Chapter 4

(AST405) Lifetime data analysis

Md Rasel Biswas

Lecture Outline

1. 4. Inference procedures for parametric models
   - 4.1 Exponential distribution
   - 4.2 Gamma distribution

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

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

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(t_i; \theta) \tag{2}$$

- Maximum likelihood estimator (MLE)

$$\hat{\theta} = arg\, max_{\theta \in \Theta} \ell(\theta) \tag{3}$$

Statistical inference (Complete data)

- Large sample property of MLE

$$\hat{\theta} \sim \mathcal{N}(\theta, [\mathcal{J}(\theta)]^{-1}) \tag{4}$$

  - Fisher expected information matrix

$$\mathcal{J}(\theta) = E\Big[-\frac{\partial^2 \ell(\theta)}{\partial\theta\,\partial\theta'}\Big] \tag{5}$$

  - Observed information matrix

$$I(\hat{\theta}) = -\frac{\partial^2 \ell(\theta)}{\partial\theta\,\partial\theta'}\bigg|_{\theta=\hat{\theta}} \tag{6}$$

Likelihood function (Type I or random censoring)

- Data: $\{(t_i, \delta_i),\ i = 1, \ldots, n\}$
  - 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

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

- Likelihood function

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

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

$$t_{(1)} \leqslant \cdots \leqslant t_{(r)}$$

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

$$L(\theta) = \left[\prod_{i=1}^{r} f(t_{(i)}; \theta) \right] \left[S(t_{(r)}; \theta) \right]^{n-r} \tag{8}$$

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

$$f(t;\theta) = (1/\theta)\ \exp(-t/\theta) \ \ t \geqslant 0,\ \ \theta > 0 \tag{9}$$

- Hazard function

$$h(t;\theta) = (1/\theta) \tag{10}$$

- Survivor function

$$S(t;\theta) = \exp(-t/\theta) \tag{11}$$

- pth 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

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \left[f(t_i;\theta)\right]^{\delta_i} \left[S(t_i;\theta)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} S(t_i;\theta) \left[h(t_i;\theta)\right]^{\delta_i} \\ &= \prod_{i=1}^{n} \exp(-t_i/\theta)\,(1/\theta)^{\delta_i} \end{aligned} \tag{12}$$

Type I or random censoring

- Likelihood function

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

- Log-likelihood function

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{n} \left[-(t_i/\theta) - \delta_i \log(\theta) \right] \\ &= -\frac{1}{\theta} \sum_i t_i - r \log(\theta) \qquad (13)\end{aligned}$$

  - $r = \sum_i \delta_i \;\rightarrow$ 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

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

$$\hat{\theta} = \sum_{i=1}^{n} t_i / r \tag{14}$$

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

Type I or random censoring

- Information matrix

$$\begin{aligned} I(\theta) &= -\frac{\partial^2 \ell(\theta)}{\partial \theta^2} \\ &= -\frac{\partial}{\partial \theta}\left[\frac{1}{\theta^2}\sum_i t_i - \frac{r}{\theta}\right] \\ &= \frac{2}{\theta^3}\sum_i t_i - \frac{r}{\theta^2} \end{aligned} \tag{15}$$

- Replacing the parameter θ 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

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

- 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 asymmeteric

Confidence interval (Method II)

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

$$\ell(\theta) = -(1/\theta)\sum_i t_i - r\log\theta \tag{16}$$

$$\ell_1(\phi) = -\phi^3 \sum_i t_i + 3r\log(\phi) \tag{17}$$

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

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

Confidence interval (Method II)

- Pivotal quantity

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

  - 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 θ, where

$$\begin{aligned} \ell(\theta) &= -(1/\theta)\sum_i t_i - r\log(\theta) \\ &= -r(\hat{\theta}/\theta) - r\log(\theta) \qquad (18) \\ \ell(\hat{\theta}) &= -(1/\hat{\theta})\sum_i t_i - r\log(\hat{\theta}) \\ &= -r - r\log(\hat{\theta}) \qquad (19) \end{aligned}$$

Confidence interval (Method III)

- LRT statistic

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

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

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

Confidence interval of survivor function

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

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

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

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

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

Confidence interval of survivor function

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

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

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

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

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

Confidence interval of hazard function

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

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

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 θ (Method I)

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

Example 4.1.1

- 95% confidence interval of θ (Method II)
  - $\hat{\phi} = \hat{\theta}^{-1/3} = (44.0)^{-1/3} = 0.283$
  - Confidence interval for ϕ

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

  - Confidence interval for θ

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

Example 4.1.1

- Likelihood ratio statistic

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

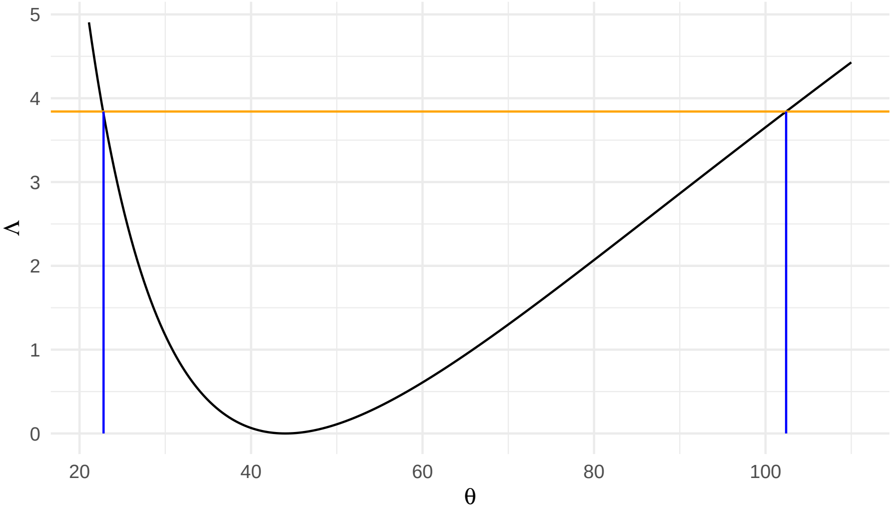
  - 95% CI for θ can be obtained from the set of θ's such that

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

  - 95% CI for θ

$$22.8 \text{ to } 102.4$$

Example 4.1.1



Example 4.1.1

Table 2: A comparison of 95% CI for θ with the data on lifetimes of 10 piees 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

