Unit 8: Maximum Likelihood for Location-Scale Distributions

Ramón V. León


Unit 8 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
Refresher

Note: If $T \sim \text{WEIB}(\mu, \sigma)$ then $Y = \log(T) \sim \text{SEV}(\mu, \sigma)$.

For $T \sim \text{WEIB}(\mu, \sigma)$ then

\begin{align*}
F(t; \mu, \sigma) &= 1 - \exp \left[ - \left( \frac{t}{\eta} \right)^\beta \right] = \Phi_{\text{SEV}} \left[ \frac{\log(t) - \mu}{\sigma} \right] \\
f(t; \mu, \sigma) &= \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta - 1} \exp \left[ - \left( \frac{t}{\eta} \right)^\beta \right] = \frac{1}{\sigma} \phi_{\text{SEV}} \left[ \frac{\log(t) - \mu}{\sigma} \right]
\end{align*}

where $\sigma = 1/\beta$, $\mu = \log(\eta)$, and

\begin{align*}
\phi_{\text{SEV}}(z) &= \exp[z - \exp(z)] \\
\Phi_{\text{SEV}}(z) &= 1 - \exp[-\exp(z)].
\end{align*}

Quantiles:

$t_p = \eta \left[ - \log(1 - p) \right]^{1/\beta} = \exp \left[ \mu + \sigma \phi_{\text{SEV}}^{-1}(p) \right]

$\phi_{\text{SEV}}^{-1}(p)$ is the $p$ quantile for a standardized SEV (i.e., $\mu = 0, \sigma = 1$).
Weibull Distribution Likelihood for Right Censored Data

- The Weibull model is
  \[ \Pr(T \leq t) = F(t; \mu, \sigma) = \Phi_{\text{sev}} \left( \frac{\log(t) - \mu}{\sigma} \right). \]

- The likelihood has the form
  \[ L(\mu, \sigma) = \prod_{i=1}^{n} L_i(\mu, \sigma; \text{data}_i) \]
  \[ = \prod_{i=1}^{n} \left[ f(t_i; \mu, \sigma) \left[ 1 - F(t_i; \mu, \sigma) \right] \right]^{1 - \delta_i} \]
  \[ = \prod_{i=1}^{n} \left[ \frac{1}{\sigma} \Phi_{\text{sev}} \left( \frac{\log(t_i) - \mu}{\sigma} \right) \right]^{\delta_i} \times \left[ 1 - \Phi_{\text{sev}} \left( \frac{\log(t_i) - \mu}{\sigma} \right) \right]^{1 - \delta_i} \]
  \[ \delta_i = \begin{cases} 1 & \text{if } t_i \text{ is an exact observation} \\ 0 & \text{if } t_i \text{ is a right censored observation} \end{cases} \]
  \( \phi_{\text{sev}}(z) \) is the standardized smallest extreme value density.

Lognormal Distribution Likelihood for Right Censored Data

- The lognormal model is
  \[ \Pr(T \leq t) = F(t; \mu, \sigma) = \Phi_{\text{nor}} \left( \frac{\log(t) - \mu}{\sigma} \right). \]

- The likelihood has the form
  \[ L(\mu, \sigma) = \prod_{i=1}^{n} L_i(\mu, \sigma; \text{data}_i) \]
  \[ = \prod_{i=1}^{n} \left[ f(t_i; \mu, \sigma) \left[ 1 - F(t_i; \mu, \sigma) \right] \right]^{1 - \delta_i} \]
  \[ = \prod_{i=1}^{n} \left[ \frac{1}{\sigma} \Phi_{\text{nor}} \left( \frac{\log(t_i) - \mu}{\sigma} \right) \right]^{\delta_i} \times \left[ 1 - \Phi_{\text{nor}} \left( \frac{\log(t_i) - \mu}{\sigma} \right) \right]^{1 - \delta_i} \]
  \[ \delta_i = \begin{cases} 1 & \text{if } t_i \text{ is an exact observation} \\ 0 & \text{if } t_i \text{ is a right censored observation} \end{cases} \]
  \( \phi_{\text{nor}}(z) \) is the standardized normal density.
Weibull Relative Likelihood for the Shock Absorber Data

