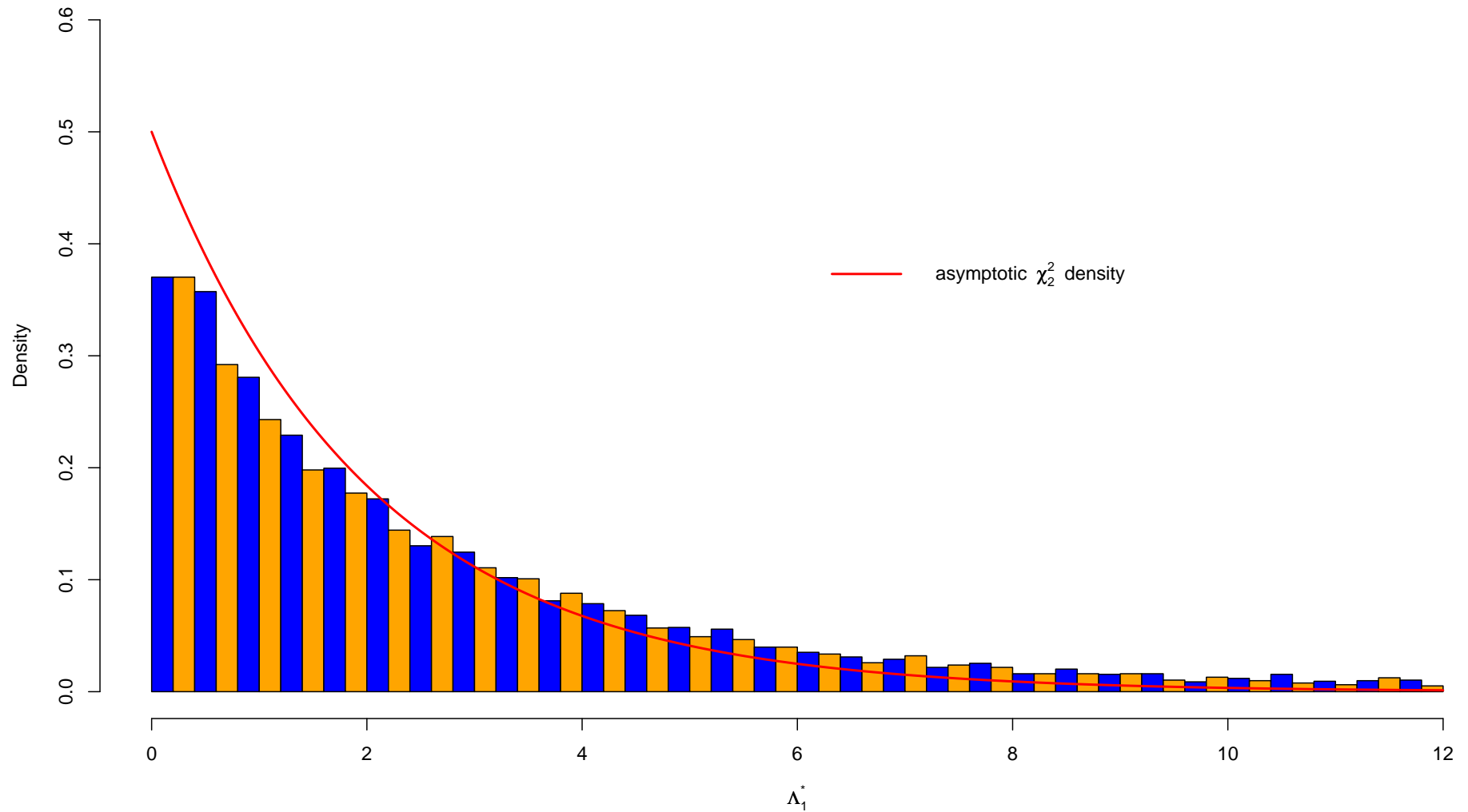
The histogram (next slide) is based on $N_{\text{sim}} = 10000$ simulated values of $\Lambda_1(Z^*)$.

The χ_2^2 density is similar in character but there are severe differences.

