Chapter 5B

(AST405) Lifetime data analysis

Md Rasel Biswas

Lecture Outline

1. 5. Inference Procedures for Log-location-scale Distributions
   - 5.1 Log-normal and normal distributions
   - 5.2 Log-logistic and logistic distributions
   - 5.3 Comparison of distributions

Section 1

5. Inference Procedures for Log-location-scale Distributions

Subsection 1

5.1 Log-normal and normal distributions

Log-normal distribution

- T follows a log-normal distribution with location parameter μ and scale parameter σ if $Y = \log T \sim \mathcal{N}(\mu, \sigma^2)$
- The pdf and survivor function of log-normal distribution

$$f(t; \mu, \sigma) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma} \right)^2 \right]$$

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

  - μ and σ are the parameters of both normal and log-normal distributions
  - $\Phi(\cdot) \rightarrow$ cumulative distribution function of standard normal distribution

Log-normal distribution

- Log-normal distribution is a member of the log-location-scale family of distributions and the corresponding location-scale distribution is normal with

$$S_0(z) = 1 - \Phi(z)$$

$$f_0(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \phi(z)$$

  - $\phi(\cdot) \rightarrow$ pdf of standard normal distribution
  - $z = (y - \mu)/\sigma$

Log-normal distribution

- Density function of log-lifetime

$$
\begin{aligned}
f(y;\mu,\sigma) &= \frac{1}{\sigma} f_0\Big(\frac{y-\mu}{\sigma}\Big) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\frac{y-\mu}{\sigma}\Big)^2\Big]
\end{aligned}
$$

- Survivor function of log-lifetime

$$
S(y;\mu,\sigma) = S_0\left(\frac{y-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{y-\mu}{\sigma}\right)
$$

Likelihood function normal distribution

- Data

$$\{(t_i, \delta_i),\ i = 1, \dots, n\}$$

- Log-likelihood function

$$\begin{aligned}
\ell(\mu, \sigma) &= \log \prod_{i=1}^{n} \left[(1/\sigma)\, f_0(z_i) \right]^{\delta_i} \left[S_0(z_i) \right]^{1-\delta_i} \\
&= -r \log \sigma + \sum_{i=1}^{n} \delta_i \log f_0(z_i) + \sum_{i=1}^{n} (1 - \delta_i) \log S_0(z_i) \\
&= -r \log \sigma - \frac{1}{2} \sum_{i=1}^{n} \delta_i z_i^2 + \sum_{i=1}^{n} (1 - \delta_i) \log S_0(z_i)
\end{aligned}$$

  - $z_i = (y_i - \mu)/\sigma$ and $y_i = \log t_i$
  - $r = \sum_{i=1}^{n} \delta_i$

Likelihood function normal distribution

- Elements of hessian matrix and score function depend on the followings

$$\frac{\partial \log f_0(z)}{\partial z} = -z$$

$$\frac{\partial^2 \log f_0(z)}{\partial z^2} = -1$$

$$\frac{\partial \log S_0(z)}{\partial z} = -\frac{f_0(z)}{S_0(z)}$$

$$\frac{\partial^2 \log S_0(z)}{\partial z^2} = \frac{z f_0(z)}{S_0(z)} - \left[\frac{f_0(z)}{S_0(z)}\right]^2$$

Likelihood function normal distribution

- MLEs

$$(\hat{\mu}, \hat{\sigma})' = \arg\max_{\Theta} \ell(\mu, \sigma)$$

  - Sampling distribution

$$(\hat{\mu}, \hat{\sigma})' \sim \mathcal{N}\Big((\mu, \sigma)', V\Big)$$

  where

$$\hat{V} = \Big[-H(\hat{\mu}, \hat{\sigma})\Big]^{-1}$$

- Confidence intervals of parameters, quantiles, and survival probabilities can be obtained using the methods described for Weibull models

Likelihood function normal distribution

- Estimate of survivor function (Log-normal distribution)

$$S(t;\hat{\mu},\hat{\sigma}) = 1 - \Phi\Big(\frac{\log t - \hat{\mu}}{\hat{\sigma}}\Big)$$
$$= 1 - \Phi(\hat{\psi})$$

where

$$\hat{\psi} = \Phi^{-1}\Big(1 - S(t;\hat{\mu},\hat{\sigma})\Big) = \frac{\log t - \hat{\mu}}{\hat{\sigma}}$$

  - Standard error of $\hat{\psi}$

$$se(\hat{\psi}) = \sqrt{\mathbf{a}'\hat{V}\mathbf{a}}$$

where

$$\mathbf{a} = (-1/\hat{\sigma}, -\hat{\psi}/\hat{\sigma})'$$

Estimate of survivor function

- $(1-\alpha)100\%$ CI of $S(t)$

$$
\begin{aligned}
L < \psi < U \\
L < \Phi^{-1}(1 - S(t;\mu,\sigma)) < U \\
\Phi(L) < 1 - S(t;\mu,\sigma) < \Phi(U) \\
1 - \Phi(U) < S(t;\mu,\sigma) < 1 - \Phi(L)
\end{aligned}
$$

where

$$
\begin{aligned}
L &= \hat{\psi} - z_{1-\alpha/2}\, se(\hat{\psi}) \\
U &= \hat{\psi} + z_{1-\alpha/2}\, se(\hat{\psi})
\end{aligned}
$$

Estimate of survivor function

- LRT statistics based method of obtaining CI for survivor function is described with

$$H_0 : S(y_0) = S(\log t_0) = s_0$$

- The $100(1-\alpha)\%$ CI for $S(t)$ can be obtained from the values of s_0 that satisfy

$$\Lambda(s_0) = 2\ell(\hat{\mu}, \hat{\sigma}) - 2\ell(\tilde{\mu}, \tilde{\sigma}) \leq \chi^2_{(1),1-\alpha}$$

