

截尾数据下 ZZ 分布的参数估计

张国志, 张 宁

(哈尔滨理工大学 应用数学系, 哈尔滨 150080)

摘 要: 在可靠性统计中, 寿命分布通常有 4 种: 指数分布、韦布尔分布、极值分布、对数正态分布。但在有些实际问题中, 一些元件的寿命用上述 4 种分布来刻画往往与实际相差甚远, 这说明该元件寿命分布并不属于熟知的这 4 种分布。由此给出了一种新型 ZZ 分布, 该分布较好地刻画了这类元件的寿命, 并在截尾数据下对参数作出最佳线性无偏估计与简单线性无偏估计。为了便于使用附录中给出了估计所需要的数表。

关键词: ZZ 分布; 截尾数据; 参数估计; 简单线性无偏估计; 最佳线性无偏估计

DOI: 10.15938/j.jhust.2020.01.023

中图分类号: O231 **文献标志码:** A **文章编号:** 1007-2683(2020)01-0149-05

Parameter Estimation of ZZ Distribution Based on Censored Data

ZHANG Guo-zhi, ZHANG Ning

(Department of Applied Mathematics, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: In reliability statistics, there are usually four types of life distributions: exponential distribution, Weibull distribution, extreme-value distribution and log-normal distribution. But in some real life applications, the results of statistical inference were found to be far from the truth if using the four life distributions to describe the life of some components. It shows that the component life distribution is not anyone of the four distributions. This paper proposes a new type of life distribution called ZZ distribution. This distribution is a better description of the life for such components. And based on censored data, the best linear unbiased estimation and the simple linear unbiased estimation of parameters were given. The required tables were given in the appendix.

Keywords: ZZ distribution; censored data; parameter estimation; simple linear unbiased estimation; best linear unbiased estimation

0 引 言

众所周知常见的寿命分布有 4 种, 指数分布、韦布尔分布、极值分布、对数正态分布。很多元件的寿命分布属于这四种分布, 因此在各种样本形式下, 关

于这四种分布的可靠性统计推断的研究非常之多。比如: Lee J. Bain^[1] 给出了截尾样本下极值分布尺度参数的简单线性无偏估计, 在此基础上 Max Engelhardt 和 Lee J. Bain^[2] 考虑了极值分布次序统计量的影响, 给出了更精确的简单线性无偏估计, 并与此样本下的极大似然估计进行了对比。W. J. Sza-

收稿日期: 2018-03-08

基金项目: 黑龙江省教育厅科学技术研究项目(11551094)。

作者简介: 张 宁(1994—), 女, 硕士研究生。

通信作者: 张国志(1961—), 男, 博士, 教授, 硕士研究生导师, E-mail: zhangguozhi@hrbust.edu.cn.

jnowski^[3] 研究了对数正态分布参数估计量,并用蒙特卡洛法模拟检验其合理性。A. Clifford Cohen 和 Betty Jones Whitten^[4] 对三参数对数正态分布的参数极大似然估计和矩估计做了修正。N. R. Farnum 和 P. Booth^[5] 给出了两参数韦布尔分布的参数极大似然估计并对其唯一性进行了验证。Zheng R^[6] 研究了三参数韦布尔分布参数估计方法并探究了其在可靠性分析中的应用。Davies I J^[7] 用最小二乘法估计了三参数韦布尔分布的尺度参数。J. William Shelnutt, Albert H. Moore 和 H. Leon Harter^[8] 研究了韦布尔分布线性估计。对于存储产品,早期失效数据对寿命分布有较大影响,K. Muralidharan 和 P. Lathika^[9] 对韦布尔分布的早期失效数据做了分析。Wayne Nelson^[10] 研究了步进应力加速寿命试验模型。N. Balakrishnan, Qihao Xie 和 D. Kundu^[11] 研究了指数分布在定时截尾数据形式下的简单步加模型。C. Xiong^[12] 研究了简单步加模型在 II 型数据截尾下参数统计推断方法。在对参数做出估计后,对这些估计的分布进行研究非常必要。Thoman, D. R. ,L. J. Bain 和 C. E. Antle^[13] 研究了韦布尔分布参数在截尾数据形式下的简单线性估计及极大似然估计的近似分布推断方法。对参数近似分布及置信限^[14] 的研究也为参数置信区间的确定^[15-17] 奠定了基础。

