

# Fitting of Twofold Mixed Model for Two-Parameter Weibull Distribution\*

Xiaoling Xu<sup>1</sup>, Ronghua Wang<sup>2</sup>, Beiqing Gu<sup>1</sup>, Shengrong Wu<sup>3</sup>

<sup>1</sup>Business Information Management School, Shanghai Institute of Foreign Trade, Shanghai

<sup>2</sup>Mathematics and Science College, Shanghai Normal University, Shanghai

<sup>3</sup>Yixing Import and Export Inspection and Quarantine Bureau, Yixing

Email: xlxu@shift.edu.cn

Received: Nov. 6<sup>th</sup>, 2012; revised: Nov. 14<sup>th</sup>, 2012; accepted: Nov. 27<sup>th</sup>, 2012

**Abstract:** Considering the data of survival time of raying mice, we can know the reason of death by dissection after the mice died. One reason is thymus lymphoma, and the other is reticulum cell sarcoma. In this paper, we give the statistical method to separate the survival data caused by two reasons without dissection. It is also considered that this kind of mice survival time can be fitted by twofold mixed model of two-parameter Weibull distribution. Besides, the optimum selection criterion is proposed, and the inverse moment estimates and interval estimates of parameters are obtained. The calculated results indicate that the method is simple and feasible.

**Keywords:** Weibull Distribution; Twofold Mixed Model; Thymus Lymphoma; Reticulum Cell Sarcoma; Inverse Moment Estimate; Interval Estimate

## 两重两参数威布尔分布混合模型的拟合\*

徐晓岭<sup>1</sup>, 王蓉华<sup>2</sup>, 顾蓓青<sup>1</sup>, 吴生荣<sup>3</sup>

<sup>1</sup>上海对外贸易学院商务信息学院, 上海

<sup>2</sup>上海师范大学数理学院, 上海

<sup>3</sup>宜兴出入境检验检疫局, 宜兴

Email: xlxu@shift.edu.cn

收稿日期: 2012年11月6日; 修回日期: 2012年11月14日; 录用日期: 2012年11月27日

**摘要:** 考虑被辐射的老鼠的存活时间的数据, 当老鼠死之后, 可以经解剖知道其死亡原因, 一类原因为胸腺淋巴瘤, 另一类原因为网状细胞肉瘤。本文给出在不经解剖的情形下如何将两种原因引起的存活数据进行分离的统计方法, 认为这类老鼠存活时间可用两重两参数威布尔分布混合模型来拟合, 同时提出了一个优选准则, 并得到了参数的逆矩估计和区间估计。计算结果表明本文方法的简便易行。

**关键词:** 威布尔分布; 两重混合模型; 胸腺淋巴瘤; 网状细胞肉瘤; 逆矩估计; 区间估计

### 1. 引言

\*资助信息: 本项目得到国家自然科学基金(11141002); 上海对外贸易学院 2011~2012 年度“085”重点学科建设项目(Z08511009); 上海对外贸易学院 2011~2012 年度“085”重点专业建设——《全球通用商科人才培养本科专业建设》项目; 上海市星系与宇宙学半解析研究重点实验室(SKLA1101); 上海高校选拔培养优秀青年教师科研专项基金(swm10010)资助。

文献[1]给出了一个有关于 60 只被辐射的老鼠的存活时间的数据。当老鼠死之后, 经解剖可以知道其死亡原因。其考虑两类死亡原因所引起的死亡数据。一类原因为胸腺淋巴瘤(包含 22 个数据), 另一类原因为网状细胞肉瘤(含有 38 个数据), 数据见表 1、表 2。

**Table 1. Thymus lymphoma (22 data)**  
**表 1. 胸腺淋巴瘤(22 个数据)**

159	189	191	198	200	207	220	235	245
250	256	261	265	266	280	343	356	383
403	414	428	432					

**Table 2. Reticulum cell sarcoma (38 data)**  
**表 2. 网状细胞肉瘤(38 个数据)**

317	318	399	495	525	536	549	552	554	557
558	571	586	594	596	605	612	621	628	631
636	643	647	648	649	661	663	666	670	695
697	700	705	712	713	738	748	753		