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

- Sprott's method

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

- LRT

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

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

Methods for CI

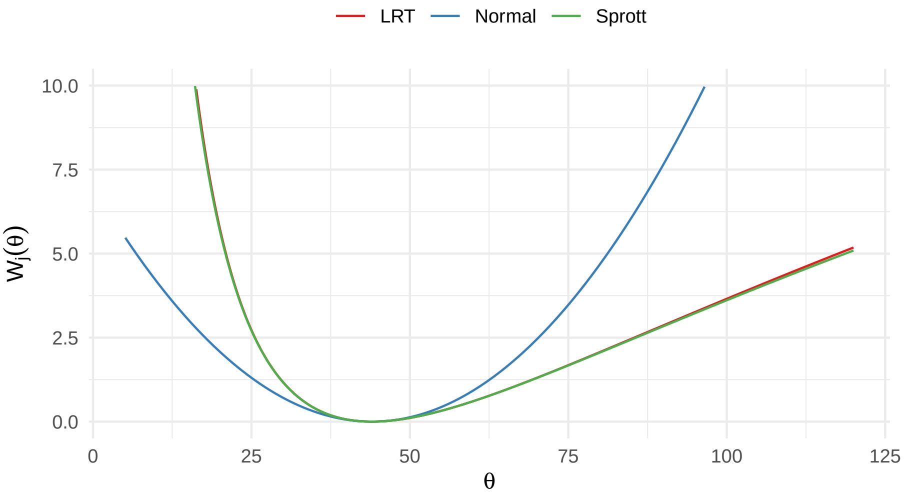


Figure 1: Comparisons between normal approximations and LRT

Subsection 4

Type II censoring plan

Type II censoring plan

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

$$
\begin{aligned}
f_T(t_{(1)}, \ldots, t_{(r)}) &= \frac{n!}{(n-r)!}\Big[\prod_{i=1}^{r}(1/\theta)e^{-t_{(i)}/\theta}\Big]\Big[e^{-t_{(r)}/\theta)}\Big]^{n-r} \\
&= \frac{n!}{(n-r)!}\,(1/\theta)^r\; e^{-\frac{1}{\theta}\left[\sum_{i=1}^{r} t_{(i)}+(n-r)t_{(r)}\right]} \\
&= \frac{n!}{(n-r)!}\,(1/\theta)^r\; e^{-T_0/\theta} \qquad (20)
\end{aligned}
$$

Type II censoring plan

- The log-likelihood function

$$\begin{aligned} \ell(\theta) &= -\frac{1}{\theta}\sum_i t_{(i)} - r\log(\theta) - (1/\theta)(n-r)t_{(r)} + \text{Const.} \\ &= -\frac{1}{\theta}\Big[\sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)}\Big] - r\log(\theta) + \text{Const.} \\ &= -\frac{T_0}{\theta} - r\log(\theta) + \text{Const.} \qquad (21) \end{aligned}$$

Type II censoring plan

- Maximum likelihood estimator

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

  - 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 θ

$$\hat{\theta} \pm z_{1-\alpha/2}\Big[I(\hat{\theta})\Big]^{-1/2} \tag{22}$$

Exact confidence interval

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

$$\begin{aligned} W_1 &= nt_{(1)} \\ W_2 &= (n-1)\big(t_{(2)} - t_{(1)}\big) \\ &\vdots \\ W_r &= (n-r+1)\big(t_{(r)} - t_{(r-1)}\big) \end{aligned}$$

- In general

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

Exact confidence interval

- It can be shown that

$$\sum_{i=1}^{r} W_i = \sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)} = T_0 \tag{23}$$

Exact confidence interval

- The joint distribution of
  $W_1 = g_1(t_{(1)}, \dots, t_{(r)}), \dots, W_r = g_r(t_{(1)}, \dots, t_{(r)})$

$$f_W(w_1, \dots, w_r) = f_T(g_1^{-1}(w_1, \dots, w_r), \dots, g_r^{-1}(w_1, \dots, w_r))\left|J\right| \quad (24)$$

where

$$J = \frac{\partial(g_1^{-1}(w_1, \dots, w_r), \dots, g_r^{-1}(w_1, \dots, w_r))}{\partial(w_1, \dots, w_r)}$$

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

Exact confidence interval

$$
\begin{aligned}
w_1 &= g_1(t_{(1)}, \dots, t_{(r)}) = nt_{(1)} \quad \Rightarrow \quad t_{(1)} = g_1^{-1}(w_1, \dots, w_r) = \frac{w_1}{n} \\
w_2 &= g_2(t_{(1)}, \dots, t_{(r)}) = (n-1)(t_{(2)} - t_{(1)}) \\
&\Rightarrow \; t_{(2)} = g_2^{-1}(w_1, \dots, w_r) = \frac{w_2}{n-1} + \frac{w_1}{n} \\
&\vdots \\
w_r &= g_r(t_{(1)}, \dots, t_{(r)}) = (n-r+1)(t_{(r)} - t_{(r-1)}) \\
&\Rightarrow \; t_{(r)} = g_r^{-1}(w_1, \dots, w_r) = \frac{w_r}{n-r+1} + \cdots + \frac{w_2}{n-1} + \frac{w_1}{n}
\end{aligned}
$$

Exact confidence interval

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

$$= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n-r+1} \end{bmatrix}$$

- We can write

$$|J| = \frac{1}{n \cdot (n-1) \cdots (n-r+1)} = \frac{(n-r)!}{n!} \tag{25}$$

Exact confidence interval

- The joint distribution of
  $W_1 = g_1(t_{(1)}, \dots, t_{(r)}), \dots, W_r = g_r(t_{(1)}, \dots, t_{(r)})$

$$
\begin{aligned}
f_W(w_1, \dots, w_r) &= f_T(g_1^{-1}(w_1, \dots, w_r), \dots, g_r^{-1}(w_1, \dots, w_r))\big|J\big| \\
&= \frac{n!}{(n-r)!} \frac{1}{\theta^r}\ e^{-\frac{1}{\theta}\left\{\sum_i g_i^{-1}(w_i) + (n-r) g_r^{-1}(w_r)\right\}} \frac{(n-r)!}{n!} \\
&= \frac{1}{\theta^r}\ e^{-\sum_i w_i/\theta} \qquad (26)
\end{aligned}
$$

  - Using (23)
- Equation (26) 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 scale parameter θ and shape parameter r

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

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

Exact confidence interval

Review

- Moment generating functions for different distributions

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

Exact confidence interval

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

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

  - $\chi^2_{(2r),\,p} \rightarrow pth$ quantile of $\chi^2_{(2r)}$ distribution

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

$$\frac{2T_0}{\chi^2_{(2r),\,1-\alpha/2}} \leq \theta \leq \frac{2T_0}{\chi^2_{(2r),\,\alpha/2}} \tag{27}$$

