Stat 486 / 886 Survival Analysis

Chapter 3. Parametric Analysis of Univariate Lifetime

Professor: Wenyu Jiang

Department of Mathematics and Statistics
Queen's University

Acknowledgement: Profs. Jerry Lawless, David Matthews (Waterloo),
Profs. Paul Peng, Dongsheng Tu (Queen's)

Sections of Chapter 3

3.1 Likelihood Methods

- A quick review

Model: Failure time T has pdf/pmf $f(t;\theta)$, and survival function $S(t;\theta)$. Dimension of θ is p, θ takes values in Θ.

Likelihood: $L(\theta) \propto P(\text{observed data; } \theta)$.
Log-likelihood: $l(\theta) = \log L(\theta)$.
Maximum-likelihood estimator (MLE): $\hat{\theta}$, the point in Θ that maximizes $l(\theta)$, or $L(\theta)$.

Score eq'n: $U(\theta) = 0$

Score function: $U(\theta) = \frac{\partial l}{\partial \theta} = (\frac{\partial l}{\partial \theta_1}, \frac{\partial l}{\partial \theta_2}, \dots, \frac{\partial l}{\partial \theta_p})^{\mathsf{T}}$.
Information matrix: $I(\theta) = -\frac{\partial^2 l}{\partial \theta \partial \theta^{\mathsf{T}}} = -\left(\frac{\partial^2 l}{\partial \theta_r \partial \theta_c}\right)_{p \times p}$, where $r = 1, \dots, p$, $c = 1, \dots, p$.

symmetric.

Fisher (expected) information matrix: $\mathscr{I}(\theta) = \mathrm{E}[I(\theta)]$. symmetric.
Observed information matrix: $I(\hat{\theta})$. symmetric.

Statistics and Large-Sample Results for Inference about θ:

1. Score statistic: $U(\theta)$, the parameter $\theta \in \Theta$.

Large-sample distribution for the score and related statistics:

- Convenient notation,

$$U(\theta_0) \approx \mathrm{N}_p(0, \mathscr{I}(\theta_0))$$

  where θ_0 is the true value of θ (in the true model of the failure time T).

- Statement about convergence in distribution.
  Under mild "regularity" conditions on the model,

$$\mathscr{I}^{-\frac{1}{2}}(\theta_0)U(\theta_0) \xrightarrow{D} \mathrm{N}_p(0, 1_{p\times p}), \text{ or } U^{\mathsf{T}}(\theta_0)\mathscr{I}^{-1}(\theta_0)U(\theta_0) \xrightarrow{D} \chi^2_p.$$

$$U^T(\theta_0)\,\mathcal{I}^{-\frac{1}{2}}(\theta_0)\,\mathcal{I}^{-\frac{1}{2}}(\theta_0)\,U(\theta_0)$$

$$= W^{\mathsf{T}}(\theta)\, W(\theta).$$

2. Wald statistic:

$W(\theta) = I^{\frac{1}{2}}(\hat{\theta})(\hat{\theta} - \theta)$, or $W^2(\theta) = (\hat{\theta} - \theta)^{\mathsf{T}} I(\hat{\theta})(\hat{\theta} - \theta)$.

Large-sample distribution for MLE $\hat{\theta}$ and related statistics:

- Convenient notation, $\hat{\theta} \approx \mathrm{N}_p(\theta_0, \mathscr{I}^{-1}(\theta_0))$.
- Useful estimator, $\hat{\mathrm{Var}}(\hat{\theta}) = I^{-1}(\hat{\theta})$.
- Statement about convergence in distribution.
  Under mild "regularity" conditions on the model,

$$W(\theta_0) = \underbrace{I^{\frac{1}{2}}(\hat{\theta})}_{[\hat{\mathrm{Var}}(\hat{\theta})]^{-\frac{1}{2}}}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathrm{N}_p(\underset{\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_p}{\underset{\|}{0}}, \underbrace{1_{p\times p}}_{= \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}}), \text{ or } \underbrace{W^2(\theta_0)}_{1 \text{ dimensional.}} \xrightarrow{D} \chi^2_p.$$

$$[\hat{\mathrm{Var}}(\hat{\theta})]^{-\frac{1}{2}}(\hat{\theta} - \theta_0).$$

Wald statistic based inference for a component of θ

$\theta = (\mu, \sigma)$.

$H_0: \mu = 0$?

Suppose $I^{-1}(\hat\theta) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ & \dots & \dots & \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix}$.

Recall that $\widehat{\mathrm{Var}}(\hat\theta) = I^{-1}(\hat\theta)$.

Let $\hat\theta_i$ be the MLE of θ_i, the ith component of θ, $i = 1, \dots, p$.
Then $\widehat{\mathrm{Var}}(\hat\theta_i) = \left[I^{-1}(\hat\theta)\right]_{i,i} = a_{ii}$ and $\widehat{\mathrm{Cov}}(\hat\theta_i, \hat\theta_j) = \left[I^{-1}(\hat\theta)\right]_{i,j} = a_{ij}$. $(i \neq j)$

Inference about θ_i can be based on the (component-wise) Wald statistic

$$\frac{\hat\theta_i - \theta_i}{\sqrt{a_{ii}}}, \text{ and the large-sample result that } \frac{\hat\theta_i - \theta_{0i}}{\sqrt{a_{ii}}} \approx \mathrm{N}(0,1),$$

where θ_{0i} is the true value of θ_i.

3. Likelihood ratio (LR) statistic:

$$\Lambda(\theta) = -2\log\frac{L(\theta)}{L(\hat\theta)} = 2[l(\hat\theta) - l(\theta)].$$

Large-sample distribution for the LR statistic:

- Statement about convergence in distribution.
  Under mild "regularity" conditions on the model,

$$\Lambda(\theta_0) = -2\log\frac{L(\theta_0)}{L(\hat\theta)} = \underline{2[l(\hat\theta) - l(\theta_0)]} \xrightarrow{D} \chi^2_p.$$

Test. $H_0: \theta = \theta_0$ vs. $H_a: \theta \neq \theta_0$.

$\Lambda(\theta_0) = 2[\ell(\hat\theta) - \ell(\theta_0)] \xrightarrow{D} \chi^2_p$.

Use this result for testing H_0.

$\hat\theta$: mle obtained for all θ.

