

逐次定数截尾数据下 ZZ 分布的参数估计

赵屹愷, 张国志, 王 萍

(哈尔滨理工大学 理学院, 哈尔滨 150080)

摘 要: 对于设计寿命之前很少失效, 设计寿命之后失效比例大幅增加的一类存储产品, 其寿命变量用 ZZ 分布来描述较为合适。基于逐次定数截尾样本, 首先利用泰勒展式将似然函数中的非线性部分转化为线性表达, 使似然方程可解, 进而得到参数的近似极大似然估计。其次, 运用最佳线性无偏估计法, 给出了参数无偏估计的一般解析表达式, 同时针对两种特殊逐次定数截尾样本, 进一步化简了估计的表达形式, 且为便于使用, 给出了计算所需数表的构造公式。最后对两种估计做了数值模拟比较, 结果表明 ZZ 分布的最佳线性无偏估计优于近似极大似然估计。

关键词: ZZ 分布; 极大似然估计; 近似极大似然估计; 最佳线性无偏估计

DOI: 10.15938/j.jhust.2021.03.023

中图分类号: O231 文献标志码: A 文章编号: 1007-2683(2021)03-0153-07

Parameter Estimation of ZZ Distribution under Progressive Type-II Censored Data

ZHAO Hong-kai, ZHANG Guo-zhi, WANG Ping

(School of Science, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: ZZ distribution is suitable to describe the life variable of a type of storage products that rarely fail before a given design life, and the proportion of failures has increased significantly after the design life. Based on a progressive type-II censored sample, firstly using Taylor's formula to transform the the nonlinear part of the likelihood function into a linear expression, the likelihood equation is solvable, and the approximate maximum likelihood estimation (AMLE) of the parameter is given. Secondly, using the best linear unbiased estimation (BLUE) method, a general analytical expression of the parameter is given. At the same time, the expression form of the estimation is further simplified for two kinds of special progressive type-II censored sample and the required number table is constructed for the convenience of use. Finally, a numerical comparison of the two methods is performed, and the results show that the best linear unbiased estimation of ZZ distribution is better than the approximate maximum likelihood estimation.

Keywords: ZZ distribution; maximum likelihood estimation; approximate maximum likelihood estimation; best linear unbiased estimation

收稿日期: 2020-03-21

基金项目: 黑龙江省自然科学基金(A2018006)。

作者简介: 张国志(1961—), 男, 博士, 教授;

王 萍(1963—), 女, 硕士, 教授。

通信作者: 赵屹愷(1996—), 女, 硕士研究生, E-mail: 1163735976@qq.com.

0 引言

传统的定时和定数截尾数据有一个共同的特点:在截尾时间点和截尾数上缺少灵活性,即在寿命试验中不允许有样品在试验中途退出试验。然而有些试验样品非常昂贵或者在研究耐久性试验时,如果试验人员为了减少试验时间希望在一些时间点将一些未失效的产品移出试验,传统的截尾样本数据就不适用了。随着基础理论的发展和实际应用中的需要,逐次定数截尾数据近年来受到广泛关注。

有关逐次定数截尾数据的研究,在国外,最早研究这类数据问题的文献可以追溯到1963年,文[1-3]详细的讨论了逐次截尾数据下针对常见几种分布的推断问题。1995年后,Sandhu等[4]给出了获得逐次定数截尾样本数据的抽样方法。Balasooryia等[5]和Tse等[6]分别在两参数指数分布和韦布尔分布下基于逐次定数截尾数据推导出了诸可靠性指标。Balakrishnan等[7]的著作对逐次定数截尾寿命试验进行了深入研究,给出了相关模型较为详细的阐述。Balakrishnan等[8-12]讨论了逐次定数截尾样本下正态分布、指数分布和极值分布参数的极大似然估计、区间估计和假设检验问题。Soliman[13]研究了服从Burr-XII分布的极大似然估计和Bayes估计。Mahmoud[14]推导出了逐次定数截尾样本下来自韦布尔-伽玛分布的顺序统计量的近似矩,并利用近似矩得到了参数的最佳线性无偏估计和极大似然估计。Singh等[15]研究了对数正态分布模型的统计分析问题,得到了参数的极大似然估计和Bayes估计。Mahmoud等[16]得到指数-帕累托参数的极大似然估计,并利用近似矩求得参数的最佳线性无偏估计。

国内对逐次定数截尾数据的研究始于上世纪九十年代,贺国芳等[17]说明了在实际问题中经常遇到这类数据,因此有必要讨论这类数据的统计分析方法。徐晓玲等[18]和王炳兴[19]讨论了Weibull分布基于逐次定数截尾寿命数据的参数估计,分别得到了参数的极大似然估计、逆矩估计以及区间估计。杨君慧等[20]给出了广义指数分布参数和可靠度函数的极大似然估计,并在熵损失和加权平方损失函数下,给出参数和可靠度函数的Bayes估计。杨敏[21]基于平方损失函数和LINEX损失函数,讨论了双参数Rayleigh分布参数的贝叶斯估计。罗嘉成

等[22]探讨了Lomax分布的形状参数和可靠性指标的Bayes估计。

由于上述给出的研究均基于常见的几种寿命分布,然而有些存储产品,在给定的设计寿命之前很少失效,过了设计寿命之后失效的比例大幅增加,再用常见一些分布进行统计推断得到的结果与实际明显不符。张宁[23]给出一个较好解决上述存储产品寿命的ZZ分布,并对这类产品的可靠性指标在截尾数据情况下进行了相关的统计分析。