上述数据是通过解剖得到的。如果要做很多次这样的试验，那么解剖老鼠的工作将耗费较大的人力、物力和财力。那么有没有不经过解剖，就能确定这些存活时间数据中哪些是由胸腺淋巴瘤引起的，哪些是由网状细胞肉瘤引起的呢？这正是本文所要研究的问题。考虑受辐射老鼠的存活时间可用两重两参数威布尔分布混合模型来拟合。关于混合模型的表述如下：

当一个总体中每一个个体都是  $k$  个不同类个体中的某一个，第  $i$  类个体在总体中所占比例为  $p_i$ ，诸  $p_i$  满足  $0 < p_i < 1$  和  $\sum_{i=1}^k p_i = 1$ ，这时就产生了混合模型。

假定第  $i$  类的个体服从生存函数为  $R_i(x)$  的分布，从这总体中任选一个个体，其生存函数为：

$$R(x) = \sum_{i=1}^k p_i R_i(x),$$

这种模型被称为“混合模型”。在不能区分不同类的个体时，常常用此模型。诸  $R_i(t)$  常取自同一参数族，当然这非必需的。

令  $F_i(x) = 1 - R_i(x), i = 1, 2, \dots, k$ ，即第  $i$  类的个体的分布函数，则混合模型的分布函数  $F(x)$  为：

$$F(x) = 1 - \sum_{i=1}^k p_i R_i(x) = \sum_{i=1}^k p_i [1 - R_i(x)] = \sum_{i=1}^k p_i F_i(x)$$

在此，称其为  $k$  重混合模型。如果  $k = 2$ ，则称为两重混合模型。

混合模型是一个很灵活而且是强有力的统计建模工具。在实际中混合模型已成为分析复杂现象的一个重要工具，它几乎涵盖了各个学科，其广泛应用于

证券分析、声学研究、生物学、医学、经济、金融、环境科学、植物学、农业和可靠性等领域。有关混合模型的研究历史，第一篇关于混合模型的文章是 Pearson (1894)用两重正态混合分布拟合 Weldon 的数据，Pearson 使用了矩方法，上世纪七十年代前对混合模型的研究大都是使用矩方法及图形方法。从七十年代开始，对混合模型的研究更多地使用了极大似然方法，虽然第一篇使用极大似然方法的是 Rao(1948)用 Fisher 得分方法研究两重等方差正态混合模型的参数估计。随着高速计算机的发展以及迅速普及，混合模型的研究大量运用极大似然方法，尤其是 EM 算法的出现真正促使极大似然方法更广泛地应用于混合模型。Dempster(1977)为 EM 算法建立了一个一般的理论，从而解决了应用极大似然方法在混合模型计算上的障碍。混合模型的前期理论研究成果可从文献[2-6]中找到。

混合模型研究已经成为统计中的一个热点。主要有两方面的原因：一是由于从数学上看，混合模型有许多新的性质，从而带来了新的理论问题；二是在众多应用问题中，混合模型提供了一种方便灵活的统计建模方法，成为应用研究者建模时一种方便的选择，具体地说混合模型比单一模型能更好地拟合现实数据，再者，正如 Pearson(1894)所遇到的问题那样，现实大量问题本身就是一个有明确物理意义的混合模型，这样混合模型中各参数都有明确的物理意义，这时混合模型便是最自然的选择。在实际中，混合模型有其物理背景。譬如，某些产品有缺陷，那就容易引起早期失效，而没有缺陷的产品是逐渐地走向失效。有关混合模型的理论 and 实际应用已有大量文献做了研究。

指数分布的混合模型的研究见 Mendenhall 和 Hader(1958)即文献[7]，Tallis 和 Light(1968)即文献[8]。威布尔分布的混合模型的研究见 Kao(1959)即文献[9]、Falls(1970)即文献[10]。伽玛分布的混合模型的研究见 Dickinson(1974)即文献[11]，蒋仁言和左明健在文献[12,13]较为系统地总结了威布尔混合模型的应用实例，其应用包括：电子管、大气数据、电器设备、老化数据、互补型金属氧化物半导体试验、作物产量估计、仪器仪表、机械零件、品面型二氧化硅击穿数据、筛选等等。Cohen 在文献[14]所研究过的混合正态分