Another version of LR statistic: inference about components of θ

Suppose for $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(1)}$ $\cancel{\in \mathbb{R}^q}$ is q dimensional and $\theta^{(2)}$ $\cancel{\in \mathbb{R}^{p-q}}$ is p-q dimensional, with $q \le p$, we are interested in testing

$$H_0 : \theta^{(2)} = \theta_0^{(2)} \text{ versus } H_a : \theta^{(2)} \neq \theta_0^{(2)}.$$

Let Θ be the entire parameter space of θ, Θ is in $\mathbb{R}^p$.
Let Θ_0 be the restricted parameter space under H_0, Θ_0 is in $\mathbb{R}^q$.
Under H_0, the LR statistic

$= 2[\ell(\tilde{\theta}) - \ell(\hat{\theta}^{(1)}, \theta_0^{(2)})]$

under H_0.

$$\Lambda(\theta_0^{(2)}) = -2\log \frac{L(\hat{\theta}^{(1)}, \theta_0^{(2)})}{L(\tilde{\theta})} \approx \chi^2_{p-q},$$

where $\tilde{\theta}$ is the MLE of θ in Θ, the entire parameter space,
and $\hat{\theta}^{(1)}$ is the MLE of $\theta^{(1)}$ in Θ_0, the parameter space under H_0.
This result can be used to test H_0 about the components $\theta^{(2)}$ of θ.

Example 3.1.1 Equipment Field Failure

See data given in Example 1.2.1., and also refer to Example 1.4.1.

Modeled the data by an exponential distribution, $\underbrace{T \sim \text{Exp}(\theta)}$.
Pdf of T: $f(t;\theta) = \frac{1}{\theta} e^{-t/\theta}$.

$D = \{(x_i, \delta_i) : i = 1, \cdots, n\}$

Find the 95% confidence interval for θ.

$$L(\theta) = \prod_{i=1}^{n} [f(x_i)]^{\delta_i} [S(x_i)]^{1-\delta_i}$$

$$= \prod_{i=1}^{n} \left[\frac{1}{\theta} e^{-\frac{x_i}{\theta}}\right]^{\delta_i} \left[e^{-\frac{x_i}{\theta}}\right]^{1-\delta_i}$$

$$= \frac{1}{\theta^{\Sigma\delta_i}} e^{-\frac{\Sigma x_i}{\theta}}$$

$$\ell(\theta) = \log L(\theta) = -\sum_{i=1}^{n} \delta_i \log\theta - \frac{\Sigma x_i}{\theta}.$$

$$u(\theta) = \ell'(\theta) = -\frac{\Sigma\delta_i}{\theta} + \frac{\Sigma x_i}{\theta^2}$$

Solve $u(\theta) = 0$

$$\Rightarrow \text{MLE } \hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \delta_i}$$

Pay attention to Fisher information, and observed information.

Information: $I(\theta) = -\ell''(\theta) = -\dfrac{\sum \delta_i}{\theta^2} + \dfrac{2\sum x_i}{\theta^3}$.

$$\left[\ell''(\theta) = \frac{\sum \delta_i}{\theta^2} - \frac{2\sum x_i}{\theta^3} \right]$$

Fisher information: $\mathcal{I}(\theta) = E[I(\theta)] = -\dfrac{E(\sum \Delta_i)}{\theta^2} + \dfrac{2E(\sum X_i)}{\theta^3}$.

In most cases, these E( ) cannot be obtained.

Observed information: $I(\hat\theta) = -\dfrac{\sum \delta_i}{\hat\theta^2} + \dfrac{2\sum x_i}{\hat\theta^3} = \underbrace{\dfrac{\sum x_i}{\hat\theta^3}}$.

$$\left(\hat\theta = \frac{\sum x_i}{\sum \delta_i}\right).$$

• In R, the exponential failure time model is parameterized as the log location-scale model with location parameter μ, while the scale parameter is fixed, $\sigma = 1$. Too see this, verify the following result.

Exercise: For $T \sim \text{Exp}(\overset{\text{hazard rate}}{\lambda})$ with $S_T(t) = e^{-\lambda t}$, $t > 0$, an equivalent model is to describe $\log T = \mu + W$ with $W \sim \text{EV}(0,1)$ whose survival function is $S_W(w) = e^{-e^w}$, $-\infty < w < \infty$.

λ μ relationship.

$\underbrace{\beta_0 + \beta_1 z_1 + \beta_2 z_2}$

Recall (for large n):

$$\hat{\theta} \approx N(\underset{\substack{\uparrow \\ \text{true value of } \theta.}}{\theta_0}, \ \mathcal{I}^{-1}(\theta_0))$$

Estimate the var of $\hat{\theta}$ by

$$\widehat{Var}(\hat{\theta}) = \underbrace{I^{-1}(\hat{\theta})} = [I(\hat{\theta})]^{-1}.$$

An approx. $100(1-\alpha)\%$ C.I. for θ:

$$\hat{\theta} \pm \underbrace{z_{1-\frac{\alpha}{2}}}_{\text{(Lower) quantile of } Z \sim N(0,1).} \ \underbrace{I^{-\frac{1}{2}}(\hat{\theta})}_{s.e.(\hat{\theta}) = \sqrt{\widehat{Var}(\hat{\theta})}.} .$$

std. err.

Apply to the data: $n=10$, $\sum_{i=1}^{10} \delta_i = 7$, $\sum_{i=1}^{10} x_i = 308$.

$$\hat{\theta} = \frac{308}{7} = 44 \text{ days.} \qquad I(\hat{\theta}) = \frac{308}{44^3}$$

$$s.e.(\hat{\theta}) = \sqrt{I^{-1}(\hat{\theta})} = 16.63.$$

95% CI for θ:

$$44 \pm \underset{z_{0.975}}{1.96} \times 16.63 \Rightarrow 11.4 < \theta < 76.6 \text{ days.}$$

Fit the exponential model to the data of Example 3.1.1 using R.

- Model for the data: $T \sim$ Exp (θ), where $\theta = \mathrm{E}(T)$. $\theta = e^{\mu}$.
  From the fit of the exponential model, the estimated location parameter $\hat{\mu} = 3.784$.
  The model fitted to the data has mean $\hat{\theta} = e^{\hat{\mu}} = 44$ days.
- Does the exponential model fit the data well?
  Compare the fit of the exponential model (for $S(t)$) to the KM estimate.

- Comment: The exponential model agrees well with the KM estimate in capturing the overall trend of the survival function.
  The exponential model fits the data very well.