此外,上述对逐次定数截尾数据的研究,所获得的大多是参数极大似然估计和贝叶斯估计的数值解,并未给出具体的解析表达式,且研究多基于几种常见分布,并不能涵盖所有产品的寿命类型,因此本文在逐次定数截尾样本下,试给出ZZ分布参数估计的解析表达式,并配以数表以便使用。

1 ZZ分布与逐次定数截尾试验

1.1 ZZ分布

ZZ分布的分布函数 $F(t)$,密度函数 $f(t)$,失效率函数分别是 $\lambda(t)$ [23]:

$$\begin{aligned} F(t) &= 1 - \exp\left\{1 - e^{\left(\frac{t}{\eta}\right)^m}\right\}, t > 0 \\ f(t) &= \frac{mt^{m-1}}{\eta^m} \cdot e^{\left(\frac{t}{\eta}\right)^m} \cdot e^{1-e^{\left(\frac{t}{\eta}\right)^m}}, t > 0 \\ \lambda(t) &= \frac{mt^{m-1}}{\eta^m} \cdot e^{\left(\frac{t}{\eta}\right)^m}, t > 0 \end{aligned}$$

它含有两个参数 $\eta > 0$ 与 $m > 0$,记为 $ZZ(m, \eta)$,其中 η 称为特征寿命, m 称为形状参数。

1.2 逐次定数截尾试验

产品实施逐次增加定数截尾试验的操作如下:从一批产品中随机抽取 n 个独立同分布产品在相同环境和应力条件下同时进行寿命试验,当观察到第1个样品失效时刻 $X_{1:r:n}^R$ 时,在剩下的 $n-1$ 个未失效的样品中任意抽取 R_1 个撤离试验,还有 $n-1-R_1$ 个未失效的产品留下继续试验;当观察到第2个样品失效时刻 $X_{2:r:n}^R$ 时,在剩下的 $n-2-R_1$ 个未失效样品中任意抽取 R_2 个撤离试验,将余下 $n-2-R_1-R_2$ 个未失效样品留下继续进行试验;如此下去,直至观察到 r 个样品失效时刻 $X_{r:r:n}^R$ 时结束试验,此时余下的 $R_r = n - r - R_1 - R_2 - \cdots - R_{r-1}$ 个样品退出试验。则 $X_{1:r:n}^R < X_{2:r:n}^R < \cdots < X_{r:r:n}^R$ 可以看作是移出数据为 $R = [R_1, R_2, \cdots, R_r]$ 的逐次定数截尾试验数

据,其中 $1 < r \leq n$, 且 $n = r + \sum_{i=1}^r R_i$ 。

考虑到在上述失效时刻 $X_{j:r:n}^R, j=1, 2, \dots, r$, 都会随机移出 R_j 个元件退出试验, 故上述观察到的失效时间不再是容量为 n 的样本的前 r 个次序观测值, 而是 n 个次序统计量中的某 r 个顺序观测值。实际上, 可以将 $X_{1:r:n}^R$ 看成容量为 n 的寿命为 $t_1 < t_2 < \dots < t_n$ 的第 i_1 ($i_1 = 1$) 个次序统计量 t_{i_1} , 将 $X_{2:r:n}^R$ 看成容量为 n 的寿命为 $t_1 < t_2 < \dots < t_n$ 的第 i_2 个次序统计量 t_{i_2} , 直到 $X_{r:r:n}^R$ 为容量为 n 的寿命为 $t_1 < t_2 < \dots < t_n$ 的第 i_r 个次序统计量 t_{i_r} , 其中 $1 = i_1 \leq i_2 \leq \dots \leq i_r \leq n$ 。

设 $X = [X_{1:r:n}^R, X_{2:r:n}^R, \dots, X_{r:r:n}^R]$ 是来自 ZZ 分布逐次定数截尾的样本, $R = [R_1, R_2, \dots, R_r]$ 是试验中相应被移出的产品数, 为书写方便, 简记 $X_j \equiv X_{j:r:n}^R, j=1, 2, \dots, r$, 则逐次定数截尾下 ZZ 分布的极大似然函数为

$$L = c \prod_{j=1}^r [f(x_j) \{1 - F(x_j)\}]^{R_j} =$$

$$c \frac{m^r}{\eta^{mr}} \prod_{j=1}^r x_j^{m-1} \cdot \exp \left\{ \sum_{j=1}^r (R_j + 1) [1 - e^{-(\frac{x_j}{\eta})^m}] + \frac{\sum_{j=1}^r x_j^m}{\eta^m} \right\} \quad (1)$$

其中

$$c = n(n-1-R_1)(n-2-R_1-R_2) \cdots (n-r-R_1-R_2-\cdots-R_r+1)$$
 是与未知参数无关的常数因子。

2 近似极大似然估计

针对 ZZ 分布在逐次定数截尾样本下的极大似然估计没有解析解的情况, 借鉴文 [6] 的方法, 考虑近似似然函数, 并求解显示结果。

设随机变量 $X \sim ZZ(m, \eta)$, 则 $Y = \ln X$ 服从 EZ 分布 [23] 并且它的概率密度函数为

$$f_y(\mu, \sigma) = \frac{1}{\sigma} e^{\frac{y-\mu}{\sigma}} \cdot \exp\{e^{\frac{y-\mu}{\sigma}}\} \cdot \exp\{1 - \exp\{e^{\frac{y-\mu}{\sigma}}\}\}$$

