

# Chapter 4

## (AST305) Lifetime Data Analysis I

Md Rasel Biswas

# Lecture Outline

- 1 4. Inference procedures for parametric models
  - 4.1 Exponential Distribution
  - 4.2 Gamma Distribution (Uncensored Data)

## Section 1

### 4. Inference procedures for parametric models

## Subsection 1

### Introduction

# Introduction

- Likelihood methods for lifetime data were introduced in Chapter 2, which includes derivation of likelihood function for different types of censored data
  - ▶ Maximum likelihood estimator
  - ▶ Inference about parameters (hypothesis testing and confidence intervals)

## Likelihood function (Complete data)

- Let  $t_1, \dots, t_n$  be a random sample from a distribution  $f(t; \theta)$ , where  $\theta$  is a  $p$ -dimensional vector of parameters
- Likelihood and log-likelihood function

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta)$$

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(t_i; \theta)$$

- Maximum likelihood estimator (MLE)

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ell(\theta) \tag{1}$$

## Statistical inference (Complete data)

- Large sample property of MLE

$$\hat{\theta} \sim \mathcal{N}(\theta, [\mathcal{J}(\theta)]^{-1}) \quad (2)$$

- ▶ Fisher expected information matrix

$$\mathcal{J}(\theta) = E \left[ - \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right] \quad (3)$$

- ▶ Observed information matrix

$$I(\hat{\theta}) = - \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \Bigg|_{\theta=\hat{\theta}} \quad (4)$$

# Likelihood function (Type I or random censoring)

- Data:  $\{(t_i, \delta_i), i = 1, \dots, n\}$ 
  - ▶  $t_i$  is a sample realization of  $\tilde{T}_i = \min(T_i, C_i)$
  - ▶ lifetime  $T_i$  follows a distribution with pdf  $f(t_i; \theta)$  and the corresponding survivor function  $S(t_i; \theta)$
  - ▶ censoring time  $C_i$  could be random or fixed depending on the censoring mechanism
  - ▶  $\delta_i = I(T_i \leq C_i)$ , censoring indicator

## Likelihood function (Type I or random censoring)

- Data for type I or random censoring

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

- Likelihood function

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i} \\ &= \prod_{i=1}^n [S(t_i; \theta) h(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i} \\ &= \prod_{i=1}^n [S(t_i; \theta)] [h(t_i; \theta)]^{\delta_i} \end{aligned} \tag{5}$$

## Likelihood function (Type II censoring)

- Let lifetime  $T_i$  ( $i = 1, \dots, n$ ) follows a distribution with pdf  $f(t_i; \theta)$  and the corresponding survivor function  $S(t_i; \theta)$
- The experiment was terminated after observing  $r$  smallest lifetimes

$$t_{(1)} \leq \dots \leq t_{(r)}$$

- The remaining  $(n - r)$  observations are considered as censored at  $t_{(r)}$
- Likelihood function

$$L(\theta) = \left[ \prod_{i=1}^r f(t_{(i)}; \theta) \right] \left[ S(t_{(r)}; \theta) \right]^{n-r} \quad (6)$$

## Statistical inference (censored sample)

- For censored samples, the result “asymptotic distribution of MLE is normal” is still valid
- The expression of the Fisher expected information matrix  $\mathcal{J}(\theta)$  is complex for censored data, observed information matrix  $I(\hat{\theta})$  is used in making inference with censored sample

## Subsection 2

### 4.1 Exponential Distribution

## 4.1 Exponential Distribution

- The exponential distribution is the first lifetime model for which statistical methodology were extensively developed
- Exact tests can be developed for exponential distribution for certain type of censoring mechanism
- Exponential distribution assume constant hazard and its use is limited for analyzing real life problems

## 4.1 Exponential Distribution

- Probability density function

$$f(t; \theta) = (1/\theta) \exp(-t/\theta) \quad t \geq 0, \theta > 0$$

- Hazard function

$$h(t; \theta) = (1/\theta)$$

- Survivor function

$$S(t; \theta) = \exp(-t/\theta)$$

- $p$ th quantile

$$S(t_p; \theta) = 1 - p \Rightarrow \exp(-t_p/\theta) = 1 - p \Rightarrow t_p = -\theta \log(1 - p)$$

## 4.1 Exponential Distribution

### Homework

- Estimation and related inference of exponential distribution when the sample has no censored observations

## Subsection 3

### Type I or random censoring

## Type I or random censoring

- Lifetime  $T \sim \text{Exp}(\theta)$ ,  $\theta > 0$
- Data:  $\{(t_i, \delta_i), i = 1, \dots, n\}$
- Likelihood function

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i} \\ &= \prod_{i=1}^n S(t_i; \theta) [h(t_i; \theta)]^{\delta_i} \\ &= \prod_{i=1}^n \exp(-t_i/\theta) (1/\theta)^{\delta_i} \end{aligned} \tag{7}$$

## Type I or random censoring

- Likelihood function

$$L(\theta) = \prod_{i=1}^n \exp(-t_i/\theta) (1/\theta)^{\delta_i}$$

- Log-likelihood function

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n [- (t_i/\theta) - \delta_i \log(\theta)] \\ &= -\frac{1}{\theta} \sum_i t_i - r \log(\theta) \end{aligned} \quad (8)$$

- ▶  $r = \sum_i \delta_i \rightarrow$  the number of failures observed in the sample

## Type I or random censoring

- Log-likelihood function

$$\ell(\theta) = -\frac{1}{\theta} \sum_i t_i - r \log(\theta)$$

- MLE

$$\left. \frac{\partial \ell(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \Rightarrow \frac{1}{\hat{\theta}^2} \sum_i t_i - \frac{r}{\hat{\theta}} = 0$$

$$\hat{\theta} = \sum_{i=1}^n t_i / r \tag{9}$$

- ▶ Assuming  $r > 0$  and no finite MLE exists for  $r = 0$

## Type I or random censoring

- Information matrix

$$\begin{aligned} I(\theta) &= -\frac{\partial^2 \ell(\theta)}{\partial \theta^2} \\ &= -\frac{\partial}{\partial \theta} \left[ \frac{1}{\theta^2} \sum_i t_i - \frac{r}{\theta} \right] \\ &= \frac{2}{\theta^3} \sum_i t_i - \frac{r}{\theta^2} \end{aligned} \tag{10}$$

- Replacing the parameter  $\theta$  by its MLE  $\hat{\theta} = \sum_i t_i / r$ , the observed information matrix becomes

$$I(\hat{\theta}) = \frac{r}{\hat{\theta}^2}$$

## Confidence interval (Method I)

- Using  $\hat{\theta}$ 's asymptotic distribution

$$Z_1 = \frac{\theta - \hat{\theta}}{[I(\hat{\theta})]^{-1/2}} = \frac{\theta - \hat{\theta}}{\hat{\theta}/\sqrt{r}} \sim \mathcal{N}(0, 1)$$

- For a small sample,  $Z_1$  does not approximate the standard normal distribution very accurately
- For a sample with a small number of uncensored observations,  $\ell(\theta)$  tends to be asymmetric

## Confidence interval (Method II)

- Sprott (1980) showed that  $\ell_1(\phi) = \ell(\phi^{-3})$  is more closer to symmetric compared to  $\ell(\theta)$ , where  $\phi = \theta^{-1/3} \Rightarrow \phi^3 = 1/\theta$

$$\ell(\theta) = -(1/\theta) \sum_i t_i - r \log \theta \quad (11)$$

$$\ell_1(\phi) = -\phi^3 \sum_i t_i + 3r \log(\phi) \quad (12)$$

▶ MLE  $\hat{\phi} = \hat{\theta}^{-1/3} = \left[ \sum_i t_i / r \right]^{-1/3}$

- ▶ Observed information matrix

$$I_1(\phi) = 6\phi \sum_i t_i + \frac{3r}{\phi^2} \Rightarrow I_1(\hat{\phi}) = \frac{9r}{\hat{\phi}^2}$$

## Confidence interval (Method II)

- Pivotal quantity

$$Z_2 = \frac{\phi - \hat{\phi}}{[I_1(\hat{\phi})]^{-1/2}} = \frac{\phi - \hat{\phi}}{\hat{\phi}/\sqrt{9r}} \sim \mathcal{N}(0, 1)$$

- ▶ Approximation of  $Z_2$  is quite accurate compared to that of  $Z_1$

## Confidence interval (Method III)

- Likelihood ratio statistic  $\Lambda(\theta) = 2\ell(\hat{\theta}) - 2\ell(\theta)$  can also be used to obtain confidence interval for  $\theta$ , where

$$\begin{aligned}\ell(\theta) &= -(1/\theta) \sum_i t_i - r \log(\theta) \\ &= -r(\hat{\theta}/\theta) - r \log(\theta)\end{aligned}\tag{13}$$

$$\begin{aligned}\ell(\hat{\theta}) &= -(1/\hat{\theta}) \sum_i t_i - r \log(\hat{\theta}) \\ &= -r - r \log(\hat{\theta})\end{aligned}\tag{14}$$

## Confidence interval (Method III)

- LRT statistic

$$\begin{aligned}\Lambda(\theta) &= 2\ell(\hat{\theta}) - 2\ell(\theta) \\ &= 2r[(\hat{\theta}/\theta) - 1 + \log(\theta/\hat{\theta})]\end{aligned}$$

- Two-sided  $(1 - \alpha)100\%$  confidence intervals are obtained as the set of  $\theta$  values that satisfy

$$\Lambda(\theta) \leq \chi_{(1),1-\alpha}^2$$

## Confidence interval of survivor function

- Confidence interval for the parameter  $\theta$  can also be used to obtain confidence intervals for a monotone function of  $\theta$ , such as

$$S(t; \theta) = \exp(-t/\theta) \text{ or } h(t; \theta) = 1/\theta$$

- Let  $100(1 - \alpha)\%$  confidence interval for  $\theta$

$$L(\text{Data}) \leq \theta \leq U(\text{Data}),$$

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

## Confidence interval of survivor function

- $100(1 - \alpha)\%$  confidence interval for  $S(t_0; \theta) = \exp(-t_0/\theta)$

$$L(\text{Data}) \leq \theta \leq U(\text{Data})$$

$$t_0/U(\text{Data}) \leq (t_0/\theta) \leq t_0/L(\text{Data})$$

$$-t_0/L(\text{Data}) \leq (-t_0/\theta) \leq -t_0/U(\text{Data})$$

$$\exp(-t_0/L(\text{Data})) \leq S(t_0; \theta) \leq \exp(-t_0/U(\text{Data}))$$

## Confidence interval of hazard function

- $100(1 - \alpha)\%$  confidence interval for  $h(t_0; \theta) = 1/\theta$

$$(1/U(\text{Data})) \leq h(t_0; \theta) \leq (1/L(\text{Data}))$$

## Example 4.1.1

- Lifetimes (in days) of 10 pieces of equipment

time	status
2	1
72	0
51	1
60	0
33	1
27	1
14	1
24	1
4	1
21	0

- Assume lifetimes follow exponential distribution with parameter, i.e.  $T_i \sim \text{Exp}(\theta)$