Estimate of survivor function

- Unrestricted and unrestricted MLEs are obtained as

$$\begin{aligned} \text{unrestricted} \quad & (\hat{\mu}, \hat{\sigma})' = \arg\max_{\Theta} \ell(\mu, \sigma) \\ \text{restricted} \quad & (\tilde{\mu}, \tilde{\sigma})' = \arg\max_{\Theta} \ell(y_0 - \sigma\Phi^{-1}(1 - s_0), \sigma) \end{aligned}$$

where under H_0, we can show

$$S(y_0) = 1 - \Phi\left(\frac{y_0 - \mu}{\sigma}\right) = s_0 \quad \Rightarrow \quad \mu = y_0 - \sigma\Phi^{-1}(1 - s_0)$$

Quantiles

- The expression of estimate of y_p

$$\hat{y}_p = \hat{\mu} + \hat{\sigma} w_p$$

where for normal distribution

$$w_p = S_0^{-1}(1-p) = \Phi^{-1}(p)$$

- Standard error of $\hat{y}_p$

$$se(\hat{y}_p) = \sqrt{\mathbf{a}' \hat{V} \mathbf{a}}$$

where

$$\mathbf{a} = (1, w_p)'$$

Homework

- Obtain the expressions of Wald-type and LRT based $100(1-\alpha)\%$ confidence intervals of y_p

Example 5.3.1

- Data are available on lifetimes (in thousand miles) of 96 locomotive controls, of which were failed.
- The test was terminated after $135K$ miles, so 59 lifetimes were censored at $135K$.

Example 5.3.1

```
dat_ex531
```

```
# A tibble: 96 x 2
    time status
   <dbl>  <int>
 1  22.5      1
 2  37.5      1
 3  46        1
 4  48.5      1
 5  51.5      1
 6  53        1
 7  54.5      1
 8  57.5      1
 9  66.5      1
10  68        1
# i 86 more rows
```

Example 5.3.1

```
dat_ex531 %>%
  count(status)
```

```
# A tibble: 2 x 2
  status     n
   <int> <int>
1      0    59
2      1    37
```

Example 5.3.1

Log-normal and normal model fit

```
mod_LN <- survreg(Surv(time, status) ~ 1,
                  dist = "lognormal",
                  data = dat_ex531)
```

```
mod_N <- survreg(Surv(log(time), status) ~ 1,
                 dist = "gaussian",
                 data = dat_ex531)
```

Example 5.3.1

MLEs $(\hat{\mu}, \log \hat{\sigma})$ **and corresponding standard errors**

```
tidy(mod_LN)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    5.19      0.129     40.3    0
2 Log(scale)    -0.136     0.131     -1.04   0.297
```

Example 5.3.1

Estimated variance of $(\hat{\mu}, \log \hat{\sigma})$

```
mod_LN$var
```

```
            (Intercept) Log(scale)
(Intercept)  0.01657557 0.00983969
Log(scale)   0.00983969 0.01703353
```

Estimated variance $\mathrm{V}(\hat{\mu}, \hat{\sigma})$ from $\mathrm{V}(\hat{\mu}, \log\hat{\sigma})$

- $G\,V(\hat{\mu}, \log\hat{\sigma})\,G'$

```
           [,1]       [,2]
[1,] 0.01657557 0.00858735
[2,] 0.00858735 0.01297359
```

$$G = \begin{bmatrix} 1 & 0 \\ 0 & \exp(\sigma) \end{bmatrix}$$

Estimated variance $\mathrm{V}(\hat{\mu}, \hat{\sigma})$ from $\mathrm{V}(\hat{\mu}, \log \hat{\sigma})$

Table 1: 95% Confidence intervals for μ and σ

par	lower	upper	lower	upper
μ	4.942	5.447	5.000	5.400
σ	0.676	1.127	0.709	1.109

Estimated variance $\mathrm{V}(\hat{\mu}, \hat{\sigma})$ from $\mathrm{V}(\hat{\mu}, \log \hat{\sigma})$

- Estimate of $S(80)$

$$\begin{aligned} S(80; \hat{\mu}, \hat{\sigma}) &= 1 - \Phi\Big(\frac{\log 80 - \hat{\mu}}{\hat{\sigma}}\Big) \\ &= 0.824 \end{aligned}$$

  - $\hat{\mu} = 5.195$ and $\hat{\sigma} = 0.873$

Estimated variance $\mathrm{V}(\hat{\mu}, \hat{\sigma})$ from $\mathrm{V}(\hat{\mu}, \log \hat{\sigma})$

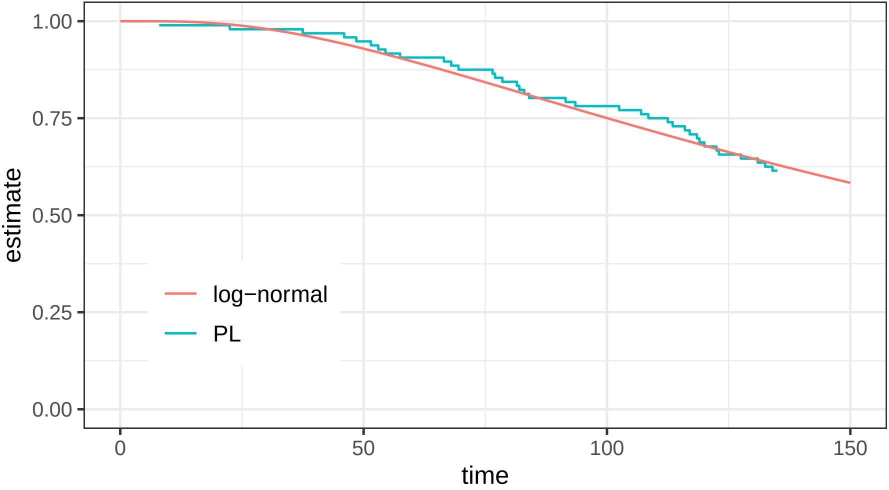

Figure 1: Comparison of the estimates of survivor function

Estimated variance $\mathrm{V}(\hat{\mu}, \hat{\sigma})$ from $\mathrm{V}(\hat{\mu}, \log \hat{\sigma})$