ML Estimates: \[ \hat{\mu} = 10.23 \text{ and } \hat{\sigma} = .3164 \]

\[ R(\mu, \log(\sigma)) = \frac{L(\mu, \log(\sigma))}{L(\hat{\mu}, \log(\hat{\sigma}))} \]

The approximate 95% likelihood confidence interval for \( \mu \) is: 10.06 10.54
The approximate 95% likelihood confidence interval for \( \beta \) is: 1.897 4.772
The approximate 95% likelihood confidence interval for \( \mu \) is: 10.06, 10.54
The approximate 95% likelihood confidence interval for \( \beta \) is: 1.897, 4.772
Six-Distribution ML Probability Plot of the Shock Absorber Data

Probability Plots with ML Estimates and Pointwise 95% Confidence Intervals

Fraction Failing

9/24/2004 Stat 567: Unit 8 - Ramón V. León
Weibull Probability Plot of Shock Absorber Failure Times (Both Failure Modes) with Maximum Likelihood Estimates and Normal-Approximation 95% Pointwise Confidence Intervals for $F(t)$

\begin{align*}
\text{Distance} &: 0.001, 0.003, 0.005, 0.01, 0.02, 0.03, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9 \\
\text{Kilometers} &: 6000, 8000, 10000, 12000, 14000, 18000, 22000, 26000
\end{align*}
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 
\( \mu \) and \( \sigma \) 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 \((\mu, \sigma)\) is

\[
R(\mu, \sigma) = \frac{L(\mu, \sigma)}{L(\hat{\mu}, \hat{\sigma})}
\]

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

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

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

Profile Likelihood

9/24/2004 Stat 567: Unit 8 - Ramón V. León

9/24/2004 Stat 567: Unit 8 - Ramón V. León
Weibull Profile Likelihood $R(\sigma)$ ($\sigma = 1/\beta$)
for the Shock Absorber Data
\[
R(\sigma) = \max_{\mu} \frac{L(\mu, \sigma)}{L(\hat{\mu}, \sigma)}
\]

Profile Likelihood

Sat Sep 18 22:00:18 EDT 2004

ShockAbsorber data
Profile Likelihood and 95% Confidence Interval
for beta from the Weibull Distribution
\[
R(\beta) = \max_{\mu} \frac{L(\mu, \beta)}{L(\hat{\mu}, \hat{\beta})}
\]
Large-Sample Approximate Theory for Likelihood Ratios for Parameter Vector Subset

**Need:** Inferences on subset $\theta_1$, from the partition $\theta = (\theta_1, \theta_2)'$.

- $k_1 = \text{length}(\theta_1)$.
- When $(\theta_1, \theta_2)' = (\mu, \sigma)$, profile likelihood for $\theta_1 = \mu$ is
  $$R(\mu) = \max_{\sigma} \left[ \frac{L(\mu, \sigma)}{L(\hat{\mu}, \hat{\sigma})} \right].$$
- If evaluated at the true $\theta_1 = \mu$, 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 $\theta_1$ is the set of all values of $\theta_1$ such that
  $$-2 \log[R(\theta_1)] < \chi^2_{1 - \alpha; k_1}$$
  or, equivalently, the set defined by
  $$R(\theta_1) > \exp\left[-\chi^2_{1 - \alpha; k_1}/2\right].$$
- Transformation of $\theta_1$ 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 $\sigma$)

$R(t_1, \sigma) = \frac{L(t_1, \sigma)}{L(\hat{t}_1, \hat{\sigma})}$
Confidence Regions and Intervals for Functions of $\mu$ and $\sigma$

- 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 $g(\mu, \sigma) = [g_1(\mu, \sigma), g_2(\mu, \sigma)]$.