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 θ

Subsection 5

Comparison of distributions

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 jth subject of the ith group $(i = 1, \dots, m;\ j = 1, \dots, n_i)$
- Assume $T_{ij} \sim \text{Exp}(\theta_i)$ and the null hypothesis of interest

$$H_0 : \theta_1 = \cdots = \theta_m$$

- Data:

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

Likelihood ratio tests

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

$$L(\theta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \Big[f(t_{ij}; \theta_i) \Big]^{\delta_{ij}} \Big[S(t_{ij}; \theta_i \Big]^{1-\delta_{ij}}$$

$$= \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[\frac{1}{\theta_i} \right]^{\delta_{ij}} e^{-t_{ij}/\theta_i}$$

- The corresponding log-likelihood function

$$\ell(\theta) = -\sum_{i} \sum_{j} \Big[(t_{ij}/\theta_i) + \delta_{ij} \log(\theta_i) \Big] \tag{28}$$

Likelihood ratio tests

- The log-likelihood function

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

  - $T_i = \sum_j t_{ij}$ and $r_i = \sum_j \delta_{ij}$
- Maximum likelihood estimator of θ_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

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

log-likelihood function

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

  - Under H_0, the MLE of θ

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

Likelihood ratio tests

- Likelihood ratio test statistic

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

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

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 $H_0 : \theta_1 = \cdots = \theta_4$

Confidence intervals for θ_1/θ_2

- Comparison between two exponential distributions

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

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

$$\hat{\theta}_i \sim \mathcal{N}(\theta_i, \theta_i^2/r_i) \tag{29}$$

$$\log \hat{\theta}_i \sim \mathcal{N}(\log \theta_i, 1/r_i) \tag{30}$$

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

$$\log \hat{\theta}_1 - \log \hat{\theta}_2 = \log(\hat{\theta}_1/\hat{\theta}_2) \sim \mathcal{N}\Big(\log(\theta_1/\theta_2), (r_1^{-1} + r_2^{-1})\Big) \tag{31}$$

Confidence intervals for θ_1/θ_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

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

Confidence intervals for θ_1/θ_2

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

$$\log(\hat{\theta}_1/\hat{\theta}_2) \pm z_{1-\alpha/2}\left(r_1^{-1} + r_2^{-1}\right)^{1/2} \tag{32}$$

- 95% CI for (θ_1/θ_2)

$$(\hat{\theta}_1/\theta_2)\exp\left(\pm z_{1-\alpha/2}\left(r_1^{-1} + r_2^{-1}\right)^{1/2}\right) \tag{33}$$

Confidence intervals for θ_1/θ_2

- LRT statistics can also be used to obtain confidence interval of (θ_1/θ_2), consider the null hypothesis for $a > 0$

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

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

$$\ell(\theta_1, \theta_2) = -\sum_i \left[(T_i/\theta_i) + r_i \log(\theta_i) \right]$$

  - MLE $\hat{\theta}_i = (T_i/r_i) \;\; i = 1, 2$

Confidence intervals for θ_1/θ_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 θ_1/θ_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

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

Confidence intervals for θ_1/θ_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

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

Confidence intervals for θ_1/θ_2

$$\ell(\hat{\theta}_1, \hat{\theta}_2) = -(r_1 + r_2) - r_1 \log(\hat{\theta}_1) - r_2 \log(\hat{\theta}_2)$$
$$\ell(\tilde{\theta}_1, \tilde{\theta}_2) = -(r_1 + r_2) - r_1 \log(\tilde{\theta}_1) - r_2 \log(\tilde{\theta}_2)$$

- 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 θ_1/θ_2

- $(1-\alpha)100\%$ confidence interval for (θ_1/θ_2) can be construed from the values a that satisfy

$$\Lambda(a) \leq \chi^2_{(1),1-\alpha} \tag{34}$$

- For type I censored sample, LRT statistics based confidence interval for (θ_1/θ_2) (34) is more accurate than that of normal approximation (32) 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 (θ_1/θ_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 $\log(\theta_1/\theta_2)$

$$\log(\hat{\theta}_1/\hat{\theta}_2) \pm z_{.975}\left(r_1^{-1} + r_2^{-1}\right)^{1/2}$$

$$\log(70/54) \pm (1.96)(10^{-1} + 10^{-1})^{1/2}$$

$$0.26 \pm 0.877$$

Example 4.1.5

- 95% approximate CI (normal distribution based) for (θ_1/θ_2)

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

Is there any significant difference between θ_1 and θ_2?

Example 4.1.5

- Likelihood ratio statistic based confidence interval for (θ_1/θ_2) requires estimate of the parameters under the null hypothesis

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

- Estimates under H_0

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

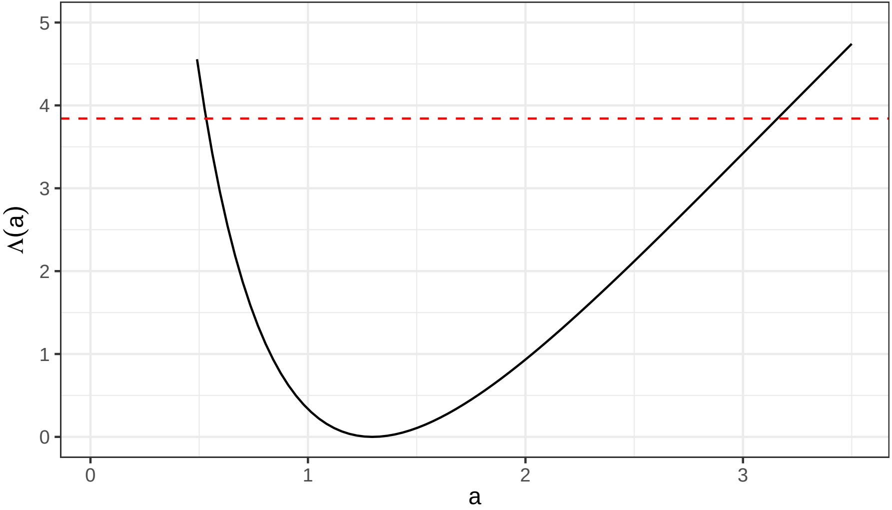
- Likelihood ratio statistic

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

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 (θ_1/θ_2) is the range of values of $a = (\theta_1/\theta_2)$, such that

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

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 (θ_1/θ_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 θ_1/θ_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 θ_1/θ_2)

