Chapter 8

Maximum Likelihood for Location-Scale Based Distributions

William Q. Meeker and Luis A. Escobar
Iowa State University and Louisiana State University

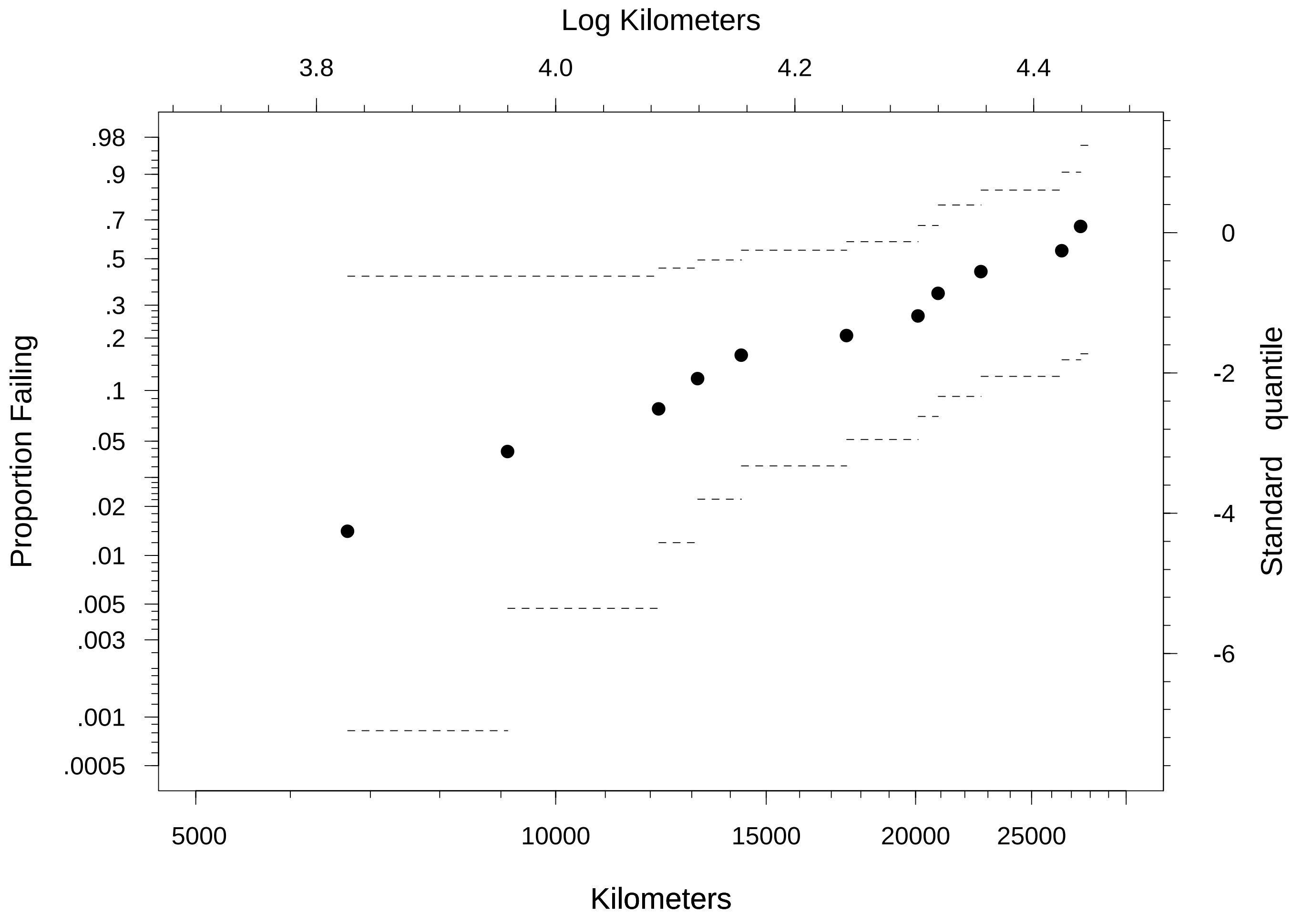
Based on the authors' text *Statistical Methods for Reliability Data*, John Wiley & Sons Inc. 1998.

July 18, 2002
12h 25min

Chapter 8
Maximum Likelihood
for Location-Scale Based Distributions
Objectives

- Illustrate likelihood-based methods for parametric models based on log-location-scale distributions (especially Weibull and Lognormal).
- Construct and interpret likelihood-ratio-based confidence intervals/regions for model parameters and for **functions** of model parameters.
- Construct and interpret normal-approximation confidence intervals/regions.
- Describe the advantages and pitfalls of assuming that the log-location-scale distribution shape parameter is known.

Weibull Probability Plot of the Shock Absorber Data



Weibull Distribution Likelihood for Right Censored Data

- The Weibull model is

$$\Pr(T \le t) = F(t; \mu, \sigma) = \Phi_{\text{sev}}\left\{[\log(t) - \mu]/\sigma\right\}.$$

- The likelihood has the form

$$\begin{aligned}
L(\mu, \sigma) &= \prod_{i=1}^{n} L_i(\mu, \sigma; \text{data}_i) \\
&= \prod_{i=1}^{n} \left[f(t_i; \mu, \sigma)\right]^{\delta_i} \left[1 - F(t_i; \mu, \sigma)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{1}{\sigma t_i} \phi_{\text{sev}}\left(\frac{\log(t_i) - \mu}{\sigma}\right)\right]^{\delta_i} \times \left[1 - \Phi_{\text{sev}}\left(\frac{\log(t_i) - \mu}{\sigma}\right)\right]^{1-\delta_i}
\end{aligned}$$

$$\delta_i = \begin{cases} 1 & \text{if } t_i \text{ is an exact observation} \\ 0 & \text{if } t_i \text{ is a right censored observation} \end{cases}$$

$\phi_{\text{sev}}(z)$ is the standardized smallest extreme value density.

Lognormal Distribution Likelihood for Right Censored Data

- The lognormal model is

$$\Pr(T \le t) = F(t;\mu,\sigma) = \Phi_{\text{nor}}\left\{[\log(t) - \mu]/\sigma\right\}.$$

- The likelihood has the form

$$\begin{aligned}
L(\mu,\sigma) &= \prod_{i=1}^{n} L_i(\mu,\sigma;\text{data}_i) \\
&= \prod_{i=1}^{n} \left[f(t_i;\mu,\sigma)\right]^{\delta_i}\left[1 - F(t_i;\mu,\sigma)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{1}{\sigma t_i}\phi_{\text{nor}}\left(\frac{\log(t_i)-\mu}{\sigma}\right)\right]^{\delta_i} \times \left[1 - \Phi_{\text{nor}}\left(\frac{\log(t_i)-\mu}{\sigma}\right)\right]^{1-\delta_i}
\end{aligned}$$

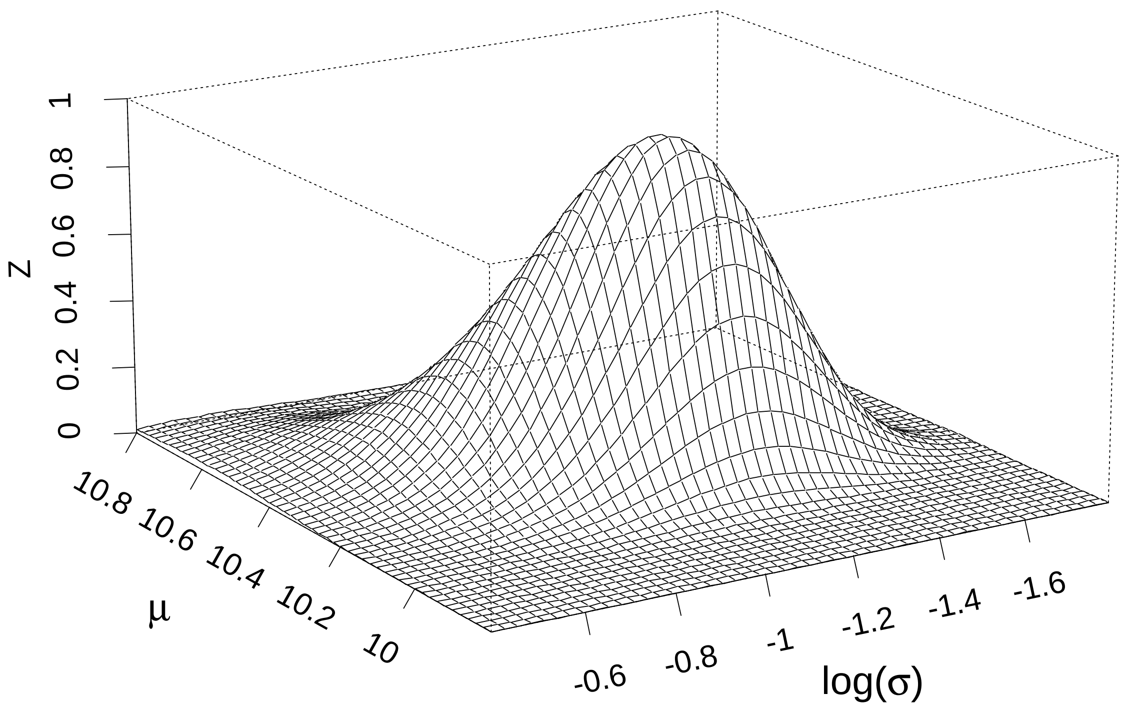
$$\delta_i = \begin{cases} 1 & \text{if } t_i \text{ is an exact observation} \\ 0 & \text{if } t_i \text{ is a right censored observation} \end{cases}$$

$\phi_{\text{nor}}(z)$ is the standardized normal density.

Weibull Relative Likelihood for the Shock Absorber Data

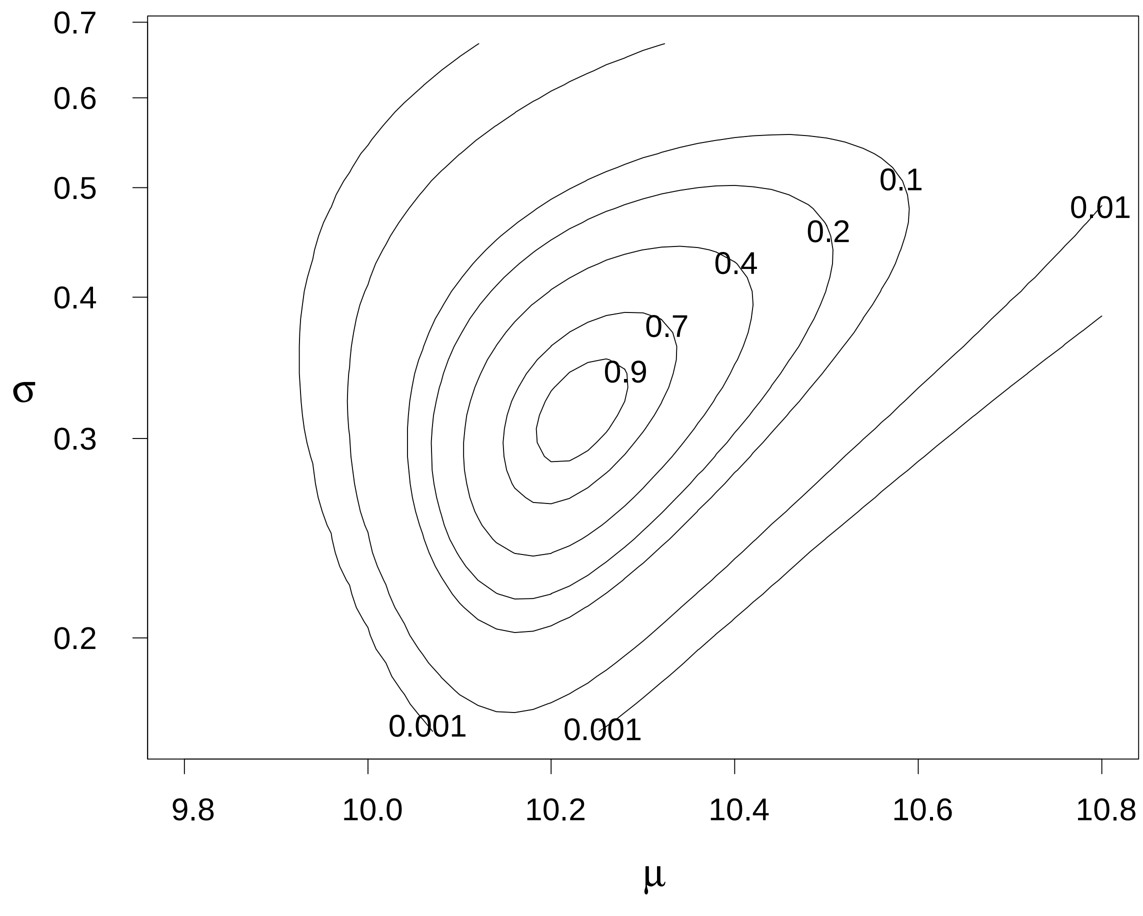
ML Estimates: $\widehat{\mu} = 10.23$ and $\widehat{\sigma} = .3164$

$$R(\mu, \log(\sigma)) = L(\mu, \log(\sigma))/L(\widehat{\mu}, \log(\widehat{\sigma}))$$