- 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
\[ R(t_{1}) = \max_{\sigma} \frac{L(t_{1}, \sigma)}{L(t_{1}, \bar{\sigma})} \]

Profile Likelihood and 95% Confidence Interval
for 0.1 Quantile Distance from the Weibull Distribution

Shock Absorber data
Profile Likelihood and 95% Confidence Interval
for 0.1 Quantile Distance from the Weibull Distribution
Maximum Likelihood Estimation in JMP 5: Shock Absorber Data

(Compare results to those of Table 8.1, Page 178 of your textbook)
### ShockAbsorber Table

<table>
<thead>
<tr>
<th>Distance</th>
<th>Mode</th>
<th>Failure</th>
<th>Censor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mode1</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>Mode2</td>
<td>Failure</td>
<td>0</td>
</tr>
</tbody>
</table>

### Survival/Reliability Window

- **Distribution**
- **Model**
- **Survival and Reliability**

**Select Columns**
- Distance
- Mode
- Failure
- Censor

**Survival**
- Time to Event
- Optional Names
- Censor
- Censor

**Reliability**
- Freq
- By
- Censor

**Options**
- Include Censor Time
- Cancel
- OK

**Help**
Compare estimates here with those on Table 8.1 of the textbook.
Maximum Likelihood in JMP

Let $t_1, t_2, ..., t_n$ be either failure times or right-censored times (removal times).
Define censoring codes $c_1, c_2, ..., c_n$ as follows:

$$c_i = \begin{cases} 
0 & \text{if } t_i \text{ is a failure time} \\
1 & \text{if } t_i \text{ is a removal time}
\end{cases}$$

Then the likelihood can be written as:

$$L(\theta) = \prod_{i=1}^{n} \left[ f(t_i, \theta) \right]^{1-c_i} \left[ S(t_i, \theta) \right]^{c_i}$$

Maximum Likelihood in JMP: Loss

The negative of the loglikelihood is:

$$-l(\theta) = \sum_{i=1}^{n} \left[ (1-c_i)(-\log f(t_i, \theta) + c_i (-\log S(t_i, \theta)) \right]$$

JMP calls this kernel the “loss” function associated with maximum likelihood estimation

$$loss = \begin{cases} 
-\log f(t_i, \theta) & \text{if } c_i = 0 \\
-\log S(t_i, \theta) & \text{otherwise}
\end{cases}$$
JMP Loss Examples: Exponential Case

\[\text{loss} = \begin{cases} 
-\log \left( \frac{e^{-\frac{t_i}{\theta}}}{\theta} \right) = \frac{t_i}{\theta} + \log \theta & \text{if } c_i = 0 \\
-\log \left( e^{-\frac{t_i}{\theta}} \right) = \frac{t_i}{\theta} & \text{otherwise}
\end{cases}\]

JMP Loss Example: Weibull/Gumbel Case

The Gumbel survival function is:

\[S(x) = \exp \left[ -\exp \left( \frac{x-\mu}{\sigma} \right) \right] \quad -\infty < x < \infty\]

The Gumbel density function is:

\[f(x) = \frac{\exp \left[ \frac{x-\mu}{\sigma} \right]}{\sigma} \times S(x)\]
JMP Loss Example: Weibull/Gumbel Case

\[ \text{loss} = \begin{cases} 
- \log f(t_i, \theta) & \text{if } c_i = 0 \\
- \log S(t_i, \theta) & \text{otherwise} 
\end{cases} \]

\[ \text{loss} = \begin{cases} 
\log \sigma - \left( \frac{x - \mu}{\sigma} \right) + \exp \left( \frac{x - \mu}{\sigma} \right) & \text{if } c_i = 0 \\
\exp \left( \frac{x - \mu}{\sigma} \right) & \text{otherwise} 
\end{cases} \]
Compare with results on Table 8.1 of your textbook.

\[ .90 = S(b10) = \exp \left[ -\frac{b10}{\eta} \right] \]

\[ \Rightarrow -0.1054 = \log(.90) = -\frac{b10}{\eta} \]

\[ \Rightarrow \mu = \log \eta = \log b10 - \left( \frac{1}{\beta} \right) \log(0.1054) \]

\[ \Rightarrow \mu = \log b10 + 2.25\sigma \]
JMP Loss for Weibull B10

\[ \mu = \log b_{10} + 2.25 \sigma \]

\[ \text{loss} = \begin{cases} 
\log \sigma - \left( \frac{x - \mu}{\sigma} \right) + \exp \left( \frac{x - \mu}{\sigma} \right) & \text{if } c_i = 0 \\
\exp \left( \frac{x - \mu}{\sigma} \right) & \text{otherwise}
\end{cases} \]