- Likelihood function

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

Type II censored sample (CI for θ_1/θ_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 θ_1/θ_2)

- We can write

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

- $(1-\alpha)100\%$ confidence interval for (θ_1/θ_2)

$$\frac{(\hat{\theta}_1/\hat{\theta}_2)}{F_{(2r_1,2r_2),1-\alpha/2}} \le (\theta_1/\theta_2) \le \frac{(\hat{\theta}_1/\hat{\theta}_2)}{F_{(2r_1,2r_2),\alpha/2}} \tag{35}$$

Subsection 6

4.2 Gamma distribution

4.2 Gamma distribution

- The pdf of two-parameter gamma distribution

$$f(t;\alpha,k) = \frac{1}{\alpha\Gamma k}\left(\frac{t}{\alpha}\right)^{k-1}\exp\left(-t/\alpha\right), \quad t>0 \tag{36}$$

  - $\alpha > 0$ and $k > 0$

4.2 Gamma distribution

- Survivor function

$$\begin{aligned} S(t;\alpha,k) &= \int_t^\infty f(u;\alpha,k)\,du \\ &= \int_t^\infty \frac{1}{\alpha\Gamma k}\left(\frac{u}{\alpha}\right)^{k-1}\exp\left(-u/\alpha\right)du \\ &= 1 - I(k, t/\alpha) \end{aligned} \tag{37}$$

- Incomplete gamma function

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

Uncensored data

- Let $t_1, \ldots, t_n$ be a random sample from Gamma(α, k), the log-likelihood function

$$\begin{aligned}
\ell(\alpha, k) &= \sum_{i=1}^{n} \log f(t_i; \alpha, k) \\
&= \sum_{i=1}^{n} \log \left[\frac{1}{\alpha \Gamma k} \left(\frac{t_i}{\alpha} \right)^{k-1} \exp\left(-t_i/\alpha \right) \right] \\
&= -n \log \Gamma k - nk \log \alpha + n(k-1) \log \tilde{t} - n\bar{t}/\alpha \qquad (38)
\end{aligned}$$

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

Uncensored data

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

- Score functions

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

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

Uncensored data

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

$$\frac{-n\hat{k}}{\hat{\alpha}} + \frac{n\bar{t}}{\hat{\alpha}^2} = 0 \tag{39}$$

$$-n\psi(\hat{k}) - n\log\hat{\alpha} + n\log\tilde{t} = 0 \tag{40}$$

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

Uncensored data

Substitution method

$$\frac{-n\hat{k}}{\hat{\alpha}} + \frac{n\bar{t}}{\hat{\alpha}^2} = 0 \;\Rightarrow\; \hat{\alpha} = \frac{\bar{t}}{\hat{k}} \tag{41}$$

$$-n\psi(\hat{k}) - n\log\hat{\alpha} + n\log\tilde{t} = 0 \;\Rightarrow\; \psi(\hat{k}) - \log\hat{k} = \log(\tilde{t}/\bar{t}) \tag{42}$$

- Solve equation (42) using a suitable optimization technique (e.g. graphical method, Newton-Raphson method, etc.) to obtain $\hat{k}$
- Then $\hat{\alpha}$ can be obtained from (41)

Newton-Raphson method

- Score functions

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

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

Newton-Raphson method

- Elements of Hessian matrix

$$
\begin{aligned}
H_{11}(\alpha, k) &= \frac{\partial^2 \ell(\alpha, k)}{\partial \alpha^2} = \frac{nk}{\alpha^2} - \frac{2n\bar{t}}{\alpha^3} \\
H_{12}(\alpha, k) &= \frac{\partial^2 \ell(\alpha, k)}{\partial \alpha \, \partial k} = \frac{-n}{\alpha} = I_{21}(\alpha, k) \\
H_{22}(\alpha, k) &= \frac{\partial^2 \ell(\alpha, k)}{\partial k^2} = -n\psi'(k)
\end{aligned}
\tag{43}
$$

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

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

$$\theta^{(1)} = \theta^{(0)} - \left[H(\theta^{(0)})\right]^{-1} U(\theta^{(0)}) \tag{44}$$

- The estimate $\theta^{(2)}$ can be obtained by using $\theta^{(1)}$ as input in equation (44)
- Repeating the procedure of evaluating the equation (44) using the current estimate, the following sequence of estimates can be obtained

$$\left\{\theta^{(j)}, j = 3, 4, 5, ...\right\}$$

Newton-Raphson method

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

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

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
  - $\left|\theta^{(j)} - \theta^{(j-1)}\right|$ is very small
  - $\left|\ell(\theta^{(j)}) - \ell(\theta^{(j-1)})\right|$ is very small
  - $U(\theta_j) \approx 0$

Newton-Raphson method

- Estimated variance-variance 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

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

Newton-Raphson method

Newton-Raphson method: Pseudo code

```
theta0 <- `initial value of the parameter`
eps <- 1
while (eps > 1e-5) {
  u0 <- U(theta0)
  h0 <- H(theta0)
  theat1 <- theta0 - inv(h0) * u0
  eps <- max(abs(theta1 - theta0))
  if (eps < 1e-5) break
  else theta0 <- theta1
}
return (list(theta0, h0))
```

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-Rapson, Mead-Nelder, simulated annealing, etc. are implemented in optimization routines
- Users may provide the expressions of score and information matrix as arguments of the routine
  - If expressions of information matrix and score functions are not provided, the routines calculate those numerically

Statistical inference

- Asymptotic distributions of MLEs, i.e.

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

- 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

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

  - $\mathsf{var}(\log \hat{\alpha}) = (1/\hat{\alpha})^2 \, \mathsf{var}(\hat{\alpha})$

Statistical inference

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

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

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

$$\hat{\alpha}\exp\big(\pm z_{1-\alpha/2}\,\mathsf{SE}(\log\hat{\alpha})\big)$$

Example 4.2.1

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

```
rtime
```

```
152 152 115 109 137 88 94 77 160 165

125 40 128 123 136 101 62 153 83 69
```

- Assume the lifetimes follow a gamma distribution with parameters α and k

Example 4.2.1

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

$$\hat{\alpha} = (\bar{t}/\hat{k}) = 113.45/\hat{k} \tag{45}$$

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

$$\psi(\hat{k}) - \log \hat{k} + 0.058 = 0 \tag{46}$$

- R function `uniroot()` can be used to obtain the value of $\hat{k}$ by solving the equation (46)