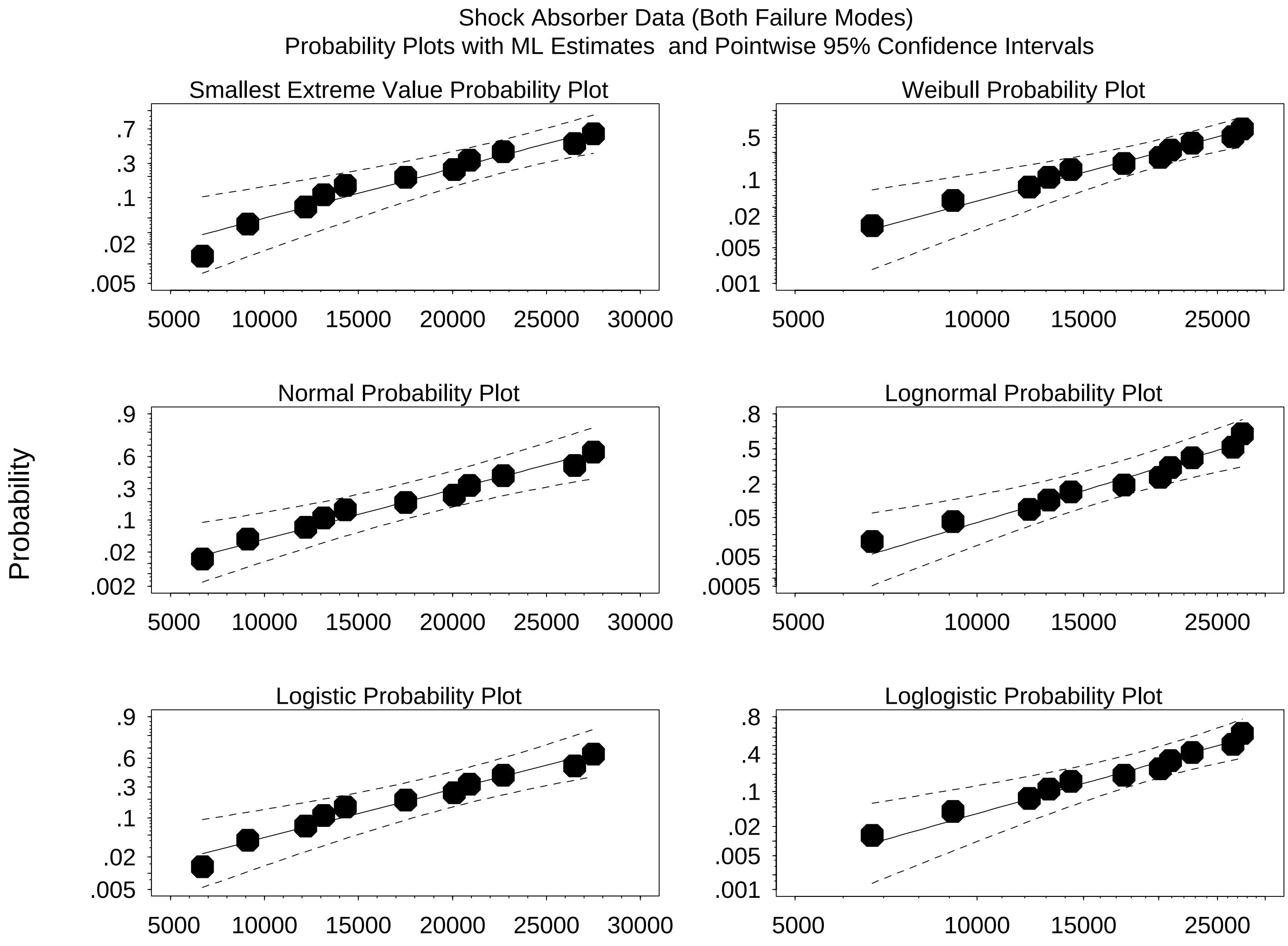


Weibull Relative Likelihood for the Shock Absorber Data ML Estimates: $\widehat{\mu} = 10.23$ and $\widehat{\sigma} = .3164$

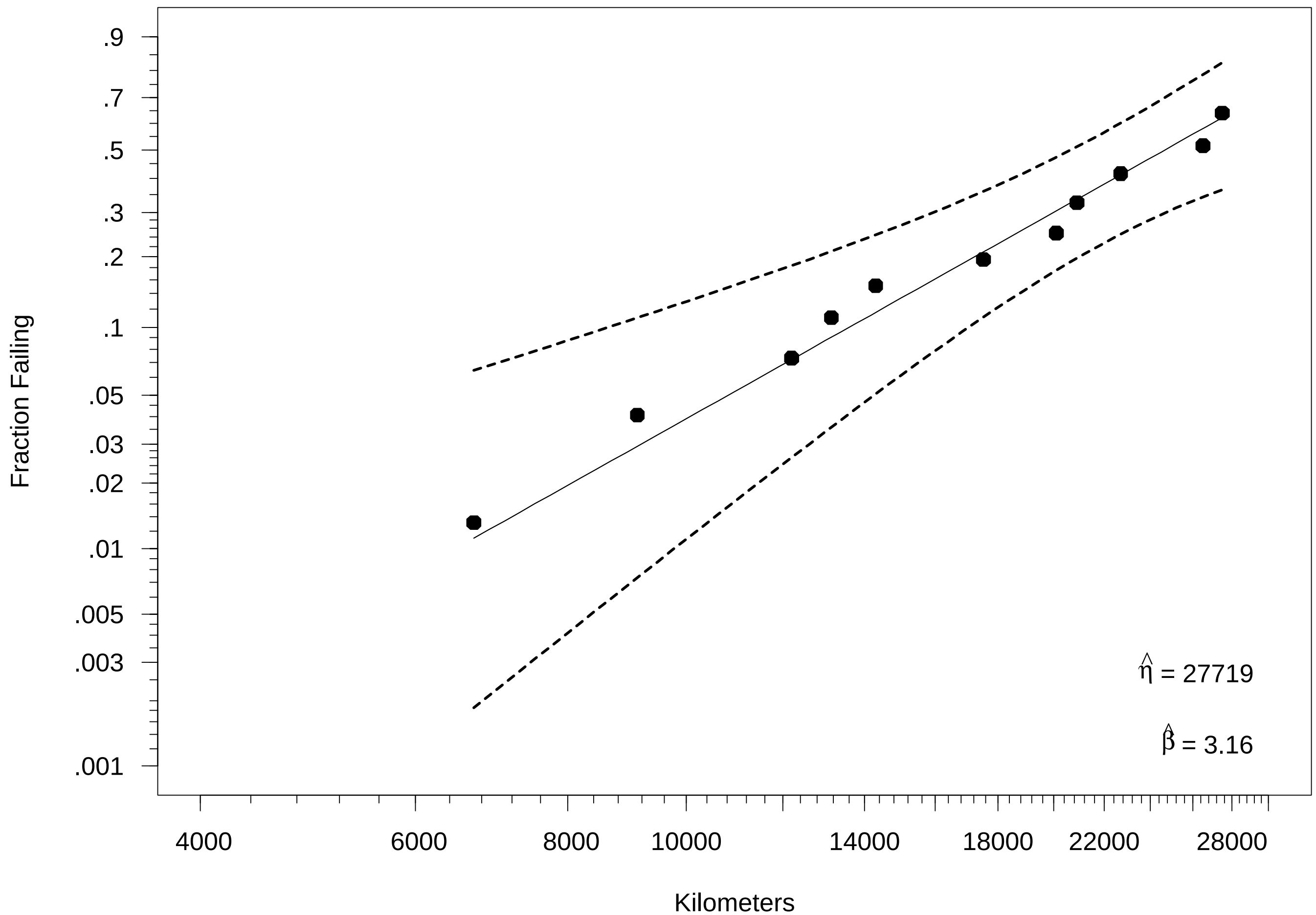
$$R(\mu, \sigma) = L(\mu, \sigma)/L(\widehat{\mu}, \widehat{\sigma})$$



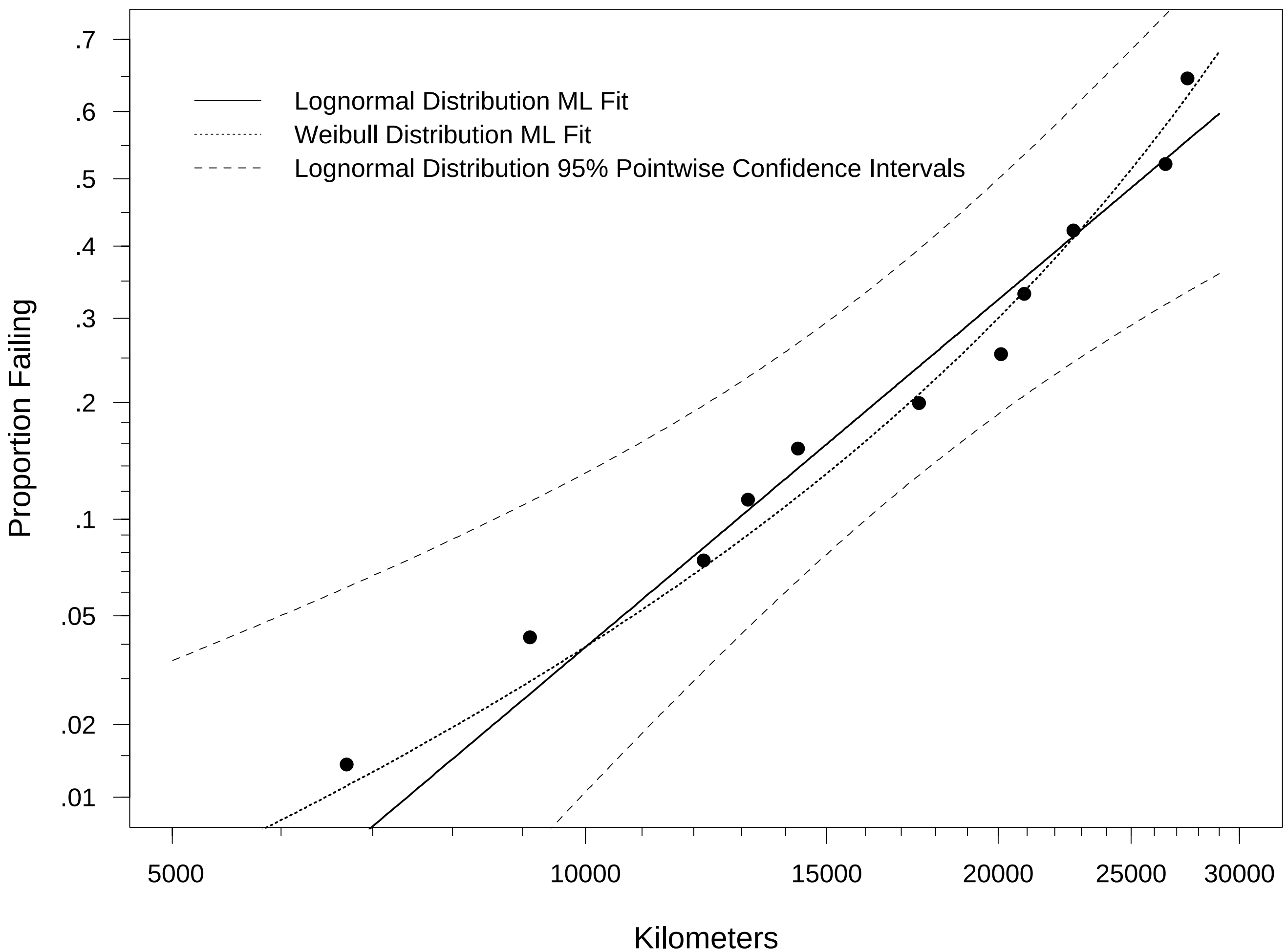
Six-Distribution ML Probability Plot of the Shock Absorber Data