Table 2: Estimate and confidence interval of $S(80)$

parameter	est	lower	upper
$S(80)$	0.824	0.667	0.924

- Obtain LRT based 95% CI for $S(80)$

Quantiles

- General expression of pth quantile of log-lifetime ($\hat{\mu} = 5.195$ and $\hat{\sigma} = 0.873$)

$$\hat{y}_p = \hat{\mu} + \hat{\sigma} w_p$$

  - $w_p = \Phi^{-1}(p)$

Quantiles

Table 3: Estimate and confidence intervals of different quantiles of locomotive controls lifetime (normal distribution)

p	w_p	$\hat{y}_p$	se	lower	upper
0.25	-0.674	4.606	0.105	NA	NA
0.50	0.000	5.195	0.129	NA	NA
0.75	0.674	5.783	0.194	NA	NA

Subsection 2

5.2 Log-logistic and logistic distributions

Log-logistic distribution

- T follows a log-logistic distribution with parameters α (scale) and β (shape) if $Y = \log T$ follows a logistic distribution with parameters u (location) and b (scale)

Log-logistic distribution

- The pdf, survivor, and hazard function of log-logistic distribution

$$f(t;\alpha,\beta) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{\left[1+(t/\alpha)^{\beta}\right]^{2}}$$

$$S(t;\alpha,\beta) = \left[1+(t/\alpha)^{\beta}\right]^{-1}$$

$$h(t;\alpha,\beta) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{\left[1+(t/\alpha)^{\beta}\right]}$$

Logistic distribution

- Log-logistic distribution is a member of the log-location-scale family of distributions and the corresponding location-scale distribution is logistic with

$$S_0(z) = \frac{1}{1 + e^z}$$

$$f_0(z) = \frac{e^z}{(1 + e^z)^2}$$

  - $z = (y - u)/b$

Logistic distribution

- Density function of log-lifetime

$$
\begin{aligned}
f(y;u,b) &= \frac{1}{b} f_0\Big(\frac{y-u}{b}\Big) \\
&= \frac{(1/b)\ \exp\left[(y-u)/b\right]}{\left\{1+\exp\left[(y-u)/b\right]\right\}^2}
\end{aligned}
$$

- Survivor function of log-lifetime

$$
\begin{aligned}
S(y;u,b) &= S_0\left(\frac{y-u}{b}\right) \\
&= \frac{1}{1+\exp\left[(y-u)/b\right]}
\end{aligned}
$$

Logistic distribution

- Data: $\{(t_i, \delta_i),\ i = 1, \dots, n\}$
- Log-likelihood function

$$
\begin{aligned}
\ell(\mu, \sigma) &= \log \prod_{i=1}^{n} \Big[(1/b)\, f_0(z_i)\Big]^{\delta_i} \Big[S_0(z_i)\Big]^{1-\delta_i} \\
&= -r \log b + \sum_{i=1}^{n} \delta_i \log f_0(z_i) + \sum_{i=1}^{n} (1-\delta_i) \log S_0(z_i) \\
&= -r \log b + \sum_{i=1}^{n} \Big[\delta_i \{z_i - \log(1+e^{z_i})\} - \log(1+e^{z_i})\Big]
\end{aligned}
$$

  - $z_i = (y_i - u)/b$ and $y_i = \log t_i$
  - $r = \sum_{i=1}^{n} \delta_i$

Logistic distribution

- Elements of hessian matrix and score function depend on the followings

$$\frac{\partial \log f_0(z)}{\partial z} = 1 - \frac{2e^z}{1+e^z}$$
$$\frac{\partial^2 \log f_0(z)}{\partial z^2} = -2f_0(z)$$
$$\frac{\partial \log S_0(z)}{\partial z} = \frac{-e^z}{1+e^z}$$
$$\frac{\partial^2 \log S_0(z)}{\partial z^2} = \frac{-e^z}{(1+e^z)^2}$$

Logistic distribution

- MLEs

$$(\hat{u}, \hat{b})' = \arg\,\max_{\Theta}\, \ell(u, b)$$

- Sampling distribution

$$(\hat{u}, \hat{b})' \sim \mathcal{N}\Big((u, b)', V\Big)$$

where

$$\hat{V} = \Big[-H(\hat{u}, \hat{b})\Big]^{-1}$$

- Confidence intervals of parameters, quantiles, and survival probabilities can be obtained using the methods described for Weibull models

Logistic distribution

- Estimate of survivor function (logistic distribution)

$$S_0\Big(\frac{y-\hat{u}}{\hat{b}}\Big) = S(y;\hat{u},\hat{b}) = \frac{1}{1+\exp\left[(y-\hat{u})/\hat{b}\right]}$$

$$\log\left[\frac{1-S(y)}{S(y)}\right] = \frac{y-\hat{u}}{\hat{b}} = \hat{\psi} = S_0^{-1}\Big(S(y)\Big)$$

  - Standard error of $\hat{\psi}$

$$se(\hat{\psi}) = \sqrt{\mathbf{a}'\hat{V}\mathbf{a}}, \quad \text{where } \mathbf{a} = (-1/\hat{b}, -\hat{\psi}/\hat{b})'$$

Logistic distribution

$(1-\alpha)100\%$ **CI of** $S(t)$

$$
\begin{aligned}
L < \psi < U \\
L < \log \frac{1-S(y)}{S(y)} < U \\
\exp(L) < \frac{1-S(y)}{S(y)} < \exp(U) \\
1+\exp(L) < 1 + \frac{1-S(y)}{S(y)} < 1+\exp(L) \\
\frac{1}{1+\exp(U)} < S(y) < \frac{1}{1+\exp(L)}
\end{aligned}
$$

where

$$
\begin{aligned}
L &= \hat{\psi} - z_{1-\alpha/2}\, se(\hat{\psi}) \\
U &= \hat{\psi} + z_{1-\alpha/2}\, se(\hat{\psi})
\end{aligned}
$$