## Example 4.1.1

- MLE

$$\hat{\theta} = \frac{\sum_i t_i}{r} = \frac{308}{7} = 44.0$$

- 95% confidence interval of  $\theta$  (Method I)

$$\begin{aligned}\hat{\theta} \pm z_{.975}[I(\hat{\theta})]^{-1/2} &\Rightarrow \hat{\theta} \pm z_{.975}(\hat{\theta}/\sqrt{r}) \\ &\Rightarrow 44.0 \pm (1.96)(44.0/\sqrt{7}) \\ &\Rightarrow 11.4 \leq \theta \leq 76.6\end{aligned}$$

## Example 4.1.1

- 95% confidence interval of  $\theta$  (Method II)

- ▶  $\hat{\phi} = \hat{\theta}^{-1/3} = (44.0)^{-1/3} = 0.283$

- ▶ Confidence interval for  $\phi$

$$\begin{aligned}\hat{\phi} \pm z_{.975}[I_1(\hat{\phi})]^{-1/2} &\Rightarrow \hat{\phi} \pm z_{.975}(\hat{\phi}/\sqrt{9r}) \\ &\Rightarrow 0.28 \pm (1.96)(0.28/\sqrt{63}) \\ &\Rightarrow 0.21 \leq \phi \leq 0.35\end{aligned}$$

- ▶ Confidence interval for  $\theta$

$$\begin{aligned}0.21 \leq \phi \leq 0.35 &\Rightarrow 0.21 \leq \theta^{-1/3} \leq 0.35 \\ &\Rightarrow (0.35)^{-3} \leq \theta \leq (0.21)^{-3} \\ &\Rightarrow 22.69 \leq \theta \leq 103.03\end{aligned}$$

## Example 4.1.1

- Likelihood ratio statistic

$$\Lambda(\theta) = 2r[(\hat{\theta}/\theta) - 1 - \log(\hat{\theta}/\theta)]$$

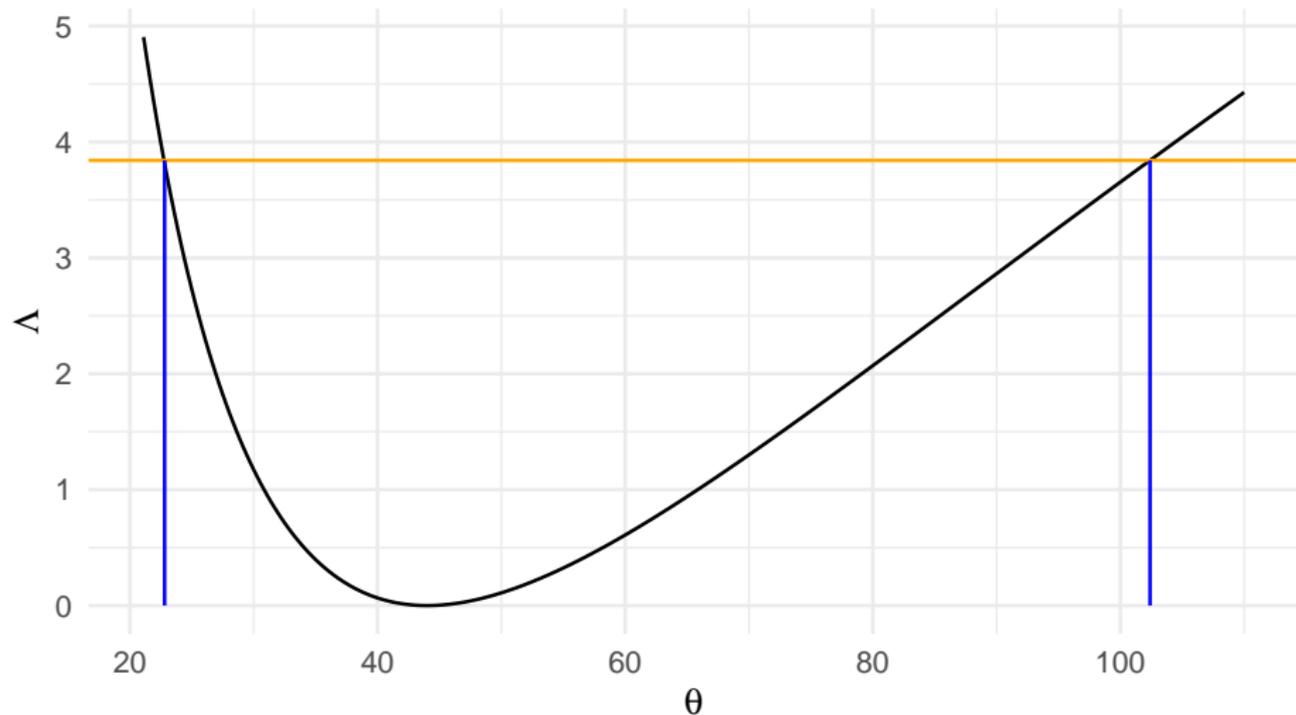
- ▶ 95% CI for  $\theta$  can be obtained from the set of  $\theta$ 's such that

$$\Lambda(\theta) \leq \chi_{(1),.95}^2 = 3.84$$

- ▶ 95% CI for  $\theta$

22.8 to 102.4

## Example 4.1.1



## Example 4.1.1

Table 2: A comparison of 95% CI for  $\theta$  with the data on lifetimes of 10 pieces of equipment (Example 4.1.1)

method	lower	upper
Normal approx.	11.40	76.60
Sprott	22.69	103.03
LRT	22.80	102.40

## Methods for CI

- Normal approximation

$$W_1(\theta) = \left[ \frac{(\hat{\theta} - \theta)}{[I(\hat{\theta})]^{-1/2}} \right]^2 = (\hat{\theta} - \theta)^2 I(\hat{\theta})$$

- Sprott's method

$$W_2(\theta) = \left[ \frac{(\hat{\phi} - \phi)}{[I_1(\hat{\phi})]^{-1/2}} \right]^2 = (\hat{\theta}^{-1/3} - \theta^{-1/3})^2 I_1(\hat{\theta}^{-1/3})$$

- LRT

$$W_3(\theta) = 2\ell(\hat{\theta}) - 2\ell(\theta) = 2r \left[ (\hat{\theta}/\theta) - 1 - \log(\hat{\theta}/\theta) \right]$$

- $W_j(\theta) \sim \chi_{(1)}^2$   $j = 1, 2, 3$  under  $H_0$

# Methods for CI

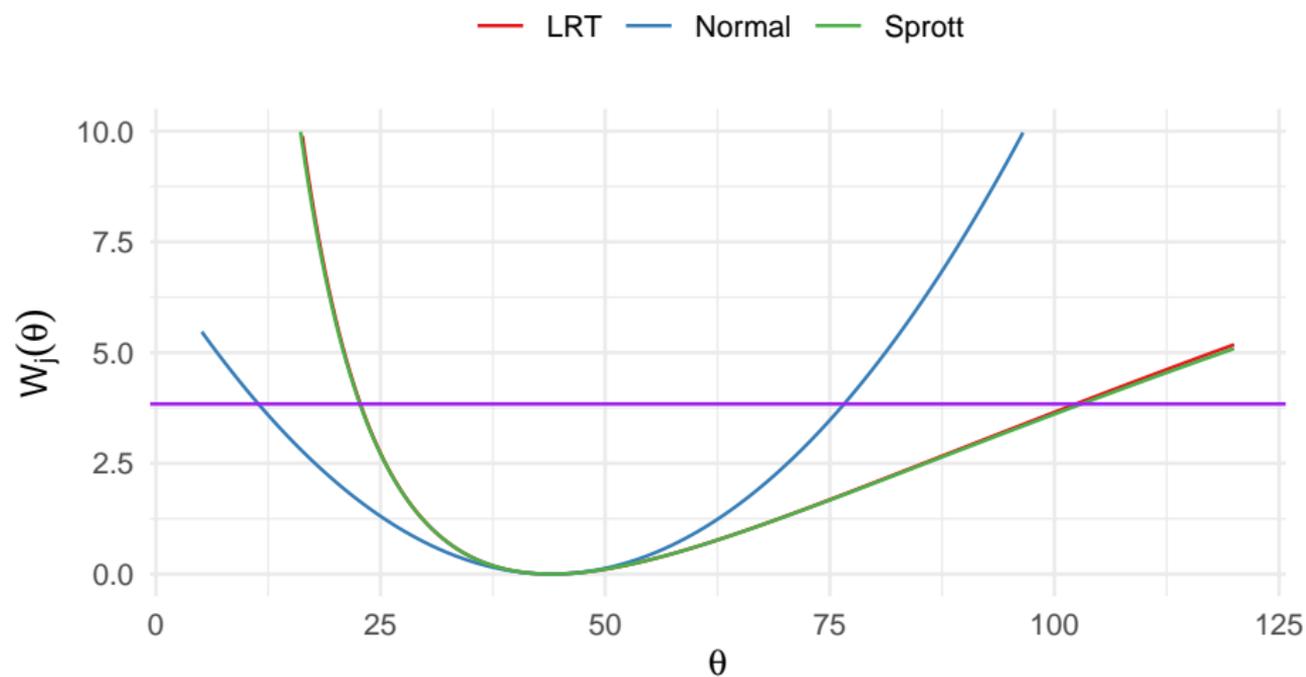


Figure 1: Comparisons between normal approximations and LRT

## Subsection 4

### Type II censoring plan

## Type II censoring plan

- Assume lifetime  $T \sim \text{Exp}(\theta)$
- Let  $t_{(1)} < \dots < t_{(r)}$  be the  $r$  smallest lifetimes observed from an experiment with  $n$  ( $\geq r$ ) subjects
- The joint distribution of  $t_{(1)}, \dots, t_{(r)}$

$$\begin{aligned} f_T(t_{(1)}, \dots, t_{(r)}) &= \frac{n!}{(n-r)!} \left[ \prod_{i=1}^r (1/\theta) e^{-t_{(i)}/\theta} \right] \left[ e^{-t_{(r)}/\theta} \right]^{n-r} \\ &= \frac{n!}{(n-r)!} (1/\theta)^r e^{-\frac{1}{\theta} [\sum_{i=1}^r t_{(i)} + (n-r)t_{(r)}]} \\ &= \frac{n!}{(n-r)!} (1/\theta)^r e^{-T_0/\theta} \end{aligned} \tag{15}$$

## Type II censoring plan

- The log-likelihood function

$$\begin{aligned}\ell(\theta) &= -\frac{1}{\theta} \sum_i t_{(i)} - r \log(\theta) - (1/\theta)(n-r)t_{(r)} + \text{Const.} \\ &= -\frac{1}{\theta} \left[ \sum_{i=1}^r t_{(i)} + (n-r)t_{(r)} \right] - r \log(\theta) + \text{Const.} \\ &= -\frac{T_0}{\theta} - r \log(\theta) + \text{Const.}\end{aligned}\tag{16}$$