其中 $\mu = \ln \eta, \sigma = 1/m$ 。

$\mu = 0, \sigma = 1$ 时标准的 EZ 分布的密度函数和分布函数有以下的形式:

$$f_y(0, 1) = e^y \cdot \exp\{e^y\} \exp\{1 - \exp\{e^y\}\}$$

$$F(y) = 1 - \exp\{1 - \exp\{e^y\}\}$$

$X = [X_{1:r:n}^R, X_{2:r:n}^R, \dots, X_{r:r:n}^R]$ 是来自 ZZ 分布逐次定数截尾的随机样本, 令 $Y_{j:r:n}^R = \ln X_{j:r:n}^R, z_{j:r:n}^R =$

$\frac{Y_{j:r:n}^R - \mu}{\sigma}$, 则 $z = [z_{1:r:n}^R, z_{2:r:n}^R, \dots, z_{r:r:n}^R]$ 可看成来自标准 EZ 分布逐次定数截尾的随机样本, 简记 $z \equiv [z_1, z_2, \dots, z_r]$, 那么由式(1)得到的似然函数可表示成以下式子:

$$L(\mu, \sigma) = C \frac{1}{\sigma^r} \prod_{j=1}^r w(z_j) [W(z_j)]^{R_j}$$

其中: $w(z) = \exp\{z + e^z + [1 - \exp(e^z)]\}, W(z) = \exp\{1 - \exp(e^z)\}$ 。

从而, 对数似然函数为

$$\ln L = -r \ln \sigma + \sum_{j=1}^r \ln(w(z_j)) + \sum_{j=1}^r R_j \cdot \ln(W(z_j))$$

上式对 μ, σ 分别求偏导数, 有

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{\sigma} \sum_{j=1}^r \frac{w'(z_j)}{w(z_j)} - \sum_{j=1}^r \frac{R_j}{\sigma} \frac{W'(z_j)}{W(z_j)} =$$

$$-\frac{1}{\sigma} \sum_{j=1}^r \frac{w'(z_j)}{w(z_j)} + \sum_{j=1}^r \frac{R_j w(z_j)}{\sigma W(z_j)}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{m}{\sigma} - \sum_{j=1}^r \frac{z_j w'(z_j)}{\sigma w(z_j)} + \sum_{j=1}^r \frac{R_j z_j w(z_j)}{\sigma W(z_j)}$$

得到下列的似然方程组:

$$-\sum_{j=1}^r \frac{w'(z_j)}{w(z_j)} + \sum_{j=1}^r \frac{R_j w(z_j)}{W(z_j)} = 0 \quad (2)$$

$$-m - \sum_{j=1}^r \frac{z_j w'(z_j)}{w(z_j)} + \sum_{j=1}^r \frac{R_j z_j w(z_j)}{W(z_j)} = 0 \quad (3)$$

由式(2)和式(3)组成的似然方程组不能解出 $\hat{\mu}$ 和 $\hat{\sigma}$ 的明显表达式。由于式(2)和式(3)不能直接解出是 $w'(z_j)/w(z_j)$ 和 $w(z_j)/W(z_j)$ 引起的, 根据文 [6] 中给出的方法, 利用泰勒展式在点 $c_j = E(z_j)$ 处将 $w'(z_j)/w(z_j)$ 和 $w(z_j)/W(z_j)$ 近似展开。其中, $z_j = F^{-1}(U_j), U_j, j=1, 2, \dots, r$ 是来自于总体为 $U(0, 1)$ 均匀分布的逐次定数截尾样本。由于 $c_j = E(z_j)$ 直接求解比较困难, 而 $E(U_j)$ 在文 [6] 中已有结论, 故有如下的近似式

$$c_j = E(z_j) \approx F^{-1}(E(U_j))$$

其中 $F^{-1}(U_j) = \ln \ln(1 - \ln(1 - U_j))$, 且

$$p_j = E(U_j) = 1 - \prod_{l=r-j+1}^r$$

$$\frac{l + R_r + R_{r-1} + \dots + R_{r-l+1}}{1 + l + R_r + R_{r-1} + \dots + R_{r-l+1}}, q_j = 1 - p_j, j=1, 2, \dots, r。$$

如此, 能得到如下的展开式:

$$\frac{w'(z_j)}{w(z_j)} \approx a_j - b_j z_j$$

$$\frac{w(z_j)}{W(z_j)} \approx 1 - a_j + b_j z_j$$

其中,

$$\begin{aligned} a_j &= \frac{w'(c_j)}{w(c_j)} - c_j \left[\frac{w''(c_j)}{w(c_j)} - \left(\frac{w'(c_j)}{w(c_j)} \right)^2 \right] = \\ &1 + e^{c_j} [1 - c_j - (1 + c_j) \exp(-e^{c_j}) - c_j \exp(-e^{c_j}) e^{c_j}] = \\ &1 + \ln(1 - \ln q_j) [1 - \ln \ln(1 - \ln q_j) - \\ &(1 + \ln \ln(1 - \ln q_j))(1 - \ln q_j) - \\ &\ln \ln(1 - \ln q_j) \ln(1 - \ln q_j)(1 - \ln q_j)] \\ b_j &= \left(\frac{w'(c_j)}{w(c_j)} \right)^2 - \frac{w''(c_j)}{w(c_j)} = \\ &e^{c_j} [\exp(-e^{c_j}) \cdot e^{c_j} + \exp(-e^{c_j}) - 1] = \\ &\ln(1 - \ln q_j) [1 - \ln q_j \cdot \ln(1 - \ln q_j) - \ln q_j] \end{aligned}$$