Estimate of survivor function

- LRT statistics based method of obtaining CI for survivor function is described with

$$H_0 : S(y_0) = S(\log t_0) = s_0$$

- The $100(1 - \alpha)\%$ CI for $S(t)$ can be obtained from the values of s_0 that satisfy

$$\Lambda(s_0) \leq \chi^2_{(1),1-\alpha}$$

where

$$\Lambda(s_0) = 2\ell(\hat{u}, \hat{b}) - 2\ell(\tilde{u}, \tilde{b})$$

Estimate of survivor function

- Unrestricted and unrestricted MLEs are obtained as

$$\text{unrestricted} \quad (\hat{u}, \hat{b})' = \arg\max_{\Theta} \ell(u, b)$$

$$\text{restricted} \quad (\tilde{u}, \tilde{b})' = \arg\max_{\Theta} \ell(y_0 - b \log\{(1 - s_0)/s_0\}, b)$$

where under H_0, we can show

$$S(y_0) = s_0 \quad \Rightarrow \quad u = y_0 - b \log \frac{1 - s_0}{s_0}$$

Quantiles

- The expression of estimate of y_p

$$\hat{y}_p = \hat{u} + \hat{b} w_p$$

where for normal distribution

$$w_p = S_0^{-1}(1-p) = \log \frac{p}{1-p}$$

- Standard error of $\hat{y}_p$

$$se(\hat{y}_p) = \sqrt{\mathbf{a}' \hat{V} \mathbf{a}}$$

where

$$\mathbf{a} = (1, w_p)'$$

Homework

- Obtain the expressions of Wald-type and LRT based $100(1-\alpha)\%$ confidence intervals of y_p

Example 5.3.1

- Data are available on lifetimes (in thousand miles) of 96 locomotive controls, of which were failed.
- The test was terminated after $135K$ miles, so 59 lifetimes were censored at $135K$.

Example 5.3.1

```
dat_ex531
```

```
# A tibble: 96 x 2
    time status
   <dbl>  <int>
 1  22.5      1
 2  37.5      1
 3  46        1
 4  48.5      1
 5  51.5      1
 6  53        1
 7  54.5      1
 8  57.5      1
 9  66.5      1
10  68        1
# i 86 more rows
```

Example 5.3.1

```
dat_ex531 %>%
  count(status)
```

```
# A tibble: 2 x 2
  status     n
   <int> <int>
1      0    59
2      1    37
```

Example 5.3.1

Log-logistic and logistic model fit

```
mod_LL <- survreg(Surv(time, status) ~ 1,
                  dist = "loglogistic",
                  data = dat_ex531)
```

```
mod_L <- survreg(Surv(log(time), status) ~ 1,
                 dist = "logistic",
                 data = dat_ex531)
```

Example 5.3.1

MLEs $(\hat{u}, \log \hat{b})$

```
[1]  5.1206418 -0.8266704
```

Estimated variance of $(\hat{u}, \log \hat{b})$

```
            (Intercept)  Log(scale)
(Intercept) 0.010490062 0.007837215
Log(scale)  0.007837215 0.022515937
```

Example 5.3.1

MLEs of $(\hat{u}, \hat{b})$

```
[1] 5.1206418 0.4375036
```

Estimated variance of $(\hat{u}, \hat{b})$

```
            [,1]        [,2]
[1,] 0.010490062 0.003428809
[2,] 0.003428809 0.004309761
```

Example 5.3.1

Table 4: 95% Confidence intervals for location and scale parameters

dist	par	est	lower	upper	lower	upper
Logistic	u	5.121	4.920	5.321	5.000	5.300
NA	b	0.438	0.326	0.587	0.360	0.559
Gaussian	μ	5.195	4.942	5.447	5.000	5.400
NA	σ	0.873	0.676	1.127	0.709	1.109

Example 5.3.1

- Estimate of $S(80)$ (log-logistic distribution)

$$S(80; \hat{u}, \hat{b}) = \frac{1}{1 + \exp\left[(\log 80 - \hat{u})/\hat{b}\right]}$$
$$= 0.844$$

  - $\hat{u} = 5.121$ and $\hat{b} = 0.438$

Example 5.3.1

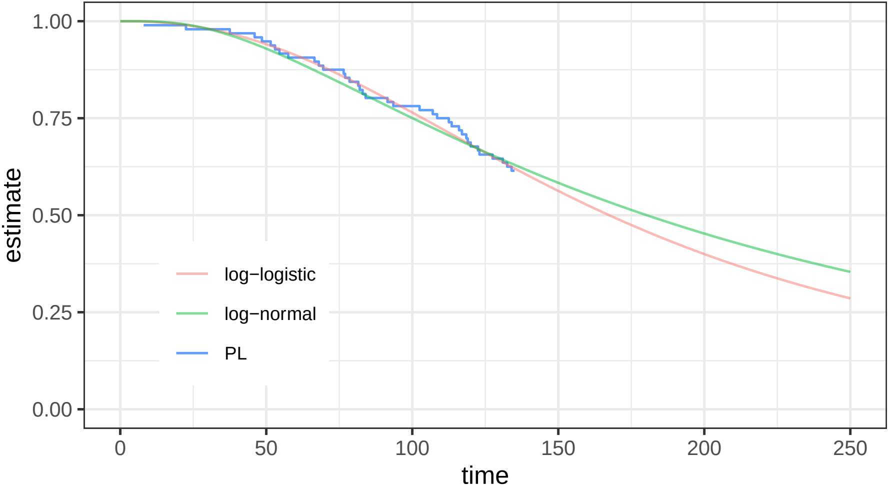

Figure 2: Comparison of the estimates of survivor function

Example 5.3.1

Table 5: Estimate and corresponding Wald-type confidence interval of the survival probability $S(80)$

dist	est	lower	upper
Log-logistic	0.844	0.566	0.957
Log-normal	0.824	0.667	0.924

- Obtain LRT based 95% CI for $S(80)$

Quantiles

- General expression of pth quantile of log-lifetime ($\hat{u} = 5.121$ and $\hat{b} = 0.438$)

$$\hat{y}_p = \hat{u} + \hat{b} w_p$$

  - $w_p = \log \frac{p}{1-p}$

Quantiles

Table 6: Estimate and confidence intervals of different quantiles

dist	p	w_p	$\widehat{y}_p$	se	lower	upper
Logistic	0.25	-1.099	4.640	0.143	NA	NA
NA	0.50	0.000	5.121	0.102	NA	NA
NA	0.75	1.099	5.601	0.234	NA	NA
Gaussian	0.25	-0.674	4.826	0.101	NA	NA
NA	0.50	0.000	5.121	0.102	NA	NA
NA	0.75	0.674	5.416	0.177	NA	NA