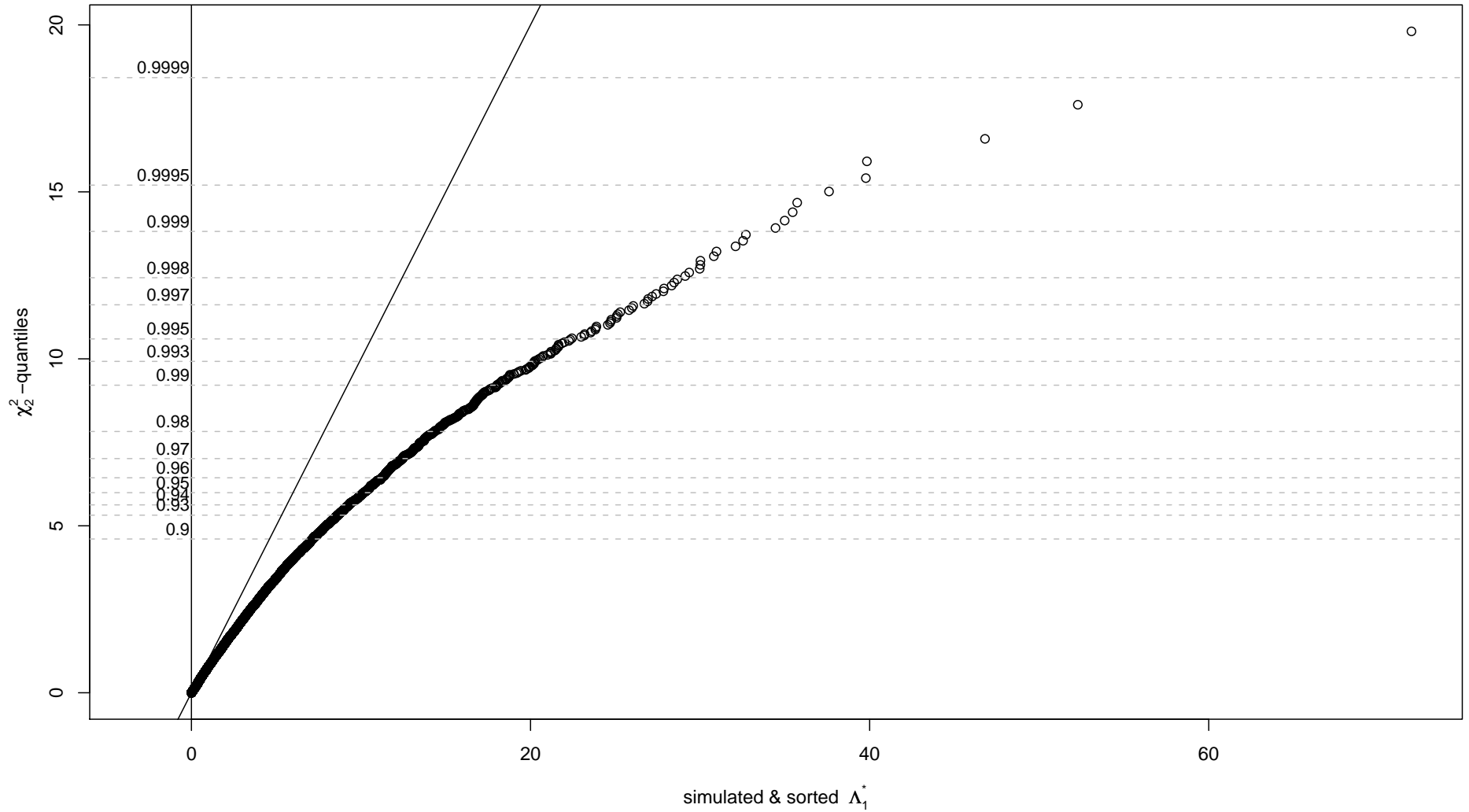
Using the χ_2^2 distribution would result in much smaller p -values than appropriate when these are on the low side.

We would strongly overstate the significance of some test results when using χ_2^2 .