## Type II censoring plan

- Maximum likelihood estimator

$$\left. \frac{\partial \ell(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \Rightarrow \hat{\theta} = (1/r) \left[ \sum_{i=1}^r t_{(i)} + (n-r)t_{(r)} \right]$$

- ▶ Observed information matrix

$$I(\hat{\theta}) = \left. \frac{-\partial^2 \ell(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}} = \frac{r}{\hat{\theta}^2}$$

- $(1 - \alpha)100\%$  CI for  $\theta$  (by Normal approximation)

$$\hat{\theta} \pm z_{1-\alpha/2} \left[ I(\hat{\theta}) \right]^{-1/2} \quad (17)$$

## Exact confidence interval

- Exact confidence interval for  $\theta$  can be derived for uncensored and Type II censored samples, define

$$W_1 = nt_{(1)}$$

$$W_2 = (n - 1)(t_{(2)} - t_{(1)})$$

$\vdots$

$$W_r = (n - r + 1)(t_{(r)} - t_{(r-1)})$$

- In general

$$W_i = (n - i + 1)(t_{(i)} - t_{(i-1)}) \quad i = 1, \dots, r$$

## Exact confidence interval

- It can be shown that

$$\sum_{i=1}^r W_i = \sum_{i=1}^r t_{(i)} + (n - r)t_{(r)} = T_0 \quad (18)$$

## Exact confidence interval

- The joint distribution of

$W_1 = g_1(t_{(1)}, \dots, t_{(r)}), \dots, W_r = g_r(t_{(1)}, \dots, t_{(r)})$  is:

$$f_W(w_1, \dots, w_r) = f_T(g_1^{-1}(w_1, \dots, w_r), \dots, g_r^{-1}(w_1, \dots, w_r)) \times |J|$$

where

$$J = \frac{\partial(g_1^{-1}(w_1, \dots, w_r), \dots, g_r^{-1}(w_1, \dots, w_r))}{\partial(w_1, \dots, w_r)}$$
$$= \begin{bmatrix} \frac{\partial g_1^{-1}(w_1, \dots, w_r)}{\partial w_1} & \frac{\partial g_2^{-1}(w_1, \dots, w_r)}{\partial w_1} & \dots & \frac{\partial g_r^{-1}(w_1, \dots, w_r)}{\partial w_1} \\ \frac{\partial g_1^{-1}(w_1, \dots, w_r)}{\partial w_2} & \frac{\partial g_2^{-1}(w_1, \dots, w_r)}{\partial w_2} & \dots & \frac{\partial g_r^{-1}(w_1, \dots, w_r)}{\partial w_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1^{-1}(w_1, \dots, w_r)}{\partial w_r} & \frac{\partial g_2^{-1}(w_1, \dots, w_r)}{\partial w_r} & \dots & \frac{\partial g_r^{-1}(w_1, \dots, w_r)}{\partial w_r} \end{bmatrix}$$

## Exact confidence interval

$$w_1 = g_1(t_{(1)}, \dots, t_{(r)}) = nt_{(1)} \Rightarrow t_{(1)} = g_1^{-1}(w_1, \dots, w_r) = \frac{w_1}{n}$$

$$w_2 = g_2(t_{(1)}, \dots, t_{(r)}) = (n-1)(t_{(2)} - t_{(1)})$$

$$\Rightarrow t_{(2)} = g_2^{-1}(w_1, \dots, w_r) = \frac{w_2}{n-1} + \frac{w_1}{n}$$

⋮

$$w_r = g_r(t_{(1)}, \dots, t_{(r)}) = (n-r+1)(t_{(r)} - t_{(r-1)})$$

$$\Rightarrow t_{(r)} = g_r^{-1}(w_1, \dots, w_r) = \frac{w_r}{n-r+1} + \dots + \frac{w_2}{n-1} + \frac{w_1}{n}$$

## Exact confidence interval

$$J = \begin{bmatrix} \frac{\partial g_1^{-1}(w_1, \dots, w_r)}{\partial w_1} & \frac{\partial g_2^{-1}(w_1, \dots, w_r)}{\partial w_1} & \dots & \frac{\partial g_r^{-1}(w_1, \dots, w_r)}{\partial w_1} \\ \frac{\partial g_1^{-1}(w_1, \dots, w_r)}{\partial w_2} & \frac{\partial g_2^{-1}(w_1, \dots, w_r)}{\partial w_2} & \dots & \frac{\partial g_r^{-1}(w_1, \dots, w_r)}{\partial w_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1^{-1}(w_1, \dots, w_r)}{\partial w_r} & \frac{\partial g_2^{-1}(w_1, \dots, w_r)}{\partial w_r} & \dots & \frac{\partial g_r^{-1}(w_1, \dots, w_r)}{\partial w_r} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ 0 & \frac{1}{n-1} & \dots & \frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n-r+1} \end{bmatrix}$$

- We can write

$$|J| = \frac{1}{n \cdot (n-1) \cdots (n-r+1)} = \frac{(n-r)!}{n!} \quad (19)$$

## Exact confidence interval

- The joint distribution of

$W_1 = g_1(t_{(1)}, \dots, t_{(r)}), \dots, W_r = g_r(t_{(1)}, \dots, t_{(r)})$  is

$$\begin{aligned} f_W(w_1, \dots, w_r) &= f_T(g_1^{-1}(w_1, \dots, w_r), \dots, g_r^{-1}(w_1, \dots, w_r)) |J| \\ &= \frac{n!}{(n-r)!} \frac{1}{\theta^r} e^{-\frac{1}{\theta} \{ \sum_i g_i^{-1}(w_i) + (n-r)g_r^{-1}(w_r) \}} \frac{(n-r)!}{n!} \\ &= \frac{1}{\theta^r} e^{-\sum_i w_i / \theta} \end{aligned} \quad (20)$$

▶ Using Equation 18

- Equation 20 shows that

▶  $W_1, \dots, W_r$  are independent and  $W_i \sim \text{Exp}(\theta)$

## Exact confidence interval

- Since  $W_i$ ' are independent and  $W_i \sim \text{Exp}(\theta)$ , so

$$T_0 = \sum_i W_i = \sum_i t_{(i)} + (n - r)t_{(r)}$$

follows a gamma distribution with shape parameter  $r$  and scale parameter  $\theta$

- Then using the relationship between gamma and chi-square distribution

$$\frac{2T_0}{\theta} \sim \chi_{(2r)}^2$$

# Exact confidence interval

## Review

- Moment generating functions for different distributions

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \begin{cases} \frac{1}{1-\theta t} & \text{for } X \sim \text{Exp}(\theta) \\ \left(\frac{1}{1-\theta t}\right)^r & \text{for } X \sim \text{Gamma}(r, \theta) \\ \left(\frac{1}{1-2t}\right)^{r/2} & \text{for } X \sim \chi_{(r)}^2 \end{cases} \end{aligned}$$

## Exact confidence interval

- Since  $\frac{2T_0}{\theta} \sim \chi^2_{(2r)}$ , we can write

$$P\left(\chi^2_{(2r), \alpha/2} \leq \frac{2T_0}{\theta} \leq \chi^2_{(2r), 1-\alpha/2}\right) = 1 - \alpha$$

- ▶  $\chi^2_{(2r), p} \rightarrow p$ th quantile of  $\chi^2_{(2r)}$  distribution
- $(1 - \alpha)100\%$  exact confidence interval for  $\theta$

$$\frac{2T_0}{\chi^2_{(2r), 1-\alpha/2}} \leq \theta \leq \frac{2T_0}{\chi^2_{(2r), \alpha/2}} \quad (21)$$

## Example 4.1.3

- The first 8 observations in a random sample of 12 lifetimes (in hours) from an assumed exponential distribution are

31, 58, 157, 185, 300, 470, 497, 673

- *Homework*
  - ▶ Obtain 95% exact and approximate confidence intervals for  $\theta$

## Subsection 5

### 4.1.3 Comparison of distributions

## 4.1.3 Comparison of distributions

- Comparison of two or more lifetime distributions is often of interest in practice
- For comparing two or more independent exponential distributions, different methods of hypothesis tests and confidence intervals are available

## Likelihood ratio tests

- Let  $T_{ij}$  be the lifetime corresponding to the  $j$ th subject of the  $i$ th group ( $i = 1, \dots, m; j = 1, \dots, n_i$ )
- Assume  $T_{ij} \sim \text{Exp}(\theta_i)$  and the null hypothesis of interest

$$H_0 : \theta_1 = \dots = \theta_m$$

- Data:

$$\{(t_{ij}, \delta_{ij}), i = 1, \dots, m; j = 1, \dots, n_i\}$$

## Likelihood ratio tests

- The likelihood function for  $\theta = (\theta_1, \dots, \theta_m)'$

$$\begin{aligned}L(\theta) &= \prod_{i=1}^m \prod_{j=1}^{n_i} [f(t_{ij}; \theta_i)]^{\delta_{ij}} [S(t_{ij}; \theta_i)]^{1-\delta_{ij}} \\ &= \prod_{i=1}^m \prod_{j=1}^{n_i} \left[\frac{1}{\theta_i}\right]^{\delta_{ij}} e^{-t_{ij}/\theta_i}\end{aligned}$$

- The corresponding log-likelihood function

$$\ell(\theta) = - \sum_i \sum_j \left[ (t_{ij}/\theta_i) + \delta_{ij} \log(\theta_i) \right] \quad (22)$$

# Likelihood ratio tests

- The log-likelihood function

$$\begin{aligned}\ell(\theta) &= - \sum_i \sum_j \left[ (t_{ij}/\theta_i) + \delta_{ij} \log(\theta_i) \right] \\ &= - \sum_i \left[ (T_i/\theta_i) + r_i \log(\theta_i) \right]\end{aligned}$$

▶  $T_i = \sum_j t_{ij}$  and  $r_i = \sum_j \delta_{ij}$

- Maximum likelihood estimator of  $\theta_i$  ( $i = 1, \dots, m$ )

$$\left. \frac{\partial \ell(\theta)}{\partial \theta_i} \right|_{\theta_i = \hat{\theta}_i} = 0 \Rightarrow \frac{T_i}{\hat{\theta}_i^2} - \frac{r_i}{\hat{\theta}_i} = 0 \Rightarrow \hat{\theta}_i = (T_i/r_i)$$

## Likelihood ratio tests

- Under

$$H_0 : \theta_1 = \dots = \theta_m = \theta \text{ (say)}$$

log-likelihood function

$$\ell(\theta) = - \sum_i \left[ (T_i/\theta) + r_i \log(\theta) \right]$$

- ▶ Under  $H_0$ , the MLE of  $\theta$

$$\tilde{\theta} = \frac{\sum_i T_i}{\sum_i r_i}$$

# Likelihood ratio tests

- Likelihood ratio test statistic

$$\begin{aligned}\Lambda &= 2\ell(\hat{\theta}) - 2\ell(\tilde{\theta}) \\ &= 2 \sum_i \left[ - (T_i/\hat{\theta}_i) - r_i \log(\hat{\theta}_i) + (T_i/\tilde{\theta}) + r_i \log(\tilde{\theta}) \right] \\ &= 2 \sum_i \left[ - r_i \log(\hat{\theta}_i) + r_i \log(\tilde{\theta}) \right]\end{aligned}$$