Weibull Probability Plot of Shock Absorber Failure Times (Both Failure Modes) with Maximum Likelihood Estimates and Normal-Approximation 95% Pointwise Confidence Intervals for $F(t)$

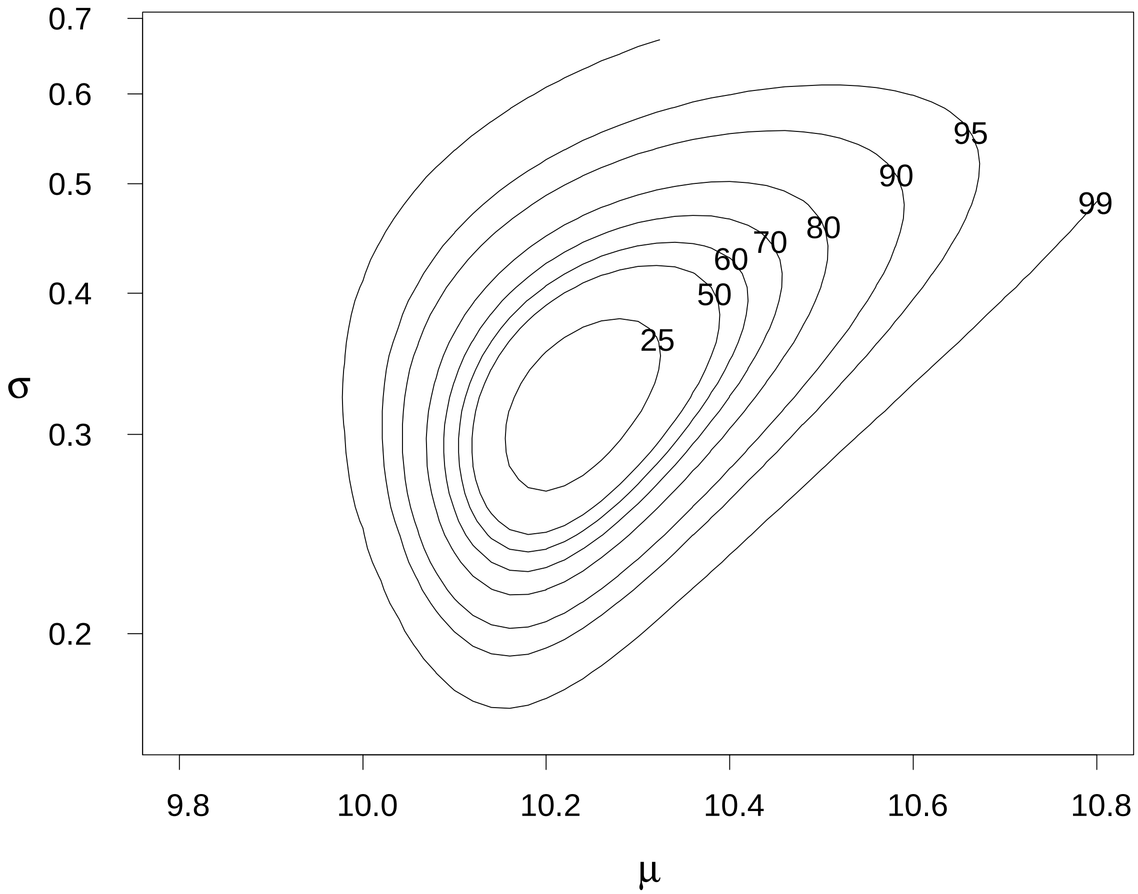


Lognormal Probability Plots of Shock Absorber Data with ML Estimates and Normal-Approximation 95% Pointwise Confidence Intervals for $F(t)$. The Curved Line is the Weibull ML Estimate.



Weibull Likelihood-Based Joint Confidence Regions for μ and σ for the Shock Absorber Data

$$R(\mu, \sigma) > \exp\left[-\chi^2_{(1-\alpha;2)}/2\right] = 100\alpha\%$$



Large-Sample Approximate Theory for Likelihood Ratios for Parameter Vector

- Relative likelihood for (μ, σ) is

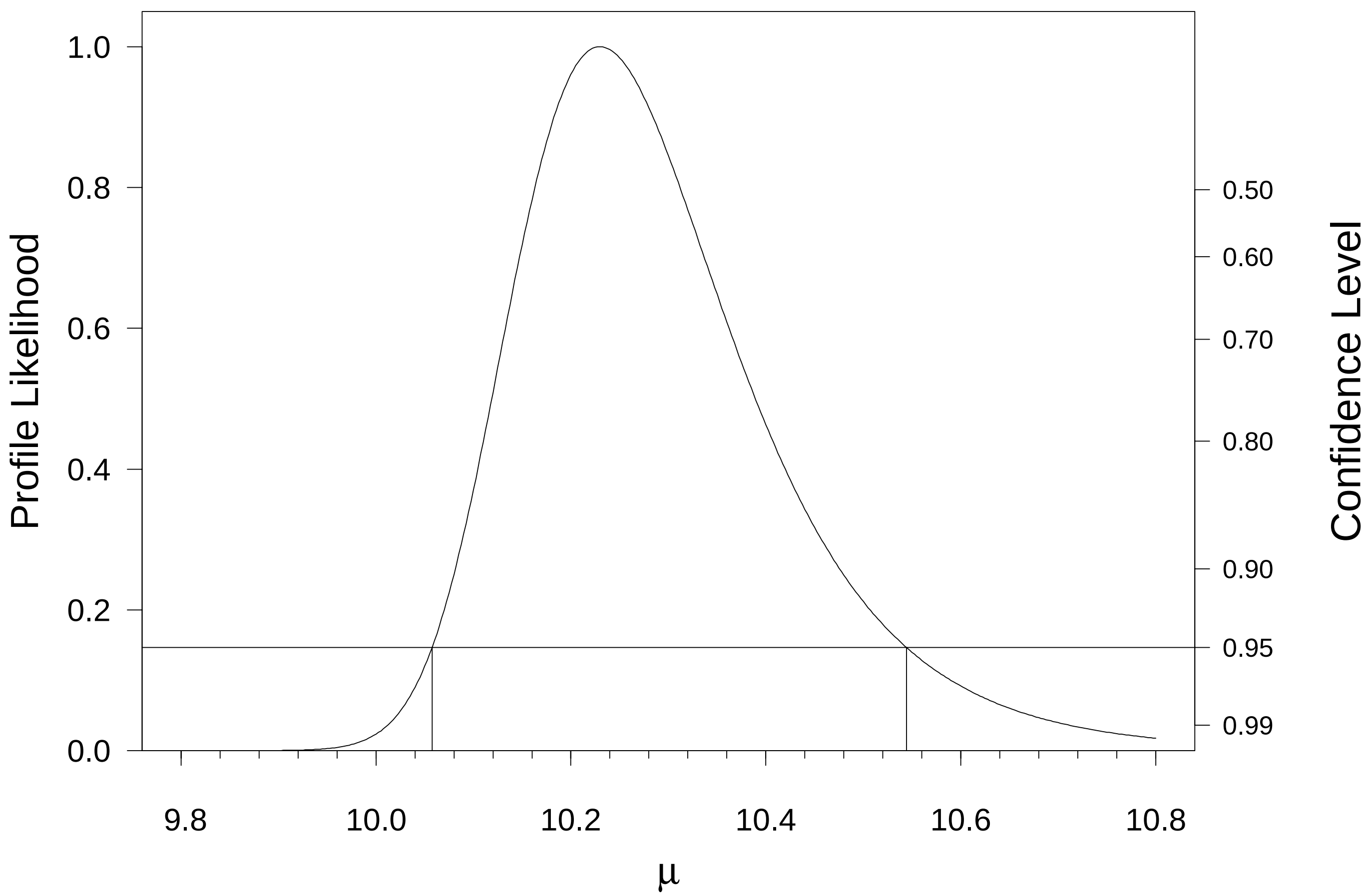
$$R(\mu, \sigma) = \frac{L(\mu, \sigma)}{L(\widehat{\mu}, \widehat{\sigma})}.$$

- If evaluated at the true (μ, σ), then, asymptotically, $-2\,\log[R(\mu, \sigma)]$ follows, a chisquare distribution with 2 degrees of freedom.

- General theory in the Appendix.

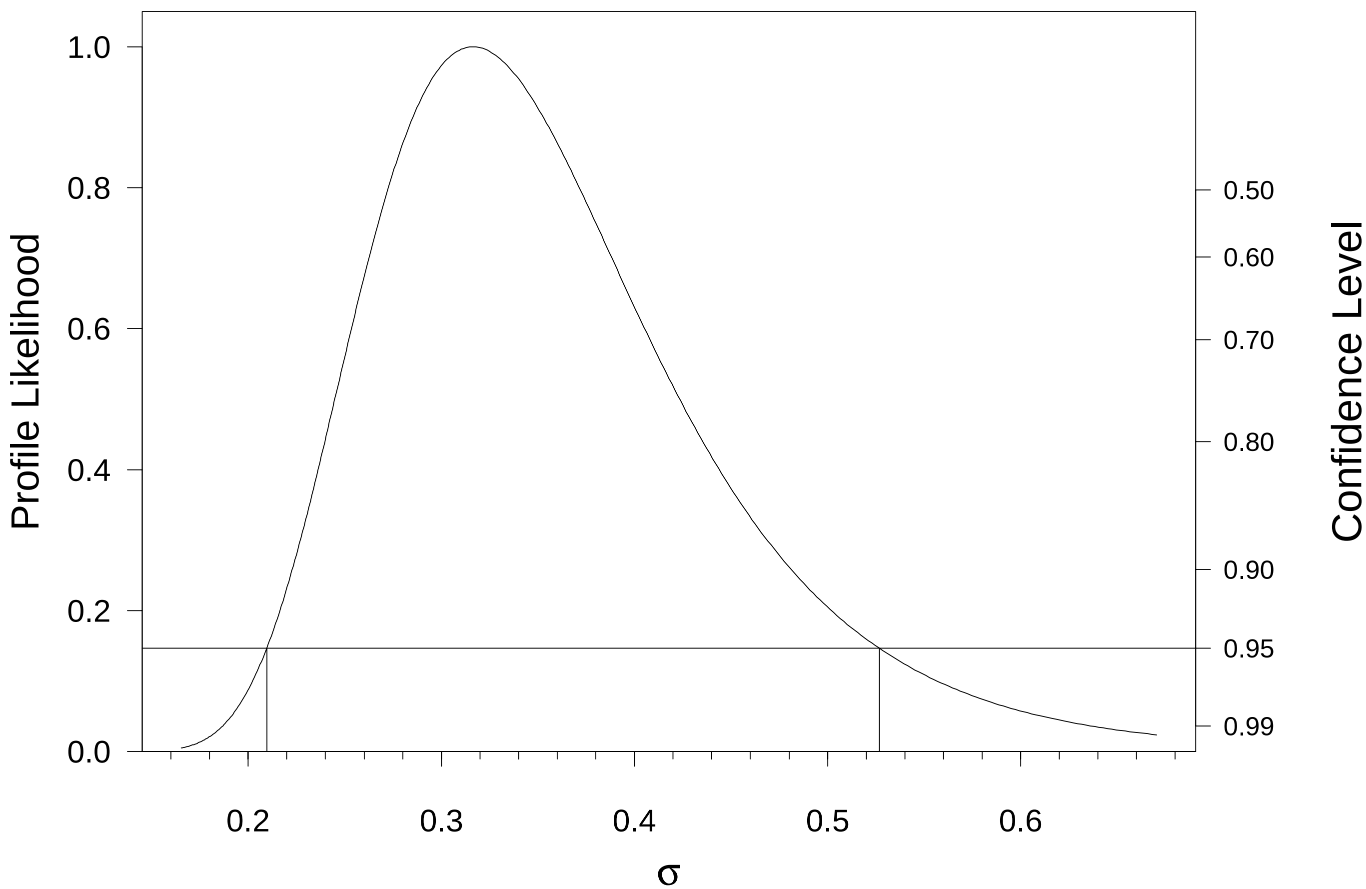
Weibull Profile Likelihood $R(\mu)$ ($\exp(\mu) \approx t_{.63}$) for the Shock Absorber Data

$$R(\mu) = \max_{\sigma} \left[\frac{L(\mu,\sigma)}{L(\widehat{\mu},\widehat{\sigma})}\right]$$



Weibull Profile Likelihood $R(\sigma)$ ($\sigma = 1/\beta$) for the Shock Absorber Data

$$R(\sigma) = \max_{\mu} \left[\frac{L(\mu,\sigma)}{L(\widehat{\mu},\widehat{\sigma})}\right]$$



Large-Sample Approximate Theory for Likelihood Ratios for Parameter Vector Subset

Need: Inferences on subset $\boldsymbol{\theta}_1$, from the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$.