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

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
```

Log-likelihood function of gamma distribution

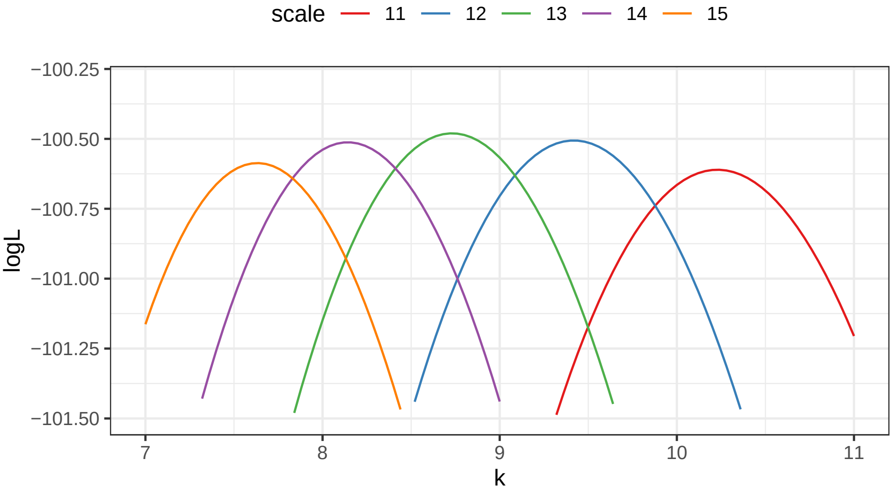


Figure 2: Plot of log-likelihood function against shape parameter k for different values of scale parameter α

Log-likelihood function of gamma distribution

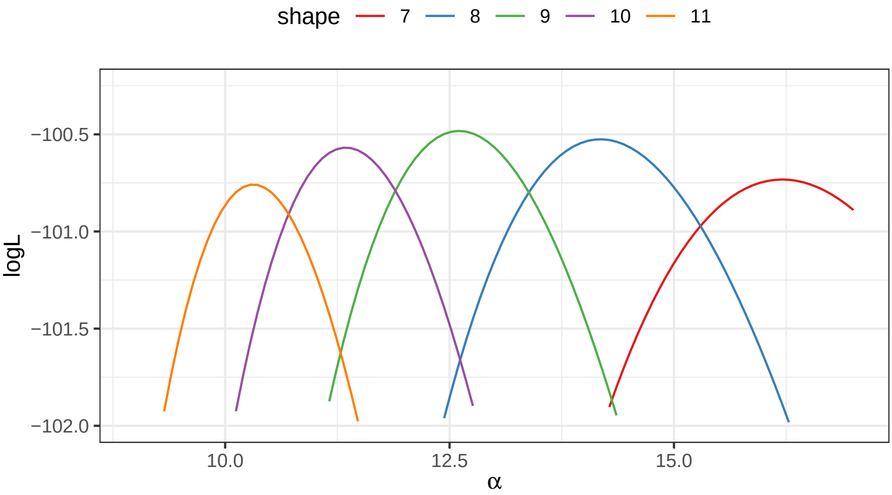


Figure 3: Plot of log-likelihood function against scale parameter α for different values of shape parameter k

Log-likelihood function of gamma distribution

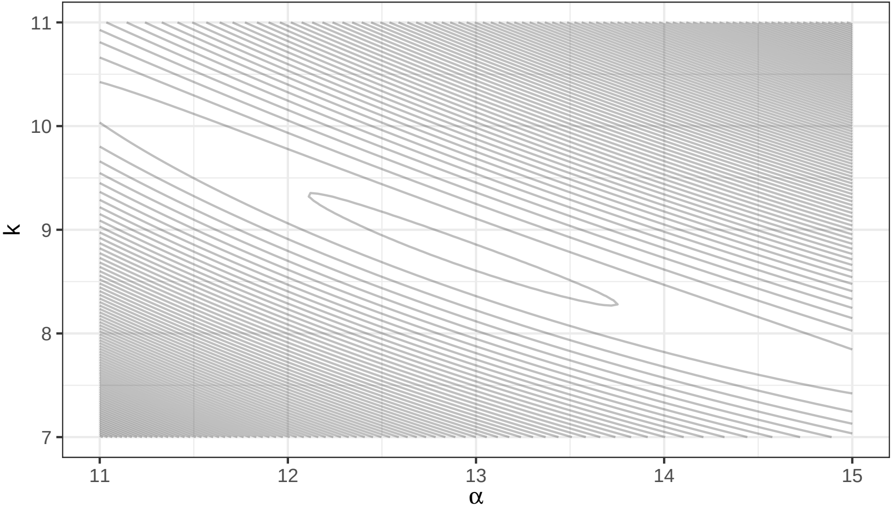


Figure 4: Contour plot of log-likelihood function of gamma distribution for rat survival data

Log-likelihood function of gamma distribution

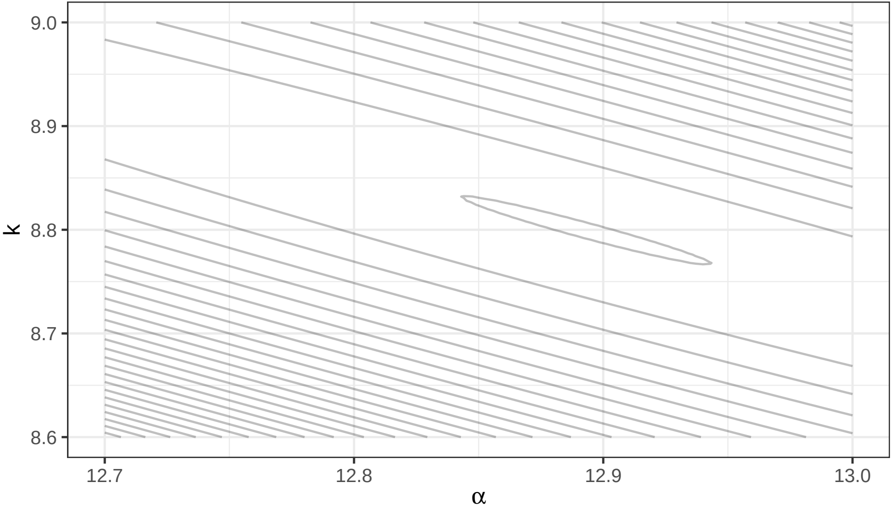


Figure 5: Contour plot of log-likelihood function of gamma distribution for rat survival data

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

```
[1] -2269
```

`optim()` function

- Since both sets of initial values van 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

- Fit of the gamma distribution to the rat data

```
> gamma_fit <- optim(
+     par = c(12.9, 8.8), fn = gamma_loglk,
+     control = list(fnscale = -1), hessian  = T,
+     time = rtime
+     )
```

Example 4.2.1

- List of objects in gamma_fit

```
names(gamma_fit)
```

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

- Converged?