那么极大似然方程组(2)和式(3)能近似表示如下

$$-\sum_{j=1}^r (a_j - b_j z_j) + \sum_{j=1}^r R_j (1 - a_j + b_j z_j) = 0 \quad (4)$$

$$-r - \sum_{j=1}^r (a_j - b_j z_j) z_j + \sum_{j=1}^r R_j (1 - a_j + b_j z_j) z_j = 0 \quad (5)$$

将 $z_j = \frac{y_j - \mu}{\sigma}$ 代入式(4),化简得:

$$\hat{\mu} = \frac{\sum_{j=1}^r (1 + R_j) b_j y_j}{\sum_{j=1}^r (1 + R_j) b_j} - \frac{\sum_{j=1}^r (1 + R_j) a_j - \sum_{j=1}^r R_j}{\sum_{j=1}^r (1 + R_j) b_j} \hat{\sigma} \quad (6)$$

$$\text{记 } A = \frac{\sum_{j=1}^r (1 + R_j) b_j y_j}{\sum_{j=1}^r (1 + R_j) b_j}, B = \frac{\sum_{j=1}^r (1 + R_j) a_j - \sum_{j=1}^r R_j}{\sum_{j=1}^r (1 + R_j) b_j},$$

则式(6)可以表示为

$$\hat{\mu} = A - B \hat{\sigma} \quad (7)$$

然后,再将 $z_j = \frac{y_j - \mu}{\sigma}$ 及式(7)代入式(5),得到

关于 $\hat{\sigma}$ 的二次方程:

$$D \hat{\sigma}^2 + E \hat{\sigma} - H = 0 \quad (8)$$

其中,

$$\begin{aligned} D &= r + B \sum_{j=1}^r a_j - B \sum_{j=1}^r R_j (1 - a_j) - B^2 \sum_{j=1}^r b_j - B^2 \sum_{j=1}^r R_j b_j = \\ &r + B \left\{ \sum_{j=1}^r (1 + R_j) a_j - \sum_{j=1}^r R_j - \left[\sum_{j=1}^r (1 + R_j) a_j - \sum_{j=1}^r R_j \right] \right\} = r \\ E &= \sum_{j=1}^r (a_j + R_j a_j - R_j) (y_j - A) - 2B \sum_{j=1}^r (1 + R_j) b_j (y_j - A) \\ H &= \sum_{j=1}^r (1 + R_j) b_j (y_j - A)^2 > 0 \end{aligned}$$

解方程式(8),得到关于 $\hat{\sigma}$ 的值:

$$\hat{\sigma} = \frac{-E + \sqrt{E^2 + 4rH}}{2r}$$

将求出的 $\hat{\sigma}$ 的值代入式(7)即可求出 $\hat{\mu}$ 的值。

3 最佳线性无偏估计

设在逐次截尾 $R = [R_1, R_2, \dots, R_r]$ 的试验中,失效元件的寿命记为 $t_{i_1} \leq t_{i_2} \leq \dots \leq t_{i_r}$,其下标看做是随机向量 $(I_1, I_2, \dots, I_r)$ 的一组观测值,该随机向量所有可能取值的集合记为 G ,设其分布律为 $P\{I_1 = i_1, I_2 = i_2, \dots, I_r = i_r\} = P_{i_1, i_2, \dots, i_r}$,其中 $[i_1, i_2, \dots, i_r] \in G$ 且 $1 = i_1 \leq i_2 \leq \dots \leq i_r \leq n$ 。则在 $\{I_1 = i_1, I_2 = i_2, \dots, I_r = i_r\}$ 条件下,有

$$F(t_{i_j}) = 1 - \exp\{1 - e^{\left(\frac{t_{i_j}}{\eta}\right)^m}\}, j = 1, 2, \dots, r$$

移项,并取三重对数后可得

$$\ln \ln(1 - \ln(1 - F(t_{i_j}))) = m \ln t_{i_j} - m \ln \eta \quad (9)$$

令 $T_{i_1, i_2, \dots, i_r} = [T_{i_1}, T_{i_2}, \dots, T_{i_r}]^T$, 其中 $T_{i_j} = \ln \ln(1 - \ln(1 - F(t_{i_j})))$, $j = 1, 2, \dots, r$ 。显然 $F(t_{i_1}) < F(t_{i_2}) < \dots < F(t_{i_r})$ 是来自 $U(0, 1)$ 均匀分布的统计量,设

$E T_{i_1, i_2, \dots, i_r} = a_{i_1, i_2, \dots, i_r} = [a_{i_1}, a_{i_2}, \dots, a_{i_r}]^T$, $\Sigma_{i_1, i_2, \dots, i_r} = E [T_{i_1, i_2, \dots, i_r} - a_{i_1, i_2, \dots, i_r}] [T_{i_1, i_2, \dots, i_r} - a_{i_1, i_2, \dots, i_r}]^T$ 则 $T_{i_1, i_2, \dots, i_r}$ 均值和方差只与 n 和 r 有关,而不依赖于其它参数。那么式(9)可表示为

$$\ln t_{i_j} = \ln \eta + \frac{1}{m} a_{i_j} + \frac{1}{m} \delta_{i_j}, j = 1, 2, \dots, r \quad (10)$$