- ▶ Under  $H_0$ ,  $\Lambda \sim \chi_{(m-1)}^2$  for a large sample size

# Likelihood ratio tests

## Example 4.1.4

- Four independent samples of size 10 each had 7 failures
- MLE under exponential model

$$\hat{\theta}_1 = 106, \hat{\theta}_2 = 80, \hat{\theta}_3 = 140, \hat{\theta}_4 = 158$$

- *Homework*
  - ▶ Test  $H_0 : \theta_1 = \dots = \theta_4$

## Confidence intervals for $\theta_1/\theta_2$

- Comparison between two exponential distributions

$$T_i \sim \text{Exp}(\theta_i), \quad i = 1, 2$$

- It can be shown that MLE  $\hat{\theta}_i$  approximately follows normal distribution

$$\hat{\theta}_i \sim \mathcal{N}(\theta_i, \theta_i^2/r_i) \quad (23)$$

$$\log \hat{\theta}_i \sim \mathcal{N}(\log \theta_i, 1/r_i) \quad (24)$$

- Distribution of  $\log(\hat{\theta}_1/\hat{\theta}_2)$

$$\log \hat{\theta}_1 - \log \hat{\theta}_2 = \log(\hat{\theta}_1/\hat{\theta}_2) \sim \mathcal{N}\left(\log(\theta_1/\theta_2), (r_1^{-1} + r_2^{-1})\right) \quad (25)$$

## Confidence intervals for $\theta_1/\theta_2$

- Null hypothesis

$$H_0 : \theta_1 = \theta_2 \Rightarrow H_0 : \theta_1/\theta_2 = 1 \Rightarrow H_0 : \log \theta_1 = \log \theta_2$$

- Test statistic

$$Z = \frac{\log(\hat{\theta}_1/\hat{\theta}_2) - \log(\theta_1/\theta_2)}{(r_1^{-1} + r_2^{-1})^{1/2}} \sim \mathcal{N}(0, 1)$$

## Confidence intervals for $\theta_1/\theta_2$

- 95% CI for  $\log(\theta_1/\theta_2)$

$$\log(\hat{\theta}_1/\hat{\theta}_2) \pm z_{1-\alpha/2} (r_1^{-1} + r_2^{-1})^{1/2} \quad (26)$$

- 95% CI for  $(\theta_1/\theta_2)$

$$(\hat{\theta}_1/\hat{\theta}_2) \exp \left( \pm z_{1-\alpha/2} (r_1^{-1} + r_2^{-1})^{1/2} \right)$$

## Confidence intervals for $\theta_1/\theta_2$

- Confidence intervals can also be found by inverting the likelihood ratio test for a hypothesis of the form

$$H_0 : \theta_1 = a\theta_2$$

where  $a > 0$

- For Type I sample  $\{(t_{ij}, \delta_{ij}), i = 1, 2; j = 1, \dots, n_i\}$ , the log-likelihood function

$$\ell(\theta_1, \theta_2) = - \sum_{i=1}^2 \left[ (T_i/\theta_i) + r_i \log(\theta_i) \right]$$

- MLE  $\hat{\theta}_1 = (T_1/r_1)$  and  $\hat{\theta}_2 = (T_2/r_2)$

## Confidence intervals for $\theta_1/\theta_2$

- Under  $H_0 : \theta_1 = a\theta_2$ , the log-likelihood function

$$\begin{aligned}\ell(a\theta_2, \theta_2) &= -r_1 \log(a\theta_2) - (T_1/(a\theta_2)) - r_2 \log(\theta_2) - (T_2/\theta_2) \\ &= -(r_1 + r_2) \log(\theta_2) - (T_1/(a\theta_2)) - (T_2/\theta_2) - r_1 \log(a)\end{aligned}$$

- ▶ MLE under  $H_0$

$$\tilde{\theta}_2 = \frac{T_1 + aT_2}{a(r_1 + r_2)} \Rightarrow \tilde{\theta}_1 = a\tilde{\theta}_2 = \frac{T_1 + aT_2}{(r_1 + r_2)}$$

## Confidence intervals for $\theta_1/\theta_2$

- The likelihood ratio test statistic

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

where

$$\begin{aligned}\ell(\hat{\theta}_1, \hat{\theta}_2) &= -\sum_{i=1}^2 \left[ (T_i/\hat{\theta}_i) + r_i \log(\hat{\theta}_i) \right] \\ &= -(r_1 + r_2) - r_1 \log(\hat{\theta}_1) - r_2 \log(\hat{\theta}_2)\end{aligned}$$

## Confidence intervals for $\theta_1/\theta_2$

and

$$\begin{aligned}\ell(\tilde{\theta}_1, \tilde{\theta}_2) &= -\frac{T_1}{\tilde{\theta}_1} - \frac{T_2}{\tilde{\theta}_2} - r_1 \log(\tilde{\theta}_1) - r_2 \log(\tilde{\theta}_2) \\ &= -(r_1 + r_2) - r_1 \log(\tilde{\theta}_1) - r_2 \log(\tilde{\theta}_2)\end{aligned}$$

- The likelihood ratio test statistic

$$\begin{aligned}\Lambda(a) &= 2\ell(\hat{\theta}_1, \hat{\theta}_2) - 2\ell(\tilde{\theta}_1, \tilde{\theta}_2) \\ &= 2r_1 \log(\tilde{\theta}_1/\hat{\theta}_1) + 2r_2 \log(\tilde{\theta}_2/\hat{\theta}_2) \\ &= 2r_1 \log(a\tilde{\theta}_2/\hat{\theta}_1) + 2r_2 \log(\tilde{\theta}_2/\hat{\theta}_2)\end{aligned}$$

## Confidence intervals for $\theta_1/\theta_2$

- $(1 - \alpha)100\%$  confidence interval for  $(\theta_1/\theta_2)$  can be construed from the values  $a$  that satisfy

$$\Lambda(a) \leq \chi_{(1),1-\alpha}^2 \quad (27)$$

- For Type I censored sample, LRT statistics based confidence interval for  $(\theta_1/\theta_2)$  Equation 27 is more accurate than that of normal approximation Equation 26 for small samples

## Example 4.1.5

- A small clinical trial was conducted to compare the duration of remission achieved by two drugs used in the treatment of leukemia.
- Duration of remission is assumed to follow an exponential distribution and two groups of 20 patients produced the followings under a Type I censoring mechanism

$$r_1 = 10, T_1 = 700, r_2 = 10, T_2 = 540$$

- Obtain 95% approximate and exact CI for  $(\theta_1/\theta_2)$

## Example 4.1.5

- Unrestricted MLEs

$$\hat{\theta}_1 = \frac{T_1}{r_1} = 70 \text{ and } \hat{\theta}_2 = \frac{T_2}{r_2} = 54$$

- 95% approximate CI for  $\log(\theta_1/\theta_2)$

$$\begin{aligned} & \log(\hat{\theta}_1/\hat{\theta}_2) \pm z_{.975}(r_1^{-1} + r_2^{-1})^{1/2} \\ & \log(70/54) \pm (1.96)(10^{-1} + 10^{-1})^{1/2} \\ & 0.26 \pm 0.877 \end{aligned}$$

## Example 4.1.5

- 95% approximate CI (normal distribution based) for  $(\theta_1/\theta_2)$

$$-0.617 < \log(\theta_1/\theta_2) < 1.136 \Rightarrow 0.54 < (\theta_1/\theta_2) < 3.114$$

**Is there any significant difference between  $\theta_1$  and  $\theta_2$ ?**

## Example 4.1.5

- Likelihood ratio statistic based confidence interval for  $(\theta_1/\theta_2)$  requires estimate of the parameters under the null hypothesis

$$H_0 : \theta_1 = a\theta_2$$

- Estimates under  $H_0$

$$\tilde{\theta}_2 = \frac{T_1 + aT_2}{a(r_1 + r_2)} \Rightarrow \tilde{\theta}_1 = a\tilde{\theta}_2 = \frac{T_1 + aT_2}{(r_1 + r_2)}$$

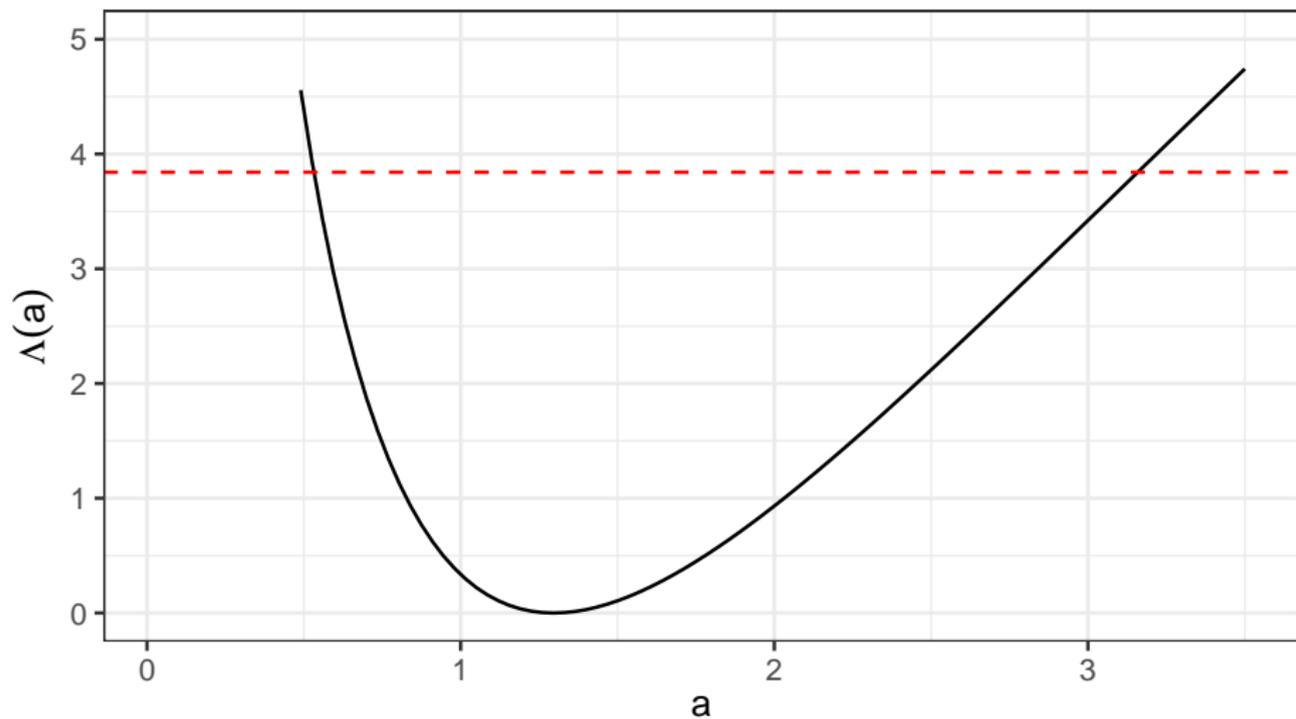
- Likelihood ratio statistic

$$\Lambda(a) = 2r_1 \log(a\tilde{\theta}_2/\hat{\theta}_1) + 2r_2 \log(\tilde{\theta}_2/\hat{\theta}_2)$$