Homework

- Analyze the locomotive control lifetimes using Weibull model and compare the results

Subsection 3

5.3 Comparison of distributions

5.3 Comparison of distributions

- Let T_{ji} be the lifetime of ith subject of the jth group ($i = 1, \dots, n_j$, $j = 1, \dots, m$)
- Assume T_{ji} follows a distribution of log-location-scale family with parameters α_j (scale) and β_j (shape)
- The corresponding distribution of log-lifetime $Y_{ji} = \log T_{ji}$ is of a location-scale family distribution with parameters u_j (location) and b_j (scale)

$$u_j = \log \alpha_j \text{ and } b_j = (1/\beta_j)$$

Survivor functions

- The survivor function of $Y_{ji} = \log T_{ji}$

$$S_j(y) = S_0\left(\frac{y - u_j}{b_j}\right)$$

- The survivor function of T_{ji}

$$S_j(t) = S_0^{\star}\left[(t/\alpha_j)^{\beta_j}\right]$$

  - $S_0^{\star}(x) = S_0(\log x)$
  - $u_j = \log \alpha_j$
  - $b_j = (1/\beta_j)$

Survivor functions

- Comparison of several normal populations is a well-known problem in statistics, where equal population variances are assumed, and the comparisons are performed on the basis of equality of population means

Quantile

- General expression of the pth quantile of the jth population takes the form
$$y_{jp} = u_j + b_j w_p, \quad j = 1, \dots, m$$
  - $w_p = S_0^{-1}(1-p)$

Equality of two populations

- When the scales are **not equal** (i.e. $b_1 \neq b_2$), the difference between the pth quantiles does depend on the probability p

$$y_{1p} - y_{2p} = u_1 - u_2 + w_p(b_1 - b_2)$$

- Under the assumption of equality of the scales (i.e. $b_1 = b_2$), difference between pth (log-lifetime) quantile of a pair of populations (say 1 and 2) is constant, i.e. it does not depend on the probability $p \in (0, 1)$

$$y_{1p} - y_{2p} = u_1 - u_2$$

Equality of two populations

- The difference between two log-lifetime quantiles can be expressed in terms of the ratio of lifetime quantiles

$$
\begin{aligned}
y_{1p} - y_{2p} &= u_1 - u_2 \\
\log t_{1p} - \log t_{2p} &= \log \alpha_1 - \log \alpha_2 \\
t_{1p}/t_{2p} &= \alpha_1/\alpha_2
\end{aligned}
$$

- The ratio of the pth quantiles of two lifetime distributions does not depend on the probability p when the corresponding shape parameters are equal $(\beta_1 = \beta_2)$

Equality of two populations

- Equality of all quantiles of two distributions, i.e.

$$y_{1p} = y_{2p} \ \ \forall \ p \in (0, 1),$$

corresponds to equality of two distributions, i.e.

$$S_1(y) = S_2(y)$$

- Under the assumption of common scale (shape for lifetime) parameter, the null hypothesis of equality of two distributions can be expressed as

$$H_0 : u_1 - u_2 = 0 \quad \text{or} \quad H_0 : (\alpha_1/\alpha_2) = 1$$

Equality of two populations

- Equality of two populations with survivor functions (say S_1 and S_2) can be expressed in terms of survivor functions
- Since

$$y_{1p} = y_{2p} + u_1 - u_2 \text{ or } t_{1p} = t_{2p}(\alpha_1/\alpha_2),$$

the corresponding survivor functions can be expressed as

$$S_1(y + u_1 - u_2) = S_2(y)$$
$$S_1(t(\alpha_1/\alpha_2)) = S_2(t)$$

- That is, the survivor functions for Y are translations of one another by an amount $(u_1 - u_2)$ along the y-axis

Wald-type statistic

- Data $\{(t_{ji}, \delta_{ji}), i = 1, 2\}$ and $y_{ji} = \log t_{ji}$
- Two populations can be compared in terms of pth quantile

$$H_0 : y_{1p} = y_{2p}$$

- Corresponding pivotal quantity

$$Z_p = \frac{(\hat{y}_{1p} - \hat{y}_{2p}) - (y_{1p} - y_{2p})}{\left[\operatorname{var}(\hat{y}_{1p}) + \operatorname{var}(y_{2p})\right]^{1/2}} \sim \mathcal{N}(0, 1) \text{ under } H_0$$

  - The statistic Z_p can be used to obtain confidence interval for $(y_{1p} - y_{2p})$

Wald-type statistic

- To test $H_0 : b_1 = b_2$, the following pivotal quantity can be considered

$$Z_b = \frac{(\log \hat{b}_1 - \log \hat{b}_2) - (\log b_1 - \log b_2)}{[\mathrm{var}(\log \hat{b}_1) + \mathrm{var}(\log \hat{b}_2)]^{1/2}} \sim \mathcal{N}(0, 1) \text{ under } H_0$$

  - The statistic Z_b can be used to obtain confidence interval for (b_1/b_2)

Wald-type statistic

- When scales are equal, two populations can be compared with respect their location parameter $H_0 : u_1 = u_2$
- The corresponding pivotal quantity

$$Z_u = \frac{(\hat{u}_1 - \hat{u}_2) - (u_1 - u_2)}{[\mathrm{var}(\hat{u}_1) + \mathrm{var}(\hat{u}_2)]^{1/2}} \sim \mathcal{N}(0,1) \text{ under } H_0$$

  - The statistic Z_u can be used to obtain confidence interval for $(u_1 - u_2)$

Wald-type statistic

- Wald statistic cannot be used to test

$$H_0 : u_1 = u_2, b_1 = b_2$$

LRT based inference

- Data $\{(t_{ji}, \delta_{ji}), j = 1, \ldots, m, i = 1, \ldots, n_j\}$ and $y_{ji} = \log t_{ji}$
- Different tests and confidence intervals of interest