布可应用于各种风速和大量生产的器件的物理尺寸的研究。Kao 在文献[9]所研究过的混合威布尔分布在可靠性研究中是有用的，特别在那些涉及电子管的可靠性问题的研究。金国驊、胡文忠在文献[15]中研究了风速频率分布的混合模型——指数分布和威布尔分布的混合，即在零风速和低风速时指数分布起主要作用；在中风速段和高风速段威布尔分布应起主要作用。文献[16]研究了混合指数分布的参数估计问题。文献[17]给出了一个双胞胎数据的有限混合模型。文献[18]给出了有限泊松混合的稳健统计推断。

## 2. 两重两参数威布尔分布混合模型的参数的逆矩估计

设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本容量为  $n$  的一个简单随机样本， $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  为其次序统计量。总体  $X$  的分布函数为  $F(x)$ ，密度函数为  $f(x)$ ，其中

$$F(x) = pF_1(x) + (1-p)F_2(x),$$

$$f(x) = pf_1(x) + (1-p)f_2(x), \quad 0 < p < 1$$

而  $F_i(x)$ ,  $f_i(x)$ ,  $i=1,2$  分别为两参数威布尔分布的分布函数及密度函数，即

$$F_i(x) = 1 - \exp\left\{-\left(\frac{x}{\eta_i}\right)^{m_i}\right\}$$

$$f_i(x) = \frac{m_i x^{m_i-1}}{\eta_i^{m_i}} \exp\left\{-\left(\frac{x}{\eta_i}\right)^{m_i}\right\}, \quad i=1,2$$

其中  $m_i > 0$  称为形状参数， $\eta_i > 0$  称为刻度参数。在此假定  $\eta_1 < \eta_2$ 。

上述即为两重两参数威布尔混合模型。所谓混合威布尔分布表示一个总体由几个威布尔子总体组成。早在 30 多年前，人们就认识到以混合威布尔分布作为具有一个以上失效起因的单元的寿命模型是合适的。蒋仁言在文献[12]、蒋仁言和左明健在文献[13]中采用 WPP 图较为详细地研究了二重、三重两参数威布尔混合模型的点估计。

在两重两参数威布尔混合模型中，可以认为  $X_1, X_2, \dots, X_n$  中有  $[np]$  个是来自总体分布函数  $F_1(x)$  的一个样本，有  $n-[np]$  个是来自总体分布函数  $F_2(x)$  的一个样本。于是只要能估计出  $p$ ，其它参数的估计

就很容易得到。

首先，假定  $p$  的估计，取  $\hat{p} = \frac{j}{n}$ ,  $j=2,3,\dots,n-2$  作为参数  $p$  的估计，于是， $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(j)}$  为来自总体分布函数为  $F_1(x)$  的容量为  $j$  的前  $j$  个次序统计量，而  $X_{(j+1)} \leq X_{(j+2)} \leq \dots \leq X_{(n)}$  为来自总体分布函数为  $F_2(x)$  的容量为  $n-j$  的前  $n-j$  个次序统计量。为保证估计尽量地精确，采用王炳兴在文献[19]提出的逆矩估计的思想可得参数  $m_1, \eta_1$ ;  $m_2, \eta_2$  的逆矩估计，其依赖于  $j$ ，记逆矩估计为  $\hat{m}_{1(j)}, \hat{\eta}_{1(j)}$ ;  $\hat{m}_{2(j)}, \hat{\eta}_{2(j)}$ 。

其中， $\hat{m}_{1(j)}$  为如下方程的根(方程有唯一正根)：

$$\sum_{i=1}^{j-1} \ln \frac{S_j}{S_i} = \begin{cases} j-2, & j \geq 3 \\ 1, & j = 2 \end{cases}$$