## Example 4.1.5

$a$	$\Lambda(a)$	$a$	$\Lambda(a)$
0.525	3.953	3.185	3.911
0.560	3.424	3.150	3.819
0.595	2.958	3.115	3.726
0.630	2.549	3.080	3.633
0.665	2.187	3.045	3.541
0.700	1.869	3.010	3.448
0.735	1.589	2.975	3.356
0.770	1.341	2.940	3.263

## Example 4.1.5



## Example 4.1.5

- 95% of  $(\theta_1/\theta_2)$  is the range of values of  $a = (\theta_1/\theta_2)$ , such that

$$\Lambda(a) \leq \chi_{(1),.95}^2 = 3.84$$

which is

$$0.56 < a < 3.15 \Rightarrow 0.56 < (\theta_1/\theta_2) < 3.15$$

## Example 4.1.5

Table 4: 95% confidence interval of  $(\theta_1/\theta_2)$  for Type I censored sample

method	lower	upper
Normal approximation	0.54	3.114
LRT	0.56	3.150

## Type II censored sample (CI for $\theta_1/\theta_2$ )

- Lifetime distributions  $T_i \sim \text{Exp}(\theta_i)$   $i = 1, 2$
- Type II censored sample for the group  $i$ , which has  $r_i$  number of failures and  $(n_i - r_i)$  subjects are censored at  $t_{(ir_i)}$

$$t_{(i1)} < \cdots < t_{(ir_i)}$$

## Type II censored sample (CI for $\theta_1/\theta_2$ )

- Likelihood function

$$\begin{aligned}L(\theta_1, \theta_2) &= \prod_{i=1}^2 (1/\theta_i)^{r_i} e^{-\sum_j t_{(ij)}/\theta_i} e^{-(n_i-r_i)t_{(ir_i)}/\theta_i} \\ &= \prod_{i=1}^2 (1/\theta_i)^{r_i} e^{-(1/\theta_i)[\sum_j t_{(ij)} + (n_i-r_i)t_{(ir_i)}]} \\ &= \prod_{i=1}^2 (1/\theta_i)^{r_i} e^{-T_i/\theta_i}\end{aligned}$$

## Type II censored sample (CI for $\theta_1/\theta_2$ )

- MLE

$$\hat{\theta}_i = \frac{T_i}{r_i}$$

- We have already shown that

$$\frac{2T_i}{\theta_i} = \frac{2r_i\hat{\theta}_i}{\theta_i} \sim \chi^2_{(2r_i)}$$

- Then we can show

$$\frac{(2r_1\hat{\theta}_1/\theta_1)/(2r_1)}{(2r_2\hat{\theta}_2/\theta_2)/(2r_2)} = \frac{\hat{\theta}_1\theta_2}{\hat{\theta}_2\theta_1} \sim F_{(2r_1, 2r_2)}$$

## Type II censored sample (CI for $\theta_1/\theta_2$ )

- We can write

$$P\left(F_{(2r_1, 2r_2), \alpha/2} \leq \frac{\hat{\theta}_1 \theta_2}{\hat{\theta}_2 \theta_1} \leq F_{(2r_1, 2r_2), 1-\alpha/2}\right) = 1 - \alpha$$

- $(1 - \alpha)100\%$  confidence interval for  $(\theta_1/\theta_2)$

$$\frac{(\hat{\theta}_1/\hat{\theta}_2)}{F_{(2r_1, 2r_2), 1-\alpha/2}} \leq (\theta_1/\theta_2) \leq \frac{(\hat{\theta}_1/\hat{\theta}_2)}{F_{(2r_1, 2r_2), \alpha/2}} \quad (28)$$

## Subsection 6

### 4.2 Gamma Distribution (Uncensored Data)

## 4.2 Gamma Distribution (Uncensored Data)

- The pdf of two-parameter gamma distribution

$$f(t; \alpha, k) = \frac{1}{\alpha \Gamma(k)} \left( \frac{t}{\alpha} \right)^{k-1} \exp(-t/\alpha), \quad t > 0$$

- ▶ Scale parameter  $\alpha > 0$  and shape parameter  $k > 0$

## 4.2 Gamma Distribution (Uncensored Data)

- Survivor function

$$\begin{aligned} S(t; \alpha, k) &= \int_t^{\infty} f(u; \alpha, k) du \\ &= \int_t^{\infty} \frac{1}{\alpha\Gamma(k)} \left(\frac{u}{\alpha}\right)^{k-1} \exp(-u/\alpha) du \\ &= 1 - I(k, t/\alpha) \end{aligned} \tag{29}$$

- Incomplete gamma function

$$I(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} e^{-u} du$$

## 4.2 Gamma Distribution (Uncensored Data)

- Let  $t_1, \dots, t_n$  be a random sample from  $\text{Gamma}(\alpha, k)$ , the log-likelihood function

$$\begin{aligned}\ell(\alpha, k) &= \sum_{i=1}^n \log f(t_i; \alpha, k) \\ &= \sum_{i=1}^n \log \left[ \frac{1}{\alpha \Gamma(k)} \left( \frac{t_i}{\alpha} \right)^{k-1} \exp(-t_i/\alpha) \right] \\ &= -n \log \Gamma k - nk \log \alpha + n(k-1) \log \tilde{t} - n\bar{t}/\alpha\end{aligned}\quad (30)$$

►  $\tilde{t} = \left( \prod_{i=1}^n t_i \right)^{1/n}$  and  $\bar{t} = (1/n) \sum_i t_i$

## 4.2 Gamma Distribution (Uncensored Data)

$$\ell(\alpha, k) = -n \log \Gamma k - nk \log \alpha + n(k-1) \log \tilde{t} - n\bar{t}/\alpha$$

- Score functions

$$U_1(\alpha, k) = \frac{\partial \ell(\alpha, k)}{\partial \alpha} = \frac{-nk}{\alpha} + \frac{n\bar{t}}{\alpha^2}$$

$$U_2(\alpha, k) = \frac{\partial \ell(\alpha, k)}{\partial k} = -n\psi(k) - n \log \alpha + n \log \tilde{t}$$

where  $\psi(k) = \Gamma'(k)/\Gamma(k)$  is the digamma function (see Appendix B of the Textbook)

## 4.2 Gamma Distribution (Uncensored Data)

- MLE  $(\hat{\alpha}, \hat{k})$  is a solution of the system of linear equations

$$\frac{-n\hat{k}}{\hat{\alpha}} + \frac{n\bar{t}}{\hat{\alpha}^2} = 0 \quad (31)$$

$$-n\psi(\hat{k}) - n \log \hat{\alpha} + n \log \bar{t} = 0 \quad (32)$$

- ▶ Two equations and two unknowns
- ▶ Equations are non-linear in terms of the variables
- ▶ No closed form solutions

# 1. Substitution method

$$\frac{-n\hat{k}}{\hat{\alpha}} + \frac{n\bar{t}}{\hat{\alpha}^2} = 0 \Rightarrow \hat{\alpha} = \frac{\bar{t}}{\hat{k}} \quad (33)$$

$$-n\psi(\hat{k}) - n \log \hat{\alpha} + n \log \bar{t} = 0 \Rightarrow \psi(\hat{k}) - \log \hat{k} = \log(\bar{t}/\bar{k}) \quad (34)$$

- Solve Equation 34 using a suitable optimization technique (e.g. graphical method, Newton-Raphson method, etc.) to obtain  $\hat{k}$
- Then  $\hat{\alpha}$  can be obtained from Equation 33

## 2. Newton-Raphson method

- Score functions

$$U_1(\alpha, k) = \frac{\partial \ell(\alpha, k)}{\partial \alpha} = \frac{-nk}{\alpha} + \frac{n\bar{t}}{\alpha^2}$$

$$U_2(\alpha, k) = \frac{\partial \ell(\alpha, k)}{\partial k} = -n\psi(k) - n \log \alpha + n \log \tilde{t}$$

## 2. Newton-Raphson method

- Elements of Hessian matrix

$$\begin{aligned}H_{11}(\alpha, k) &= \frac{\partial^2 \ell(\alpha, k)}{\partial \alpha^2} = \frac{nk}{\alpha^2} - \frac{2n\bar{t}}{\alpha^3} \\H_{12}(\alpha, k) &= \frac{\partial^2 \ell(\alpha, k)}{\partial \alpha \partial k} = \frac{-n}{\alpha} = H_{21}(\alpha, k) \\H_{22}(\alpha, k) &= \frac{\partial^2 \ell(\alpha, k)}{\partial k^2} = -n\psi'(k)\end{aligned}\tag{35}$$

## 2. Newton-Raphson method

- Score vector

$$U(\alpha, k) = \begin{bmatrix} U_1(\alpha, k) \\ U_2(\alpha, k) \end{bmatrix}$$

- Hessian matrix

$$H(\alpha, k) = \begin{bmatrix} H_{11}(\alpha, k) & H_{12}(\alpha, k) \\ H_{21}(\alpha, k) & H_{22}(\alpha, k) \end{bmatrix}$$

- Score vector and information matrix are function of parameters and data

## 2. Newton-Raphson method

- Initial values  $\theta^{(0)} = (\alpha^{(0)}, k^{(0)})'$  are chosen so that elements of score vector and hessian matrix are finite
- Updated estimate  $\theta^{(1)} = (\alpha^{(1)}, k^{(1)})'$  is obtained as

$$\theta^{(1)} = \theta^{(0)} - [H(\theta^{(0)})]^{-1} U(\theta^{(0)}) \quad (36)$$

- The estimate  $\theta^{(2)}$  can be obtained by using  $\theta^{(1)}$  as input in Equation 36
- Repeating the procedure of evaluating the Equation 36 using the current estimate, the following sequence of estimates can be obtained

$$\{\theta^{(j)}, j = 3, 4, 5, \dots\}$$

## 2. Newton-Raphson method

- A convergence criterion needs to be defined to obtain the MLE from the sequence of estimates

$$\left\{ \theta^{(j)}, j = 1, 2, 3, \dots \right\},$$

## 2. Newton-Raphson method

- Convergence criteria are defined on the basis of two successive values of the parameters estimates
- $\theta^{(j)} = (\hat{\alpha}, \hat{k})'$  is considered as MLE if one of the following criteria is satisfied
  - ▶  $|\theta^{(j)} - \theta^{(j-1)}|$  is very small
  - ▶  $|\ell(\theta^{(j)}) - \ell(\theta^{(j-1)})|$  is very small
  - ▶  $U(\theta_j) \approx 0$