其中 $\delta_{i_1, i_2, \dots, i_r} = [\delta_{i_1}, \delta_{i_2}, \dots, \delta_{i_r}]^T$, 并且

$$E(\delta_{i_1, i_2, \dots, i_r}) = 0, \text{Cov}(\delta_{i_1, i_2, \dots, i_r}) = \Sigma_{i_1, i_2, \dots, i_r}。$$

再令

$$Y_{i_1, i_2, \dots, i_r} = [Y_{i_1}, Y_{i_2}, \dots, Y_{i_r}]^T = [\ln t_{i_1}, \ln t_{i_2}, \dots, \ln t_{i_r}]^T$$

$$X_{i_1, i_2, \dots, i_r} = \begin{pmatrix} 1 & a_{i_1} \\ 1 & a_{i_2} \\ \vdots & \vdots \\ 1 & a_{i_r} \end{pmatrix}$$

$$\theta = (\mu, \sigma)^T, \text{其中 } \mu = \ln \eta, \sigma = \frac{1}{m}$$

$$\varepsilon_{i_1, i_2, \dots, i_r} = \frac{1}{m} \delta_{i_1, i_2, \dots, i_r}$$

将式(10)转化为如下矩阵形式

$$\begin{cases} Y_{i_1, i_2, \dots, i_r} = X_{i_1, i_2, \dots, i_r} \theta + \varepsilon_{i_1, i_2, \dots, i_r} \\ \text{Cov}(\varepsilon_{i_1, i_2, \dots, i_r}) = \sigma^2 \Sigma_{i_1, i_2, \dots, i_r} \end{cases}$$

根据高斯-马尔科夫定理,用加权最小二乘法求得参数的最佳线性无偏估计(BLUE)

$$\hat{\theta}_{i_1, i_2, \dots, i_r} = \left[\hat{\mu}_{i_1, i_2, \dots, i_r} \right] =$$

$$\left[X'_{i_1, i_2, \dots, i_r} \sum_{i_1, i_2, \dots, i_r}^{-1} X_{i_1, i_2, \dots, i_r} \right]^{-1} X'_{i_1, i_2, \dots, i_r} \times \sum_{i_1, i_2, \dots, i_r}^{-1} Y_{i_1, i_2, \dots, i_r}$$

当 n, r 及 $i_1, i_2, \dots, i_r$ 给定时, 矩阵 $X_{i_1, i_2, \dots, i_r}$ 与 $\sum_{i_1, i_2, \dots, i_r}^{-1}$ 都是已知的, 所以为了便于使用, 将参数的最佳线性无偏估计简化如下形式

$$\begin{aligned} \hat{\mu}(i_1, i_2, \dots, i_r) &= \sum_{j=1}^r D_0(i_1, i_2, \dots, i_r, n, r, j) \ln t_{i_j} \\ \hat{\sigma}(i_1, i_2, \dots, i_r) &= \sum_{j=1}^r D_1(i_1, i_2, \dots, i_r, n, r, j) \ln t_{i_j} \end{aligned}$$

其中: $D_0(i_1, i_2, \dots, i_r, n, r, j)$ 称为 $\mu(i_1, i_2, \dots, i_r)$ 的最佳线性无偏估计系数; $D_1(i_1, i_2, \dots, i_r, n, r, j)$ 称为 $\sigma(i_1, i_2, \dots, i_r)$ 的最佳线性无偏估计系数。

因此, 在 $R = (R_1, R_2, \dots, R_r)$ 下, ZZ 分布参数最佳线性无偏估计的一般表达形式为

$$\hat{\mu} = \sum_{(i_1, i_2, \dots, i_r) \in G} P_{i_1, i_2, \dots, i_r} \cdot \hat{\mu}(i_1, i_2, \dots, i_r) \quad (11)$$

$$\hat{\sigma} = \sum_{(i_1, i_2, \dots, i_r) \in G} P_{i_1, i_2, \dots, i_r} \cdot \hat{\sigma}(i_1, i_2, \dots, i_r) \quad (12)$$

上面的表达式涉及到系数 $D_0(i_1, i_2, \dots, i_r, n, r, j)$, $D_1(i_1, i_2, \dots, i_r, n, r, j)$ 和概率 $P\{I_1 = i_1, I_2 = i_2, \dots, I_r = i_r\} \equiv P_{i_1, i_2, \dots, i_r}$ 的计算, 本文仅对下面两种情况, 给出参数估计的简化形式, 并构造相应的数表计算公式, 便于使用。

3.1 $R = [n-r, 0, 0, \dots, 0]_{1 \times r}$ 时的参数估计

当移出数据为 $R = [n-r, 0, 0, \dots, 0]_{1 \times r}$ 时, n 个元件参加试验, 在第一个元件失效时刻一次随机移出 $n-r$ 个未失效元件, 剩下的 $r-1$ 个元件继续参加试验直至全部失效。 $t_{i_1} < t_{i_2} < \dots < t_{i_r}$ 可以看成来自容量为 n 的次序统计量 $t_1 < t_2 < \dots < t_n$ 中固定第一个次序统计量后, 从余下 $n-1$ 元件中随机抽取 $r-1$ 个的子样本, 且该子样本有 C_{n-1}^{r-1} 种等可能。所以此时有

$$P\{I_1 = i_1, I_2 = i_2, \dots, I_r = i_r\} \equiv P_{i_1, i_2, \dots, i_r} = 1/C_{n-1}^{r-1}$$