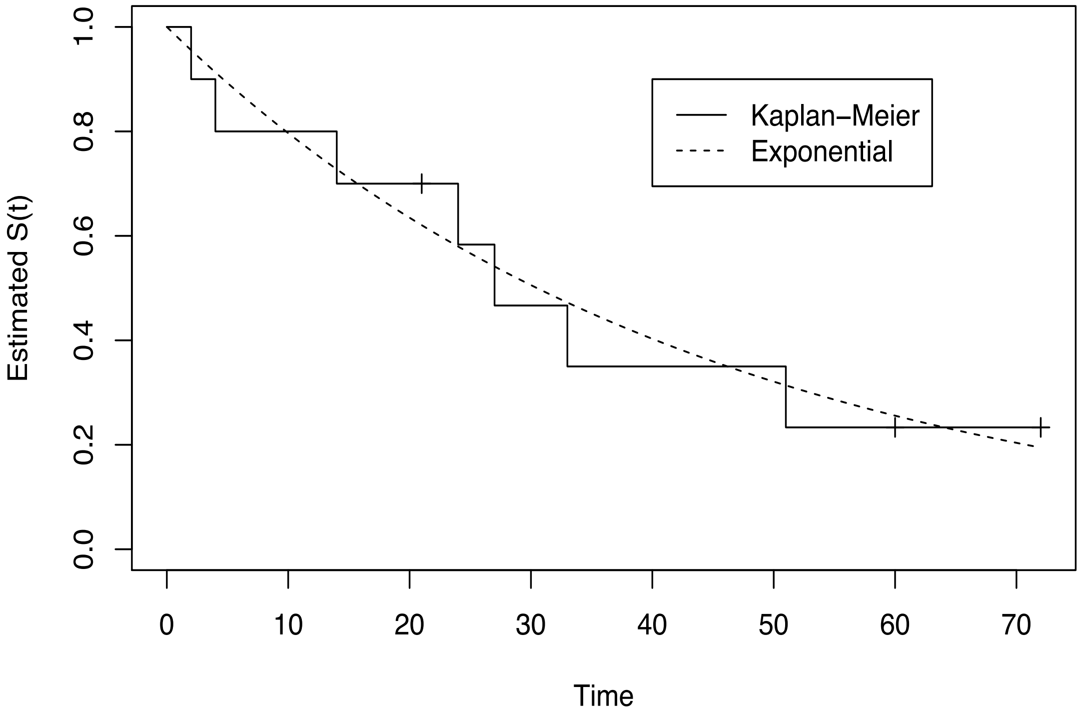


Figure: Estimated survival functions by the exponential model and Kaplan-Meier method for the time to equipment failure.

```
> library(survival)
> data<-read.table("eg121.txt", header=T)
> x<-data$time
> delta<-data$status
> #Fit univariate Exp model.
> ?survreg
>
> fit.exp<- survreg(Surv(x, delta)~1, dist = "exp")
> print(summary(fit.exp))
Call:
survreg(formula = Surv(x, delta) ~ 1, dist = "exp")
            Value Std. Error  z      p
(Intercept) 3.784      0.378 10 <2e-16
Scale fixed at 1
Exponential distribution
Loglik(model)= -33.5   Loglik(intercept only)= -33.5
Number of Newton-Raphson Iterations: 4
n= 10
```

p-val for testing this (2-sided test).

$3.784 = \hat{\mu}$.

0.378: s.e.$(\hat{\mu})$

10: Wald stat. $\dfrac{\hat{\mu}-0}{s.e(\hat{\mu})}$

$<2e{-16}$: for testing $H_0: \mu = 0$. vs. $H_a: \mu \neq 0$.

Scale fixed at 1: $\sigma = 1$

```
> fit.exp$coeff
(Intercept)
μ̂ =  3.78419
> mu<-fit.exp$coeff
> theta<-exp(mu)   # Estimated mean failure time.
> theta
(Intercept)
         44
> # Compare the exponential model to KM estimate.
> fit.km<-survfit(Surv(x, delta)~1)
> plot(fit.km, xlab = "Time",ylab = "Estimated S(t)",
conf.int=F,mark.time=T)
> t<-0:72
> # S(t) estimated from exponential model.
> st<-exp(-t/theta)
> lines(t,st,lty=2)
> legend(40,0.9,c("Kaplan-Meier","Exponential"),lty=1:2)
```

3.2 Parametric Models and Analysis

- Weibull failure time model

  Assume $T \sim$ Weibull with $S_T(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$, $t > 0$.

  Recall that we can also describe T as a log location-scale model with parameters (μ, σ). That is,

  $S_W(w) = e^{-e^{w}}$,

  $$\log T = \mu + \sigma W, \text{ with } W \sim \text{EV}(0,1), \quad -\infty < w < \infty$$

  or equivalently, $Y = \log T$ has survival function

  $S_Y(y) = e^{-e^{\frac{y-\mu}{\sigma}}}$ $\quad -\infty < y < \infty$. $\quad Y \sim EV(\mu, \sigma)$.

  With right censored data, let $X_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$.

  Data:

  $D = \{(x_i, \delta_i) : i = 1, \ldots, n\}$, or equivalently, $D = \{(y_i, \delta_i) : i = 1, \ldots, n\}$, where $y_i = \log x_i$.

Likelihood function $L(\mu, \sigma) = \overset{\prod}{\sum}_{i=1}^{n} [f_Y(y_i)]^{\delta_i} [S_Y(y_i)]^{1-\delta_i}$

$$= \prod_{i=1}^{n} \left[\frac{1}{\sigma} e^{\frac{y_i-\mu}{\sigma}} e^{-e^{\frac{y_i-\mu}{\sigma}}} \right]^{\delta_i} \left[e^{-e^{\frac{y_i-\mu}{\sigma}}} \right]^{1-\delta_i}$$

$$= \frac{1}{\sigma^{\sum \delta_i}} e^{\sum \frac{y_i-\mu}{\sigma} \delta_i} e^{-\sum_{i=1}^{n} e^{\frac{y_i-\mu}{\sigma}}}$$

Let $d = \sum_{i=1}^{n} \delta_i$; $w_i = \frac{y_i-\mu}{\sigma}$

Log-likelihood

$$= \frac{1}{\sigma^d} e^{\sum_{i=1}^{n} w_i \delta_i} \cdot e^{-\sum_{i=1}^{n} e^{w_i}}$$

$$\ell(\mu, \sigma) = \log L(\mu, \sigma) = -d \log \sigma + \sum_{i=1}^{n} w_i \delta_i - \sum_{i=1}^{n} e^{w_i}$$

Score functions

$$\frac{\partial \ell}{\partial \mu} = -\sum_{i=1}^{n} \frac{\delta_i}{\sigma} + \sum_{i=1}^{n} \frac{e^{w_i}}{\sigma} = -\frac{d}{\sigma} + \frac{1}{\sigma} \sum e^{w_i} \overset{①}{=} ; \quad \Rightarrow \frac{\partial w_i}{\partial \mu} = -\frac{1}{\sigma}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{d}{\sigma} - \frac{\sum_{i=1}^{n} w_i \delta_i}{\sigma} + \sum e^{w_i} \frac{w_i}{\sigma} \overset{②}{=} ; \quad \frac{\partial w_i}{\partial \sigma} = -\frac{y_i-\mu}{\sigma^2} = -\frac{w_i}{\sigma}$$

Solve $\frac{\partial \ell}{\partial \mu} = 0$, $\frac{\partial \ell}{\partial \sigma} = 0$ for $\mu, \sigma \Rightarrow$ mle's $\hat{\mu}, \hat{\sigma}$.