## 2. Newton-Raphson method

- Estimated variance-covariance matrix of the MLE  $(\hat{\alpha}, \hat{k})'$  can be obtained from the inverse of the negative of the hessian matrix evaluated at the MLE

$$\widehat{\text{Var}} \begin{pmatrix} \hat{\alpha} \\ \hat{k} \end{pmatrix} = - \left[ H(\hat{\alpha}, \hat{k}) \right]^{-1}$$

## 2. Newton-Raphson method

### Newton-Raphson method: Pseudo code

```
1: theta0 <- `initial value of the parameter`
2: eps <- 1
3: while (eps > 1e-5) {
4:   u0 <- U(theta0)
5:   h0 <- H(theta0)
6:   theta1 <- theta0 - inv(h0) %*% u0
7:   eps <- max(abs(theta1 - theta0))
8:   if (eps < 1e-5) break
9:   else theta0 <- theta1
10: }
11: return (list(theta0, h0))
```

## 2. Newton-Raphson method

- Statistical software have routines (such as `optim()` of R) that can optimize likelihood function to obtain MLE
  - ▶ Such routines require providing a “function” of likelihood function as an argument
  - ▶ Different optimization algorithms, such as Newton-Raphson, Mead-Nelder, simulated annealing, etc. are implemented in optimization routines

# Statistical inference

- Asymptotic distributions of MLEs, i.e.

$$\hat{\alpha} \sim \mathcal{N}(\alpha, \text{var}(\hat{\alpha})) \quad \text{and} \quad \hat{k} \sim \mathcal{N}(k, \text{var}(\hat{k}))$$

- Since  $\alpha > 0$  and  $k > 0$ , confidence intervals based on the sampling distribution of  $\log \hat{\alpha}$  and  $\log \hat{k}$  ensure non-negative lower limit of the confidence interval

$$\log \hat{\alpha} \sim \mathcal{N}(\log \alpha, \text{var}(\log \hat{\alpha})) \quad \text{and} \quad \log \hat{k} \sim \mathcal{N}(\log k, \text{var}(\log \hat{k}))$$

►  $\text{var}(\log \hat{\alpha}) = (1/\hat{\alpha})^2 \text{var}(\hat{\alpha})$

# Statistical inference

- $(1 - \alpha)100\%$  confidence interval for  $\log \alpha$

$$\log \hat{\alpha} \pm z_{1-\alpha/2} \text{SE}(\log \hat{\alpha})$$

- $(1 - \alpha)100\%$  confidence interval for  $\alpha$

$$\hat{\alpha} \exp \left( \pm z_{1-\alpha/2} \text{SE}(\log \hat{\alpha}) \right)$$

## Example 4.2.1

Following are survival time of 20 male rats that were exposed to a high level of radiation

```
rtime <- c(152, 152, 115, 109, 137, 88, 94, 77, 160, 165, 125,  
rtime
```

```
[1] 152 152 115 109 137 88 94 77 160 165 125 40 128 123 1  
[20] 69
```

- Assume the lifetimes follow a gamma distribution with scale parameter  $\alpha$  and shape parameter  $k$

## Example 4.2.1

- From the data:  $\bar{t} = 113.45$  and  $\tilde{t} = 107.07$
- Expressions of the score function

$$\hat{\alpha} = (\bar{t}/\hat{k}) = 113.45/\hat{k}$$

$$\psi(\hat{k}) - \log \hat{k} - \log(\tilde{t}/\bar{t}) = 0$$

$$\psi(\hat{k}) - \log \hat{k} + 0.058 = 0$$

## Example 4.2.1

- R function `uniroot()` can be used to obtain the value of  $\hat{k}$  by solving the Equation 37

```
kfun <- function(k0, time) {  
  t_bar <- mean(time)  
  tgbar <- exp(mean(log(time)))  
  return(digamma(k0) - log(k0) -  
         log(tgbar / t_bar))  
}  
k_hat <- uniroot(kfun, lower = 3,  
                upper = 10,  
                time = rtime)
```

$$\hat{k} = 8.799 \Rightarrow \hat{\alpha} = 12.893$$

```
k_hat  
$root  
[1] 8.799215  
  
$f.root  
[1] 8.181174e-10  
  
$iter  
[1] 7  
  
$init.it  
[1] NA  
  
$estim.prec  
[1] 6.103516e-05
```

## Log-likelihood function of gamma distribution

- R function to calculate log-likelihood function of a gamma distribution for a given sample

```
gamma_loglk <- function(par, time) {  
  sum(  
    dgamma(time, scale = par[1], shape = par[2], log = T)  
  )  
}
```

- `par` is a vector with the parameter `scale` as the first element and `shape` as the second element
- `time` is the observed failure times
- `gamma_loglk` function can be evaluated for any given valid values of `par` and `time`

## Log-likelihood function of gamma distribution

- For given values of parameters, say ( $\alpha_0 = 1, k_0 = 2.$ ), the value of log-likelihood function can be obtained for the rat data

```
gamma_loglk(par = c(1, 2), time = rtime)
```

```
[1] -2175.531
```

- For another set of values ( $\alpha_0 = 100, k_0 = 80$ ), the corresponding value of log-likelihood function

```
gamma_loglk(par = c(100, 80), time = rtime)
```

```
[1] -5392.711
```

## optim() function

- R function `optim()` is a general purpose optimization routine

```
optim(par, fn, gr = NULL, method = "Nelder-Mead",  
      lower = -Inf, upper = Inf,  
      control = list(), hessian = FALSE, ...  
      )
```

- `par` → initial values for the parameters to be optimized
- `fn` → a function to be *minimized*
- `gr` → a function to return the gradient (score function)
- `method` → a method to be used, e.g. "Nelder-Mead", "BFGS", "CG", "L-BFGS-B", etc.
- `lower` and `upper` → lower and upper limit of the parameters to be optimized
- `control` → list of arguments (e.g. `fnscale`, etc.) for controlling the iterations

## optim() function

### Initial values

- Initial values of the parameters can be selected from the exploratory plots of log-likelihood function
- The log-likelihood function must provide finite values with the initial values of the parameters

## optim() function

- As an example, assume the initial values  $\alpha_0 = 12.9$  and  $k_0 = 8.8$ , and the corresponding value of log-likelihood function

```
gamma_loglik(par = c(12.9, 8.8), time = rtime)
```

```
[1] -100.48
```

- For another set of initial values,  $\alpha_0 = 1.0$  and  $k_0 = 1.0$

```
gamma_loglik(par = c(1.0, 1.0), time = rtime)
```

```
[1] -2269
```

## `optim()` function

- Since both sets of initial values can provide finite values of log-likelihood function, we can use either of these two as an initial value for optimizing the log-likelihood function using the R function `optim()`

## Example 4.2.1 using optim()

- Fit of the gamma distribution to the rat data

```
gamma_fit <- optim(  
  par = c(1, 1), # initial values  
  fn = gamma_loglk, # function name  
  control = list(fnscale = -1), # for maximization  
  hessian = T, # will return hessian matrix  
  time = rtime  
)
```

## Example 4.2.1 using `optim()`

- List of objects in `gamma_fit`

```
names(gamma_fit)
```

```
[1] "par"          "value"        "counts"       "convergence" "n"
[6] "hessian"
```

- Converged?

```
gamma_fit$convergence
```

```
[1] 0
```

## Example 4.2.1 using `optim()`

- Estimate of scale and shape parameters

```
gamma_fit$par
```

```
[1] 12.896976  8.795312
```

- $\hat{\alpha} = 12.897$  and  $\hat{k} = 8.795$

## Example 4.2.1 using `optim()`

- Estimated variance-covariance matrix

```
cvar <- solve(-gamma_fit$hessian)
cvar
```

```
           [,1]      [,2]
[1,] 16.88167 -10.871356
[2,] -10.87136   7.416138
```

- $SE(\hat{\alpha}) = \sqrt{16.882} = 4.109$
- $SE(\hat{k}) = \sqrt{7.416} = 2.723$

## Example 4.2.1 using `optim()`

### 95% CI for $\alpha$ and $k$

- Using the sampling distribution of  $\hat{\alpha}$  and  $\hat{k}$

$$\hat{\alpha} \pm z_{1-\alpha/2} \text{SE}(\hat{\alpha}) \Rightarrow 4.84 \leq \alpha \leq 20.95$$

$$\hat{k} \pm z_{1-\alpha/2} \text{SE}(\hat{k}) \Rightarrow 3.46 \leq k \leq 14.13$$

- Using the sampling distribution of  $\log \hat{\alpha}$  and  $\log \hat{k}$

$$\hat{\alpha} \exp(\pm z_{1-\alpha/2} \text{SE}(\log \hat{\alpha})) \Rightarrow 6.91 \leq \alpha \leq 24.08$$

$$\hat{k} \exp(\pm z_{1-\alpha/2} \text{SE}(\log \hat{k})) \Rightarrow 4.79 \leq k \leq 16.14$$

## Example 4.2.1 using `optim()`

### Quantiles

- $p$ th quantile

$$\begin{aligned} S(t_p; \alpha, k) = (1 - p) &\Rightarrow \frac{1}{\Gamma(k)} \int_{t_p/\alpha}^{\infty} u^{k-1} e^{-u} du = (1 - p) \\ &\Rightarrow (t_p/\alpha) = Q(p, k) \\ &\Rightarrow t_p = \alpha Q(p, k) \end{aligned} \quad (37)$$

- ▶  $Q(p, k)$  is the  $p$ th quantile function of one-parameter gamma distribution [R function `qgamma(p, scale=1, shape)`]

- Estimate of the median for the rat data is  $\hat{t}_{.5} = \hat{\alpha} Q(.5, \hat{k})$

```
12.893951 * qgamma(.5, scale = 1, shape = 8.799089)
```

```
[1] 109.1872
```

```
qgamma(.5, scale = 12.893951, shape = 8.799089)
```

```
[1] 109.1872
```

# Likelihood ratio statistic

- Likelihood ratio statistic

$$\Lambda(\alpha, k) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\alpha, k)$$