由式(11)和式(12)得 ZZ 分布参数估计可表示为如下形式

$$\hat{\mu}_{R=(n-r, 0, 0, \dots, 0)} = \frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} \hat{\mu}(i_1, i_2, \dots, i_r)$$

$$\hat{\sigma}_{R=(n-r, 0, 0, \dots, 0)} = \frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} \hat{\sigma}(i_1, i_2, \dots, i_r)$$

令

$$\frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} D_0(i_1, i_2, \dots, i_r, n, r, j) =$$

$$D'_{0, R=(n-r, 0, 0, \dots, 0)}(n, r, j)$$

$$\frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} D_1(i_1, i_2, \dots, i_r, n, r, j) =$$

$$D'_{1, R=(n-r, 0, 0, \dots, 0)}(n, r, j)$$

则 $R = [n-r, 0, 0, \dots, 0]_{1 \times r}$ 时 ZZ 分布的最佳线性无偏估计如下

$$\hat{\mu}_{R=(n-r, 0, 0, \dots, 0)} = \frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} \sum_{j=1}^r D_0(i_1, i_2, \dots, i_r, n, r, j) \ln t_{i_j} =$$

$$\dots, i_r, n, r, j) \ln t_{i_j} =$$

$$\sum_{j=1}^r \left[\frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} D_0(i_1, i_2, \dots, i_r, n, r, j) \right] \ln t_{i_j} =$$

$$\sum_{j=1}^r D'_{0, R=(n-r, 0, 0, \dots, 0)}(n, r, j) \ln t_{i_j}$$

$$\hat{\sigma}_{R=(n-r, 0, 0, \dots, 0)} = \frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} \sum_{j=1}^r D_1(i_1, i_2, \dots, i_r, n, r, j) \ln t_{i_j} =$$

$$\dots, i_r, n, r, j) \ln t_{i_j} =$$

$$\sum_{j=1}^r \left[\frac{1}{C_{n-1}^{r-1}} \sum_{(i_1, i_2, \dots, i_r) \in G} D_1(i_1, i_2, \dots, i_r, n, r, j) \right] \ln t_{i_j} =$$

$$\sum_{j=1}^r D'_{1, R=(n-r, 0, 0, \dots, 0)}(n, r, j) \ln t_{i_j}$$

为了便于计算参数的估计, 其系数 $D'_{0, R=(n-r, 0, 0, \dots, 0)}(n, r, j)$ 和 $D'_{1, R=(n-r, 0, 0, \dots, 0)}(n, r, j)$ 在 $n=5$ 到 $n=20$ 时已由计算机给出。

3.2 $R = [1, 1, \dots, 1, 0, 0, \dots, 0]_{1 \times r}$ 时的参数估计

在每次失效时刻 $X_{j:r:n}^R, j=1, 2, \dots, r$ 分别移出 1 个未失效产品的情况: n 个元件参加试验, 设有 r 个失效, 且在每次失效时刻移出 1 个未失效元件。在 $R = [1, 1, \dots, 1, 0, 0, \dots, 0]_{1 \times r}$ 时, 每种情况可能出现的概率 $P\{I_1 = i_1, I_2 = i_2, \dots, I_r = i_r\} \equiv P_{i_1, i_2, \dots, i_r}$ 不再相等, 但其计算可借助计算机完成, 最终体现在后面的数表中。

由式(11)和式(12), 移出数据为 $R = [1, 1, \dots, 1, 0, 0, \dots, 0]_{1 \times r}$ 的 ZZ 分布的最佳线性无偏估计表示如下

$$\hat{\mu}_{R=(1, \dots, 1, 0, 0, \dots, 0)} = \sum_{j=1}^r D'_{0, R=(1, \dots, 1, 0, 0, \dots, 0)}(n, r, j) \ln t_{i_j}$$

$$\hat{\sigma}_{R=(1, \dots, 1, 0, 0, \dots, 0)} = \sum_{j=1}^r D'_{1, R=(1, \dots, 1, 0, 0, \dots, 0)}(n, r, j) \ln t_{i_j}$$

其中,

$$D_{0,R=(1,\dots,1,0,\dots,0)}(n,r,j) = \sum_{\substack{(i_1,i_2,\dots,i_r) \in G \\ 1=i_1 < i_2 < \dots < i_r \leq n}} P_{i_1,i_2,\dots,i_r} \times$$

$$D_0(i_1,i_2,\dots,i_r,n,r,j)$$

$$D_{1,R=(1,\dots,1,0,\dots,0)}(n,r,j) = \sum_{\substack{(i_1,i_2,\dots,i_r) \in G \\ 1=i_1 < i_2 < \dots < i_r \leq n}} P_{i_1,i_2,\dots,i_r} \times$$

$$D_1(i_1,i_2,\dots,i_r,n,r,j)$$

为了便于计算参数的估计,其系数

$D_{0,R=(1,1,\dots,1,0,\dots,0)}(n,r,j)$ 和 $D_{1,R=(1,1,\dots,1,0,\dots,0)}(n,r,j)$ 在 $n=5$ 到 $n=20$ 时已由计算机给出。