```
> gamma_fit$convergence
```

```
[1] 0
```

Example 4.2.1

- Estimate of scale and shape parameters

```
gamma_fit$par
```

```
[1] 12.893951  8.799089
```

- $\hat{\alpha} = 12.894$ and $\hat{k} = 8.799$

Example 4.2.1

- Estimated variance-variance matrix

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

```
          [,1]       [,2]
[1,]  16.99179 -10.949816
[2,] -10.94982   7.471712
```

- SE$(\hat{\alpha}) = \sqrt{16.992} = 4.122$
- SE$(\hat{k}) = \sqrt{7.472} = 2.733$

Example 4.2.1

95% CI for α and k

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

$$\hat{\alpha} \pm z_{1-\alpha/2}\,\mathsf{SE}(\hat{\alpha}) \;\Rightarrow\; 4.84 \le \alpha \le 20.95$$

$$\hat{k} \pm z_{1-\alpha/2}\,\mathsf{SE}(\hat{k}) \;\Rightarrow\; 3.46 \le \alpha \le 14.13$$

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

$$\hat{\alpha}\exp\big(\pm z_{1-\alpha/2}\,\mathsf{SE}(\log\hat{\alpha})\big) \;\Rightarrow\; 6.91 \le \alpha \le 24.08$$

$$\hat{k}\exp\big(\pm z_{1-\alpha/2}\,\mathsf{SE}(\hat{k})\big) \;\Rightarrow\; 4.79 \le k \le 16.14$$

Example 4.2.1

Quantiles

- pth quantile

$$
\begin{aligned}
S(t_p;\alpha,k) = (1-p) \;\; &\Rightarrow \;\; \frac{1}{\Gamma(k)} \int_{t_p/\alpha}^{\infty} u^{k-1}\, e^{-u}\, du = (1-p) \\
&\Rightarrow \;\; (t_p/\alpha) = Q(p,k) \\
&\Rightarrow \;\; t_p = \alpha Q(p,k) \qquad (47)
\end{aligned}
$$

  - $Q(p,k)$ is the pth quantile function of one-parameter gamma distribution [R function `qgamma(p, scale=1, shape)`]
- Estimate of the median for the rat data

$$\hat{t}_{.5} = \hat{\alpha}\, Q(.5, \hat{k}) = (12.9)(8.46) = 109.16$$

Likelihood ratio statistic

- Joint relative log-likelihood function

$$r(\alpha, k) = \ell(\alpha, k) - \ell(\hat{\alpha}, \hat{k}) \qquad (48)$$

- Contour plot of $r(\alpha, k)$ provides an informative picture of the information about the parameters
- Approximately ellipsoidal contours indicates
  - sampling distribution of $\hat{\alpha}$ and $\hat{k}$ is approximately normal
  - large sample based confidence intervals will be in agreement with results based on likelihood ratio procedures

Likelihood ratio statistic

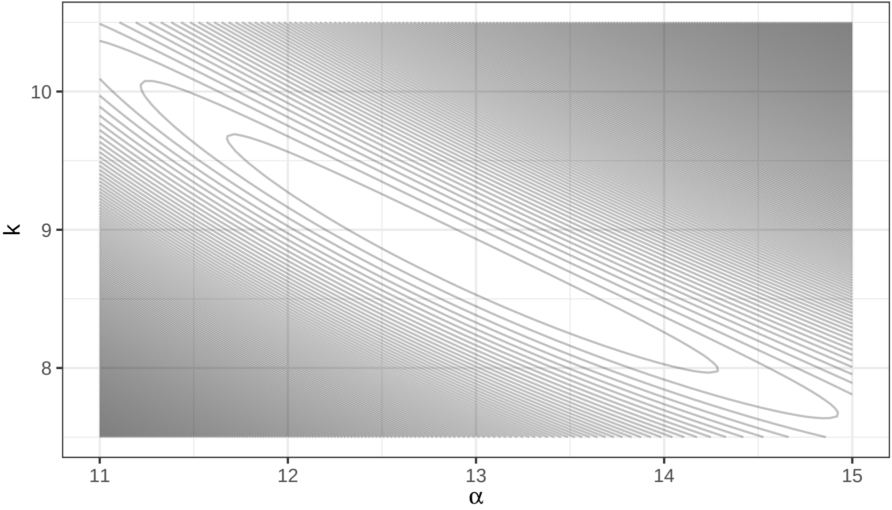


Figure 6: Contour plot of $r(\alpha, k)$ for rat survival data

Likelihood ratio statistic

- Likelihood ratio statistic

$$\Lambda(\alpha, k) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\alpha, k) = -2r(\alpha, k) \tag{49}$$

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

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

Likelihood ratio statistic

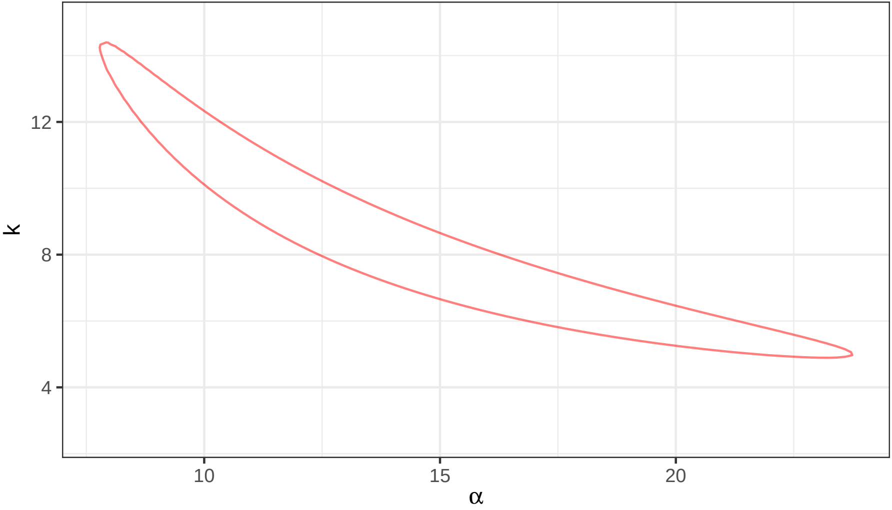


Figure 7: 95% confidence region of (α, k) for rat survival data

LRT statistic based CI for k

- Consider the following null hypothesis

$$H_0 : k = k_0$$

- MLEs (unrestricted and under restriction)

$$(\hat{\alpha}, \hat{k}) = arg\,max_{(\alpha,k)\in\Theta}\ell(\alpha, k) \quad \text{under } H_1 \tag{50}$$

$$\tilde{\alpha}(k_0) = arg\,max_{\alpha\in\Theta_\alpha}\ell(\alpha, k_0) \quad \text{under } H_0 \tag{51}$$

LRT statistic based CI for k

- Likelihood ratio statistic

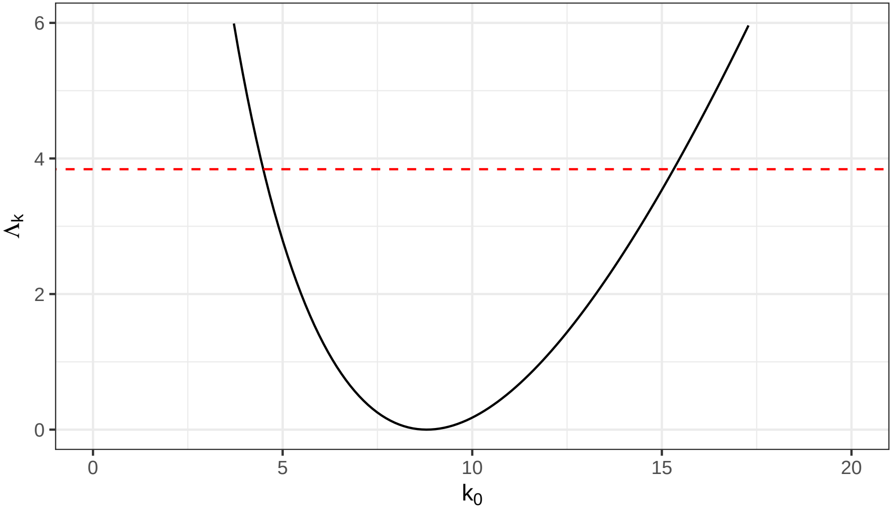
$$\Lambda_k(k_0) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\tilde{\alpha}(k_0), k_0) \tag{52}$$

  - Under $H_0 : k = k_0$, $\Lambda_k(k_0)$ approximately follows a $\chi^2_{(1)}$ distribution

- An approximate two-sided $(1-p)100\%$ confidence interval for k can be obtained from a set of values of k_0 satisfying

$$\Lambda_k(k_0) \leq \chi^2_{(1),1-p} \tag{53}$$

LRT statistic based CI for k



$$95\% \text{ CI}: \quad 4.54 \leq k \leq 15.28$$