- Approximate  $(1 - p)100\%$  confidence region for  $(\alpha, k)'$  can be obtained from the set of points  $(\alpha, k)$  satisfying

$$\Lambda(\alpha, k) \leq \chi_{(2), 1-p}^2$$

## Likelihood ratio statistic

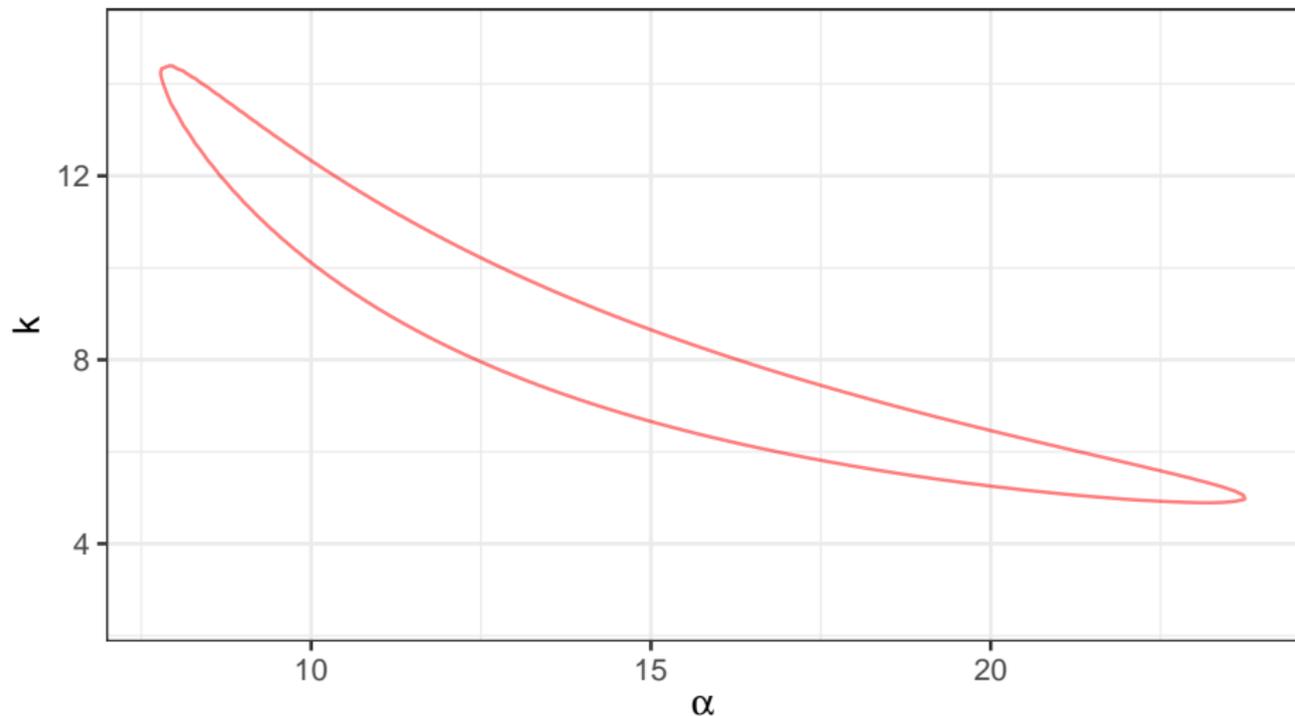


Figure 2: 95% confidence region of  $(\alpha, k)$  for rat survival data

## LRT statistic based CI for $k$

- Consider the following null hypothesis

$$H_0 : k = k_0$$

- MLEs (unrestricted and under restriction)

$$(\hat{\alpha}, \hat{k}) = \arg \max_{(\alpha, k) \in \Theta} \ell(\alpha, k) \quad \text{under } H_1 \quad (38)$$

$$\tilde{\alpha}(k_0) = \arg \max_{\alpha \in \Theta_\alpha} \ell(\alpha, k_0) \quad \text{under } H_0 \quad (39)$$

## LRT statistic based CI for $k$

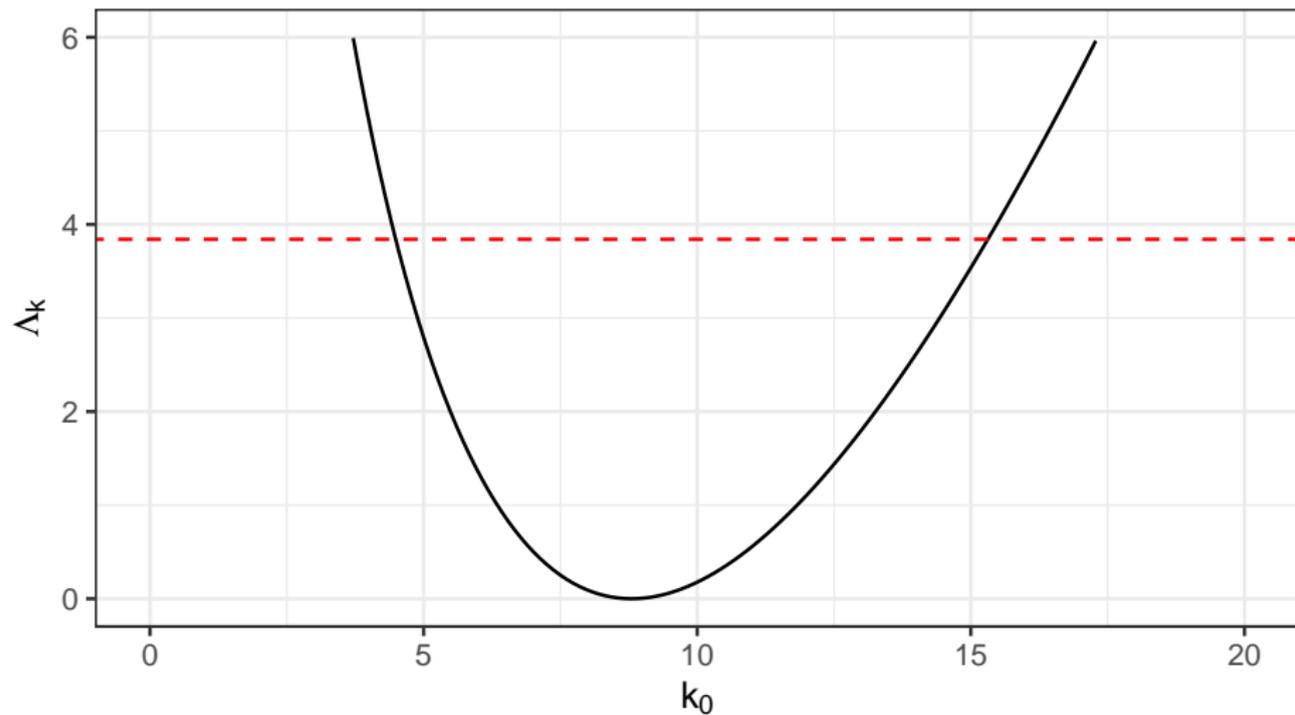
- Likelihood ratio statistic

$$\Lambda_k(k_0) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\tilde{\alpha}(k_0), k_0) \quad (40)$$

- ▶ Under  $H_0 : k = k_0$ ,  $\Lambda_k(k_0)$  approximately follows a  $\chi_{(1)}^2$  distribution
- An approximate two-sided  $(1 - p)100\%$  confidence interval for  $k$  can be obtained from a set of values of  $k_0$  satisfying

$$\Lambda_k(k_0) \leq \chi_{(1), 1-p}^2 \quad (41)$$

## LRT statistic based CI for $k$



95% CI :  $4.54 \leq k \leq 15.28$

## LRT statistic based CI for $\alpha$

- Consider the following null hypothesis

$$H_0 : \alpha = \alpha_0$$

- MLEs (unrestricted and under restriction)

$$(\hat{\alpha}, \hat{k}) = \arg \max_{(\alpha, k) \in \Theta} \ell(\alpha, k) \quad \text{under } H_1 \quad (42)$$

$$\tilde{k}(\alpha_0) = \arg \max_{k \in \Theta_k} \ell(\alpha_0, k) \quad \text{under } H_0 \quad (43)$$

## LRT statistic based CI for $\alpha$

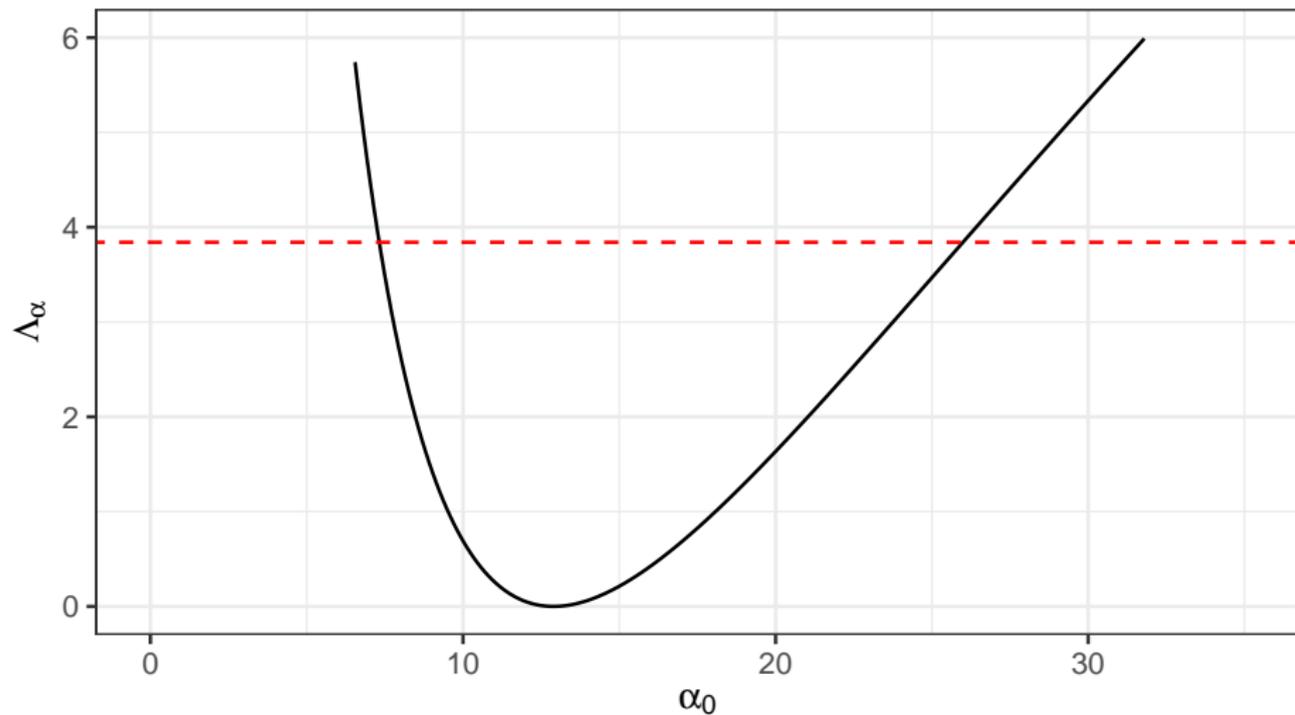
- Likelihood ratio statistic

$$\Lambda_{\alpha}(\alpha_0) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\alpha_0, \tilde{k}(\alpha_0)) \quad (44)$$