JMP Using Local Variables
Compare to Table 8.1 of your textbook.
JMP loss for Weibull \( F(10,000) \)

\[
S(t) = \exp \left[ -\left( \frac{t}{\eta} \right)^{\beta} \right]
\]

\[\log \eta = \log t - \frac{1}{\beta} \log (-\log S(t))\]

\[\mu = \log t - \sigma \log (-\log [1 - F(t)])\]

\[\mu = \log 10000 - \sigma \log (-\log [1 - F(10000)])\]

JMP loss for Weibull \( F(10,000) \)

\[
\mu = \log 10000 - \sigma \log (-\log [1 - F(10000)])
\]

\[
\text{loss} = \begin{cases} 
\log \sigma - \left( \frac{x - \mu}{\sigma} \right) + \exp \left( \frac{x - \mu}{\sigma} \right) & \text{if } c_i = 0 \\
\exp \left( \frac{x - \mu}{\sigma} \right) & \text{otherwise}
\end{cases}
\]
Formula Using Local Variables

\[ \mu = \log(10000) - \sigma \log(1 - F10000) \]

\[ z = \frac{\log(\text{Distance (km)}) - \mu}{\sigma} \]

\[ \begin{align*}
\text{if } \text{Censor} &= 0 \\
\log(\sigma) &= z + \exp(z) \\
\text{else} &\quad \exp(z)
\end{align*} \]
Exercises

- Fit the exponential model to the Shock Absorber data and calculate the MLEs for theta, b10, and F(10,000)
- Fit lognormal model to the Shock Absorber data and calculate the MLEs for theta, b10, and F(10,000)
Asymptotic Theory of ML Estimation

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

- If evaluated at the true value of $\theta$, then asymptotically, (large samples) $\hat{\theta}$ has a $\text{MVN}(\theta, \Sigma_{\theta})$ and thus the Wald statistic
  \[
  (\hat{\theta} - \theta)' [\Sigma_{\hat{\theta}}]^{-1}(\hat{\theta} - \theta)
  \]
  has a chi-square distribution with $k$ degrees of freedom, where $k$ is the length of $\theta$.

- Here, $\Sigma_{\hat{\theta}} = I_{\theta}^{-1}$ is the large sample approximate covariance matrix where
  \[
  I_{\theta} = E \left[ -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta'} \right].
  \]

---

Asymptotic Theory for Wald’s Statistic

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

- Let $\Sigma_{\hat{\theta}}$ be a consistent estimator of $\Sigma_{\theta}$, the asymptotic covariance matrix of $\hat{\theta}$. For example,
  \[
  \Sigma_{\hat{\theta}} = \left[ -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta'} \right]^{-1}
  \]
  where the derivatives are evaluated at $\hat{\theta}$.

- Asymptotically, the Wald statistic
  \[
  w(\theta) = (\hat{\theta} - \theta)' \Sigma_{\hat{\theta}}^{-1}(\hat{\theta} - \theta)
  \]
  when evaluated at the true $\theta$, follows a chi-square distribution with $k$ degrees of freedom, where $k$ is the length of $\theta$. 
Asymptotic Theory for Wald’s Statistic – Continued

- An approximate $100(1 - \alpha)\%$ confidence region for $\theta$ is the set of all values of $\theta$ in the ellipsoid
  $$(\hat{\theta} - \theta)^T \hat{\Sigma}^{-1} (\hat{\theta} - \theta) \leq \chi^2_{(1 - \alpha, k)}.$$

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

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

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

- Assuming that $Z_{\hat{\mu}} = (\hat{\mu} - \mu)/\hat{s}_\hat{\mu} \sim \text{NOR}(0, 1)$ distribution, an approximate $100(1 - \alpha)\%$ confidence interval for $\mu$ is
  $$[\hat{\mu}, \hat{\mu}] = \hat{\mu} \pm z_{(1 - \alpha/2)} \hat{s}_\hat{\mu}$$

  where $\hat{s}_\hat{\mu} = \sqrt{\text{Var}(\hat{\mu})}$.