LRT statistic based CI for α

- Consider the following null hypothesis

$$H_0 : \alpha = \alpha_0$$

- MLEs (unrestricted and under restriction)

$$(\hat{\alpha}, \hat{k}) = arg\, max_{(\alpha,k)\in\Theta} \ell(\alpha, k) \quad \text{under } H_1 \tag{54}$$

$$\tilde{k}(\alpha_0) = arg\, max_{k\in\Theta_k} \ell(\alpha_0, k) \quad \text{under } H_0 \tag{55}$$

LRT statistic based CI for α

- Likelihood ratio statistic

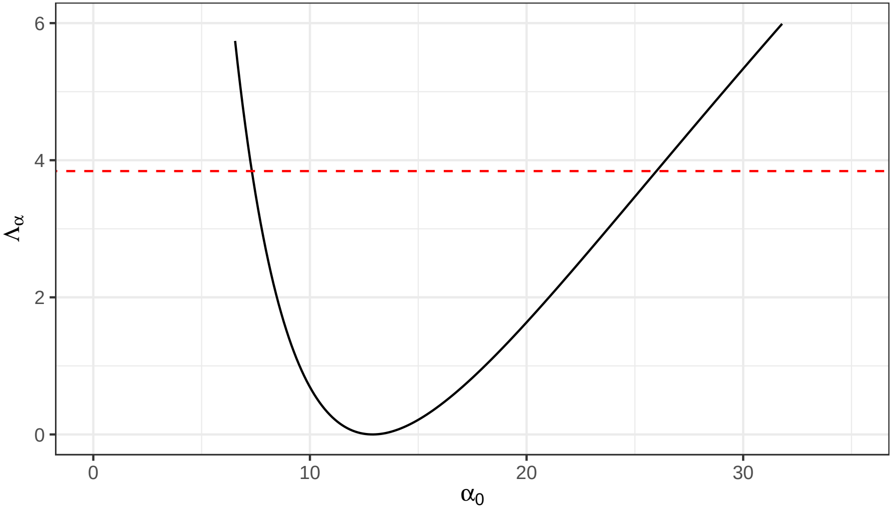
$$\Lambda_\alpha(\alpha_0) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\alpha_0, \tilde{k}(\alpha_0)) \tag{56}$$

  - Under $H_0 : \alpha = \alpha_0$, $\Lambda_\alpha(\alpha_0)$ approximately follows a $\chi^2_{(1)}$ distribution

- An approximate two-sided $(1-p)100\%$ confidence interval for α can be obtained from a set of values of α_0 satisfying

$$\Lambda_\alpha(\alpha_0) \leq \chi^2_{(1),1-p} \tag{57}$$

LRT statistic based CI for α



$$95\%\ \text{CI}: \quad 7.37 \le \alpha \le 25.9$$

LRT statistic based CI for α

Summary of 95% CI

- For α

$$\begin{aligned}\text{no-transformation}: &\ 4.84 \le \alpha \le 20.95\\ \text{log-transformation}: &\ 6.91 \le \alpha \le 24.08\\ \text{LRT}: &\ 7.37 \le \alpha \le 25.9\end{aligned}$$

- For k

$$\begin{aligned}\text{no-transformation}: &\ 3.46 \le k \le 14.13\\ \text{log-transformation}: &\ 4.79 \le k \le 16.14\\ \text{LRT}: &\ 4.54 \le k \le 15.28\end{aligned}$$

Confidence intervals of survivor function

- Consider the following null hypothesis

$$\begin{aligned} & H_0 : S(t_0; \alpha, k) = 1 - I(k, t_0/\alpha) = s_0 \\ \Rightarrow \quad & H_0 : I(k, t_0/\alpha) = 1 - s_0 \end{aligned}$$

- For a given value of t_0, $s_0 = S(t_0; \hat{\alpha}, \hat{k})$, where $\hat{\alpha}$ and $\hat{k}$ are MLEs
- LRT statistic

$$\Lambda(s_0) = 2\ell(\hat{\alpha}, \hat{k}) - 2\ell(\tilde{\alpha}, \tilde{k})$$

  - $\hat{\alpha}$ and $\hat{k}$ are unrestricted MLEs
  - $\tilde{\alpha}$ and $\tilde{k}$ are MLEs under $H_0 : I(k, t_0/\alpha) = 1 - s_0$

Confidence intervals of survivor function

- To obtain $\tilde{\alpha}$ and $\tilde{k}$, consider the function

$$M(k) = \ell(k, \alpha(k))$$

  - For any value of k, the value of α can be obtained as

$$I(k, t_0/\alpha) = (1 - s_0) \;\Rightarrow\; \alpha = t_0/Q(1 - s_0, k)$$

    where $Q(\cdot, k)$ is quantile function of a gamma distribution with scale as 1 and shape as k

- MLE under H_0

$$\tilde{k} = arg\,max_k M(k) \quad \Rightarrow \quad \tilde{\alpha} = t_0/Q(1 - s_0, \tilde{k})$$

LRT statistic based CI for $t_{.5}$

Censored sample

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

$$\begin{aligned}\ell(\alpha, k) &= \log \prod_{i=1}^{n} \left[\frac{1}{\alpha\Gamma(k)}\Big(\frac{t_i}{\alpha}\Big)^{k-1} e^{-t_i/\alpha}\right]^{\delta_i} \Big[1 - I(k, t_i/\alpha)\Big]^{1-\delta_i} \\ &= -k \log \alpha - r \log \Gamma(k) + (k-1) \sum_i \delta_i \log t_i \\ &\quad - \sum_i \delta_i t_i/\alpha + \sum_i (1-\delta_i) \log \Big[1 - I(k, t_i/\alpha)\Big] \qquad (58)\end{aligned}$$

Censored sample

- 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 (58) 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

Censored sample

- 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 ≥ 150 weeks is considered as censored, there are 5 censored observations
  - An R object `rtimec` is created, which has two columns `time` and `status`

time	status
152	0
152	0
115	1
109	1
137	1
88	1
94	1
77	1
160	0

Censored sample

Updated gamma_loglk function

- R function for evaluating gamma log-likelihood function for censored sample