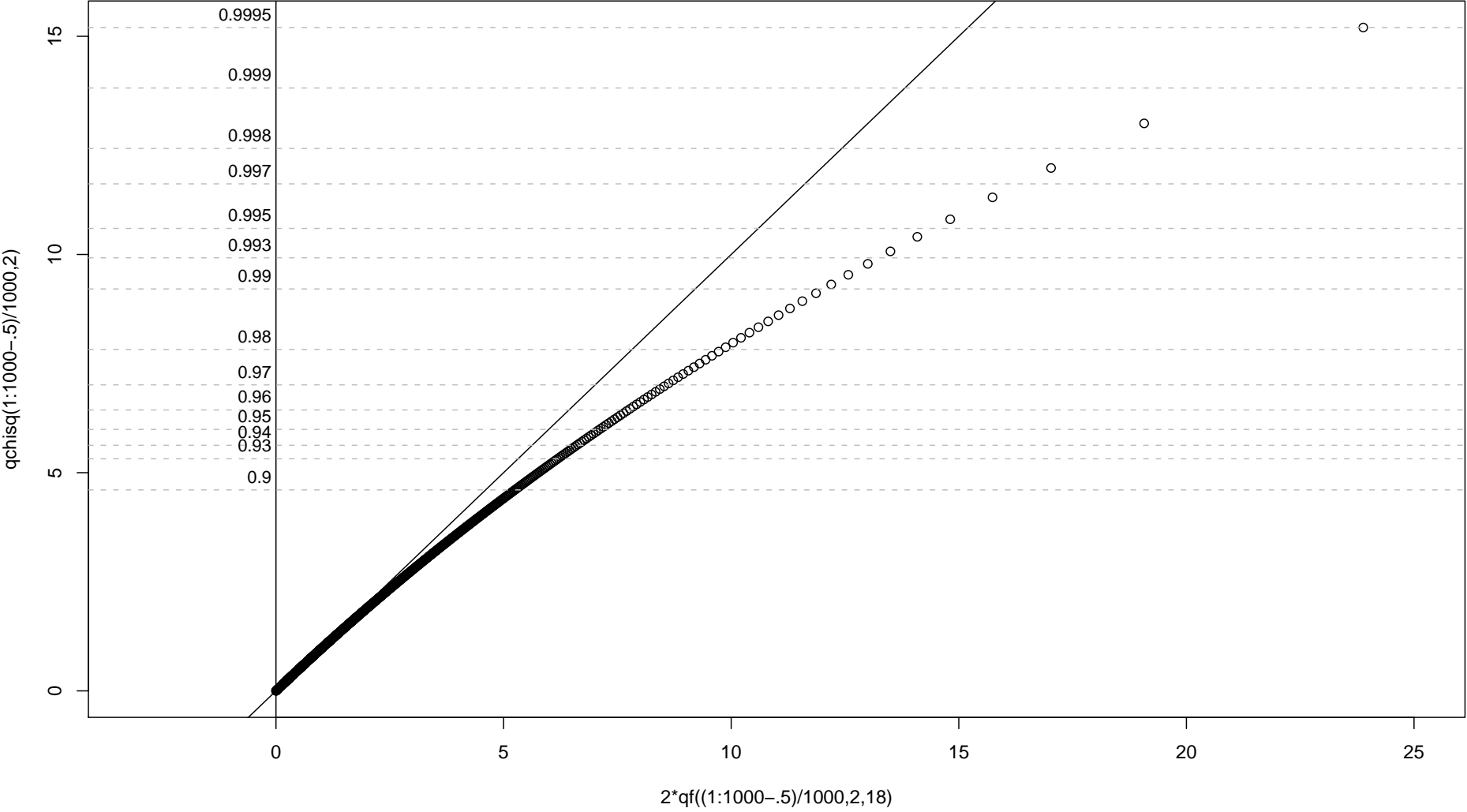
The Λ_1 Null Distribution and its χ_2^2 Approximation



QQ-Plot of Λ_1 Null Distribution and χ_2^2



QQ-Plot of Normal Theory F Distribution and its χ^2_2 Approximation



Comments on QQ-Plots

In the QQ-plots each point on the curve corresponds to a particular value of p .

The QQ-curve bends away from the main diagonal to the right.