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

  where $w = \exp\left[z_{(1 - \alpha/2)} \hat{s}_\log(\hat{\sigma})/\hat{\sigma}\right]$ and $\hat{s}_\hat{\sigma} = \sqrt{\text{Var}(\hat{\sigma})}$. 
Parameter estimate variance-covariance matrix

\[ \begin{align*}
\mu & \quad \sigma \\
0.01208 & \quad 0.003990 \\
0.003990 & \quad 0.005353 \\
\end{align*} \]

Parameter estimate correlation matrix is

\[ \begin{align*}
\mu & \quad \sigma \\
1.0000 & \quad 0.4963 \\
0.4963 & \quad 1.0000 \\
\end{align*} \]

\[ \text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{se(X)se(Y)} \]
Normal-Approximation Confidence Intervals for Function $g_1 = g_1(\mu, \sigma)$

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

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

where

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

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

$\phi = g(\theta_1, \theta_2, \theta_3)$
$\hat{\phi} = g(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$

$$\left[\text{se}(\hat{\phi})^2\right] = \left(\frac{\partial g}{\partial \theta_1}\right)^2 v_{11} + \left(\frac{\partial g}{\partial \theta_2}\right)^2 v_{22} + \left(\frac{\partial g}{\partial \theta_3}\right)^2 v_{33} +$$

$$2\left[\left(\frac{\partial g}{\partial \theta_1}\right)\left(\frac{\partial g}{\partial \theta_2}\right) v_{12} + \left(\frac{\partial g}{\partial \theta_1}\right)\left(\frac{\partial g}{\partial \theta_3}\right) v_{13} + \left(\frac{\partial g}{\partial \theta_2}\right)\left(\frac{\partial g}{\partial \theta_3}\right) v_{23}\right]$$

where

$$V = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 l}{\partial \theta^2_1} & -\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & -\frac{\partial^2 l}{\partial \theta_1 \partial \theta_3} \\ -\frac{\partial^2 l}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 l}{\partial \theta^2_2} & -\frac{\partial^2 l}{\partial \theta_2 \partial \theta_3} \\ -\frac{\partial^2 l}{\partial \theta_3 \partial \theta_1} & -\frac{\partial^2 l}{\partial \theta_3 \partial \theta_2} & -\frac{\partial^2 l}{\partial \theta^2_3} \end{bmatrix}^{-1}$$

The derivatives are evaluated at the MLEs $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$.

$V$ is the variance-covariance matrix, and $l$ is the loglikelihood.
Normal-Approximation Confidence Interval for $F(t_c; \mu, \sigma)$

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

- The ML estimates $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ and $\hat{\Sigma}_\theta$ are available.
- The ML estimate for $F(t_c; \mu, \sigma)$ is
  \[
  \hat{F} = F(t_c; \hat{\mu}, \hat{\sigma}) = \Phi(\hat{\zeta}_c)
  \]
  where $\hat{\zeta}_c = [\log(t_c) - \hat{\mu}]/\hat{\sigma}$.
- In the context of Wald’s theory, however, there are many ways to obtain a confidence interval for $F(t_c; \mu, \sigma)$.

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

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

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

- The asymptotic normality of $Z_{\hat{F}} = (\hat{F} - F)/\hat{SE}_{\hat{F}}$
  \[
  [\hat{F}, \ F] = \hat{F}(t_c) \pm z_{(1-\alpha/2)}\hat{SE}_{\hat{F}}.
  \]
- The asymptotic normality of $Z_{\logit(\hat{F})} = [\logit(\hat{F}) - \logit(F)]/\hat{SE}_{\logit(\hat{F})}$
  \[
  [\hat{F}, \ F] = \left[ \frac{\hat{F}(t_c)}{\hat{F}(t_c) + (1 - \hat{F}(t_c)) \times w}, \ \frac{\hat{F}(t_c)}{\hat{F}(t_c) + (1 - \hat{F}(t_c)) / w} \right].
  \]
  where $w = \exp\{z_{(1-\alpha/2)}\hat{SE}_{\hat{F}}/[(\hat{F}(t_c)(1 - \hat{F}(t_c)))]\}$.
Confidence Interval for $F(t; \mu, \sigma)$—Continued