Solving the score equations gives the mle's. No close form expressions.

Solved numerically, often using Newton-Raphson method.

Information matrix $I(\mu,\sigma) = - \begin{pmatrix} \frac{\partial^2 \ell}{\partial \mu^2} & \frac{\partial^2 \ell}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \sigma \partial \mu} & \frac{\partial^2 \ell}{\partial \sigma^2} \end{pmatrix}$

$$= \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^{n} e^{w_i} & -\frac{d}{\sigma^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n} e^{w_i} + \frac{1}{\sigma^2}\sum_{i=1}^{n} w_i e^{w_i} \\ & \cdots\cdots \end{pmatrix}$$

Information matrix is needed for estimating the variance of the mle's.

Known results: $\widehat{\text{Var}}\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} = I^{-1}(\hat{\mu},\hat{\sigma}).$ If using R, $\widehat{Var}\begin{pmatrix} \hat{\mu} \\ \hat{\phi} \end{pmatrix} = I^{-1}(\hat{\mu},\hat{\phi})$

What if we need to estimate the variance of a function of the mle's, $g(\hat{\mu},\hat{\sigma})$?

$g(\hat{\mu},\hat{\phi})$.

Theorem: Δ - Method (Vector Version)

Let Z_n be a random vector (in $\mathbb{R}^p$), such that $Z_n \to \theta$ in probability as $n \to \infty$, where $\theta = (\theta_1, \ldots, \theta_p)^T \in \mathbb{R}^p$. For a continuous function $g()$ from $\mathbb{R}^p$ to $\mathbb{R}$, we have

1. $g(Z_n) \to g(\theta)$ in probability;
2. $\mathrm{Var}[g(Z_n)] \approx \left(\frac{\partial g}{\partial \theta}\right)^T \mathrm{Var}(Z_n) \left(\frac{\partial g}{\partial \theta}\right)$, where $\frac{\partial g}{\partial \theta} = \left(\frac{\partial g}{\partial \theta_1}, \ldots, \frac{\partial g}{\partial \theta_p}\right)^T$.

We can then estimate $\mathrm{Var}[g(Z_n)]$ by

$$\widehat{\mathrm{Var}}[g(Z_n)] = \left(\widehat{\frac{\partial g}{\partial \theta}}\right)^T \mathrm{Var}(Z_n) \left(\widehat{\frac{\partial g}{\partial \theta}}\right).$$

$\widehat{\frac{\partial g}{\partial \theta}}$: estimating θ by Z_n in $\frac{\partial g}{\partial \theta}$.

Example 3.2.1 Ball Bearing Endurance

The data below arose from a test on the endurance of ball bearings. The data give the number of million revolutions to failure for 23 ball bearings tested under fixed conditions. One objective is to estimate the lifetime distribution of the ball bearings. (Reference: Lawless 2002)

17.88 28.92 33.00 41.52 42.12 45.60 48.40 51.84 51.96 54.12 55.56 67.80
68.64 68.64 68.88 84.12 93.12 98.64 105.12 105.84 127.92 128.04 173.40

Fit a Weibull model to the data.
Engineering theory suggests this model.

In R, "survreg" fct'n takes parameters (μ, ϕ)

where $\phi = \log \sigma$.

$l(\mu,\phi) = \log L(\mu,\phi)$.

Based on Likelihood $L(\mu, \phi)$.

Score fct'ns $\frac{\partial l(\mu,\phi)}{\partial \mu}$, and $\frac{\partial l(\mu,\phi)}{\partial \phi}$.

Statistical inference based on the Weibull model

- Confidence interval for σ

$\phi = \log\sigma, \quad \sigma = e^{\phi}.$

Method 1: 95% CI for ϕ: $\hat{\phi} \pm \underbrace{z_{1-\frac{\alpha}{2}}}_{0.975}\sqrt{\widehat{Var}(\hat{\phi})}$

$-0.743 \pm 1.96\sqrt{0.0244}$

$\hat{\phi}_L = -1.049 < \phi < -0.437 = \hat{\phi}_u$

95% CI for σ:

$e^{\hat{\phi}_L} < e^{\phi} < e^{\hat{\phi}_u}$

$0.35 < \sigma < 0.65$

Method 2. $\sigma = e^{\phi} \triangleq g(\phi) \quad \hat{\sigma} = e^{\hat{\phi}}. \quad \frac{dg}{d\phi} = e^{\phi}.$

By Δ-method,

$\widehat{Var}(\hat{\sigma}) = \left(\widehat{\frac{dg}{d\phi}}\right)^2 \widehat{Var}(\hat{\phi}) = \left(e^{\hat{\phi}}\right)^2 \widehat{Var}(\hat{\phi}) \doteq \hat{\sigma}^2 \widehat{Var}(\hat{\phi})$

$= 0.00553$

95% C.I. for σ: $\hat{\sigma} \pm Z_{0.975} \sqrt{\widehat{Var}(\hat{\sigma})}$

$$= 0.476 \pm 1.96 \sqrt{0.00553}$$

$$0.43 < \sigma < 0.52.$$

The 2 methods are asymptotically equivalent.
$(n \to \infty)$.

- Confidence interval for quantile, for example, $t_{0.1}$

$$y_{0.1} = \log t_{0.1} \qquad (Y = \log T).$$

$$\Big[\ S_T(t_{0.1}) = 0.9.$$

$$0.9 = S_T(t_{0.1}) = P(T \geq t_{0.1}) = P(\log T \geq \log t_{0.1}) = P(Y \geq \log t_{0.1})$$
$$= S_Y(\log t_{0.1}).$$

$$y_{0.1} = \log t_{0.1}\ \Big].$$

Find 95% CI for $y_{0.1}$ first.

$T \sim$ Weibull, $\quad Y = \log T \sim EV(\mu, \sigma)$

$$S_Y(y) = e^{-e^{\frac{y-\mu}{\sigma}}}$$

$$0.9 = S_Y(y_{0.1}) = e^{-e^{\frac{y_{0.1}-\mu}{\sigma}}} \Rightarrow \frac{y_{0.1}-\mu}{\sigma} = \log(-\log 0.9) = -2.25.$$

$$y_{0.1} = \mu - 2.25\sigma = \mu - 2.25e^{\phi} \triangleq g(\mu, \phi)$$

$$\hat{y}_{0.1} = \hat{\mu} - 2.25 e^{\hat{\phi}} = \hat{\mu} - 2.25\hat{\sigma} = 3.33$$