This indicates that the p -quantile for the abscissa distribution is further out than the p -quantile of the χ_2^2 ordinate distribution.

Pick an abscissa value x_0 , move up to the quantile curve and read off its probability value p on the χ_2^2 ordinate probability scale. $\implies x_0 = x_p$.

To see what probability value the approximating χ_2^2 distribution associates with x_0 we just go up to the main diagonal from the x_0 abscissa value and then read off the p from the χ_2^2 ordinate probability scale to the left.

For Λ_1^* a .95-quantile becomes a .993-quantile on the χ_2^2 scale, i.e., a p -value of .05 gets falsely turned into a much more significant p -value of .007.

Goodness of Fit Tests for Weibull Regression Data

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

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

Pierce and Kopecky (1979) showed that the asymptotic null distributions of D , W^2 and A^2 , using the sorted values $V_{(1)} \leq \dots \leq V_{(n)}$ 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.

Goodness of Fit by Simulation

However, for finite sample sizes the effects of these covariates may still be relevant.

The effect of using the small sample tables (location/scale) 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 $Y_i = c'_i \zeta + bZ_i$ with i.i.d. $Z_i \sim G(z)$, or $Y = C\zeta + bZ$ this is seen from the equivariance properties as follows

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

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

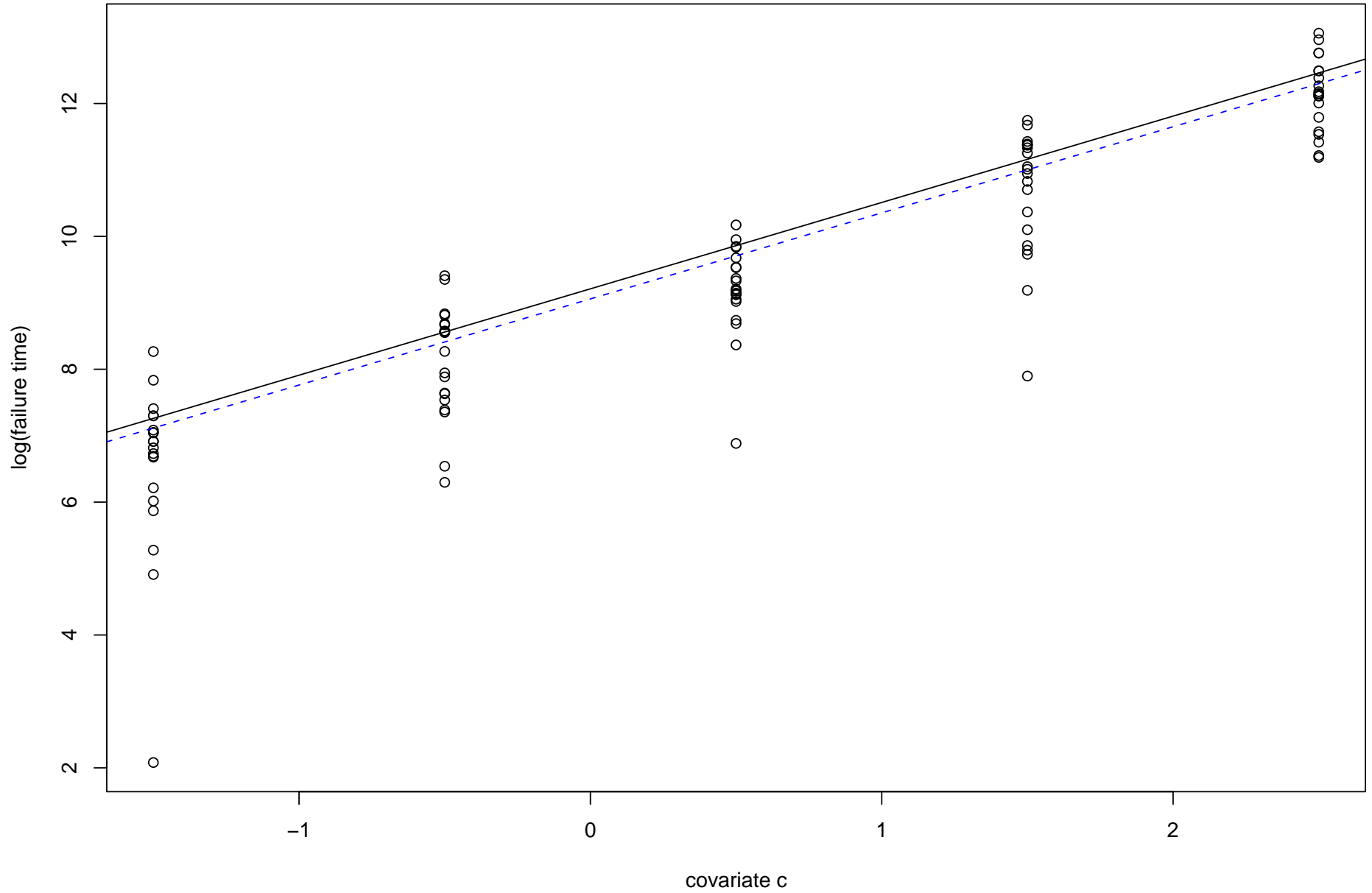
The Null Distributions of D , W^2 and A^2