注：文献[19]指出当  $j$  较小时，求解  $m_1$  的估计可用上述方程。事实上，文献[20]通过大量的 Monte-Carlo 模拟说明当  $j$  较大时也是合适的。

$$\hat{\eta}_{1(j)} = \left[ \frac{1}{j} \sum_{i=1}^j X_{(i)}^{m_{1(j)}} \right]^{\frac{1}{m_{1(j)}}}$$

$\hat{m}_{2(j)}$  为如下方程的根：

$$\sum_{i=j+1}^{n-1} \ln \frac{S_n}{S_i} = \begin{cases} n-j-2, & j \leq n-3 \\ 1, & j = n-2 \end{cases}$$

$$\hat{\eta}_{2(j)} = \left[ \frac{1}{n-j} \sum_{i=j+1}^n X_{(i)}^{m_{2(j)}} \right]^{\frac{1}{m_{2(j)}}}$$

其中

$$S_i = \sum_{l=1}^i X_{(l)}^{m_1} + (j-i) X_{(i)}^{m_1}, \quad i=1,2,\dots,j$$

$$S_{j+i} = \sum_{l=1}^i X_{(j+l)}^{m_2} + (n-j-i) X_{(j+i)}^{m_2}, \quad i=1,2,\dots,n-j$$

特别地，若假定  $m_1 = m_2$ ，并同设为  $m$ ，则参数  $m$  的逆矩估计  $\hat{m}$  为如下方程的根：

$$\sum_{i=1}^{j-1} \ln \frac{S_j}{S_i} + \sum_{i=j+1}^{n-1} \ln \frac{S_n}{S_i} = n-3$$

其中

$$S_i = \sum_{l=1}^i X_{(l)}^m + (j-i) X_{(i)}^m, \quad i=1,2,\dots,j$$

$$S_{j+i} = \sum_{l=1}^i X_{(j+l)}^m + (n-j-i) X_{(j+i)}^m, \quad i=1,2,\dots,n-j$$

### 3. 两重两参数威布尔分布混合模型的参数估计的优选准则

余下的问题是如何选择  $j, j = 2, 3, \dots, n-2$ , 下面给出一个优选准则, 称为极大似然准则。

优选准则(极大似然准则): 设似然函数为  $L$ , 其为参数的函数, 而  $L$  的依赖于  $j$  的估计记为  $\hat{L}_j$ , 选择这样的  $j_0$ , 使得:  $\hat{L}_{j_0} = \max_{2 \leq j \leq n-2} \hat{L}_j$ , 或

$$\ln \hat{L}_{j_0} = \max_{2 \leq j \leq n-2} \ln \hat{L}_j, \text{ 其中}$$

$$L = n! \prod_{i=1}^n \left[ p f_1(x_{(i)}) + (1-p) f_2(x_{(i)}) \right]$$

$$\hat{L}_j = n! \prod_{i=1}^n \left\{ \frac{j}{n} \frac{\hat{m}_{1(j)}^{x_{(i)}}}{\hat{\eta}_{1(j)}^{\hat{m}_{1(j)}}} \exp \left[ - \left( \frac{x_{(i)}}{\hat{\eta}_{1(j)}} \right)^{\hat{m}_{1(j)}} \right] + \left( 1 - \frac{j}{n} \right) \frac{\hat{m}_{2(j)}^{x_{(i)}}}{\hat{\eta}_{2(j)}^{\hat{m}_{2(j)}}} \exp \left[ - \left( \frac{x_{(i)}}{\hat{\eta}_{2(j)}} \right)^{\hat{m}_{2(j)}} \right] \right\}$$

注: 通过大量模拟发现, 在使用极大似然准则时, 为保证估计的精确性, 在  $n$  很大时, 通常  $j$  的取值不应太小或太大。另外, 应根据参数的估计值作适当甄别, 例如形状参数过大等等。