- $k_1 = \text{length}(\boldsymbol{\theta}_1)$.
- When $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)' = (\mu, \sigma)$, profile likelihood for $\boldsymbol{\theta}_1 = \mu$ is

$$R(\mu) = \max_{\sigma} \left[\frac{L(\mu, \sigma)}{L(\widehat{\mu}, \widehat{\sigma})} \right].$$

- If evaluated at the true $\boldsymbol{\theta}_1 = \mu$, then, asymptotically, $-2\log[R(\mu)]$ follows, a chisquare distribution with $k_1 = 1$ degrees of freedom.
- General theory in the Appendix.

Asymptotic Theory of Likelihood Ratios – Continued

- An approximate $100(1-\alpha)\%$ likelihood-based confidence region for $\boldsymbol{\theta}_1$ is the set of all values of $\boldsymbol{\theta}_1$ such that

$$-2\log[R(\boldsymbol{\theta}_1)] < \chi^2_{(1-\alpha;k_1)}$$

  or, equivalently, the set defined by

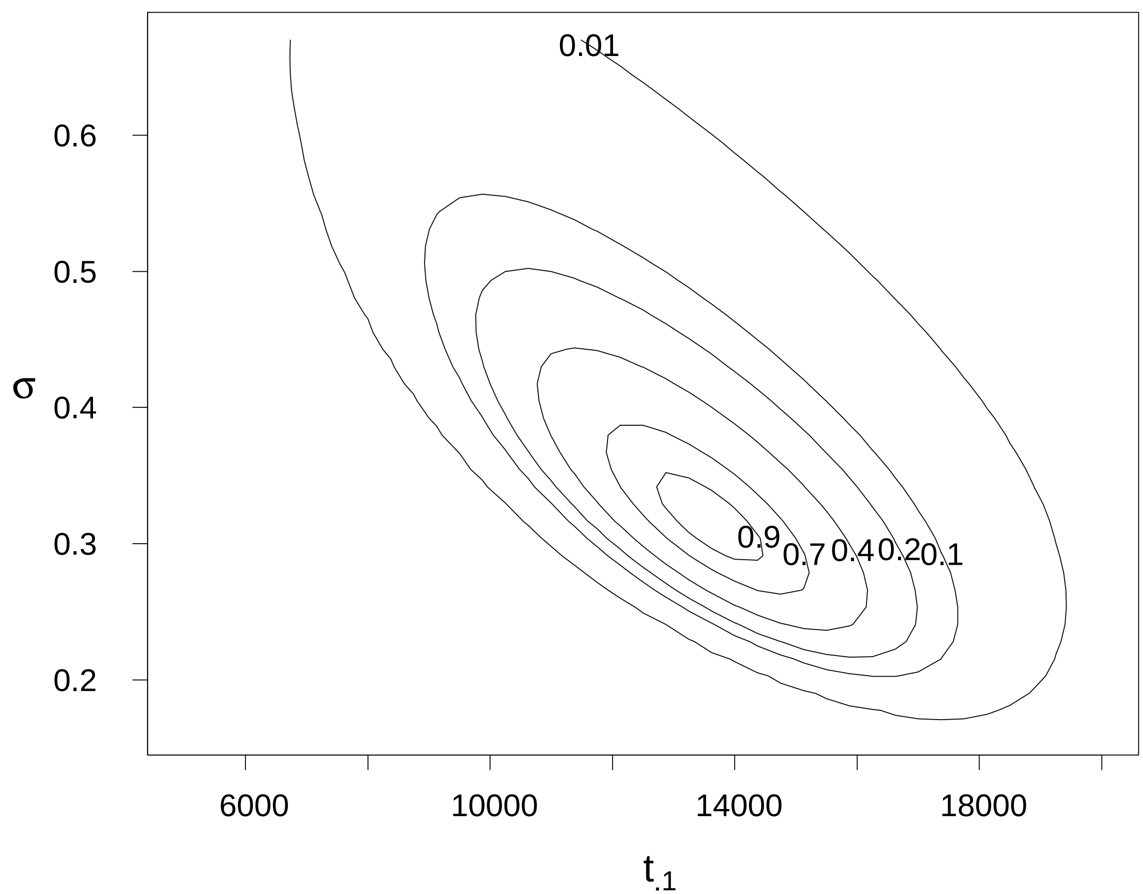
$$R(\boldsymbol{\theta}_1) > \exp\left[-\chi^2_{(1-\alpha;k_1)}/2\right].$$

- Transformation of $\boldsymbol{\theta}_1$ will not affect the confidence statement.

- Can improve the asymptotic approximation with simulation (only small effect except in very small samples).

Contour Plot of Weibull Relative Likelihood $R(t_{.1}, \sigma)$ for the Shock Absorber Data (Parameterized with $t_{.1}$ and σ)

$$R(t_{.1}, \sigma) = L(t_{.1}, \sigma)/L(\widehat{t}_{.1}, \widehat{\sigma})$$

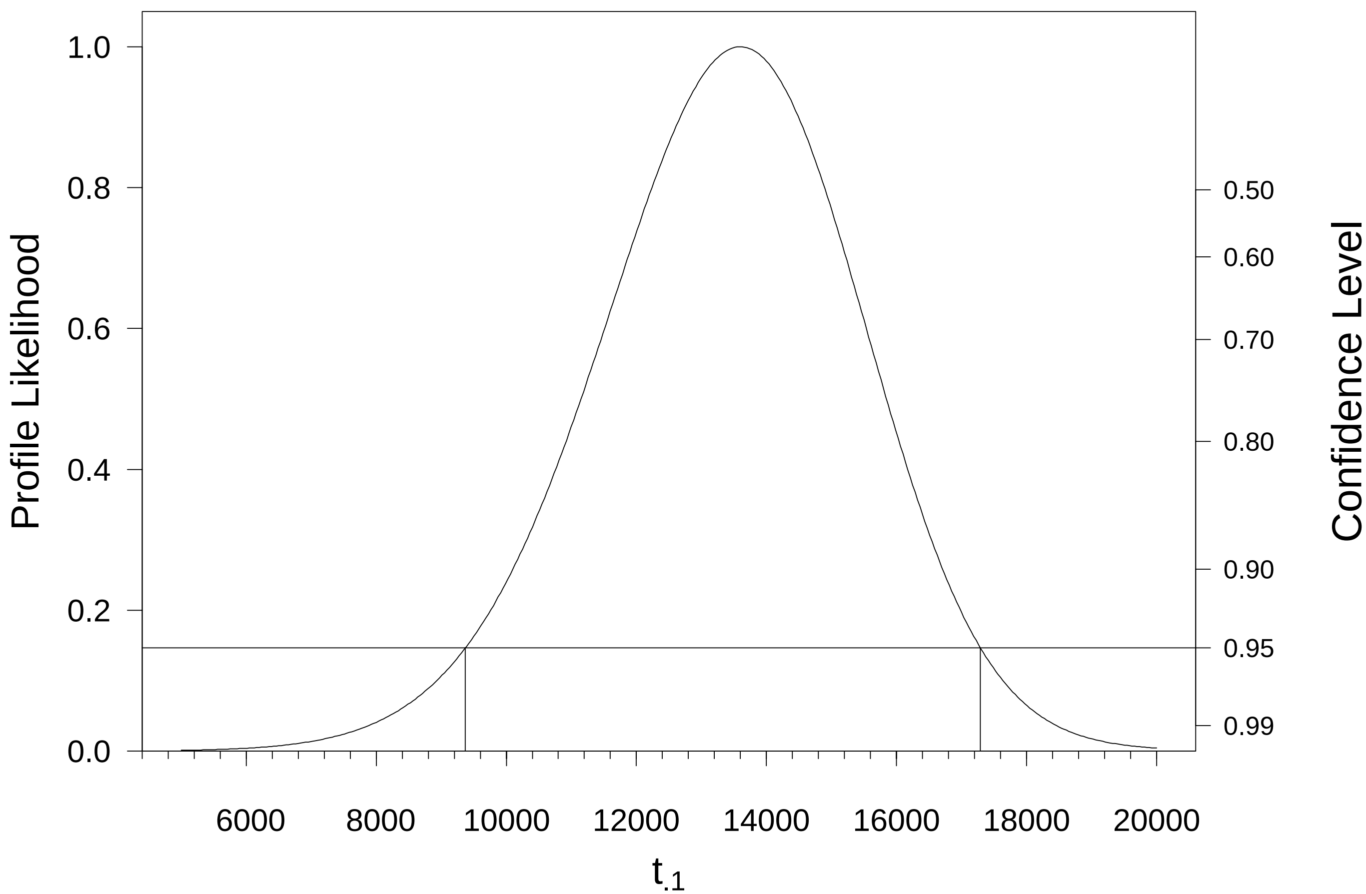


Confidence Regions and Intervals for Functions of μ and σ

- Likelihood approach can be applied to functions of parameters.
- Define the function of interest as one of the parameters, replacing one of the original parameters giving one-to-one reparameterization $\boldsymbol{g}(\mu, \sigma) = [g_1(\mu, \sigma), g_2(\mu, \sigma)]$.
- Then follow previous procedure.
- Simple to implement if function and its inverse are easy to compute.