By Δ-method,

$$\widehat{Var}(\hat{y}_{0.1}) = \underbrace{\begin{pmatrix} \widehat{\frac{\partial g}{\partial \mu}} \\ \widehat{\frac{\partial g}{\partial \phi}} \end{pmatrix}^T \underbrace{\widehat{Var}\begin{pmatrix} \hat{\mu} \\ \hat{\phi} \end{pmatrix}} \begin{pmatrix} \widehat{\frac{\partial g}{\partial \mu}} \\ \widehat{\frac{\partial g}{\partial \phi}} \end{pmatrix}},$$

$$\frac{\partial g}{\partial \mu} = 1, \quad \widehat{\frac{\partial g}{\partial \phi}} = -2.25 e^{\hat{\phi}} = -2.25\hat{\sigma}$$

$$\widehat{Var}(\hat{y}_{0.1}) = \begin{pmatrix} 1 \\ -2.25 \times \underbrace{0.476}_{\hat{\sigma}} \end{pmatrix}^T \underbrace{\begin{pmatrix} 0.0110 & -0.0059 \\ & 0.0244 \end{pmatrix}} \begin{pmatrix} 1 \\ -2.25 \times 0.476 \end{pmatrix}$$

$$= 0.051.$$

95% CI for $y_{0.1}$ is $\hat{y}_{0.1} \pm 1.96\sqrt{\widehat{Var}(\hat{y}_{0.1})}$

$$\Rightarrow \underbrace{2.89} < y_{0.1} < \underbrace{3.77}.$$

95% C.I. for $t_{0.1} = e^{y_{0.1}}$ is

$$e^{2.89} < t_{0.1} < e^{3.77}$$

$$17.94 < t_{0.1} < 43.49$$

- Confidence interval for $S(t)$, for example, $S(60)$

$[\; Y = \log T, \quad \underline{S_T(t)} = P(T \geq t) = P(Y \geq \log t) = \underline{S_Y(\log t)} \;]$.

$Y = \log T \sim EV(\mu, \sigma) \qquad [\, T \sim \text{Weibull} \,]$.

$S_Y(y) = e^{-e^{\frac{y-\mu}{\sigma}}}$

$S_T(t) = e^{-e^{\frac{\log t - \mu}{\sigma}}}, \quad S_T(60) \qquad$ 1

$*$ First find C.I. for $\log[-\log S_T(t)] = \dfrac{\log t - \mu}{\sigma} = (\log t - \mu) e^{-\phi}$

$g(\mu, \phi) = (\log t - \mu) e^{-\phi}$.

95% CI for $\log[-\log S_T(\overset{=60}{t})]$ is

$\log[-\log \hat{S}_T(t)] \pm 1.96 \sqrt{\widehat{\text{var}}\{\log[-\log \hat{S}_T(t)]\}}$

By Δ-method. $\begin{pmatrix} \frac{\partial g}{\partial \mu} \\ \frac{\partial g}{\partial \phi} \end{pmatrix} = \begin{pmatrix} -e^{-\phi} \\ -(\log t - \mu) e^{-\phi} \end{pmatrix}$

Find

$$\widehat{Var}\{\log[-\log \widehat{S_T(t)}]\} = \begin{pmatrix} \frac{\widehat{\partial g}}{\partial \mu} \\ \frac{\widehat{\partial g}}{\partial \phi} \end{pmatrix}^T \widehat{Var}\begin{pmatrix} \hat{\mu} \\ \hat{\phi} \end{pmatrix} \begin{pmatrix} \frac{\widehat{\partial g}}{\partial \mu} \\ \frac{\widehat{\partial g}}{\partial \phi} \end{pmatrix}$$

(for $t = 60$)

$$= 0.074$$

...

* 95% CI for $\underbrace{\log[-\log S_T(60)]}$ is

$$(\underbrace{-1.18}_{\hat{\varphi}_L}, \; \underbrace{-0.12}_{\hat{\varphi}_u}).$$

95% CI for $\underbrace{S_T(60)}$ is

$$\left(e^{-e^{\hat{\varphi}_u}}, \; e^{-e^{\hat{\varphi}_L}} \right)$$

$$\Rightarrow (0.41, \; 0.74).$$

[Find $\widehat{S_T(60)}$? $\hat{S}_T(60) = 0.594.$ from 1]

```
>  library(survival)
>  ballbear<-read.table("eg321.txt", header=T)
>  fit.weib<-survreg(Surv(time,status)~1,data=ballbear)
>  summary(fit.weib)
Call:
survreg(formula = Surv(time, status) ~ 1, data = ballbear)
              Value Std. Error     z       p
(Intercept)   4.405      0.105 41.93 <2e-16
Log(scale)   -0.743      0.156 -4.75  2e-06
Scale= 0.476
Weibull distribution
Loglik(model)= -113.7   Loglik(intercept only)= -113.7
Number of Newton-Raphson Iterations: 6
n= 23
>  ?survreg            # Help file for "survreg".
>  ?survreg.object     # Explains objects in output.
```

$$\log T = \underbrace{\beta_0 + \cancel{\beta_1 Z_1} + \cancel{\beta_2 Z_2}}_{\mu} + \sigma W. \qquad W \sim EV(0,1)$$

$\hat{\mu} = 4.405$

$\hat{\phi} = -0.743$

$\hat{\sigma} = e^{\hat{\phi}}$.

$$z = \frac{\hat{\phi} - 0}{s.e.(\hat{\phi})}.$$

$H_0: \phi = 0$. vs. $H_a: \phi \neq 0$.

```
>  fit.weib$coef
(Intercept)
   4.405188
>  fit.weib$icoef
(Intercept)  Log(scale)
  4.4051883  -0.7428164
>  fit.weib$scale
[1] 0.4757721
>  fit.weib$var
            (Intercept)   Log(scale)
(Intercept)  0.011035513 -0.005402699
Log(scale)  -0.005402699  0.024450317
```

In SAS, log loc-scale models are fitted by "lifereg", it take (μ, σ) as the parameters.

$\hat{\mu}$

$\hat{\mu}$ $\hat{\phi}$

$\hat{\sigma}$.

$$\widehat{Var}\begin{pmatrix}\hat{\mu}\\ \hat{\phi}\end{pmatrix} = I^{-1}(\hat{\mu}, \hat{\phi})$$