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 $Z^* = (Z_1, \dots, Z_n)'$ i.i.d. $\sim G(z)$, compute the mle's $\hat{\zeta}(Z)$, $\hat{b}(Z)$, and from that $V^* = (V_1^*, \dots, V_n^*)'$, followed by $D^* = D(V^*)$, $W^{2*} = W^2(V^*)$ and $A^{2*} = A^2(V^*)$.

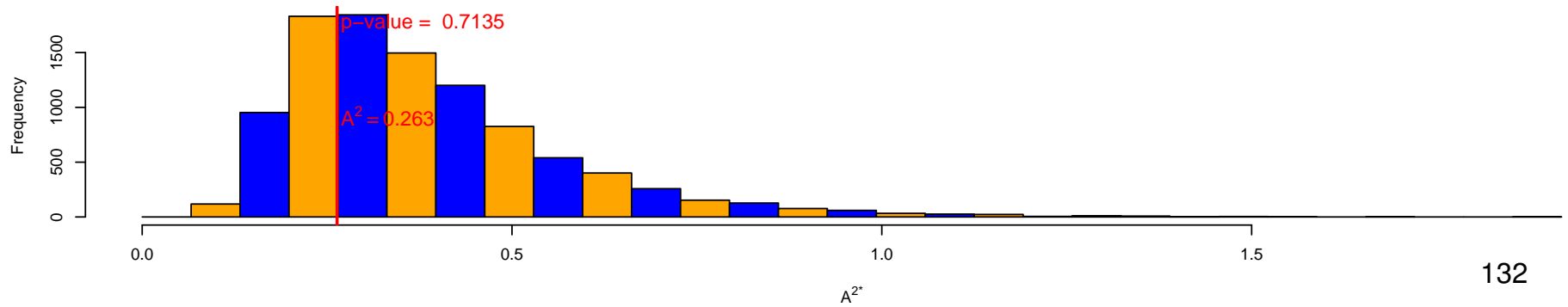
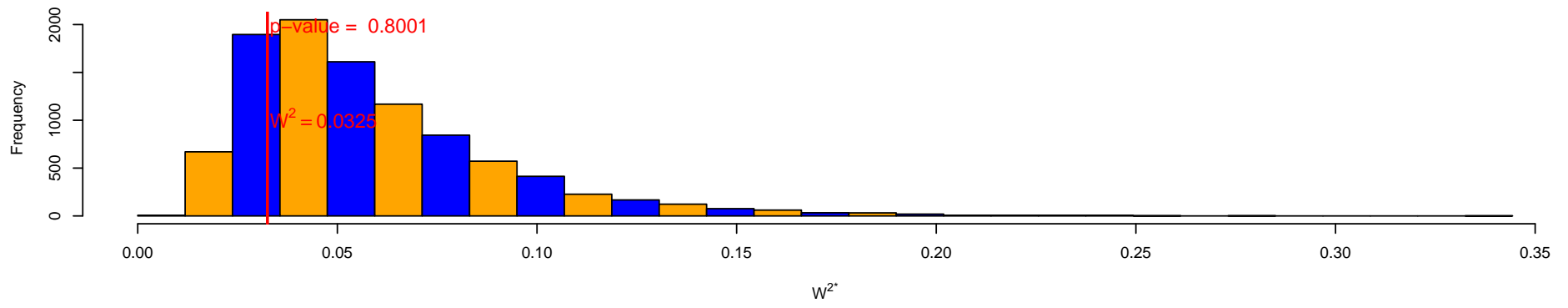
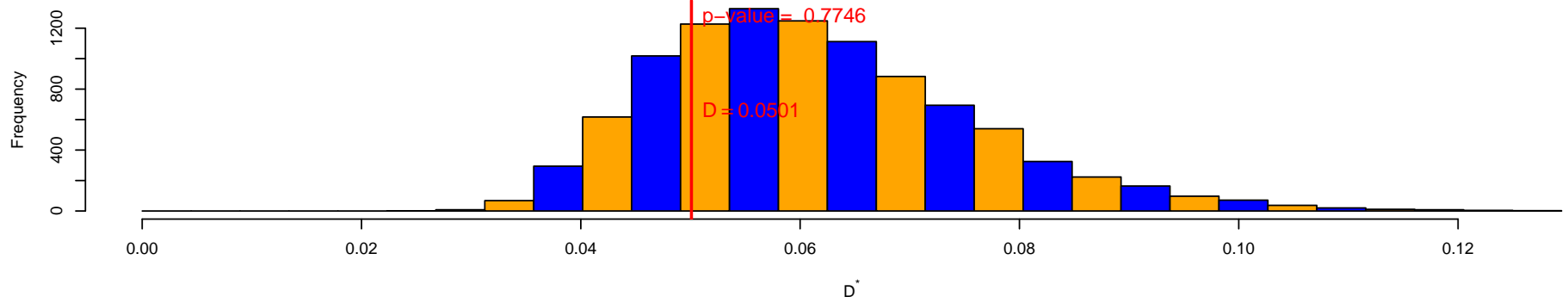
Repeating this a large number of times, say $N_{\text{sim}} = 10000$, would yield values $D_i^*, W_i^{2*}, A_i^{2*}$, $i = 1, \dots, N_{\text{sim}}$. Their respective empirical distributions would serve as excellent approximations to the desired null distributions of these test of fit criteria.