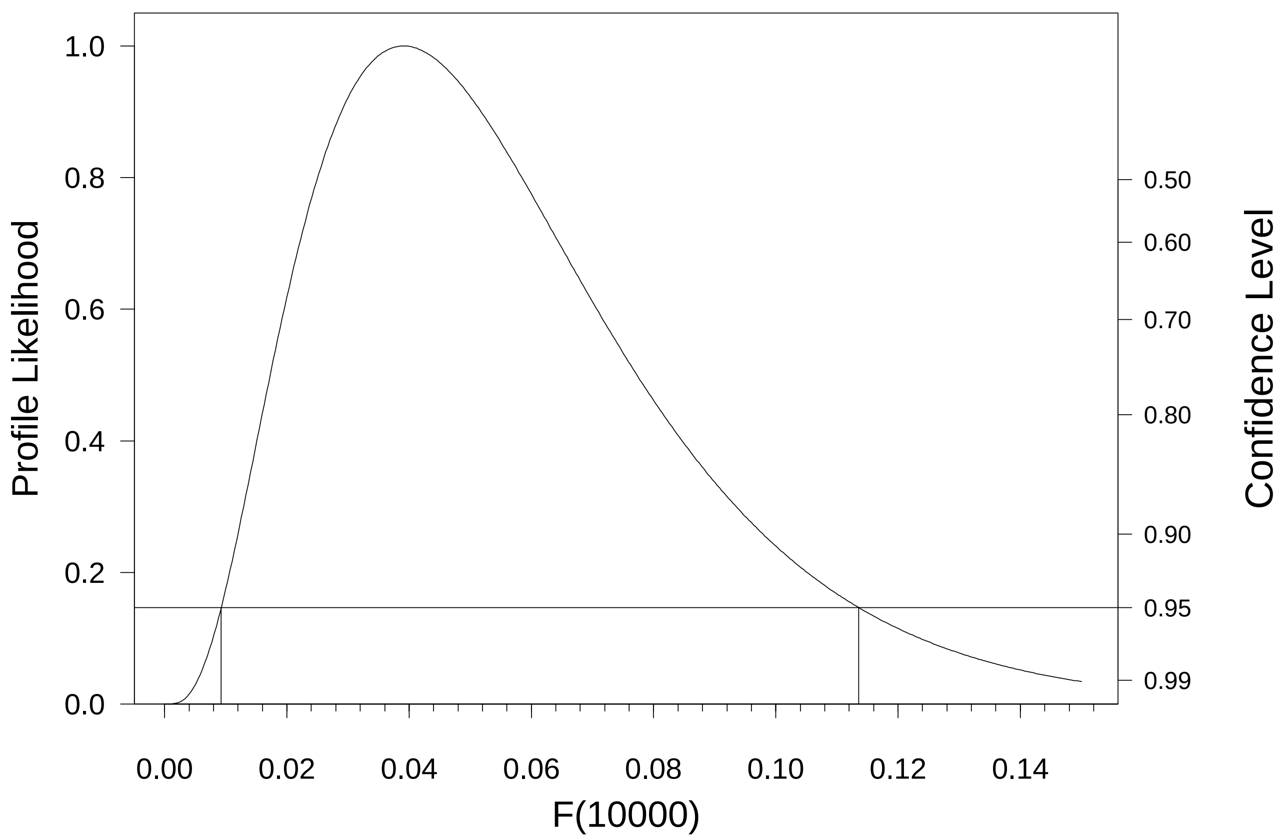
Weibull Profile Likelihood $R(t_{.1})$ for the Shock Absorber Data

$$R(t_{.1}) = \max_{\sigma} \left[\frac{L(t_{.1},\sigma)}{L(\widehat{t}_{.1},\widehat{\sigma})} \right]$$



Weibull Profile Likelihood $R[F(10000)]$ for the Shock Absorber Data

$$R\left[F(10000)\right] = \max_{\sigma} \left\{ \frac{L[F(10000),\sigma]}{L\left[\widehat{F}(10000),\widehat{\sigma}\right]} \right\}$$



Asymptotic Theory of ML Estimation

Let $\widehat{\boldsymbol{\theta}}$ denote the ML estimator of $\boldsymbol{\theta}$.

- If evaluated at the true value of $\boldsymbol{\theta}$, then asymptotically, (large samples) $\widehat{\boldsymbol{\theta}}$ has a MVN($\boldsymbol{\theta}, \Sigma_{\widehat{\boldsymbol{\theta}}}$) and thus the Wald statistic

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left[\Sigma_{\widehat{\boldsymbol{\theta}}}\right]^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

  has a chisquare distribution with k degrees of freedom, where k is the length of $\boldsymbol{\theta}$.

- Here, $\Sigma_{\widehat{\boldsymbol{\theta}}} = I_{\boldsymbol{\theta}}^{-1}$ is the large sample approximate covariance matrix where

$$I_{\boldsymbol{\theta}} = \mathsf{E}\left[-\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right].$$

Asymptotic Theory for Wald's Statistic

- Alternative asymptotic theory is based on the large-sample distribution of quadratic forms (Wald's statistic).

- Let $\hat{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ be a consistent estimator of $\Sigma_{\widehat{\boldsymbol{\theta}}}$, the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$. For example,

$$\hat{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \left[-\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1}$$

  where the derivatives are evaluated at $\widehat{\boldsymbol{\theta}}$.

- Asymptotically, the Wald statistic

$$w(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left[\hat{\Sigma}_{\widehat{\boldsymbol{\theta}}} \right]^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

  when evaluated at the true $\boldsymbol{\theta}$, follows a chisquare distribution with k degrees of freedom, where k is the length of $\boldsymbol{\theta}$.

Asymptotic Theory for Wald's Statistic – Continued

- An approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ is the set of all values of $\boldsymbol{\theta}$ in the ellipsoid

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left[\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}}\right]^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le \chi^2_{(1-\alpha;k)}.$$

- This is sometimes known as the normal-theory confidence region.

- Can specialize to functions or subsets of $\boldsymbol{\theta}$.

- Can transform to improve asymptotic approximation. Try to get a log likelihood with approximate quadratic shape.

Normal-Approximation Confidence Intervals for Model Parameters

- Estimated variance matrix for the shock absorber data

$$\widehat{\Sigma}_{\widehat{\mu},\widehat{\sigma}} = \begin{bmatrix} \widehat{\text{Var}}(\widehat{\mu}) & \widehat{\text{Cov}}(\widehat{\mu},\widehat{\sigma}) \\ \widehat{\text{Cov}}(\widehat{\mu},\widehat{\sigma}) & \widehat{\text{Var}}(\widehat{\sigma}) \end{bmatrix} = \begin{bmatrix} .01208 & .00399 \\ .00399 & .00535 \end{bmatrix}$$

- Assuming that $Z_{\widehat{\mu}} = (\widehat{\mu} - \mu)/\widehat{\text{se}}_{\widehat{\mu}} \,\dot{\sim}\, \text{NOR}(0,1)$ distribution, an approximate $100(1-\alpha)\%$ confidence interval for μ is

$$[\underset{\sim}{\mu}, \quad \tilde{\mu}] = \widehat{\mu} \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{\mu}}$$

  where $\widehat{\text{se}}_{\widehat{\mu}} = \sqrt{\widehat{\text{Var}}(\widehat{\mu})}$.

- Assuming that $Z_{\log(\widehat{\sigma})} = [\log(\widehat{\sigma}) - \log(\sigma)]/\widehat{\text{se}}_{\log(\widehat{\sigma})} \,\dot{\sim}\, \text{NOR}(0,1)$ an approximate $100(1-\alpha)\%$ confidence interval for σ is

$$[\underset{\sim}{\sigma}, \quad \tilde{\sigma}] = [\widehat{\sigma}/w, \quad \widehat{\sigma} \times w]$$

  where $w = \exp\left[z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{\sigma}}/\widehat{\sigma}\right]$ and $\widehat{\text{se}}_{\widehat{\sigma}} = \sqrt{\widehat{\text{Var}}(\widehat{\sigma})}$.

Normal-Approximation Confidence Intervals for Function $g_1 = g_1(\mu, \sigma)$

- ML estimate $\widehat{g}_1 = g_1(\widehat{\mu}, \widehat{\sigma})$.
- Assuming $Z_{\widehat{g}_1} = (\widehat{g}_1 - g_1)/\widehat{\text{se}}_{\widehat{g}_1} \dot{\sim} \text{NOR}(0,1)$, an approximate $100(1-\alpha)\%$ confidence interval for g_1 is

$$[\underset{\sim}{g_1}, \quad \tilde{g}_1] \;=\; \widehat{g}_1 \pm z_{(1-\alpha/2)} \widehat{\text{se}}_{\widehat{g}_1},$$

where

$$\widehat{\text{se}}_{\widehat{g}_1} = \sqrt{\widehat{\text{Var}}(\widehat{g}_1)} = \left[\left(\frac{\partial g_1}{\partial \mu}\right)^2 \widehat{\text{Var}}(\widehat{\mu}) + \left(\frac{\partial g_1}{\partial \sigma}\right)^2 \widehat{\text{Var}}(\widehat{\sigma}) + 2 \left(\frac{\partial g_1}{\partial \mu}\right) \left(\frac{\partial g_1}{\partial \sigma}\right) \widehat{\text{Cov}}(\widehat{\mu}, \widehat{\sigma}) \right]^{\frac{1}{2}}$$

- Partial derivatives evaluated at $\widehat{\mu}, \widehat{\sigma}$.
- General theory in the appendix.

Normal-Approximation Confidence Interval for $F(t_e;\mu,\sigma)$

Objective: Obtain a point estimate and a confidence interval for $\Pr(T \le t_e) = F(t_e;\mu,\sigma)$ at a fixed and known point t_e.

- The ML estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\mu}, \widehat{\sigma})$ and $\hat{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ are available.
- The ML estimate for $F(t_e;\mu,\sigma)$ is

$$\widehat{F} = F(t_e;\widehat{\mu},\widehat{\sigma}) = \Phi(\widehat{\zeta_e})$$

  where $\widehat{\zeta_e} = [\log(t_e) - \widehat{\mu}]/\widehat{\sigma}$.

- In the context of Wald's theory, however, there are many ways to obtain a confidence interval for $F(t_e;\mu,\sigma)$.

Confidence Interval for $F(t_e;\mu,\sigma)$–Continued

Note: Wald's confidence intervals depend on the parameterization used to derive the intervals.

For example, $100(1-\alpha)\%$ confidence interval for $F(t_e;\mu,\sigma)$ can be obtained using:

- The asymptotic normality of $Z_{\widehat{F}} = (\widehat{F} - F)/\widehat{\text{se}}_{\widehat{F}}$

$$[\underset{\sim}{F}, \quad \tilde{F}] = \widehat{F}(t_e) \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{F}}.$$

- The asymptotic normality of $Z_{\text{logit}(\widehat{F})} = [\text{logit}(\widehat{F}) - \text{logit}(F)]/\widehat{\text{se}}_{\text{logit}(\widehat{F})}$

$$[\underset{\sim}{F}, \quad \tilde{F}] = \left[\frac{\widehat{F}(t_e)}{\widehat{F}(t_e) + (1 - \widehat{F}(t_e)) \times w}, \quad \frac{\widehat{F}(t_e)}{\widehat{F}(t_e) + (1 - \widehat{F}(t_e))/w}\right]$$

where $w = \exp\{z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{F}}/[\widehat{F}(t_e)(1 - \widehat{F}(t_e))]\}$.

Confidence Interval for $F(t_e; \mu, \sigma)$—Continued

Comments:

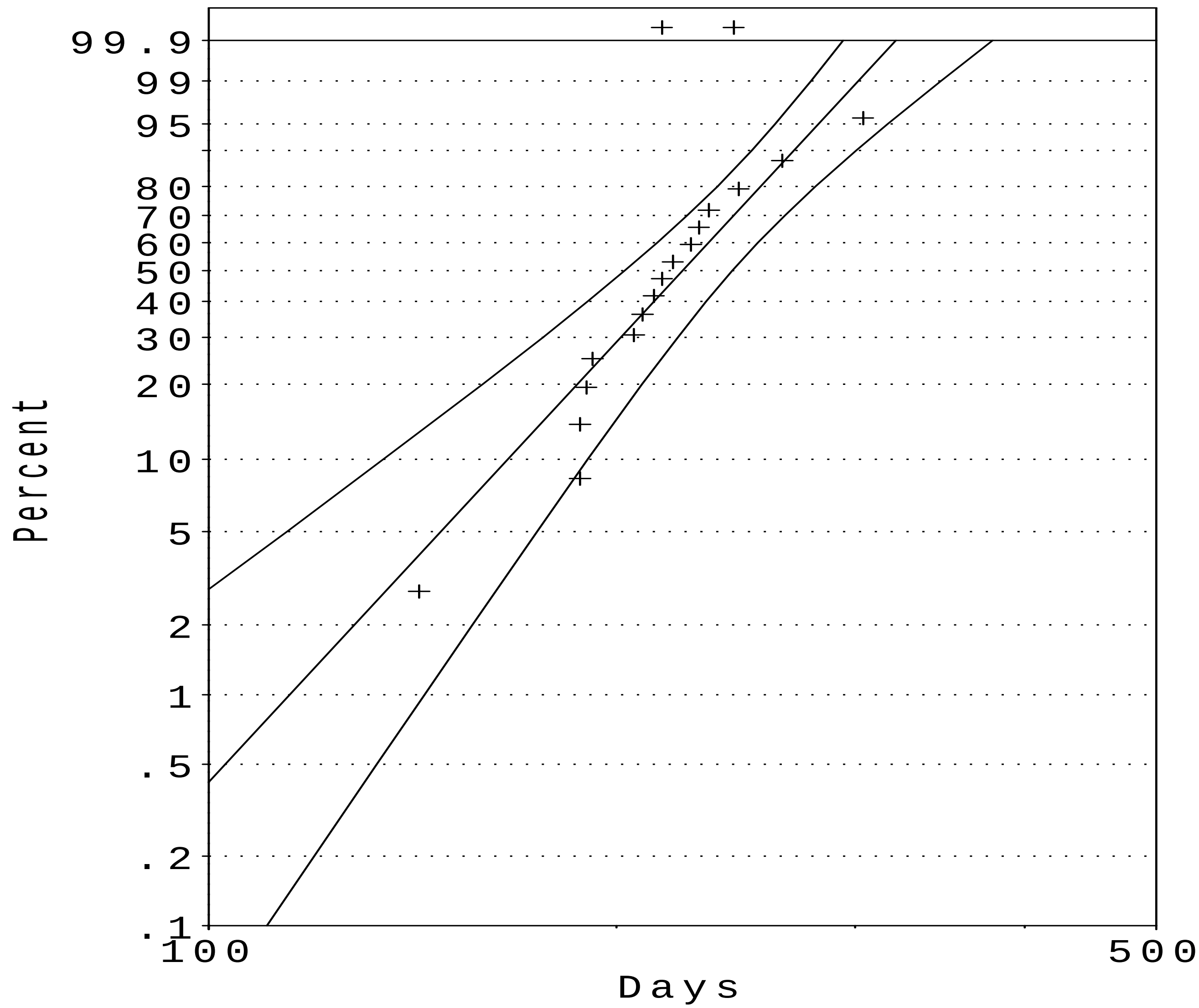
- Often the confidence interval based on the asymptotic normality of $Z_{\widehat{F}}$ has poor statistical properties caused by the slow convergence toward normality of $Z_{\widehat{F}}$.