4 数值模拟

基于 Monte Carlo 模拟,给出以上两种方法下 ZZ 分布参数的估计结果。根据文 [4] 给出的算法,产生 ZZ 分布下逐次定数截尾数据的步骤如下:

1) 产生 r 个来自均匀分布 $U(0,1)$ 的独立随机样本 $W_1, W_2, \dots, W_r$;

2) 在预先设定的逐次定数截尾移出数据

$R = [R_1, R_2, \dots, R_r]$ 下,令

$$V_j = W_j^{1/(j+R_j+R_{j-1}+\dots+R_{r-j+1})}, j=1, 2, \dots, r;$$

3) 再令 $U_j = 1 - V_r V_{r-1} \dots V_{m-i+1}, j=1, 2, \dots, r$, 则 $U_1, U_2, \dots, U_r$ 是来自均匀分布 $U(0,1)$ 的逐次定数截尾数据;

4) 令 $X_j = F^{-1}(U_j), j=1, 2, \dots, r$, 得到来自 ZZ 分布的逐次定数截尾试验数据,其中 $F^{-1}(\cdot)$ 是 ZZ 分布函数的反函数,

$$X_j = (\beta \cdot \ln(1 - \ln(1 - U_j)))^{1/m}, j=1, 2, \dots, r.$$

设定两组不同样本量和两组不同的移出数据,根据 3.1 和 3.2 节方法,构造最佳线性无偏估计的系数值分别如表 1 和表 2 所示。

表 1 $n=10, r=8$ 的 BLUE 系数值

Tab. 1 the coefficient table of BLUE

j	$R = [r-r, 0, \dots, 0]_{1 \times r}$		$R = [1, \dots, 1, 0, \dots, 0]_{1 \times r}$	
	D_0'	D_1'	D_0'	D_1'
1	0.127 8	-0.348 4	0.138 8	-0.348 4
2	-0.086 7	0.009 2	-0.102 4	0.011 4
3	0.003 6	-0.010 1	0.004 3	-0.010 5
4	0.003 9	-0.007 0	0.005 6	-0.009 8
5	-0.014 3	0.031 3	-0.016 4	0.028 4
6	-0.035 0	0.107 6	-0.043 6	0.106 0
7	0.087 0	0.135 3	0.073 4	0.132 1
8	0.913 8	0.082 1	0.940 2	0.090 7

表 2 $n=15, r=11$ 的 BLUE 系数值

Tab. 2 the coefficient table of BLUE

j	$R = [r-r, 0, \dots, 0]_{1 \times r}$		$R = [1, \dots, 1, 0, \dots, 0]_{1 \times r}$	
	D_0'	D_1'	D_0'	D_1'
1	0.346 5	-0.228 0	0.605 8	-0.155 6
2	-0.447 5	-0.073 8	-0.829 6	-0.186 9
3	-0.233 9	-0.045 2	-0.415 0	-0.088 0
4	0.008 0	0.004 1	0.014 1	0.006 7
5	0.089 3	0.020 2	0.148 4	0.030 1
6	0.170 4	0.038 2	0.275 3	0.064 0
7	0.188 8	0.038 5	0.300 4	0.073 8
8	0.132 1	0.003 8	0.210 6	0.035 1
9	0.062 4	-0.056 4	0.084 3	-0.036 0
10	0.106 6	-0.074 4	0.085 2	-0.062 6
11	0.275 2	-0.088 0	0.520 5	0.319 3

应用上述数据,模拟得参数估计的结果如表 3 所示。

表 3 $m_{\text{真}}=2, \eta_{\text{真}}=2$ 的数值模拟结果

Tab. 3 Numerical simulation results

(n, r)	$R = [R_1, R_2, \dots, R_r]$	AMLE		BLUE	
		$\hat{m}$	$\hat{\eta}$	$\hat{m}$	$\hat{\eta}$
(10, 8)	$(2, 0, 0, \dots, 0)_{1 \times 8}$	1.619 5	1.358 4	1.799 8	1.574 7
	$(1, 1, 0, \dots, 0)_{1 \times 8}$	1.663 9	1.340 2	1.799 7	1.431 8
(15, 11)	$(3, 0, 0, \dots, 0)_{1 \times 12}$	1.688 5	1.435 8	1.833 1	1.598 0
	$(1, 1, 1, 0, \dots, 0)_{1 \times 12}$	1.690 4	1.408 9	1.711 0	1.475 4

从表 3 中可以看出,最佳线性无偏估计结果和近似极大似然估计结果都较接近参数真值,且最佳线性无偏估计结果较优。因此,二者均可应用于寿命服从 ZZ 分布、逐次定数截尾寿命试验下的参数估计。

5 结 论

本文基于逐次定数截尾寿命试验模型,运用近似极大似然法和最佳线性无偏估计法,对产品寿命服从 ZZ 分布的参数进行了统计分析。主要构建了 r 维不可观测但可求分布的随机向量,据此将收集到的逐次定数截尾样本数据转化为一定概率分布条件下的顺序统计量,进而给出 ZZ 分布参数的最佳线性无偏估计,并可依据构造出的数表再进行简单计算即得到参数的估计值。且此法具有的解析表达