于是, 可得参数  $p, m_1, \eta_1, m_2, \eta_2$  的点估计  $\hat{p}, \hat{m}_1, \hat{\eta}_1, \hat{m}_2, \hat{\eta}_2$ :

$$\hat{p} = \frac{j_0}{n}, \hat{m}_1 = \hat{m}_{1(j_0)}, \hat{\eta}_1 = \hat{\eta}_{1(j_0)}, \hat{m}_2 = \hat{m}_{2(j_0)}, \hat{\eta}_2 = \hat{\eta}_{2(j_0)}$$

若给定置信水平  $1-\alpha$ , 则可得参数  $m_1, m_2$  的置信区间分别为:  $[\hat{m}_{1(j_0)1}, \hat{m}_{1(j_0)2}]$ ,  $[\hat{m}_{2(j_0)1}, \hat{m}_{2(j_0)2}]$ , 其中  $\hat{m}_{1(j_0)1}, \hat{m}_{1(j_0)2}$  分别为如下方程的根:

$$\sum_{i=1}^{j_0-1} \ln \frac{S_{j_0}}{S_i} = \frac{1}{2} \chi_{1-\frac{\alpha}{2}}^2 (2(j_0-1))$$

$$\sum_{i=1}^{j_0-1} \ln \frac{S_{j_0}}{S_i} = \frac{1}{2} \chi_{\frac{\alpha}{2}}^2 (2(j_0-1))$$

$\hat{m}_{2(j_0)1}, \hat{m}_{2(j_0)2}$  分别为如下方程的根:

$$\sum_{i=j_0+1}^{n-1} \ln \frac{S_n}{S_i} = \frac{1}{2} \chi_{1-\frac{\alpha}{2}}^2 (2(n-j_0-1))$$

$$\sum_{i=j_0+1}^{n-1} \ln \frac{S_n}{S_i} = \frac{1}{2} \chi_{\frac{\alpha}{2}}^2 (2(n-j_0-1))$$

若有  $m_1 = m_2 \triangleq m$ , 可得参数  $m$  的置信区间分别

为:  $[\hat{m}_{(j_0)1}, \hat{m}_{(j_0)2}]$ , 其中  $\hat{m}_{(j_0)1}, \hat{m}_{(j_0)2}$  分别为如下方程的根:

$$\sum_{i=1}^{j_0-1} \ln \frac{S_{j_0}}{S_i} + \sum_{i=j_0+1}^{n-1} \ln \frac{S_n}{S_i} = \frac{1}{2} \chi_{1-\frac{\alpha}{2}}^2 (2(n-2))$$

$$\sum_{i=1}^{j_0-1} \ln \frac{S_{j_0}}{S_i} + \sum_{i=j_0+1}^{n-1} \ln \frac{S_n}{S_i} = \frac{1}{2} \chi_{\frac{\alpha}{2}}^2 (2(n-2))$$

### 4. 老鼠受辐射的生存时间的实例分析

首先检验表 1 和表 2 的两组数据都服从两参数威布尔分布。检验方法采用: 范蒙特福特(Van-Montfort)检验统计量。具体步骤表述如下:

设生存时间  $X$  的分布函数为  $F(x)$ ,  $X_1, X_2, \dots, X_n$  为来自总体  $X$  的容量为  $n$  的一个简单随机样本, 而  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  为来自总体  $X$  的容量为  $n$  的前  $n$  个次序统计量。要检验假设  $H_0$ :

$$F(x) = 1 - \exp \left\{ - \left( \frac{x}{\eta} \right)^m \right\}$$

其中  $m > 0$  为形状参数,  $\eta > 0$  为刻度参数。

$$\text{令 } H_i = \frac{\ln X_{(i+1)} - \ln X_{(i)}}{EZ_{i+1} - EZ_i}, i = 1, 2, \dots, n-1, \text{ 其中}$$

$EZ_i, i = 1, 2, \dots, n$  为标准极小值总体容量为  $n$  的次序统计量的数学期望, 其值可查阅文献[21]。

构造范蒙特福特(Van-Montfort)检验统计量:

$$W = \frac{r}{n-r-1} \frac{\sum_{i=r+1}^{n-1} H_i}{\sum_{i=1}^r H_i}, \text{ 其中 } r = \left[ \frac{n}{2} \right].$$