- ▶ Under  $H_0 : \alpha = \alpha_0$ ,  $\Lambda_{\alpha}(\alpha_0)$  approximately follows a  $\chi_{(1)}^2$  distribution
- An approximate two-sided  $(1 - p)100\%$  confidence interval for  $\alpha$  can be obtained from a set of values of  $\alpha_0$  satisfying

$$\Lambda_{\alpha}(\alpha_0) \leq \chi_{(1), 1-p}^2 \quad (45)$$

## LRT statistic based CI for $\alpha$



95% CI :  $7.37 \leq \alpha \leq 25.9$

## LRT statistic based CI for $\alpha$

### Summary of 95% CI

- For  $\alpha$

no-transformation :  $4.84 \leq \alpha \leq 20.95$

log-transformation :  $6.91 \leq \alpha \leq 24.08$

LRT :  $7.37 \leq \alpha \leq 25.9$

- For  $k$

no-transformation :  $3.46 \leq k \leq 14.13$

log-transformation :  $4.79 \leq k \leq 16.14$

LRT :  $4.54 \leq k \leq 15.28$

## Subsection 7

### Gamma Distribution (Censored Data)

## Gamma Distribution (Censored Data)

Data  $\{(t_i, \delta_i), i = 1, \dots, n\}$  and the corresponding log-likelihood function

$$\begin{aligned}\ell(\alpha, k) &= \log \prod_{i=1}^n \left[ \frac{1}{\alpha \Gamma(k)} \left( \frac{t_i}{\alpha} \right)^{k-1} e^{-t_i/\alpha} \right]^{\delta_i} \left[ 1 - I(k, t_i/\alpha) \right]^{1-\delta_i} \\ &= -k \log \alpha - r \log \Gamma(k) + (k-1) \sum_i \delta_i \log t_i \\ &\quad - \sum_i \delta_i t_i/\alpha + \sum_i (1 - \delta_i) \log \left[ 1 - I(k, t_i/\alpha) \right] \quad (46)\end{aligned}$$

## Gamma Distribution (Censored Data)

- No closed form solutions are available for MLEs  $\hat{\alpha}$  and  $\hat{k}$
- Algebraic expressions of score function and information matrix for the log-likelihood function (46) are very complicated
- Score functions and information matrix can be evaluated numerically
- Optimization routines available in statistical software can be used to obtain MLEs  $\hat{\alpha}$  and  $\hat{k}$ , and the corresponding SEs
  - ▶ Different optimization algorithms such as Newton-Raphson, Nelder-Mead, etc. are available in such routines

## Gamma Distribution (Censored Data)

- As an example, we are going to use the same rat data that we have used for the analysis of complete data
- For the rat data, assume a time  $\geq 150$  weeks is considered as censored, there are 5 censored observations
  - ▶ An R object `rtimec` is created, which has two columns `time` and `status`

```
rtimec |> dplyr::glimpse()
```

```
Rows: 20
```

```
Columns: 2
```

```
$ time <dbl> 152, 152, 115, 109, 137, 88, 94, 77, 160, 165,  
$ status <dbl> 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1,
```

## Gamma Distribution (Censored Data)

gamma log-likelihood function for censored sample

```
gamma_loglik <- function(par, time, status = NULL) {  
  #  
  if (is.null(status)) status <- rep(1, length(time))  
  #  
  llk_f <- sum(status * dgamma(time, scale = par[1],  
                               shape = par[2], log = T))  
  #  
  llk_c <- sum((1 - status) * pgamma(time, scale = par[1],  
                                       shape = par[2], lower.tail = F,  
                                       log.p = T))  
  return(llk_f + llk_c)  
}
```

## Gamma Distribution (Censored Data)

- R codes to obtain MLE of parameters of a gamma distribution from a censored sample

```
gamma_out_c <- optim(  
  par = c(1, 1), fn = gamma_loglk,  
  control = list(fnscale = -1), hessian = T,  
  time = rtimec$time, status = rtimec$status  
)
```

## Gamma Distribution (Censored Data)

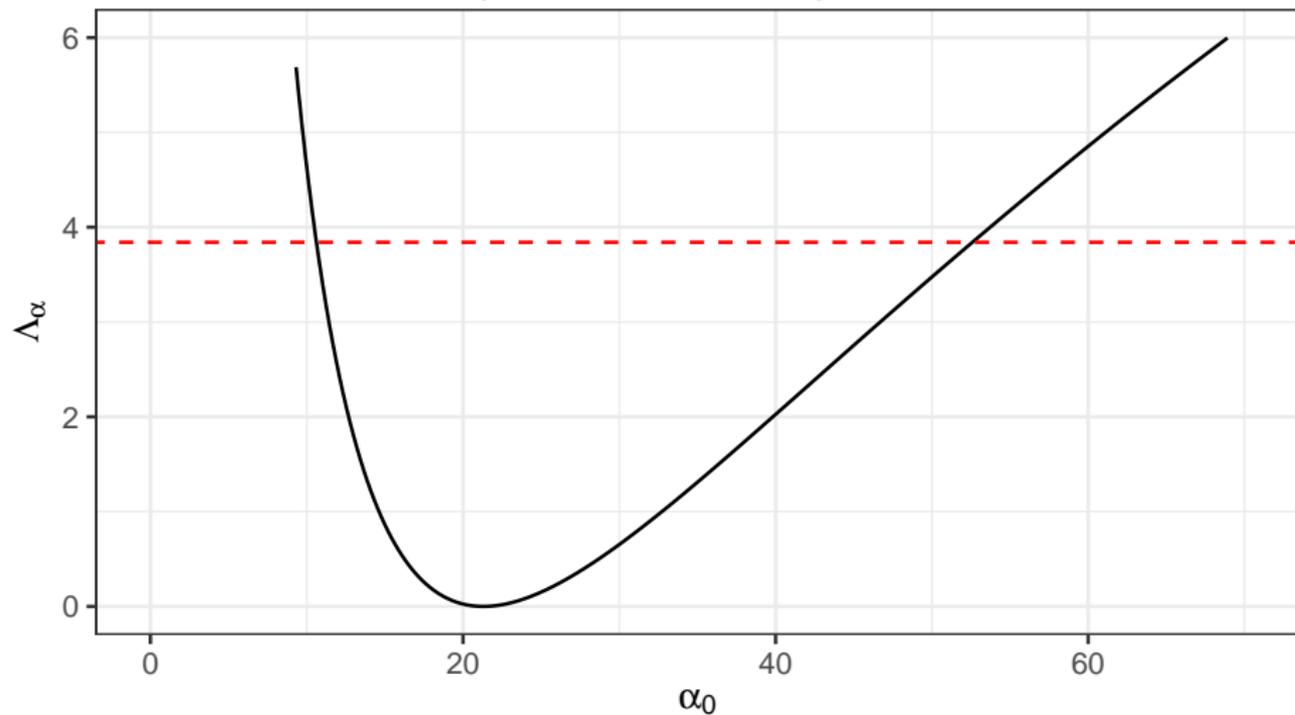
- Maximum likelihood estimators [`gamma_out_c$par`]

$$\hat{\alpha} = 21.3 \quad \text{and} \quad \hat{k} = 5.79$$

- Standard errors [`solve(-gamma_out_c$hessian)`]

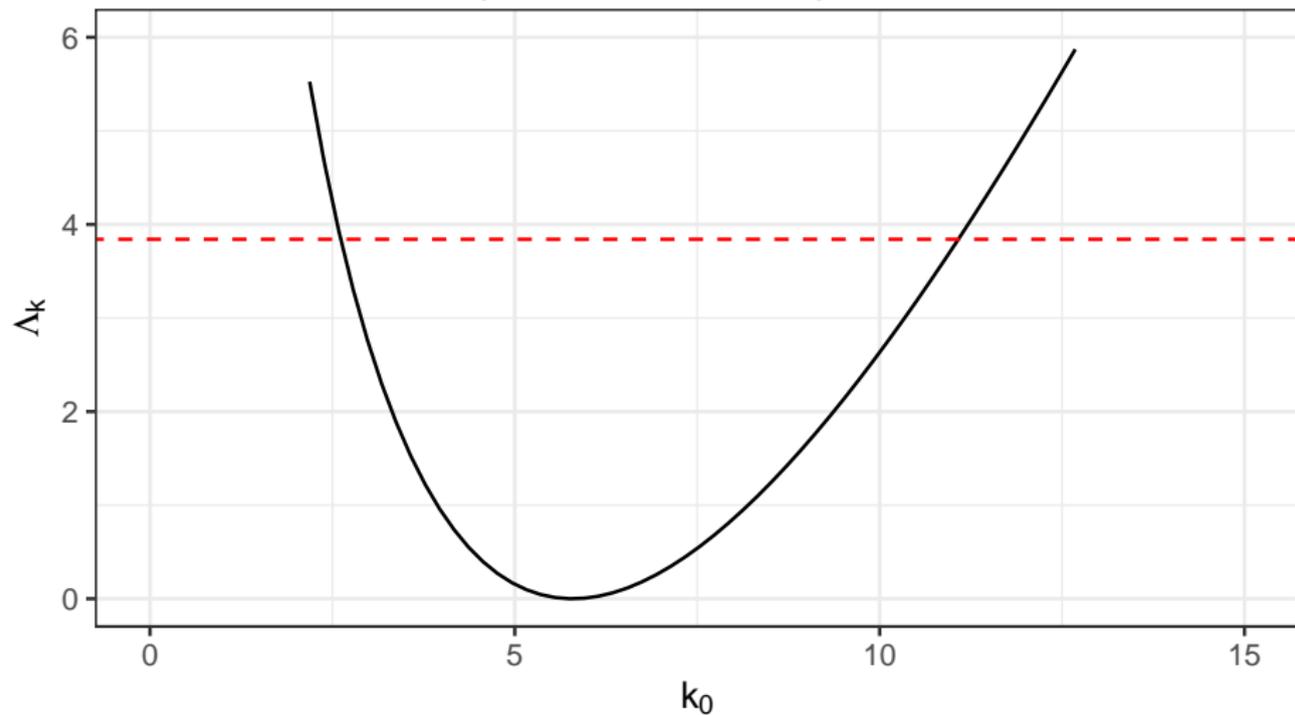
$$se(\hat{\alpha}) = 8.61 \quad \text{and} \quad se(\hat{k}) = 2.14$$

## Gamma Distribution (Censored Data)



95% CI :  $10.7 \leq \alpha \leq 52.48$

## Gamma Distribution (Censored Data)



95% CI :  $2.78 \leq k \leq 10.9$

## Gamma Distribution (Censored Data)

- 95% CI for  $\alpha$

no-transformation :  $4.44 \leq \alpha \leq 38.17$

log-transformation :  $9.65 \leq \alpha \leq 47.02$

LRT :  $10.7 \leq \alpha \leq 52.48$

- 95% CI for  $k$

no-transformation :  $3.46 \leq k \leq 14.13$

log-transformation :  $4.79 \leq k \leq 16.14$

LRT :  $2.78 \leq k \leq 10.9$

# Gamma Distribution (Censored Data)

- Estimate of median

$$\hat{t}_{.5} = \hat{\alpha} Q(.5, \hat{k}) = 116.4 \text{ weeks}$$

# Gamma Distribution (Censored Data)

## Subsection 8

### References

# References

- Sprott, David Arthur. 1980. "Maximum Likelihood in Small Samples: Estimation in the Presence of Nuisance Parameters." *Biometrika* 67 (3): 515–23.