Weibull Regression Data



Goodness of Fit Tests for Weibull Regression Data

$n = 100$, $N_{sim} = 10000$



Comments on the Weibull Regression Example

The solid sloped line in the Weibull regression data plot indicates the true $\log(\text{characteristic life})$ while the dashed line represents its estimate.

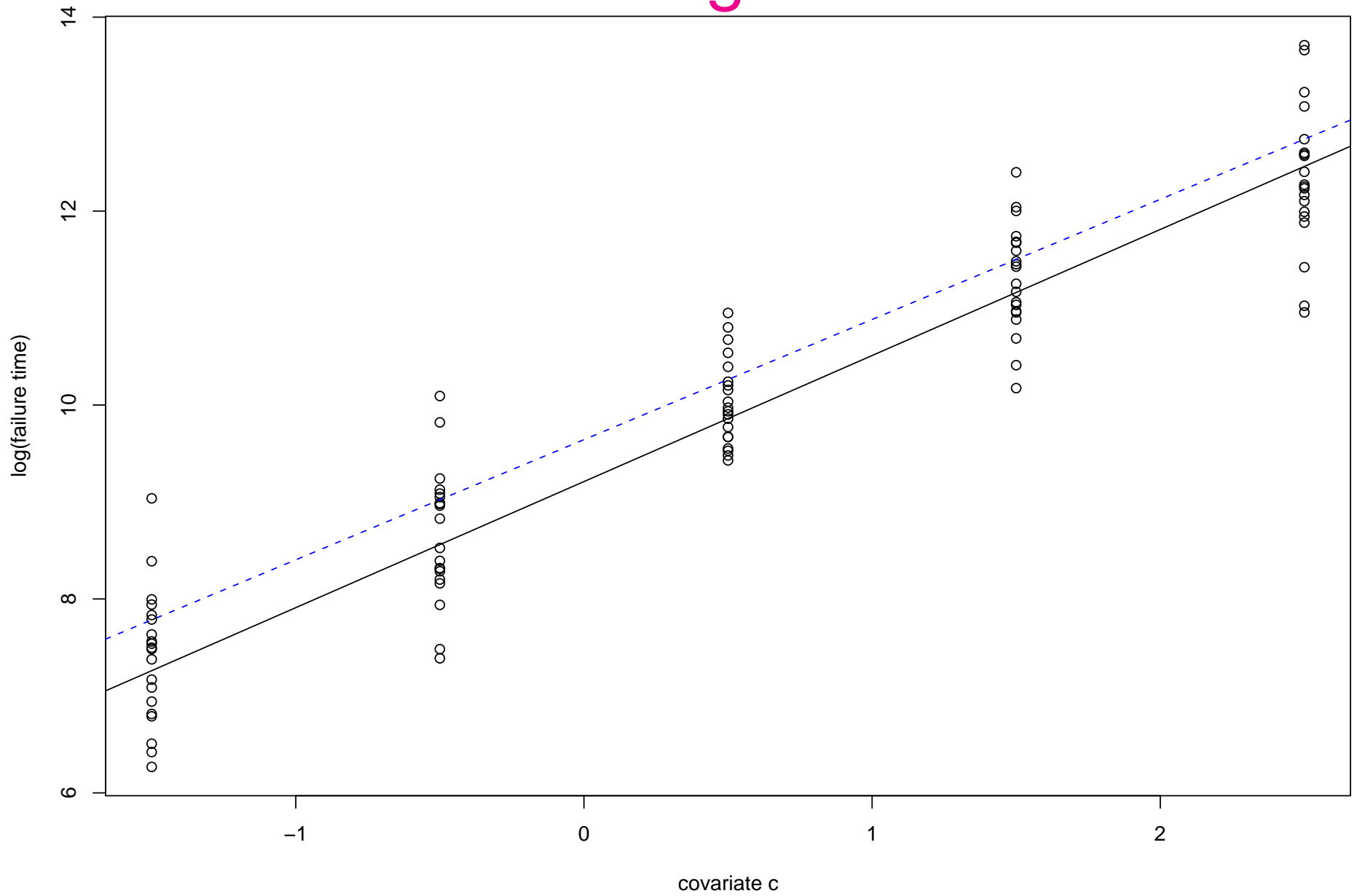
The histograms on the previous slide show the results of the Weibull goodness of fit tests in relation to the simulated null distributions of D , W^2 and A^2 .

Based on the shown p -values the hypothesis of a Weibull lifetime distribution cannot be rejected by any of the three test of fit criteria.

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.