当  $H_0$  成立时,  $W$  渐近地服从  $F(2(n-r-1), 2r)$ 。

于是给定置信水平  $1-\alpha$ , 当

$F_{1-\alpha/2}(2(n-r-1), 2r) \leq W \leq F_{\alpha/2}(2(n-r-1), 2r)$  时, 接受  $H_0$ , 即认为  $X_1, X_2, \dots, X_n$  为来自两参数威布尔分布总体。

1) 针对第一组胸腺淋巴瘤的 22 个数据:

$$n = 22, r = \left[ \frac{n}{2} \right] = 11,$$

$$W = \frac{r}{n-r-1} \frac{\sum_{i=r+1}^{n-1} H_i}{\sum_{i=1}^r H_i} = 1.0747, 1-\alpha = 0.95$$

$$F_{1-\alpha/2}(2(n-r-1), 2r) = F_{0.975}(20, 22) = 0.4109$$

$$F_{\alpha/2}(2(n-r-1), 2r) = F_{0.025}(20, 22) = 2.3890$$

由于

$$F_{1-\alpha/2}(2(n-r-1), 2r) \leq W \leq F_{\alpha/2}(2(n-r-1), 2r)$$

所以接受  $H_0$ ，即认为第一组胸腺淋巴瘤的 22 个数据来自两参数威布尔分布总体。

2) 针对第二组网状细胞肉瘤的 38 个数据:

$$n = 38, r = \left\lfloor \frac{n}{2} \right\rfloor = 19,$$

$$W = \frac{r}{n-r-1} \frac{\sum_{i=r+1}^{n-1} H_i}{\sum_{i=1}^r H_i} = 1.0372, 1-\alpha = 0.95$$

$$F_{1-\alpha/2}(2(n-r-1), 2r) = F_{0.975}(36, 38) = 0.5181$$

$$F_{\alpha/2}(2(n-r-1), 2r) = F_{0.025}(36, 38) = 1.9190$$

由于

$$F_{1-\alpha/2}(2(n-r-1), 2r) \leq W \leq F_{\alpha/2}(2(n-r-1), 2r)$$

所以接受  $H_0$ ，即认为第二组网状细胞肉瘤的 38 个数据来自两参数威布尔分布总体。

因此，由胸腺淋巴瘤引起的存活数据和由网状细胞肉瘤引起的存活时间都可认为服从两参数威布尔分布。那么受辐射老鼠的存活时间可用两重两参数威布尔分布混合模型来拟合。

由于胸腺淋巴瘤引起的生存时间共 22 个数据，网状细胞肉瘤引起的生存时间共有 38 个数据，所以参数  $p$  的真值为  $p = \frac{22}{38+22} = 0.3667$ 。

若采用胸腺淋巴瘤这 22 个数据可得参数  $m_1, \eta_1$  的逆矩估计为:  $\hat{m}_1 = 3.4748, \hat{\eta}_1 = 311.0182$ ; 若采用网状细胞肉瘤这 38 个数据可得参数  $m_2, \eta_2$  的逆矩估计为:  $\hat{m}_2 = 7.4442, \hat{\eta}_2 = 645.8610$ 。

若给定置信水平  $1-\alpha = 0.95$ ，此时，

$$\chi_{0.975}^2(42) = 25.99866, \chi_{0.025}^2(42) = 61.77676,$$

$$\chi_{0.975}^2(74) = 52.10283, \chi_{0.025}^2(74) = 99.67835,$$

则可得参数  $m_1$  的置信区间为:  $[2.8557, 5.5116]$ ; 参数  $m_2$  的置信区间为:  $[6.4330, 10.9231]$ 。

由文献[1]中得到的参数估计如下(称为E-J方法):

$$\hat{p} = 0.5, \hat{m}_1 = 5.0, \hat{\eta}_1 = 301.2; \hat{m}_2 = 11.6, \hat{\eta}_2 = 672.7$$