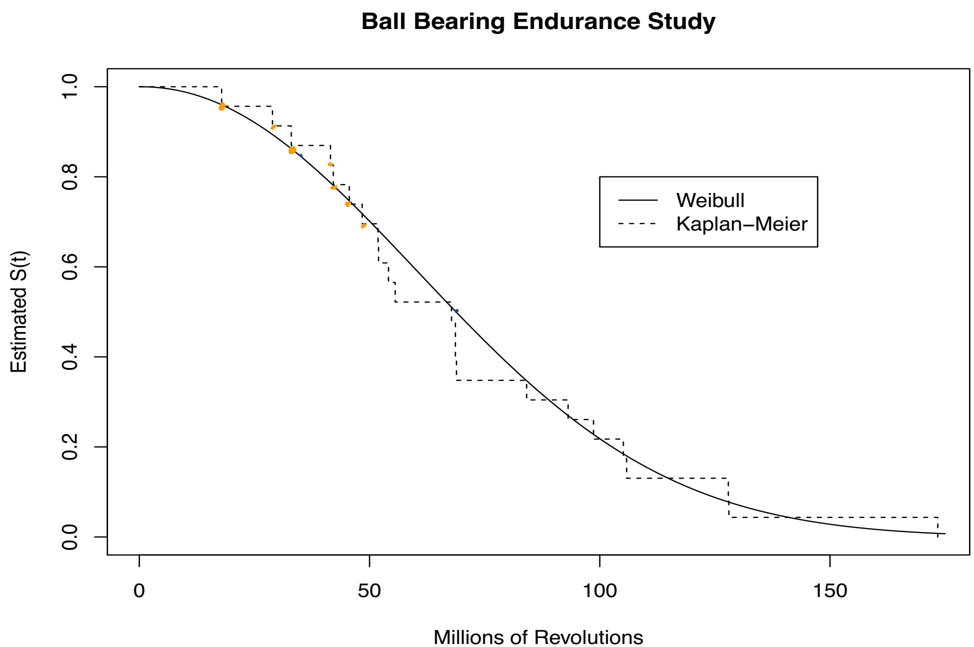
$\widehat{Var}(\hat{\mu})$

$\widehat{cov}(\hat{\mu}, \hat{\phi})$.

$\widehat{Var}(\hat{\phi})$

Model checking:

Plot the estimated survival functions based on Weibull model and Kaplan-Meier method. Comment on model fit!



```
> sigma<-fit.weib$scale
> mu<-fit.weib$coef
> fit.km<-survfit(Surv(time, status)~1,data=ballbear)
> t<-0:175
> sf.weib<-exp(-exp((log(t)-mu)/sigma))
> plot(fit.km,xlab="Millions of Revolutions",
ylab="Estimated S(t)",main="Ball Bearing Endurance Study",
conf.int=F,lty=2)
> lines(t, sf.weib,lty=1)
> legend(100, 0.8, c("Weibull","Kaplan-Meier"),lty=1:2)
```

$S_Y(y) = e^{-e^{\frac{y-\mu}{\sigma}}}$;

$$\left.\begin{aligned} S_T(t) &= P(T \geq t) = P(Y \geq \log t) \\ &= S_Y(\log t) \\ &= e^{-e^{\frac{\log t - \mu}{\sigma}}} \end{aligned}\right\}$$

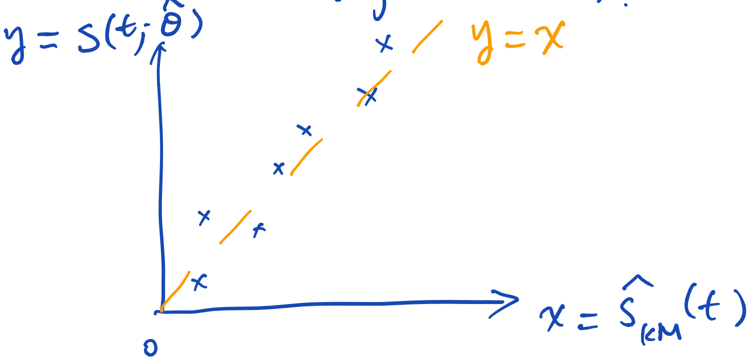
s.f. of $T \sim$ Weibull.

3.3 Model Checking

3.3.1 Graphical Methods

1. Plot $\underbrace{S(t;\hat{\theta})}_{\text{parametric model fitted to data}}$ and $\hat{S}_{KM}(t)$ together for comparison.

2. Probability-Probability (P-P) plot
   Plot $S(a_j;\hat{\theta})$ against $\underline{\hat{S}_{KM}(a_j)}$ where $\underline{a_1 < a_2 < \ldots < a_K}$ are the distinct failure times. (of the data).



If the parametric model fits the data well, P-P plot should roughly agree with the line $y = x$.

3 Standard Quantile-Quantile (Q-Q) plot
checks if log location-scale model is suitable for the survival data.

Idea: If T is from a log location-scale distribution with parameters μ, σ, then $Y = \log T$ has a location-scale distribution with (μ, σ).
Let $K()$ be the survival function of this location-scale distribution with $(\mu = 0, \sigma = 1)$. Then

$W \sim$ Standardized loc-scale dist'n.

$$K^{-1}(S(t)) = \frac{1}{\sigma}\log t - \frac{\mu}{\sigma}.$$

probability.

$$S(t) = P(T \geq t) = P(Y \geq y) = P\left(\frac{Y-\mu}{\sigma} \geq \frac{y-\mu}{\sigma}\right) = P\left(W \geq \underset{\log t}{\frac{y-\mu}{\sigma}}\right) = K\left(\underset{\log t}{\frac{y-\mu}{\sigma}}\right).$$

$$K^{-1}(S(t)) = \frac{\log t - \mu}{\sigma} = \frac{1}{\sigma}\log t - \frac{\mu}{\sigma}$$

Standard quantile.

Method: Plot $K^{-1}(\hat{S}_{KM}(t))$ against $\log t$.

If the model fits the data well, then we should see an approximately linear pattern.

Example: ① $T \sim \text{Weibull}$, i.e. $Y \sim EV(\mu, \sigma)$

$$S(t;\mu,\sigma) = e^{-e^{\frac{\log t - \mu}{\sigma}}}$$

$$\log[-\log S(t;\mu,\sigma)] = \frac{\log t - \mu}{\sigma} = \frac{1}{\sigma}\log t - \frac{\mu}{\sigma}.$$

Q-Q plot:

Plot $\log[-\log \hat{S}_{KM}(t)]$ as $\log t$. $\square$

② $T \sim \text{Log Normal}$, i.e. $Y \sim N(\mu, \sigma)$.