1. $H_0 : b_1 = \cdots = b_m$
2. Confidence interval for (b_1/b_2)
3. Equality of several location parameters when scale parameters are equal

$$H_0 : u_1 = \cdots = u_m, b_1 = \cdots = b_m$$
$$H_1 : \text{all } u_j\text{'s are not equal}, b_1 = \cdots = b_m$$

4. Confident interval for $(u_1 - u_2)$ when $b_1 = b_2$
5. Confidence interval for $(y_{1p} - y_{2p})$ when $b_1 \neq b_2$

Case 1

- Hypothesis of interest

$$H_0 : b_1 = \cdots = b_m = b \ \ \text{(say)} \tag{1}$$

- Log-likelihood function

$$\ell(u_1, \dots, u_m, b_1, \dots, b_m) = \sum_{j=1}^{m} \ell_j(u_j, b_j)$$

- Contribution to log-likelihood function for the jth population

$$\ell_j(u_j, b_j) = -r_j \log b_j + \sum_{i=1}^{n_j} \left[\delta_i \log f_0(z_{ji}) + (1 - \delta_{ji}) \log S_0(z_{ji}) \right]$$

  - $r_j = \sum_i \delta_{ji}$

Case 1

- LRT statistic

$$\Lambda = 2\ell(\hat{u}_1, \dots, \hat{u}_m, \hat{b}_1, \dots, \hat{b}_m) - 2\ell(\tilde{u}_1, \dots, \tilde{u}_m, \tilde{b}, \dots, \tilde{b})$$

  - $\Lambda \sim \chi^2_{(m-1)}$ under the null hypothesis defined in Equation 1
- MLEs
  - $(\hat{u}_j, \hat{b}_j)' = \arg\max_\Theta \ell_j(u_j, b_j), \;\; j = 1, \dots, m$
  - $(\tilde{u}_1, \dots, \tilde{u}_m, \tilde{b}, \dots, \tilde{b})' = \arg\max_\Theta \ell(u_1, \dots, u_m, b, \dots, b)$

Case 2

- To obtain confidence interval of (b_1/b_2), consider

$$H_0 : (b_1/b_2) = a \quad \Rightarrow \quad H_0 : b_1 = ab_2$$

- $100(1-\alpha)\%$ confidence interval of (b_1/b_2) can be obtained from the range of a values that satisfy

$$\Lambda(a) \leq \chi^2_{(1),1-\alpha},$$

where the LRT statistic

$$\Lambda(a) = 2\ell(\hat{u}_1, \hat{u}_2, \hat{b}_1, \hat{b}_2) - 2\ell(\tilde{u}_1, \tilde{u}_2, a\tilde{b}_2, \tilde{b}_2)$$

  - $(\hat{u}_j, \hat{b}_j)' = \arg\max_{\Theta} \ell_j(u_j, b_j), \;\; j = 1, 2$
  - $(\tilde{u}_1, \tilde{u}_2, \tilde{b}_2)' = \arg\max_{\Theta} \ell(u_1, u_2, ab_2, b_2)$

Case 3

- Test equality of several location parameters when scale parameters are equal

$$H_0 : u_1 = \cdots = u_m, \; b_1 = \cdots = b_m$$

$$H_1 : \text{all } u_j\text{'s are not equal}, \; b_1 = \cdots = b_m$$

Case 3

- MLEs
  - under H_0, $(u^\star, b^\star) = \arg\max_\Theta \ell(u, \ldots, u, b, \ldots, b)$
  - under H_1, $(\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{b}) = \arg\max_\Theta \ell(u_1, \ldots, u_m, b, \ldots, b)$
- LRT statistic

$$\Lambda = 2\ell(\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{b}, \ldots, \tilde{b}) - 2\ell(u^\star, \ldots, u^\star, b^\star, \ldots, b^\star)$$

  - Under the null hypothesis, Λ follows $\chi^2_{(m-1)}$ distribution

Case 4

- To obtain a confidence interval of $(u_1 - u_2)$ when $b_1 = b_2$, consider the null and alternative hypothesis

$$H_0 : u_1 - u_2 = \delta,\ b_1 = b_2 \quad \text{vs} \quad H_1 : u_1 - u_2 \neq \delta,\ b_1 = b_2$$

- LRT statistic

$$\Lambda(\delta) = 2\ell(\tilde{u}_1, \tilde{u}_2, \tilde{b}, \tilde{b}) - 2\ell(u_2^\star + \delta, u_2^\star, b^\star, b^\star)$$

  - under H_0, $(u^\star, b^\star) = \arg\max_\Theta \ell(u, u, b, b)$
  - under H_1, $(\tilde{u}_1, \tilde{u}_2, \tilde{b}) = \arg\max_\Theta \ell(u_1, u_2, b, b)$

- $100(1-\alpha)$ confidence interval for $(u_1 - u_2)$ can be obtained from the set of δ values that satisfy $\Lambda(\delta) \leq \chi^2_{(1),1-\alpha}$

Case 5

- When $b_1 \neq b_2$, to obtain confidence interval for $(y_{1p} - y_{2p})$ consider the following hypothesis

$$H_0 : y_{1p} - y_{2p} = \Delta \;\; \Rightarrow \;\; H_0 : u_1 - u_2 = \Delta + (b_2 - b_1) w_p$$

  - $w_p = S_0^{-1}(1-p)$
- LRT statistic

$$\Lambda(\Delta) = 2\ell(\hat{u}_1, \hat{u}_2, \hat{b}_1, \hat{b}_2) - 2\ell(\tilde{u}_1, \tilde{u}_2, \tilde{b}_1, \tilde{b}_2)$$

Case 5

- under H_0

$$(\tilde{u}_1, \tilde{u}_2, \tilde{b}_1, \tilde{b}_2) = \arg\max_{\Theta} \ell(u_2 + \Delta + (b_2 - b_1)w_p, u_2, b_1, b_2)$$

- under H_1

$$(\hat{u}_1, \hat{u}_2, \hat{b}_2, \hat{b}_1) = \arg\max_{\Theta} \ell(u_1, u_2, b_1, b_2)$$