- The confidence interval based on the transformation $Z_{\text{logit}(\widehat{F})}$ can have better statistical properties if $Z_{\text{logit}(\widehat{F})}$ converges to normality faster than $Z_{\widehat{F}}$.

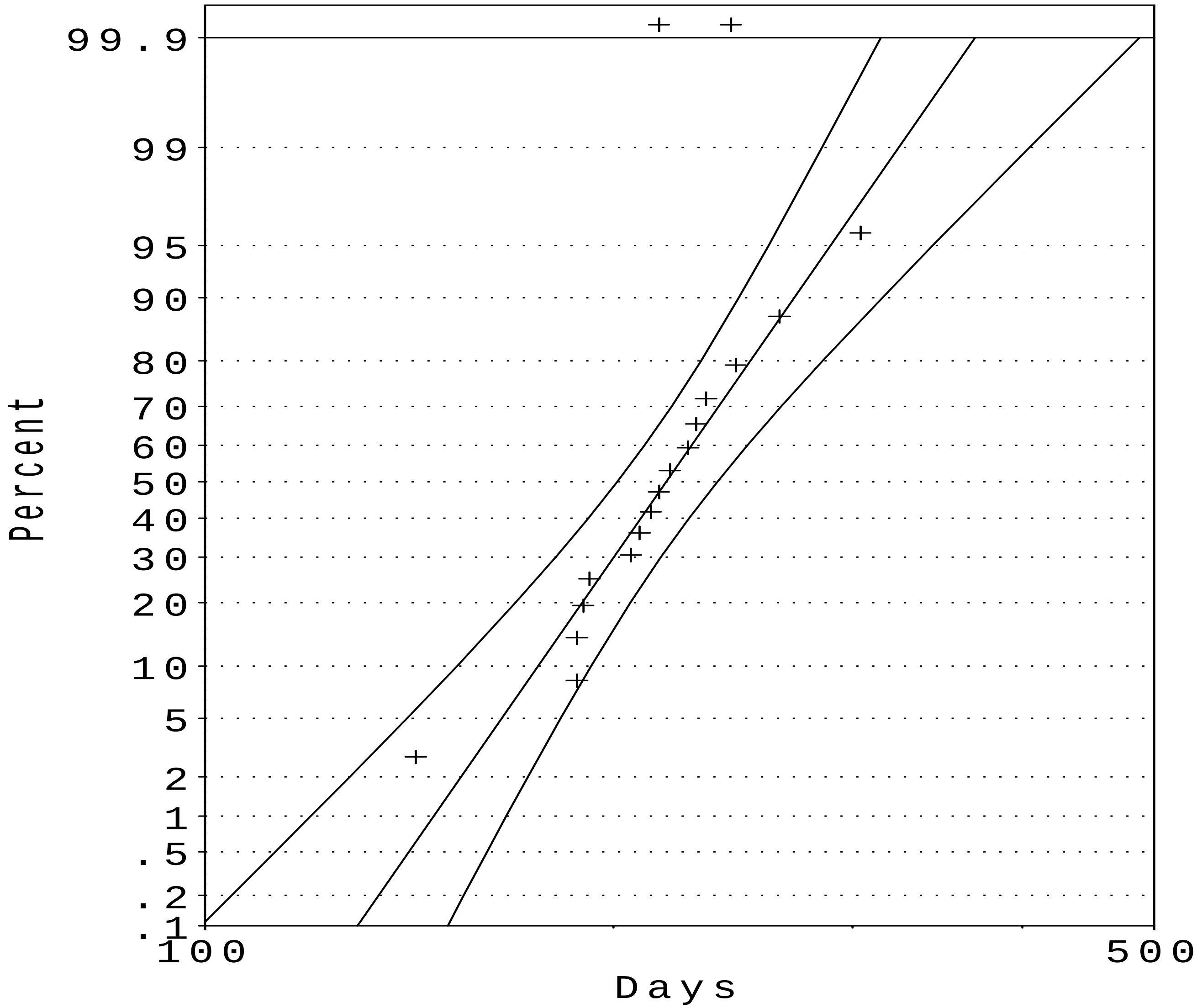
ML Estimates for Biomedical Data

Here we show ML estimates (Weibull and lognormal) for the DMBA and the IUD data.

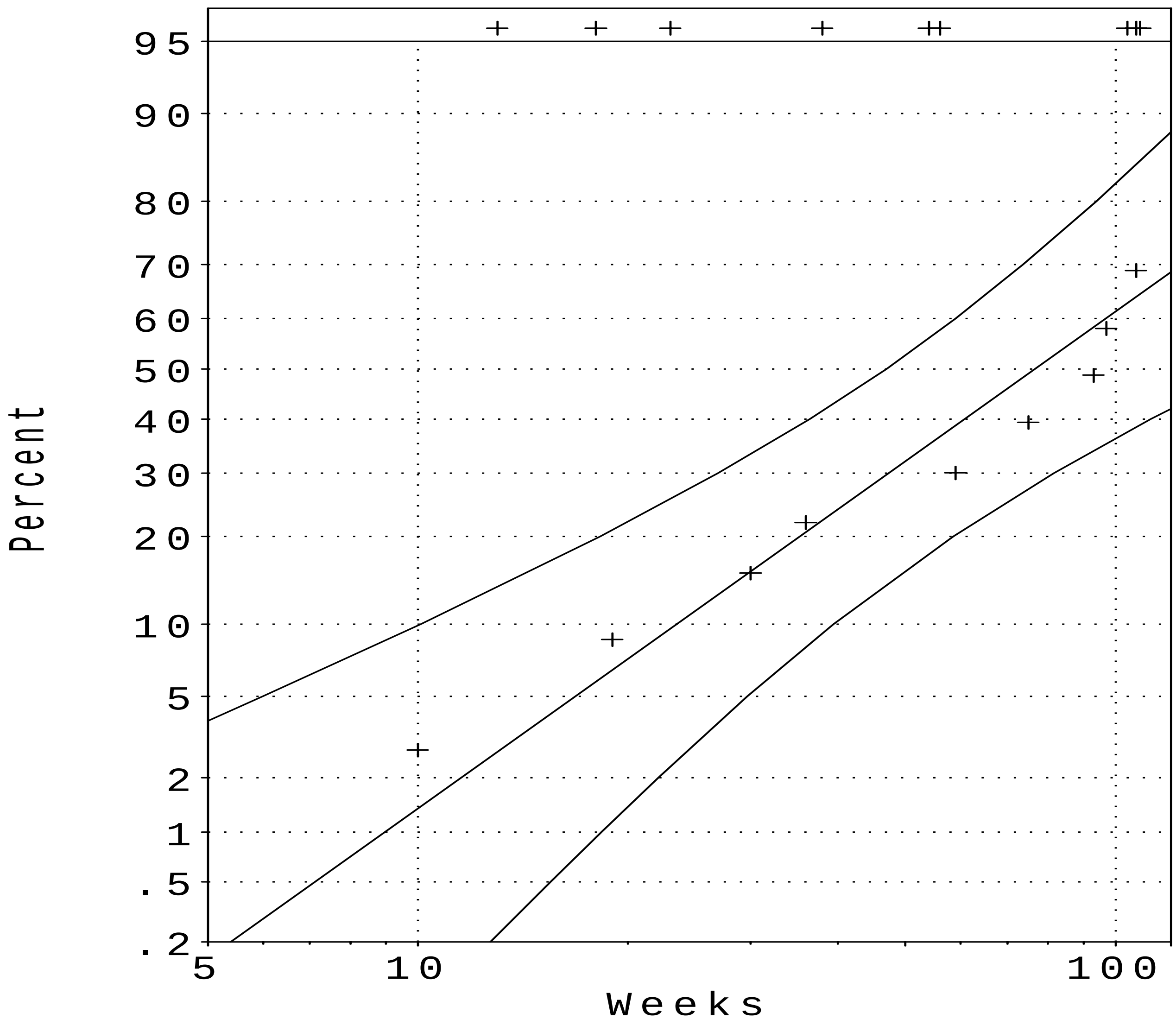
Nonparametric and Weibull ML Estimate for DMBA Data with Parametric Pointwise Approximate 95% Confidence Intervals



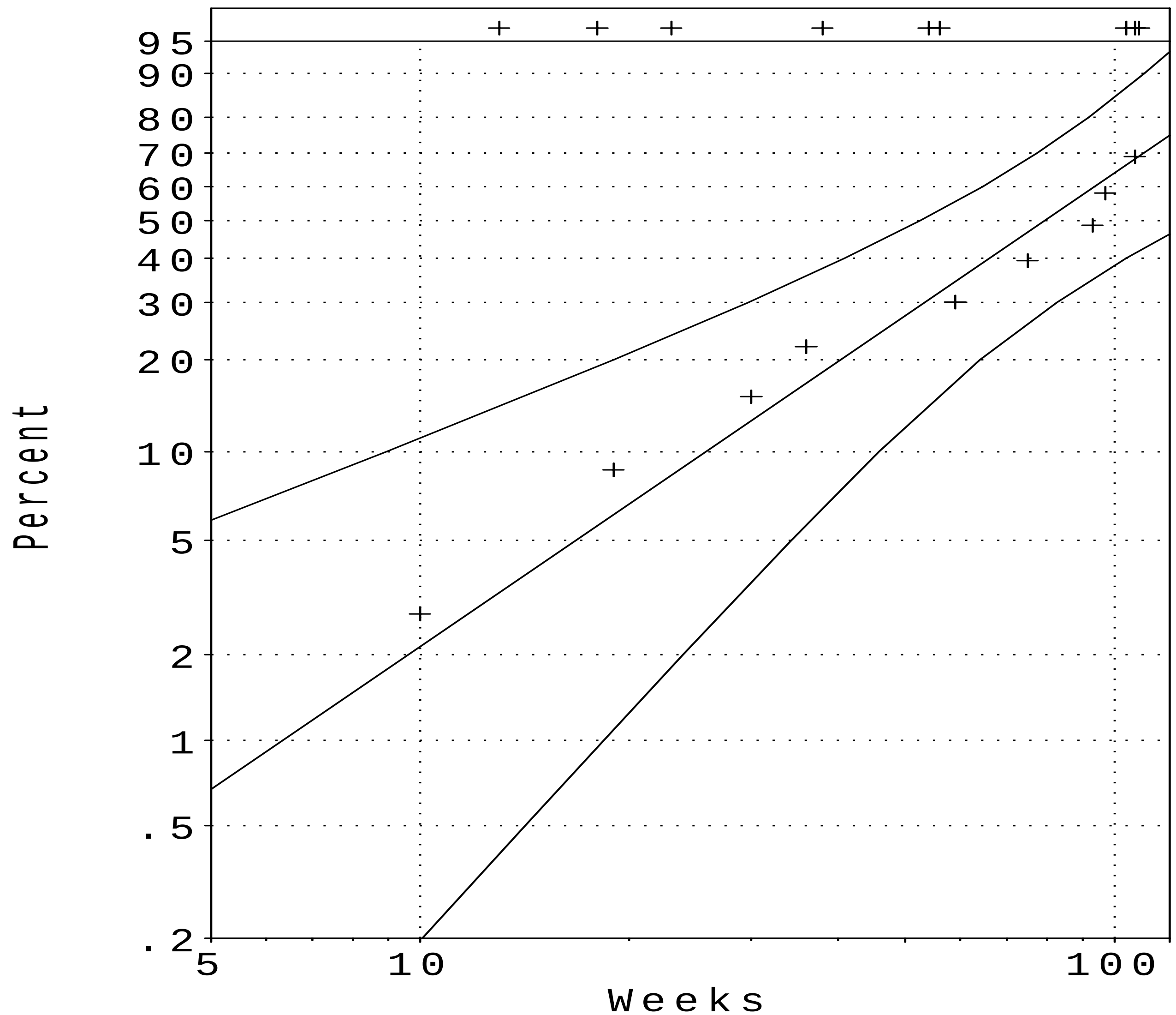
Nonparametric and Lognormal ML Estimate for DMBA Data with Parametric Pointwise Approximate 95% Confidence Intervals