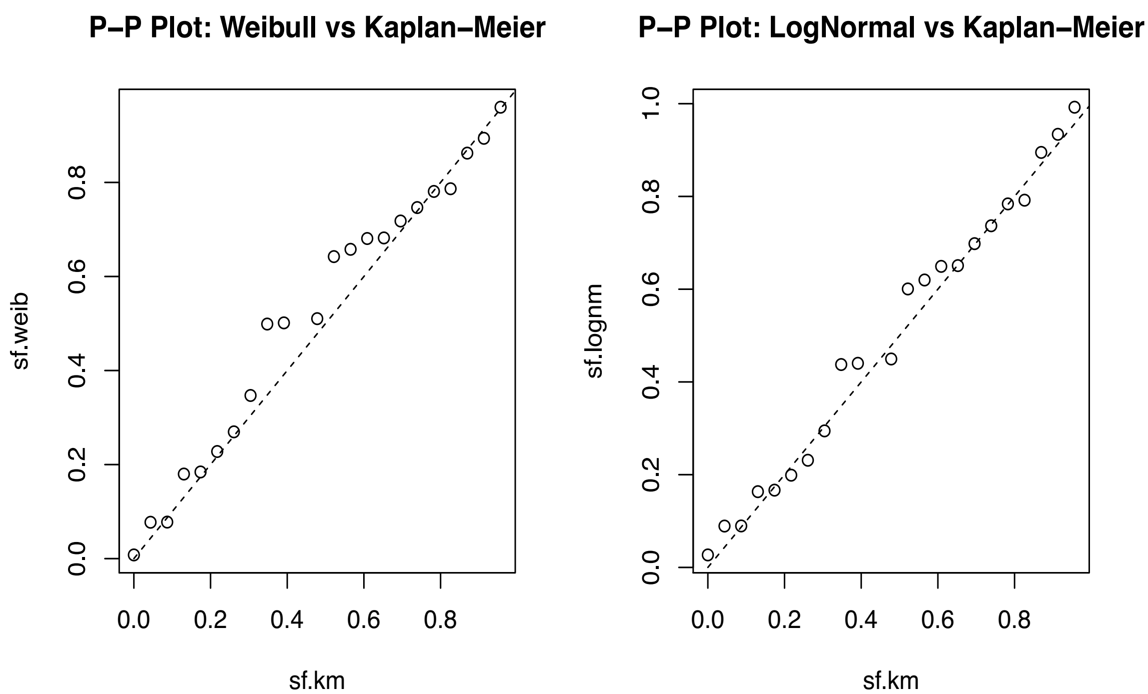
$$S(t;\mu,\sigma) = \underbrace{1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)}.$$

Q-Q plot ?

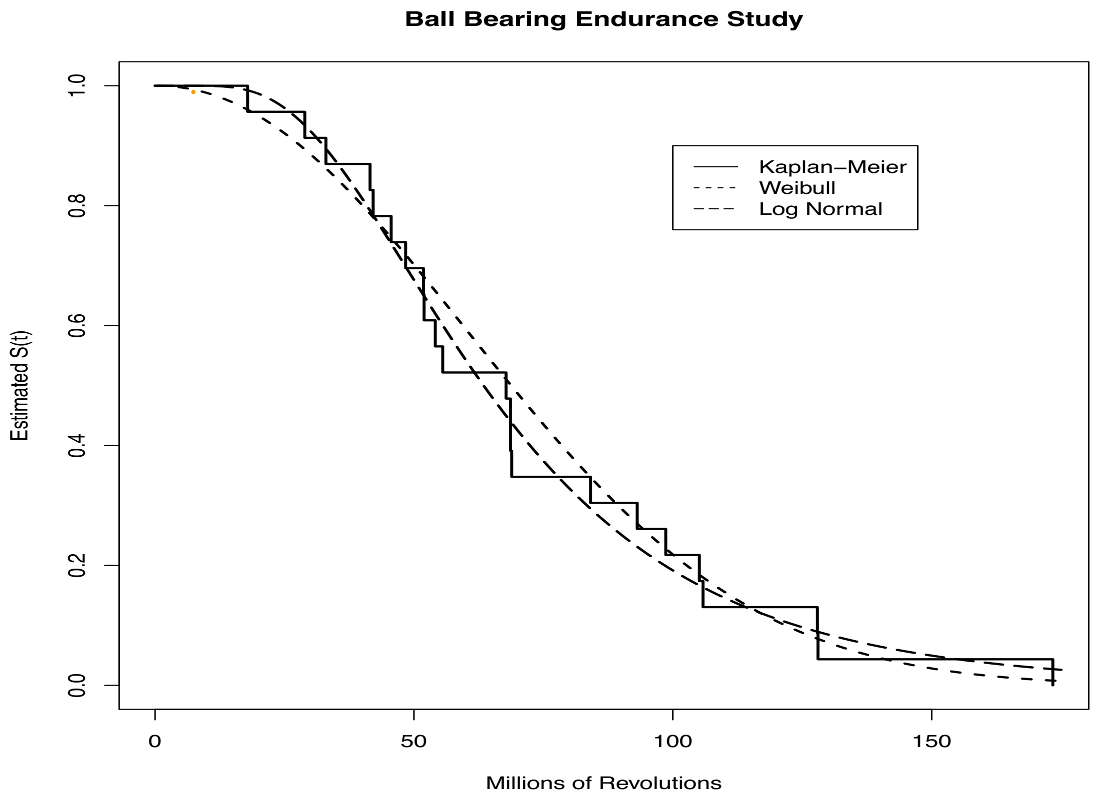
Example 3.3.1. Ball Bearing Endurance

Continuation of Example 3.2.1. More model checking...

P-P Plot: Check if the Weibull model fits the data.



Plot estimated survival functions by Weibull, log normal models, and by KM.



Discussion:

- Weibull model fits the data reasonably well for small and large survival times; it does not fit so well for intermediate survival times.
- In the middle range of time, $t \in [50, 80]$, failures occurred more densely between 67 and 69.
- Log normal model fits the data better.

```
> # Weibull fit and Kaplan-Meier in Example 3.2.1.
> sig1<-fit.weib$scale
> mu1<-fit.weib$coef
> time<-fit.km$time
> time
```

a_1 a_2 a_K

```
 [1]  17.88  28.92  33.00  41.52  42.12 ...
[21] 128.04 173.40
> sf.weib<-exp(-exp((log(time)-mu1)/sig1))
```

$S(a_j;\hat{\theta})$

$\hat{\theta} = (\hat{\mu}, \hat{\sigma})$.

```
> sf.km<-fit.km$surv
> sf.km
 [1] 0.95652174 0.91304348 0.86956522 ...
[19] 0.13043478 0.08695652 0.04347826 0.00000000
>
> # Fit a Log Normal model.
> fit.lognm<-survreg(Surv(time,status)~1,data=ballbear,
dist="lognormal")
```

```
> mu2<-fit.lognm$coeff
> sig2<-fit.lognm$scale
> mu2
(Intercept)
   4.150383
> sig2
[1] 0.5216865
> sf.lognm<-1-pnorm((log(time)-mu2)/sig2)
>
> # Construct P-P plots for Weibull fit and Log Normal fit.
> par(mfrow=c(1,2))
>  plot(sf.km,sf.weib,main="P-P Plot: Weibull vs
Kaplan-Meier")
> lines(seq(0,1,0.1),seq(0,1,0.1), lty=2)
>  plot(sf.km, sf.lognm, main="P-P Plot: LogNormal vs
Kaplan-Meier")
> lines(seq(0,1,0.1),seq(0,1,0.1), lty=2)
```

Sf of $T \sim$ log normal distn?