- $100(1-\alpha)$ confidence interval for $(y_{1p} - y_{2p})$ can be obtained from the set of Δ values that satisfy $\Lambda(\Delta) \leq \chi^2_{(1),1-\alpha}$

Comparison of Weibull or extreme value distributions

- Assume $T_{ji} \sim \text{Weibull}(\alpha_j, \beta_j)$ $(j = 1, \dots, m, \;\; i = 1, \dots, n_j)$
  - Data $\{(t_{ji}, \delta_{ji}), j = 1, \dots, m, \;\; i = 1, \dots, n_j\}$
- Survivor function of Weibull distribution

$$S_j(t) = \exp\left[\, - (t/\alpha_j)^{\beta_j} \right]$$

- Survivor function of extreme value distribution

$$S_j(y) = \exp\left[\, - e^{(y-u_j)/b_j} \right]$$

  - $u_j = \log \alpha_j$
  - $b_j = 1/\beta_j$

Example 5.4.1

- Data of the following table are on the time to breakdown of electrical insulating fluid subject to a constant voltage stress in a lifetest experiment

Table 1.1. Times to Breakdown (in minutes) at Each of Seven Voltage Levels

Voltage Level (kV)	n_i	Breakdown Times
26	3	5.79, 1579.52, 2323.7
28	5	68.85, 426.07, 110.29, 108.29, 1067.6
30	11	17.05, 22.66, 21.02, 175.88, 139.07, 144.12, 20.46, 43.40, 194.90, 47.30, 7.74
32	15	0.40, 82.85, 9.88, 89.29, 215.10, 2.75, 0.79, 15.93, 3.91, 0.27, 0.69, 100.58, 27.80, 13.95, 53.24
34	19	0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89
36	15	1.97, 0.59, 2.58, 1.69, 2.71, 25.50, 0.35, 0.99, 3.99, 3.67, 2.07, 0.96, 5.35, 2.90, 13.77
38	8	0.47, 0.73, 1.40, 0.74, 0.39, 1.13, 0.09, 2.38

Example 5.4.1

Table 7: Estimate of voltage-specific extreme value models

voltage	$\hat{u}_j \pm se(\hat{u}_j)$	$\hat{b}_j \pm se(\hat{b}_j)$
26	6.862 ± 1.104	1.834 ± 0.885
28	5.865 ± 0.486	1.022 ± 0.474
30	4.351 ± 0.302	0.944 ± 0.303
32	3.256 ± 0.486	1.781 ± 0.254
34	2.503 ± 0.315	1.297 ± 0.211
36	1.457 ± 0.309	1.125 ± 0.221
38	0.001 ± 0.273	0.734 ± 0.367

Example 5.4.1

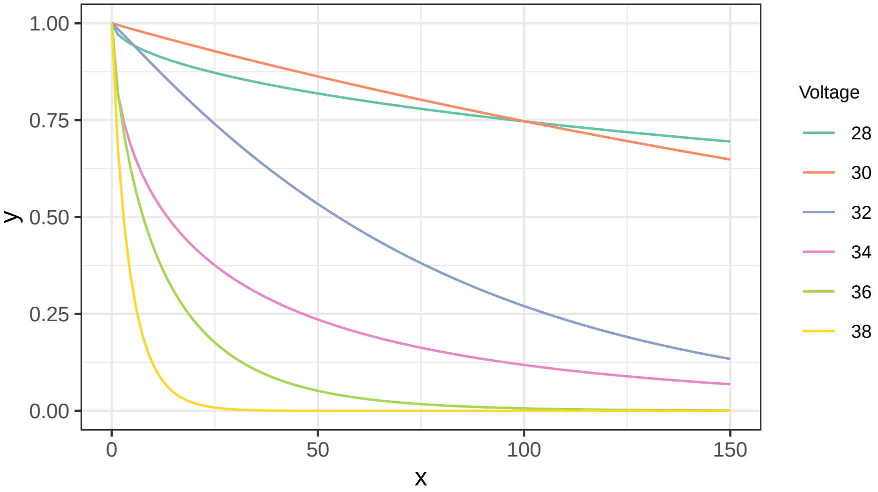

Figure 3: Comparison of estimated survivor function

LRT (Case 1)

- Null hypothesis

$$H_0 : b_1 = \cdots = b_7$$

- LRT statistic

$$\begin{aligned}\Lambda &= 2\ell(\hat{u}_1, \dots, \hat{u}_7, \hat{b}_1, \dots, \hat{b}_7) - 2\ell(\tilde{u}_1, \dots, \tilde{u}_7, \tilde{b}, \dots, \tilde{b}) \\ &= 2(-132.181) - 2(-136.578) \\ &= 8.794\end{aligned}$$

  - p-value

$$Pr(\chi^2_{(6)} \geq \Lambda) = 0.185$$

It does not provide enough evidence to reject the null hypothesis of equality of the scale parameters.

Confidence interval of (b_1/b_2) (Case 2)

- Wald-type

$$(\log \hat{b}_1 - \log \hat{b}_2) \pm z_{1-\alpha/2}\, se(\log \hat{b}_1 - \log \hat{b}_2)$$

$$(\hat{b}_1/\hat{b}_2)\; e^{\pm z_{1-\alpha/2}\, se(\log \hat{b}_1 - \log \hat{b}_2)}$$

$$(1.834/1.022)\; e^{\pm\,(1.96)(0.624)}$$

$$0.529 < (b_1/b_2) < 6.095$$

  - Similarly confidence intervals for $(b_j/b_{j'})$ $j > j'$ can be obtained

Confidence interval of (b_1/b_2) (Case 2)