Lognormal ML Estimate for IUD Data with a set of Pointwise Approximate 95% Confidence Intervals



Weibull ML Estimate for IUD Data with a set of Pointwise Approximate 95% Confidence Intervals



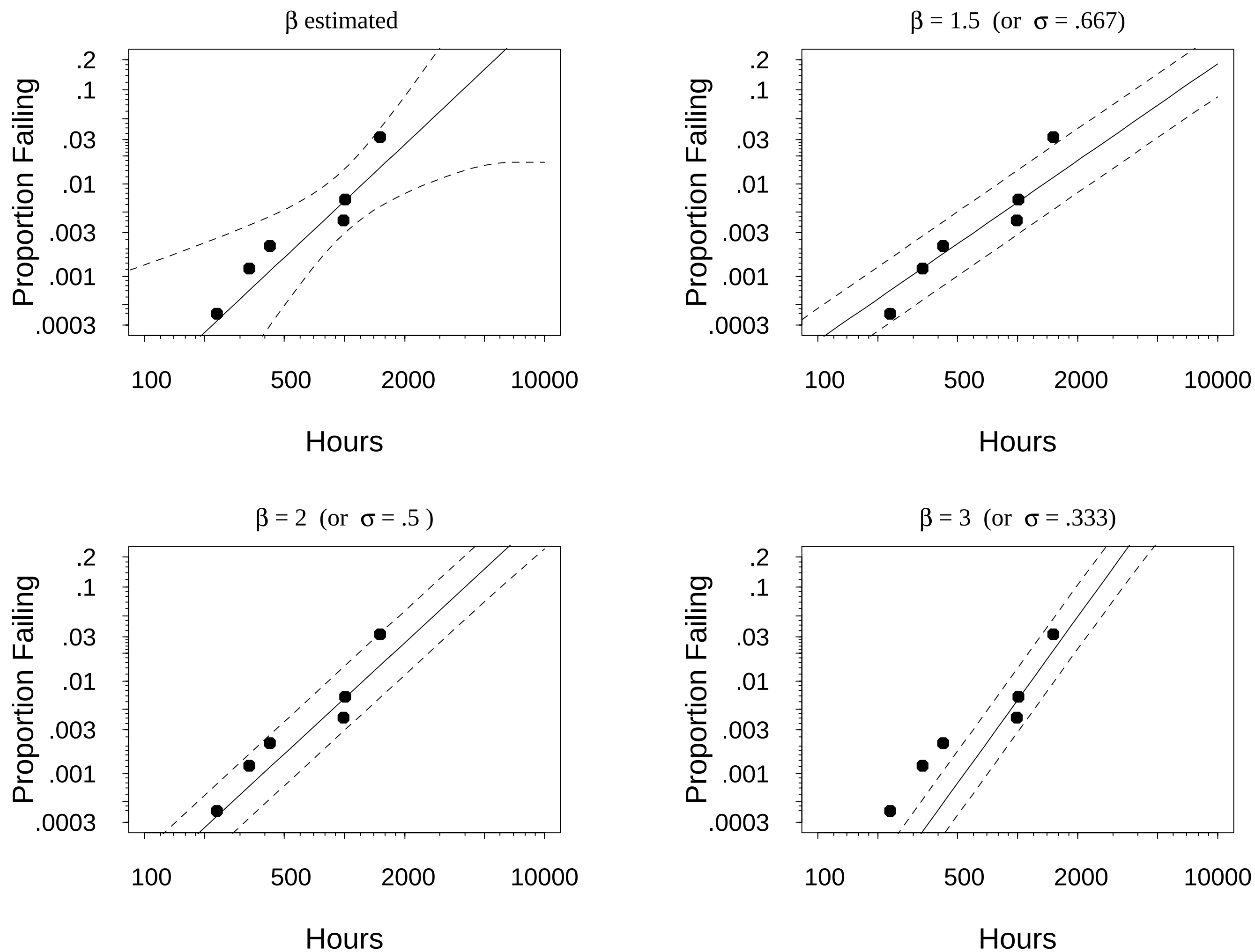
Inference when σ (or Weibull β) is Given

- Simplifies problem. Only one parameter with r failures and $t_1, \ldots, t_n$ failures and censor times

$$\widehat{\eta} = \left(\frac{\sum_{i=1}^{n} t_i^{\beta}}{r} \right)^{1/\beta}, \quad \widehat{\text{se}}_{\widehat{\eta}} = \frac{\widehat{\eta}}{\beta} \sqrt{\frac{1}{r}}.$$

- Provides much more precision, especially with small r.
- If 0 failures can provide
  - ▶ Upper confidence bound on $F(t)$.
  - ▶ Lower confidence bound on t_p.
- Requires sensitivity analysis because β is in doubt.
- Danger of misleading inferences.

Weibull Probability Plots Bearing Cage Fracture Data with Weibull ML Estimates and Sets of 95% Pointwise Confidence Intervals for $F(t)$ with β Estimated, and Assumed Known Values of β= 1.5, 2, and 3.



Bearing-Cage Fracture Field Data

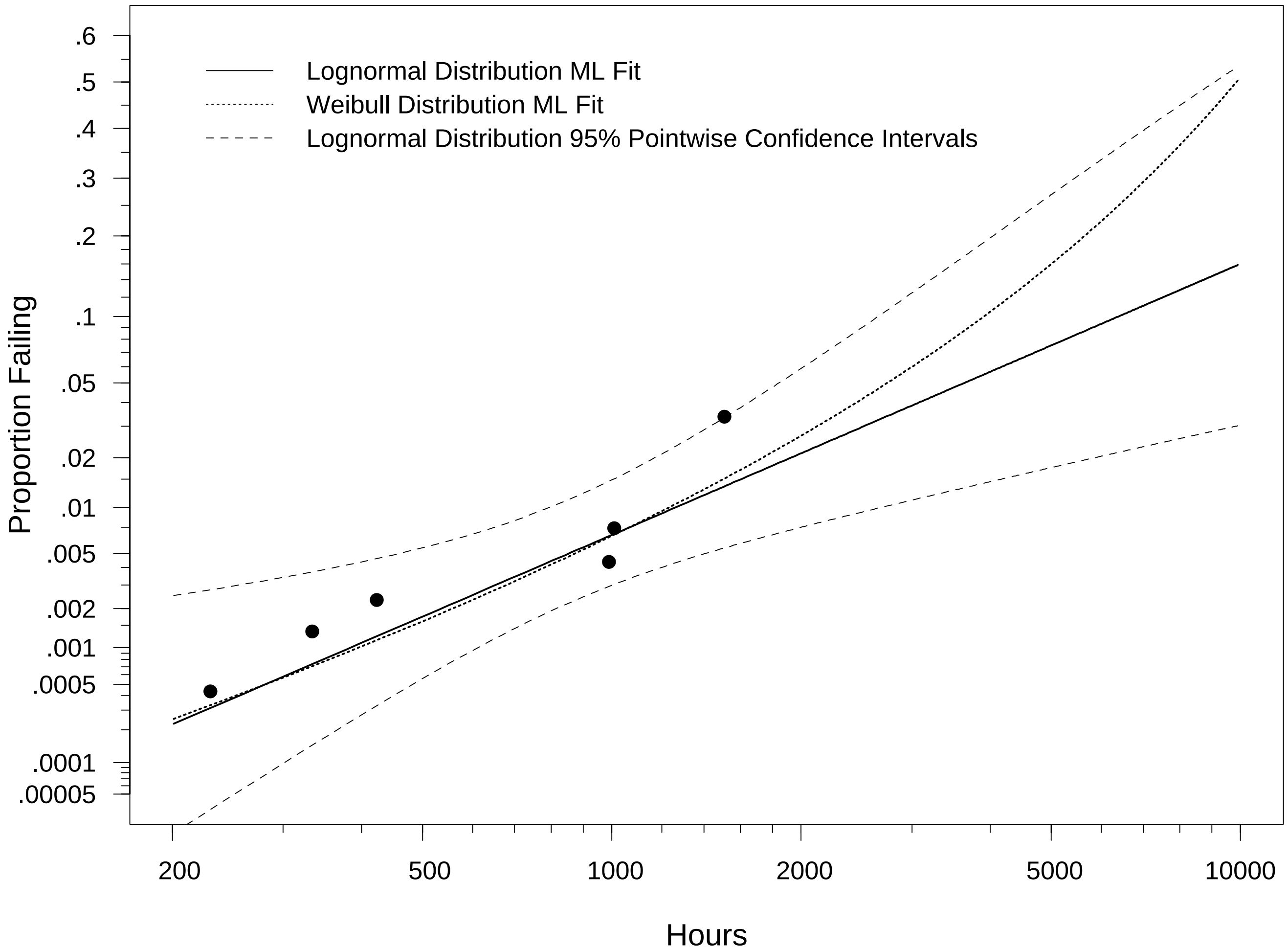
- A population of $n = 1703$ units had been introduced into service over time and 6 failures have been observed.

- There is concern that the B10 design life specification of $t_{.1} = 8$ thousand hours was not being met.

- ML estimate is $\widehat{t}_{.1} = 3.903$ thousand hours and an approximate 95% likelihood-ratio confidence interval for $t_{.1}$ is [2.093, 22.144] thousand hours.

- Management also wanted to know how many additional failures could be expected in the next year.

Comparison Between Lognormal and Weibull Distributions Fit to the Bearing-Cage Fracture Field Data



Weibull/SEV **Distribution with Given** $\beta = 1/\sigma$ **and Zero Failures**

- ML Estimate for the Weibull Scale Parameter η Cannot be Computed Unless the Available Data Contains One or More Failures.
- For a sample of n units with running times $t_1, \ldots, t_n$ and no failures, a conservative $100(1-\alpha)\%$ lower confidence bound for η is

$$\underset{\sim}{\eta} = \left(\frac{2\sum_{i=1}^{n} t_i^{\beta}}{\chi^2_{(1-\alpha;2)}} \right)^{\frac{1}{\beta}}.$$

- The lower bound $\underset{\sim}{\eta}$ can be translated into an lower confidence bound for functions like t_p for specified p or a upper confidence bound for $F(t_e)$ for a specified t_e.