Comments:

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

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

ML Estimates for Biomedical Data

Here we show ML estimates (Weibull and lognormal) for the DMBA and the IUD data.
DMBA Carcinogen Data  
(Data from Pike, 1966. See Lawless, 1982)

Number of days until the appearance of vaginal carcinoma in \( n = 19 \) female rats painted with carcinogen DMBA

\[
143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304, 216^*, 244^*
\]

By the end of the study only 17 out of the 19 rats had developed carcinoma, so that two of the times (marked \(*\)) are censoring times.
IUD Data
(Data from WHO, 1987)

Number of weeks from the moment of initial IUD use until discontinuation because of bleeding problems in $n = 18$ women

10, 13*, 18*, 19, 23*, 30, 36, 38*, 54*, 56*, 59, 75, 93, 97, 104*, 107, 107*, 107*

By the end of the study only 9 out of the 18 women had developed bleeding, so that 9 of the times (marked *) are censoring times.
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 $\sigma$ (or Weibull $\beta$) is Given

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

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

- 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 $\beta$ is in doubt.
- Danger of misleading inferences.

---

Derivation

If $T$ is $\text{Wei}(\eta,\beta)$, then $T^\beta$ is $\text{Exp}(\eta^\beta)$ since

$$P(T^\beta > t) = P(T > t^{1/\beta}) = \exp\left[-\frac{t^{1/\beta}}{\eta}\right] = \exp\left[-\frac{t}{\eta^\beta}\right].$$

It follows that if $T_1,T_2,\ldots,T_n$ are $\text{Wei}(\eta,\beta)$ failure or removal times, then $T_1^\beta,T_2^\beta,\ldots,T_n^\beta$ are $\text{Exp}(\eta^\beta)$ failure or removal times.
Now for the exponential $\frac{\sum t_i^{\theta}}{r}$ estimates $\eta^\theta$ (with
standard error $\frac{\hat{r}}{\sqrt{r}}$) so that $\hat{\eta} = \frac{\sum t_i^{\theta}}{r}$ estimates $\eta$.

Also se $\left( \frac{\sum t_i^{\theta}}{r} \right)^{\frac{1}{\theta}} = \frac{1}{\beta} \left( \frac{\sum t_i^{\theta}}{r} \right)^{\frac{1}{\theta} - 1} \frac{\hat{\beta}}{\sqrt{r}} \left( \frac{\sum t_i^{\theta}}{r} \right)^{\frac{1}{\theta}} = \frac{1}{\beta} \left( \frac{\sum t_i^{\theta}}{r} \right)^{\frac{1}{\theta} - 1} \frac{\hat{\beta}}{\sqrt{r}} \left( \frac{\sum t_i^{\theta}}{r} \right)^{\frac{1}{\theta}}$

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 $t_{0.1} = 8$ thousand hours was not being met.

- ML estimate is $\hat{t}_{0.1} = 3.903$ thousand hours and an approximate 95% likelihood-ratio confidence interval for $t_{0.1}$ is [2.093, 22.144] thousand hours.

- Management also wanted to know how many additional failures could be expected in the next year.
Weibull Probability Plots Bearing Cage Fracture Data with Weibull ML Estimates and Sets of 95% Pointwise Confidence Intervals for $F(i)$ with $\beta$ Estimated, and Assumed Known Values of $\beta = 1.5$, 2, and 3.

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

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

Component B Safe-Life 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 param-
  eter 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.

<table>
<thead>
<tr>
<th>Hours:</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
<th>3500</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Units:</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

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

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:

  $$\lim_{z \to \infty} \frac{z^2 \phi^2(z)}{\Phi(z)} = 0$$

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

- In non-regular models, asymptotic behavior is more complicated (e.g., behavior depends on $\theta$), 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. \( \theta \)).
- Bounded derivatives of likelihood.
- Can exchange the order of differentiation of log likelihood w.r.t. \( \theta \) 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.