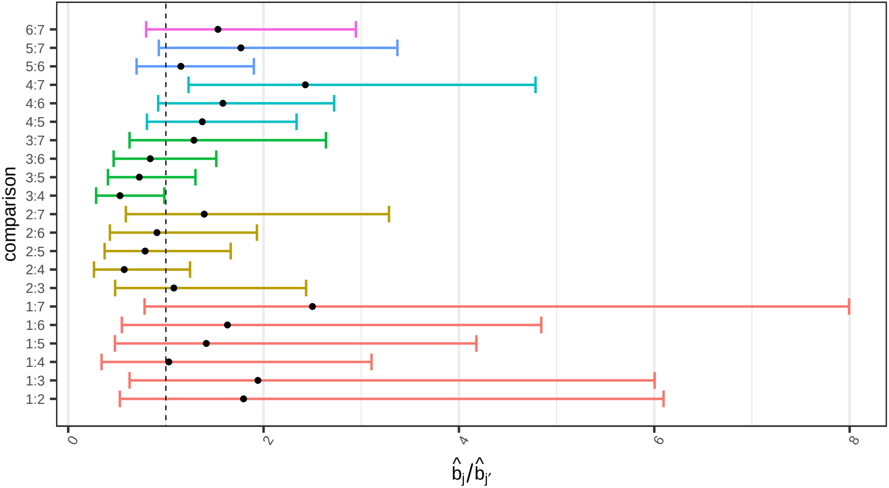

Figure 4: Estimate and 95% confidence interval of pair-wise comparisons of scale parameters $(b_j/b_{j'})$

Case 3

- Equality of all location parameters when scales are equal

$$H_0 : u_1 = \cdots = u_m,\ b_1 = \cdots = b_m$$

- LRT statistic

$$\begin{aligned}\Lambda &= 2\ell(\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{b}, \ldots, \tilde{b}) - 2\ell(u^\star, \ldots, u^\star, b^\star, \ldots, b^\star) \\ &= 2(-136.578) - 2(-176.584) \\ &= 80.013\end{aligned}$$

  - p-value $P\big(\chi^2_{(1)} > 80.013\big) < .001 \rightarrow$ There is a strong evidence against the assumption of equality of m location parameters

Case 4

- Wald-type confidence interval of $(u_1 - u_2)$

$$(\hat{u}_1 - \hat{u}_2) \pm z_{1-\alpha/2}\, se(\hat{u}_1 - \hat{u}_2)$$

$$(6.862 - 4.351) \pm (1.96)(1.206)$$

$$-1.367 < (u_1 - u_2) < 3.361$$

  - There is no significant difference between u_1 and u_2

Case 4

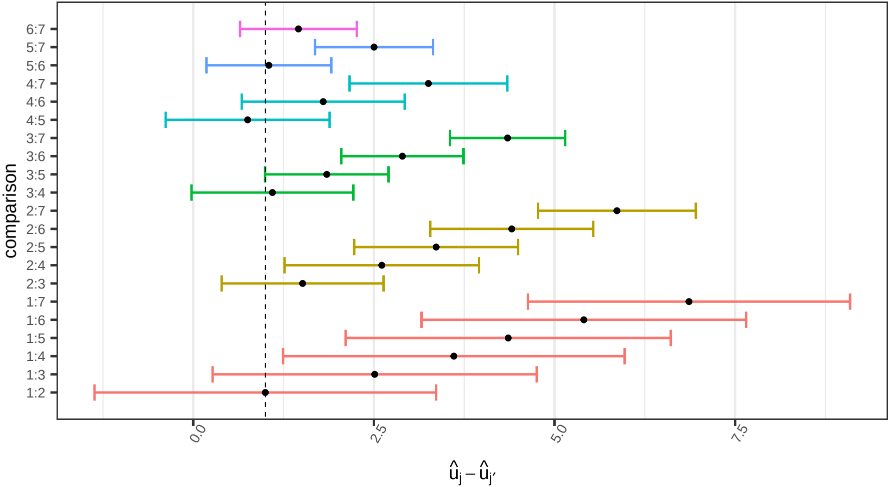

Figure 5: Estimate and 95% confidence interval of pair-wise comparisons of location parameters $(u_j - u_{j'})$

Case 5

- General expression of pth quantile of the group j

$$y_{jp} = u_j + b_j w_p,\ j = 1, \dots, m$$

- Difference of pth quantile between groups 1 and 2

$$y_{1p} - y_{2p} = u_1 - u_2 + (b_1 - b_2) w_p$$

  - 95% confidence interval for the difference of median between groups 1 and 2

$$\hat{y}_{1m} - \hat{y}_{2m} \pm z_{1-\alpha/2} se(\hat{y}_{1m} - \hat{y}_{2m})$$
$$(6.19 - 5.491) \pm (1.96)(1.291)$$
$$-1.831 < (y_{1m} - y_{2m}) < 3.231$$

Case 5

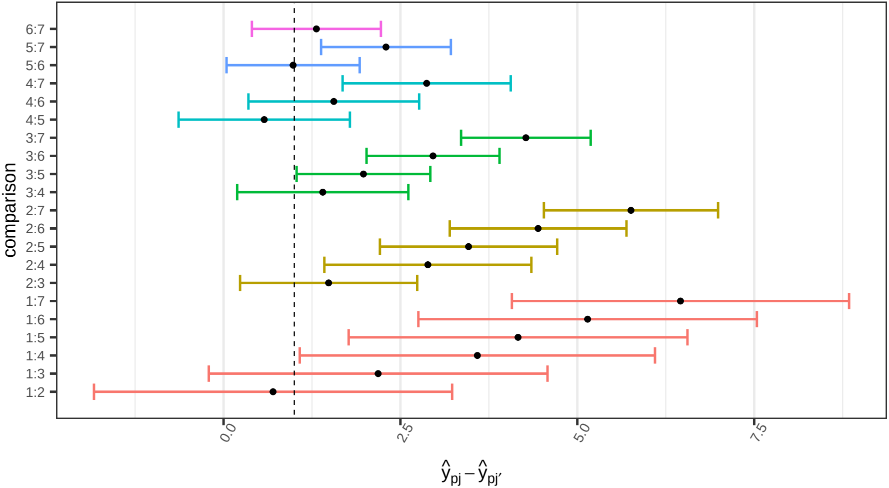

Figure 6: Estimate and 95% confidence interval of pair-wise comparisons of medians $(y_{j,.5} - y_{j',.5})$

Homework

- Analyse the breakdown time data using log-logistic and log-normal distributions and compare the results with that of Weibull distribution