```
gamma_loglk <- 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)
}
```

Censored sample

- 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
)
```

Censored sample

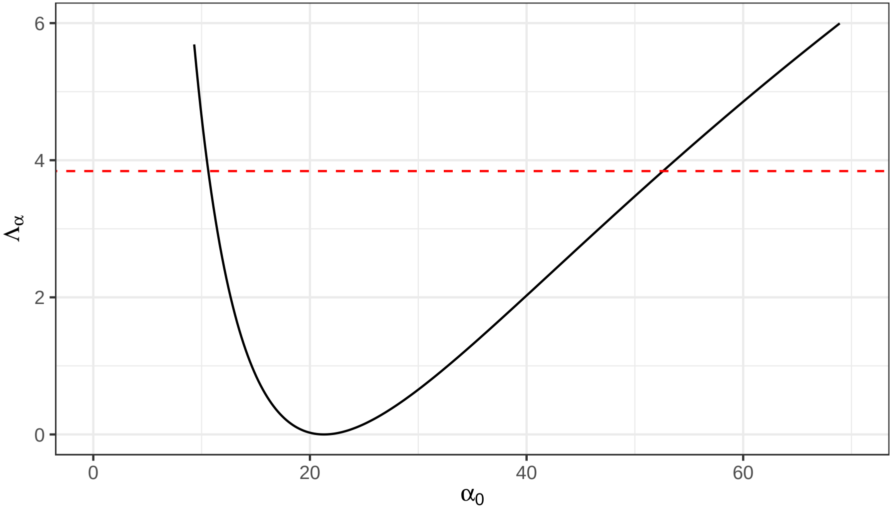
- Maximum likelihood estimators [`gamma_out_c$par`]

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

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

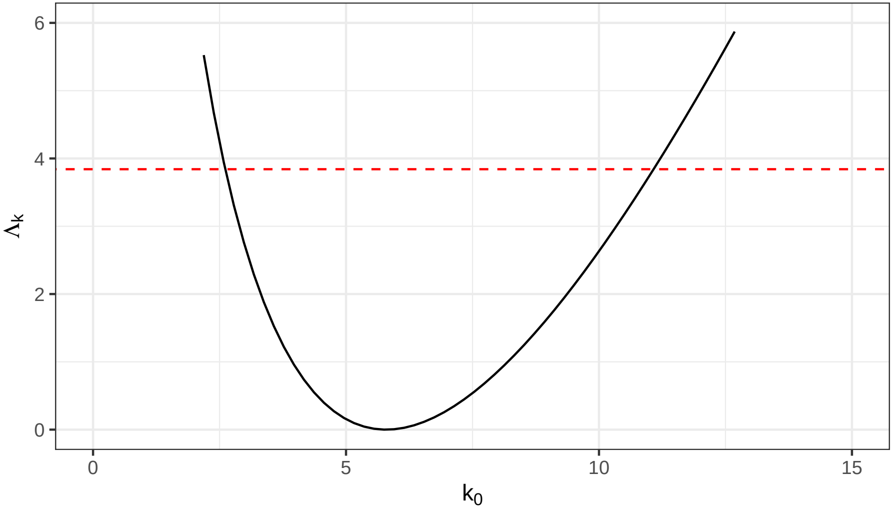
$$se(\hat{\alpha}) = 8.61 \text{ and } se(\hat{k}) = 2.14$$

Censored sample



$$95\%\ \text{CI}: \quad 10.7 \le \alpha \le 52.48$$

Censored sample



$$95\%\ \text{CI}: \quad 2.78 \le k \le 10.9$$

Censored sample

- 95% CI for α

$$\begin{aligned} \text{no-transformation}: &\ 4.44 \leq \alpha \leq 38.17 \\ \text{log-transformation}: &\ 9.65 \leq \alpha \leq 47.02 \\ \text{LRT}: &\ 10.7 \leq \alpha \leq 52.48 \end{aligned}$$

- 95% CI for k

$$\begin{aligned} \text{no-transformation}: &\ 3.46 \leq k \leq 14.13 \\ \text{log-transformation}: &\ 4.79 \leq k \leq 16.14 \\ \text{LRT}: &\ 2.78 \leq k \leq 10.9 \end{aligned}$$

Censored sample

- Estimate of median

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

Acknowledgements

This lecture is adapted from materials created by Mahbub Latif

References

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