若给定置信水平  $1-\alpha = 0.95$ ，此时，

$$\chi_{0.975}^2(58) = 38.84351, \chi_{0.025}^2(58) = 80.93559,$$

则可得参数  $m_1$  的置信区间为:  $[2.2129, 3.8913]$ ，从中可以看到区间估计没有包含其点估计，在此可以认为E-J方法得到的参数估计精度较差; 参数  $m_2$  的置信区间为:  $[9.2455, 16.2217]$ 。

在形状参数不等的假定下，利用本文方法可解得: 当  $j = 24$  时， $\hat{L}_{24}$  最大，此时， $\hat{m}_1 = 3.7305, \hat{\eta}_1 = 311.2029; \hat{m}_2 = 9.5753, \hat{\eta}_2 = 657.1054;$   
 $\hat{p} = \frac{24}{60} = 0.4$ 。

若给定置信水平  $1-\alpha = 0.95$ ，此时，

$$\chi_{0.975}^2(46) = 29.16005, \chi_{0.025}^2(46) = 66.61653,$$

$$\chi_{0.975}^2(70) = 48.75756, \chi_{0.025}^2(70) = 95.02318,$$

则可得参数  $m_1$  的置信区间为:  $[2.7085, 5.1238]$ ; 参数  $m_2$  的置信区间为:  $[7.3338, 12.4593]$ 。

文献[12]假定形状参数相等，并认为是恰当的，用WPP图方法得到的参数  $p$  的估计更接近于真值，参数估计如下:  $\hat{p} = 0.31, \hat{m} = 7.6, \hat{\eta}_1 = 250, \hat{\eta}_2 = 650$ 。

若给定置信水平  $1-\alpha = 0.95$ ，此时，

$$\chi_{0.975}^2(116) = 88.0848, \chi_{0.025}^2(116) = 147.699,$$

则可得参数  $m$  的置信区间为:  $[5.0106, 7.5743]$ ，从中可以看到区间估计没有包含其点估计，在此可以认为假定形状参数相等不一定是合适的。

在形状参数相等的假定下，利用本文方法可解得: 当  $j = 17$  时， $\hat{L}_{17}$  最大，此时， $\hat{m} = 6.1428, \hat{\eta}_1 = 257.3367, \hat{\eta}_2 = 628.8087, \hat{p} = \frac{17}{60} = 0.2833$ 。

若给定置信水平  $1-\alpha = 0.95$ ，此时，

$$\chi_{0.975}^2(116) = 88.0848, \chi_{0.025}^2(116) = 147.699,$$

则可得参数  $m$  的置信区间为:  $[4.9630, 7.5913]$ ，从中可以看到区间估计包含其点估计。

评述: 1) 若在形状参数不等的假定下，由本文方法得到的  $p$  的估计值为 0.4，与真值的误差为 0.0333;

而采用 E-J 方法得到的  $p$  的估计值为 0.5, 与真值的误差为 0.1333, 从中可以看到本文方法是比较精确的。另外, 用 E-J 方法得到的

$$\begin{aligned}\hat{m}_1 &= 5 \notin (2.2129, 3.8913), \\ \hat{m}_2 &= 11.6 \notin (6.4330, 10.9231), \\ 7.4442 &\notin (9.2455, 16.2217);\end{aligned}$$

用本文方法得到的

$$\begin{aligned}\hat{m}_1 &= 3.7305 \in (2.7085, 5.1238), \\ 3.7305 &\in (2.8557, 5.5116), \\ 3.4748 &\in (2.7085, 5.1238) \\ \hat{m}_2 &= 9.5753 \in (7.3338, 12.4593), \\ 9.5753 &\in (6.4330, 10.9231), \\ 7.4442 &\in (7.3338, 12.4593)\end{aligned}$$

由此可以看出本文的方法更合适。

2) 若假定形状参数相等, 采用 WPP 图方法得到的  $p$  的估计值为 0.31, 与真值的误差为 -0.0567, 用本文方法得到的  $p$  的估计值为 0.2833, 与真值的误差为 -0.0834; 而用 WPP 图方法得到的  $\hat{m} = 7.6 \notin (5.0106, 7.5743)$ , 用本文方法得到的  $\hat{m} = 6.1428 \in (4.9630, 7.5913)$ , 在此可以认为假定形状参数相等不一定是合适的。