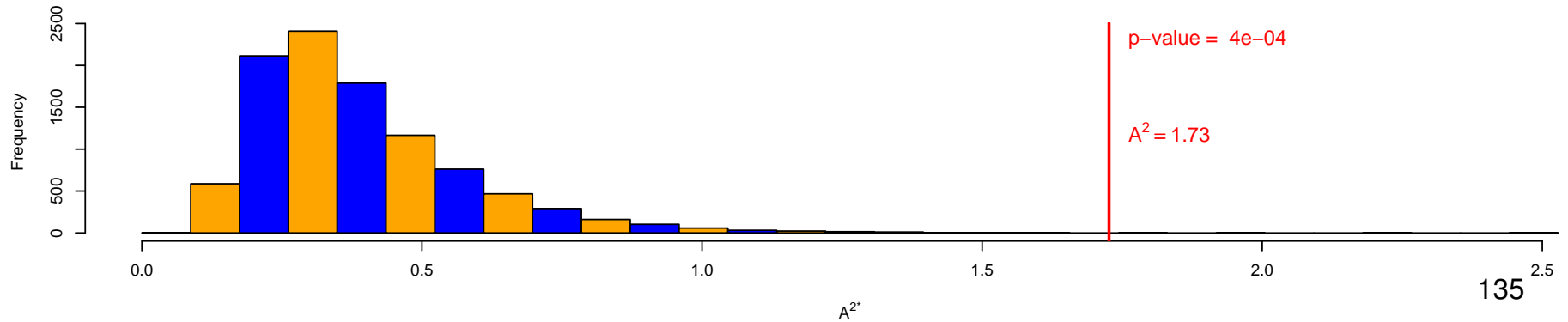
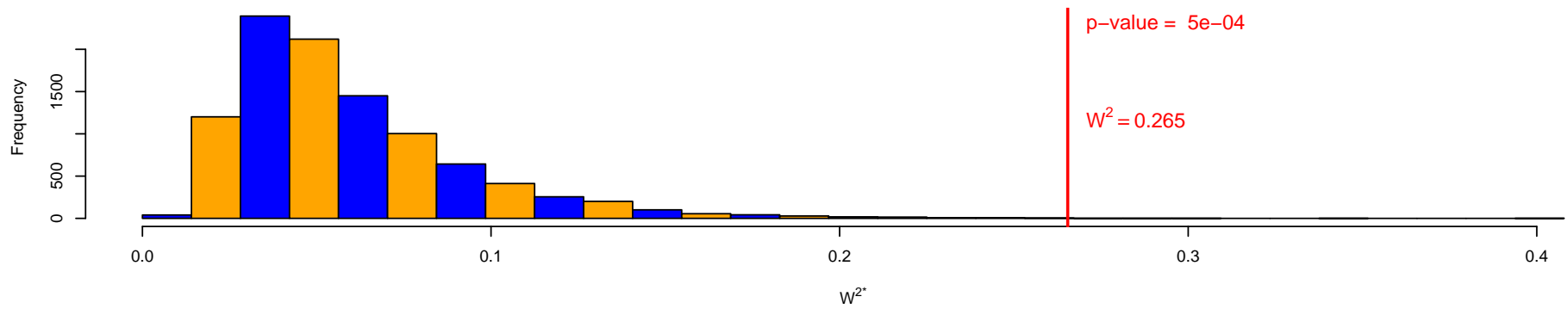
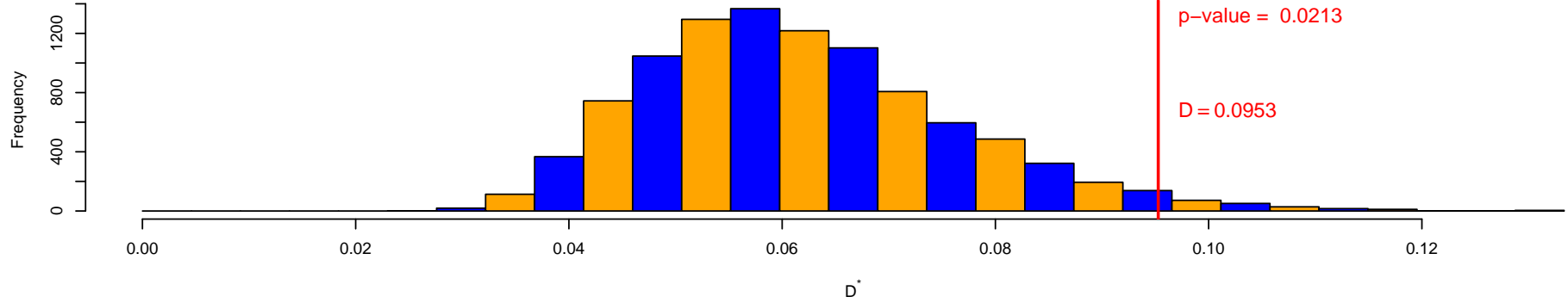
This function performs Weibull goodness of fit tests for any supplied regression data set. When this data set is missing it generates its own Weibull regression data set.

Normal Regression Data



Goodness of Fit Tests for Normal Regression Data

$n = 100$, $N_{sim} = 10000$



Comments on the Normal Regression Example

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 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)`.

The simulation of the test of fit null distributions remains essentially unchanged except that a different random number starting seed was used.

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 previous case.

The reason for this wider gap is that the fitted line aims for the .632-quantile and not the mean/median of that data set.

Comments on the Normal Regression Example (contd)

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 of a Weibull distribution.

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.

Finis