形式、运用数表计算方便等优点, 易可应用到 Weibull 分布或其他指数型寿命分布中。

参考文献:

- [1] COHEN, CLIFFORD A. Progressively Censored Samples in Life Testing [J]. Technometrics, 1963, 5(3): 327.
- [2] COHEN A C. Progressively Censored Sampling in the Three Parameter Log-normal Distribution [J]. Technometrics, 1976, 18(1): 99.
- [3] NORGAARD C N J. Progressively Censored Sampling in the Three-parameter Gamma Distribution [J]. Technometrics, 1977, 19(3): 333.
- [4] SANDHU, BALAKRISHNAN N. A Simple Simulation Algorithm for Generating Progressive Type-II Censored Samples [J]. The American Statistician, 1995, 49(2): 229.
- [5] BALASOORIYA U, SAW S L C. Reliability Sampling Plans for the Two-parameter Exponential Distribution under Progressive Censoring [J]. Journal of Applied Statistics, 1998, 25(5): 707.
- [6] TSE S K, YANG C. Reliability Sampling Plans for the Weibull Distribution under Type-II Progressive Censoring with Binomial Removals [J]. Journal of Applied Statistics, 2003, 30(6): 709.
- [7] BALAKRISHNAN N, AGGARWALA R. Progressive Censoring [M]. Birkhauser, 2000.
- [8] BALAKRISHNAN N, JIE Mi. Existence and Uniqueness of the MLEs for Normal Distribution Based on General Progressively Type-II Censored Samples [J]. Statistics and Probability Letters, 2003, 64(4): 407.
- [9] BALAKRISHNAN N, KANNAN N, LIN C T. Point and Interval Estimation for Gaussian Distribution, Based on Progressively Type-II Censored Samples [J]. IEEE Transactions on Reliability, 2003, 52(1): 90.
- [10] BALAKRISHNAN N, LIN C T. On the Distribution of a Test for Exponentiality Based on Progressively Type-II Right Censored Spacings [J]. Journal of Statistical Computation & Simulation, 2003, 73(4): 277.
- [11] BALAKRISHNAN N, NG H K T, KANNAN N. A Test of Exponentiality Based on Spacings for Progressively Type-II Censored Data [J]. Goodness-of-fit Tests and Model Validity, 2002: 89.
- [12] BALAKRISHNAN N, KANNAN N, LIN C T, et al. Inference for the Extreme Value Distribution under Progressive Type-II Censoring [J]. Journal of Statistical Computation and Simulation, 2004, 74(1): 25.
- [13] SOLIMAN, A. A. Estimation of Parameters of Life from Progressively Censored Data Using Burr-XII Model [J]. IEEE Transactions on Reliability, 2005, 54(1): 34.
- [14] MAHMOUD M A W, MOSHREF M, YHIEA N M, et al. Progressively Censored Data from the Weibull Gamma Distribution Moments and Estimations [J]. J Appl Probab Stat, 2014, 3(1): 45.
- [15] SINGH S, TRIPATHI Y M, WU S J. On Estimating Parameters of a Progressively Censored Log-normal Distribution [J]. Journal of Statistical Computation and Simulation, 2015, 85(6): 1071.
- [16] ALIM A M, JAHEEN Z F. Estimation of Parameters for the Exponentiated Pareto Distribution Based on Progressively Type-II Right Censored Data [J]. Journal of the Egyptian Mathematical Society, 2016, 24(3).
- [17] 贺国芳, 瞿荣贞编写. 可靠性数据的收集与分析 [M]. 北京: 国防工业出版社, 1995: 41.
- [18] 徐晓玲, 王蓉华. Weibull 分布逐步增加的 II 型截尾试验的统计分析 [J]. 强度与环境, 2003, 30(2): 31.
XU Xiaoling, WANG Ronghua. Statistical Analysis of Type II Censoring Test with Increasing Weibull Distribution [J]. Strength and Environment, 2003, 30(2): 31.
- [19] 王炳兴. Weibull 分布基于定数逐次截尾寿命数据的统计分析 [J]. 科技通报, 2004(6): 488.
WANG Bingxing. Statistical Analysis of Weibull Distribution Based on Constant Successive Censored Life Data [J]. Science and Technology Bulletin, 2004(6): 488.
- [20] 杨君慧, 师义民, 曹弘毅. 逐步增加 II 型截尾试验下广义指数分布的统计分析 [J]. 统计与决策, 2014(16): 28.
YANG Junhui, SHI Yimin, CAO Hongyi. Statistical Analysis of Generalized Exponential Distribution under Gradually Increasing Type II Censoring Test [J]. Statistics and Decision Making, 2014(16): 28.
- [21] 杨敏. 删失数据场合双参数 Rayleigh 分布的参数估计 [J]. 广西师范学院, 2015: 12.
- [22] 罗嘉成, 陈勇明. 逐次定数截尾下 Lomax 分布的贝叶斯分析 [J]. 成都信息工程大学学报, 2016, 31(3): 323.
LUO Jiacheng, CHEN Yongming. Bayesian Analysis for Lomax Distribution under Progressively Type-II Censored Data [J]. Journal of Chengdu University of Information Technology, 2016, 31(3): 323.
- [23] 张宁. 截尾数据下 ZZ 分布参数估计及性质 [D]. 哈尔滨: 哈尔滨理工大学, 2019.

(编辑: 温泽宇)