## 5. 结论

本文利用被辐射老鼠的存活时间的已有数据, 运用两重两参数威布尔分布混合模型来拟合该数据, 并提出极大似然优选准则对参数进行逆矩估计和区间估计, 得到的估计结果比原有方法更优越一些。另外, 本文的方法具有一般性, 还可应用于电器装备到失效时所使用的周期数<sup>[22]</sup>, 某型号继电器进行寿命试验的寿命检验数据<sup>[23]</sup>等其他实际应用问题。

## 参考文献 (References)

- [1] R. C. Elandt-Johnson, N. L. Johnson. Survival model and data analysis. New York: John Wiley & Sons, 1980: 193.
- [2] G. J. MaLachlan, D. Peel. Finite mixture models. New York: Wiley, 2000.
- [3] B. G. Lindsay, Mixture model: Theory, geometry and application. Hayward: Institute for mathematical Statistics, 1995.
- [4] G. J. MaLachlan, K. E. Basford, Mixture models, inference and application to clustering. New York and Basel: Marcel Dekker, 1988.
- [5] D. M. Titterington. Statistical analysis of finite mixture distribution. New York: John Wiley & Sons Ltd., 1985.
- [6] K. C. Lanf. Introduction to the special issue on finite mixture models. Social Research Methods, 2001, 29(3): 275-281.
- [7] W. Mendenhall, R. J. Hader. Estimation of parameters of mixed exponential distributed failure times from censored life test data. Biometrical, 1958, 45(3-4): 504-520.
- [8] G. M. Tallis, R. Light. The use of fractional moments for estimating the parameters of a mixed exponential distribution. Technometrics, 1968, 10(1): 161-175.
- [9] J. H. K. Kao. A graphical estimation of mixed Weibull parameters in life testing of electron tubes. Technometrics, 1959, 1(4): 389-407.
- [10] L. W. Falls. Estimation of parameters in compound Weibull distributions. Technometrics, 1970, 12(2): 399-407.
- [11] J. P. Dickinson. On the resolution of a mixture of observations from two gamma distributions by the method of maximum likelihood. Metrika, 1974, 21(1): 133-141.
- [12] 蒋仁言. 威布尔模型族特性、参数估计和应用[M]. 北京: 科学出版社, 1998: 62-65, 122-125, 172-174, 199-202.
- [13] 蒋仁言, 左明健. 可靠性模型与应用[M]. 北京: 机械工业出版社, 1999: 20, 124, 126-127, 169-170.
- [14] A. C. Cohen. Estimation of mixtures of poisson and mixtures of exponential distributions. NASA Technical Memorandum NASA TMX 53235, 1965.
- [15] 金国驊, 胡文忠. 风速频率分布混合模型的研究[J]. 太阳能学报, 1994, 15(4): 353-356.
- [16] 朱利平, 卢一强, 蒯诗松. 混合指数分布的参数估计[J]. 应用概率统计, 2006, 22(2): 137-150.
- [17] G. N. Michael. A finite mixture distribution model for data collected from twins. Twin Research, 2003, 6(3): 235-239.
- [18] D. Kalis, E. Xekalaki. Robust inference for finite Poisson mixtures. Journal of Statistical Planning and Inference, 2001, 93(1-2): 93-115.
- [19] 王炳兴. Weibull 分布的统计推断[J]. 应用概率统计, 1992, 8(4): 357-364.
- [20] 王蓉华, 徐晓岭. 三参数 Weibull 分布的统计推断[J]. 数理统计与应用概率, 1997, 12(1): 86-100.
- [21] 中国电子技术标准化研究所. 可靠性试验用表(增订本)[M]. 北京: 国防工业出版社, 1987.
- [22] J. F. Lawless. Statistical models and methods for lifetime data. New York: John Wiley & Sons, Inc., 1982.
- [23] D. B. Kececioglu. Reliability Engineering Handbook (Volume 1). Endlewood Cliffs: Prentice Hall Inc., 1996.