Component A Safe Data

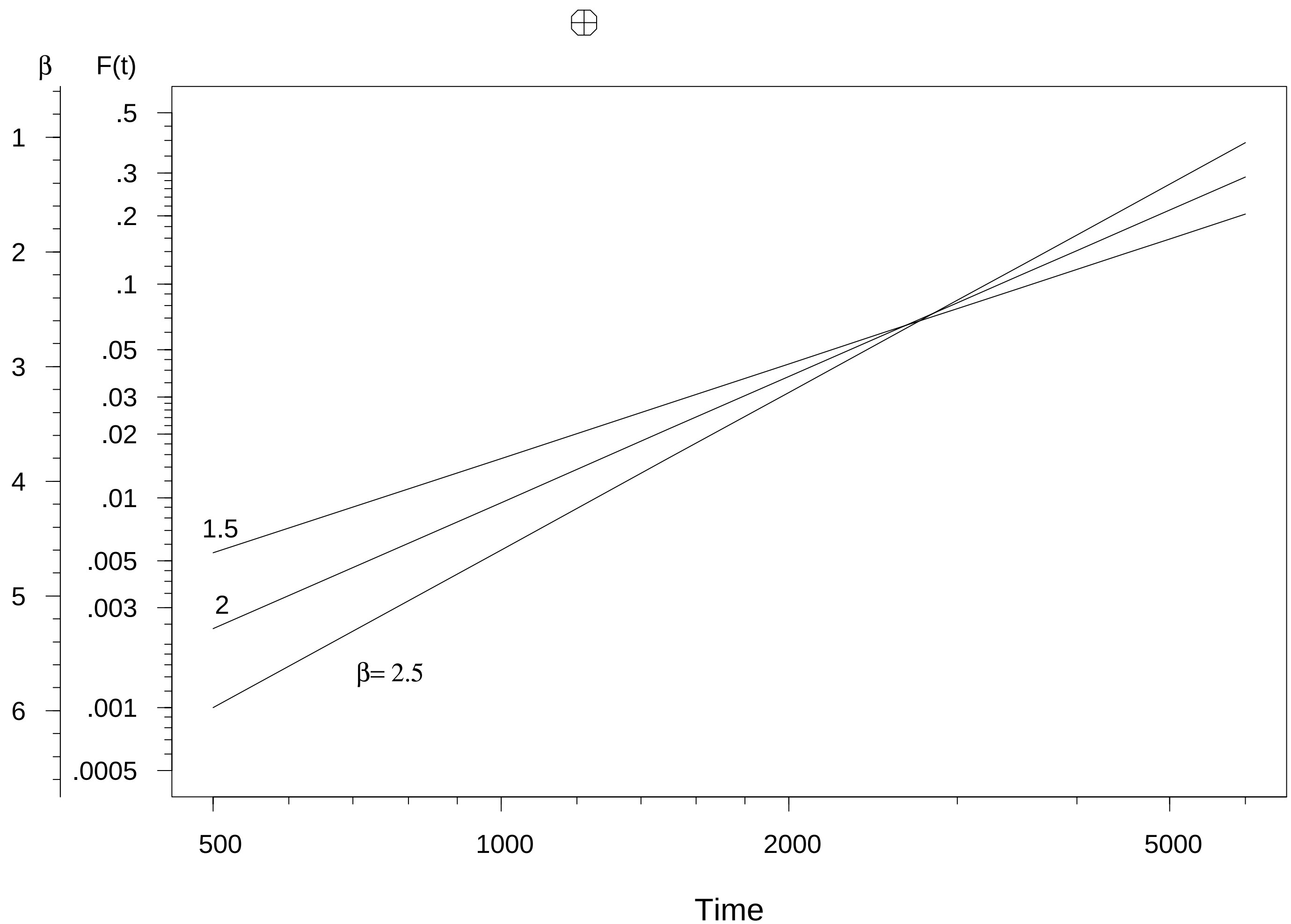
- A metal component in a ship's propulsion system fails from fatigue-caused fracture.
- Because of persistent reliability problems, the component was redesigned to have a longer service life.
- Previous experience suggests that the Weibull shape parameter is near $\beta = 2$, and almost certainly between 1.5 and 2.5.
- Newly designed components were put into service during the past year and no failures have been reported.

Hours:	500	1000	1500	2000	2500	3000	3500	4000
Number of Units:	10	12	8	9	7	9	6	3

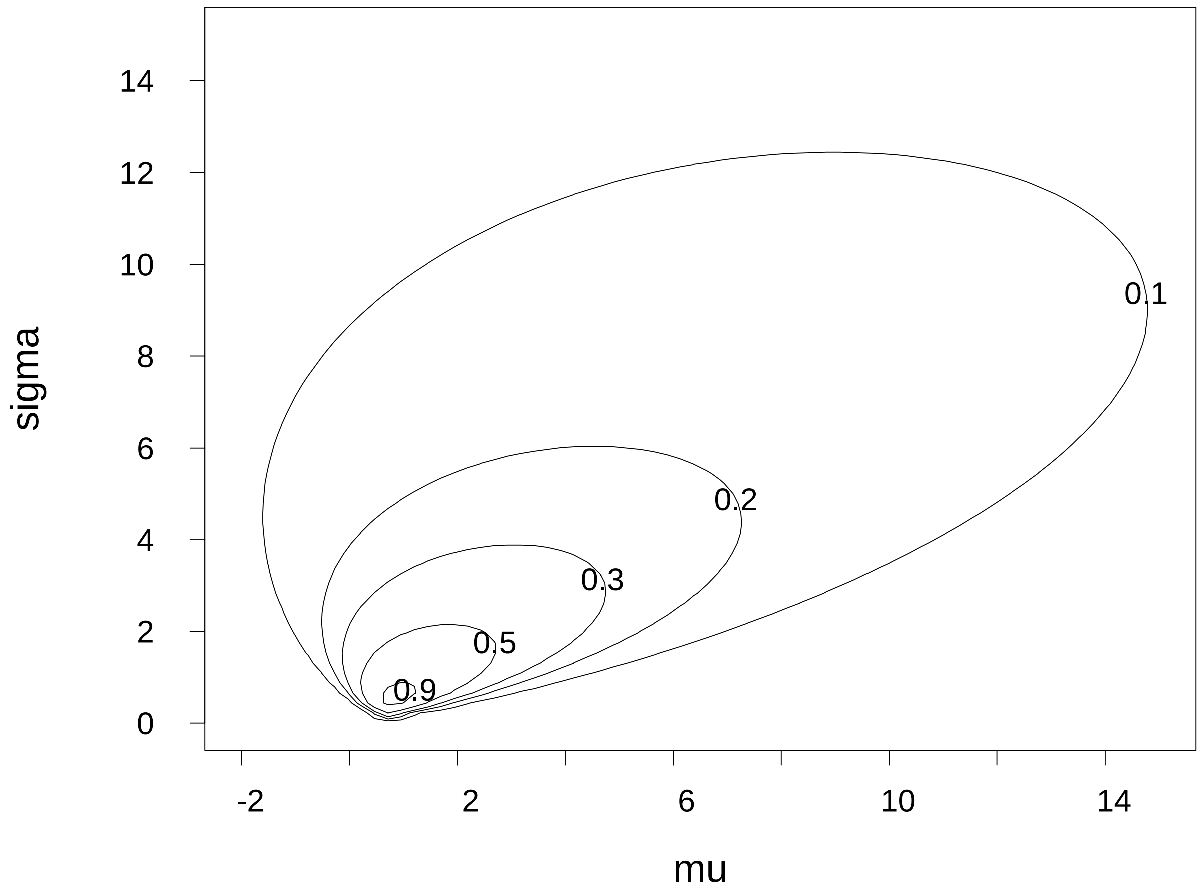
Staggered entry data, with no reported failures.

- Can replacement be increased from 2000 hours to 4000 hours?

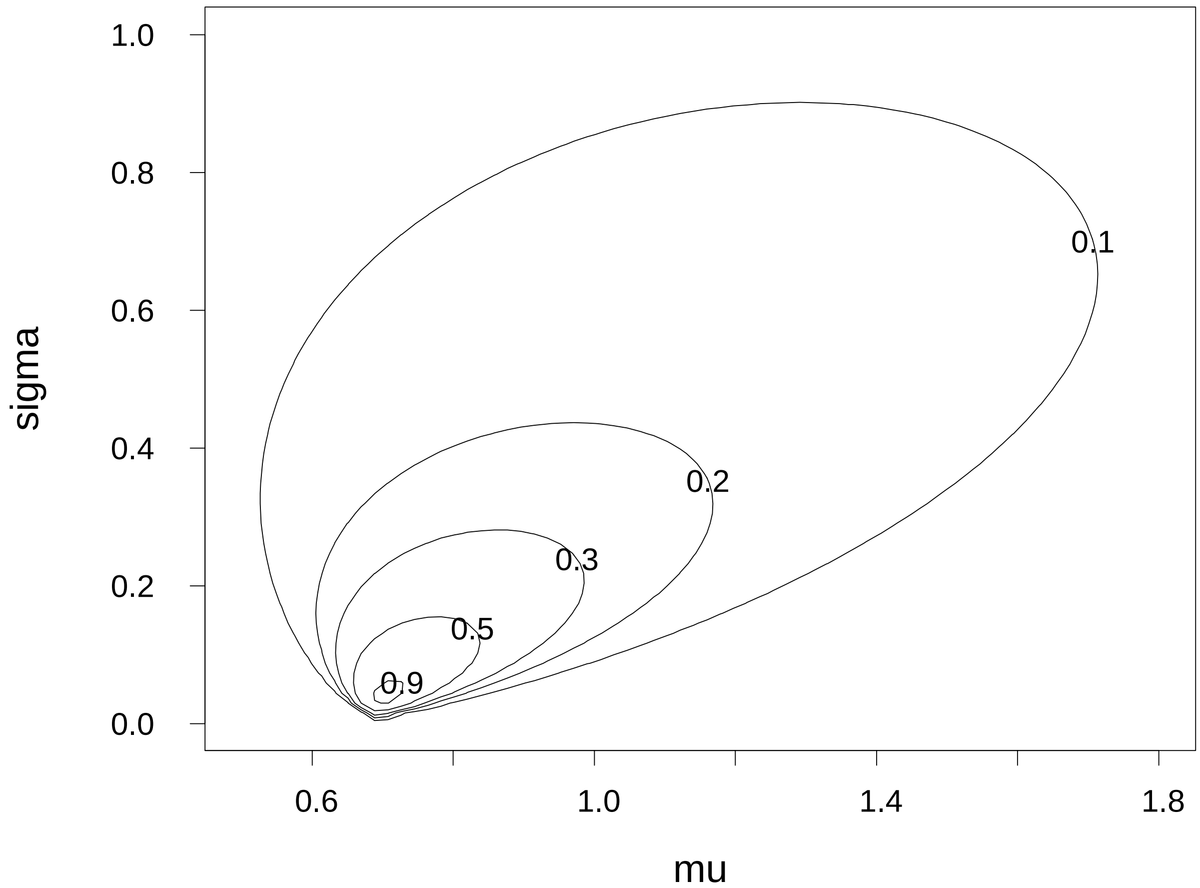
Weibull Model 95% Upper Confidence Bounds on $F(t)$ for Component-A with Different Fixed Values for the Weibull Shape Parameter



Relative Weibull Likelihood
with One Failure at 1 and One Survivor at 2



Relative Weibull Likelihood
with One Failure at 1.9 and One Survivor at 2



Regularity Conditions

- Each technical result (e.g., asymptotic distribution of an estimator) has its own set of conditions on the model (see Lehmann 1983, Rao 1973).
- Frequent reference to Regularity Conditions which give rise to simple results.
- For special cases the regularity conditions are easy to state and check. For example, for some location-scale distributions the needed conditions are:

$$
\begin{aligned}
\lim_{z \to -\infty} \frac{z^2 \phi^2(z)}{\Phi(z)} &= 0 \\
\lim_{z \to +\infty} \frac{z^2 \phi^2(z)}{1 - \Phi(z)} &= 0.
\end{aligned}
$$

- In **non-regular** models, asymptotic behavior is more complicated (e.g., behavior depends on $\boldsymbol{\theta}$), but there are still useful asymptotic results.

Regularity Conditions – Continued

Some **typical** regularity conditions include:

- Support does not depend on unknown parameters.
- Number of parameters does not grow too fast with n.
- Continuous derivatives of log likelihood (w.r.t. $\boldsymbol{\theta}$).
- Bounded derivatives of likelihood.
- Can exchange the order of differentiation of log likelihood w.r.t. $\boldsymbol{\theta}$ and integration w.r.t. data.
- Identifiability.

Other Topics Related to Parametric Likelihood Covered in Book

- Truncated data.
- Threshold parameters.
- Other distributions (e.g., gamma).
- Bayesian methods.
- Multiple failure modes.