$Y = \log T \sim N(\mu, \sigma)$, $\frac{Y-\mu}{\sigma} \sim N(0,1)$. Let $\Phi(\,)$ be cdf of $N(0,1)$.

$$S_T(t) = S_Y(\log t) = 1 - P(Y < \log t) = 1 - P\left(\frac{Y-\mu}{\sigma} < \frac{\log t - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$$

```
> # Compare S(t) estimated from K-M, Weibull
> # and Log Normal.
> t<-0:175
> st.weib<-exp(-exp((log(t)-mu1)/sig1))
> st.lognm<-1-pnorm((log(t)-mu2)/sig2)
>  plot(fit.km,xlab="Millions of Revolutions",
ylab="Estimated S(t)",main="Ball Bearing Endurance Study",
conf.int=F,lty=1,lwd=2)
>  lines(t,st.weib,lty=2,lwd=2)
>  lines(t,st.lognm,lty=5,lwd=2)
>  legend(100, 0.9, c("Kaplan-Meier", "Weibull",
"Log Normal"), lty=c(1,2,5))
```

3.3.2 Model Expansion

parameter vector.

Model to be checked: $S(t;\theta)$

Method: Embed $S(t;\theta)$ in a "larger" model $S_1(t;\theta,\lambda)$, so that $S(t;\theta) = S_1(t;\theta,\lambda_0)$. $\lambda = \lambda_0$

Test $H_0 : \lambda = \lambda_0$.

Example: Test the suitability of an exponential model, $S(t;\theta) = e^{-t/\theta}$.

Recall that it is a special case of Weibull model, $S(t;\theta,\beta) = \underline{e^{-(t/\theta)^\beta}}$, with $\beta = 1$.

Fit a Weibull model, then test $H_0: \beta = 1$. $(*)$.

For Weibull model, as log loc-scale model,

$\mu = \log\theta$, $\sigma = \frac{1}{\beta}$. $(*) \Leftrightarrow H_0: \underline{\sigma = 1}$.

Note: R takes $(\mu, \phi = \log\sigma)$ as parameters;

$(*) \Leftrightarrow \underline{H_0: \phi = 0}$.

Example 3.3.2 Advanced non-Hodgkin's Lymphoma

Continuation of Example 2.3.1.

- Is Weibull (or exponential) model suitable for the data?
  Plot nonparametric estimate of $H(t)$ (see Example 2.3.1);
  P-P plot;
  Standard Q-Q plot
- Does Weibull model fit the data better than exponential model?
  Method: Model Expansion
  Fit a Weibull model using R, with parameters (μ, ϕ) where $\phi = \log \sigma$.
  Fit an exponential model using R, with parameters μ where $\mu = \log \theta$.
  See attached for code and output. $H_0: \sigma = 1 \Leftrightarrow H_0: \phi = 0.$

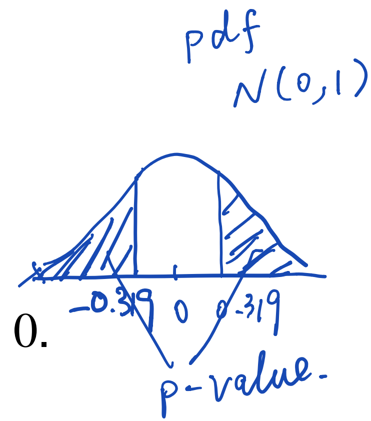
Test $H_0 : \phi = 0.$

$\theta = (\mu, \phi).$

$H_a: \phi \neq 0,$

1) Test based on Wald statistic

$$Z = \frac{\hat{\phi} - \overset{=0}{\phi_0}}{\sqrt{\hat{Var}(\hat{\phi})}} \approx \text{Normal}(0, 1),$$



where ϕ_0 is the value of ϕ under H_0. In this example, $\phi_0 = 0$.
Observed z-value is 0.319.
For a two-sided test,
p-value=$P(|Z| \geq |z_{obs}|)$=$P(Z \leq -0.319, Z \geq 0.319)$ =0.75.

Cannot reject H_0.

2) Test based on likelihood ratio (LR) statistic

$$\Lambda(\phi_0) = 2[l(\hat{\mu}, \hat{\phi}) - l(\tilde{\mu}, \phi_0)] \approx \chi^2_1. \Leftarrow 2-1$$

Weibull Exp

where $l(\hat{\mu}, \hat{\phi})$ is (the maximum) log-likelihood under Weibull model and $l(\tilde{\mu}, \phi_0)$ is (the maximum) log-likelihood under exponential model.

$H_0: \beta_2 = \beta_3 = 0.$

Model under H_0: parameter : μ.
(Exp model). Restricted parameter space $\Theta_0 = \{\mu \in \mathbb{R}\}$: 1-dim

Model not restricted under H_0: parameters : (μ, ϕ).
(Weibull). Entire parameter space Θ : 2-dim.
$\Theta = \{(\mu, \phi) : \mu \in \mathbb{R}, \phi \in \mathbb{R}\}$.

$\lambda_{obs} = 2[-94.2 - (-94.3)] = 0.2$.

p-val $= P(\Lambda \geq \underbrace{\lambda_{obs}}_{=0.2}) = 0.65$.

Conclusion:

We cannot reject $H_0 : \phi = 0$. This implies the Weibull model does not fit the data any better than the simple exponential model.

```
> library(survival)
> data<-read.table("eg231.txt", header=T)
> fit.weib<-survreg(Surv(time, status)~1,data=data)
> print(summary(fit.weib))
Call:
survreg(formula = Surv(time, status) ~ 1, data = data)
              Value Std. Error      z        p
(Intercept) 3.7245       0.241 15.443 8.38e-54
Log(scale)  0.0612       0.192  0.319 7.50e-01
Scale= 1.06
Weibull distribution
Loglik(model)= -94.2   Loglik(intercept only)= -94.2
Number of Newton-Raphson Iterations: 5
n= 31
```

$\hat{\mu}$ (Intercept) ; $\hat{\phi}$ Log(scale)

$H_0: \mu = 0$ vs. $H_a: \mu \neq 0$.

$H_0: \phi = 0$ vs $H_a: \phi \neq 0$.

$\log T = \underbrace{\mu}_{\shortparallel} + \sigma W$

$\underbrace{\beta_0} + \underbrace{\cancel{\beta_1 z_1 + \cdots \beta_{p-1} z_{p-1}}}$ intercept-only model.

```
> fit.exp<-survreg(Surv(time, status)~1,dist="exp",
data=data)
> print(summary(fit.exp))
Call:
survreg(formula = Surv(time, status) ~ 1, data = data,
dist = "exp")
            Value Std. Error    z       p
(Intercept)  3.71      0.224 16.6 6.6e-62
Scale fixed at 1
Exponential distribution
Loglik(model)= -94.3   Loglik(intercept only)= -94.3
Number of Newton-Raphson Iterations: 4
n= 31
```