以上的研究工作都是基于常见的 4 种分布展开的,然而有些存储产品,在厂家给出的设计寿命之前几乎很少失效,过了设计寿命之后失效的比例大幅

增加。这时根据样本对寿命分布进行检验,往往做出的推断是: 或者拒绝这四种分布; 或者接受某些分布,但统计推断的结论与实际相差甚远。因此,需要寻找一种新的寿命分布来刻画这类元件寿命。下面给出的例子就是一个实际遇到的问题。对于这个实际问题,自然希望: ①新分布的检验获得通过; ②新分布下平均寿命的估计接近实际。

例 1 下表是某存储产品在不同时间段内,参与检测及检测失效的数据。

表 1 某产品检测结果
Tab.1 Test result of the product

存储时间/年	试验数/个	失效数/个	失效频率
2	2 829	1	1/2 829
3	1 561	0	0/1 561
4	4 090	6	6/4 090
5	3 587	3	3/3 587
6	5 041	7	7/5 041
7	2 493	1	1/2 493
8	2 431	4	4/2 431
9	971	8	8/971

从表 1 中可以看到,失效频率并没有随着存储时间增加而呈现非降状态,这就是所谓的“数据倒挂”现象。首先根据张尧庭^[18] 对“数据倒挂”进行处理,然后利用杨振海^[19] 的经验极小卡方法对数据其来自的分布进行检验。“数据倒挂”处理以及寿命分布的检验结果汇总于表 2。

表 2 某产品寿命分布检验结果
Tab.2 Test result of life distribution

参数	指数分布 $E(\theta)$	韦布尔分布 $Weib(m, \eta)$	极值分布 $EV(\mu, \sigma)$	对数正态分布 $LN(\mu, \sigma^2)$
经验极小 χ^2 值	9. 5112	6. 9153	6. 2990	9. 3953
结论($\alpha=0. 05$)	不拒绝	不拒绝	不拒绝	不拒绝
平均寿命 $\hat{\theta}$ (经验极小方法)	5000	375	29	135
平均寿命 $\hat{\theta}$ (极大似然方法)	4167	263	36	329
可靠寿命 (经验极小方法)	$t_{0. 8}=1116$ $t_{0. 9}=527$ $t_{0. 95}=256$	$t_{0. 8}=169$ $t_{0. 9}=107$ $t_{0. 95}=69$	$t_{0. 8}=26$ $t_{0. 9}=23$ $t_{0. 95}=20$	$t_{0. 8}=45$ $t_{0. 9}=30$ $t_{0. 95}=22$

从这个例子可以看出,根据样本对产品寿命的分布进行检验,其结果是: 四种分布假设都不拒绝。

但对指数分布、韦布尔分布、对数正态分布的平均寿命的估计却与事实相差甚远; 对于极值分布,相比其

它3个分布平均寿命的估计最小,但也远远高于实际寿命,而且服从极值分布的随机变量不是非负的,这多少也是一种遗憾。自然希望找到一个新的非负随机变量来描述该产品的寿命,并且统计推断的结果与实际相符。本文的工作由此展开。

1 新型分布—ZZ(m, η) 分布

定义1 若随机变量 T 的分布函数为

$$F(t) = 1 - \exp\{1 - e^{(\frac{t}{\eta})^m}\}, t > 0 (m > 0, \eta > 0)$$

那么称 T 服从参数为 m, η 的 ZZ 分布,记为 $T \sim ZZ(m, \eta)$ 。

为了给出 $ZZ(m, \eta)$ 分布的数字特征,记 $Z(\alpha) = \int_1^{+\infty} (\ln x)^{\alpha} e^{-x} dx, \alpha > 0$ 。

性质1 若 $T \sim ZZ(m, \eta)$, 那么有

$$E(T) = \eta \cdot e \cdot Z(m^{-1})$$

$$E(T^2) = \eta^2 e \cdot Z(2m^{-1})$$

证明:

$$\begin{aligned} E(T) &= \int_0^{\infty} t \cdot f(t) dt = \\ &= \int_0^{\infty} t \cdot m \cdot \left(\frac{t}{\eta}\right)^{m-1} \cdot \frac{1}{\eta} \cdot e^{(\frac{t}{\eta})^m} \times \\ &\quad e^{1-e^{(\frac{t}{\eta})^m}} dt \end{aligned}$$

令 $y = \left(\frac{t}{\eta}\right)^m, dy = m \cdot \left(\frac{t}{\eta}\right)^{m-1} \cdot \frac{1}{\eta} dt$, 有

$$\begin{aligned} E(T) &= \int_0^{\infty} \eta \cdot y^{\frac{1}{m}} \cdot e^y \cdot e^{1-e^y} dy = \\ &= \eta \cdot \int_0^{\infty} y^{\frac{1}{m}} \cdot e^y \cdot e^{1-e^y} dy \end{aligned}$$

令 $y = \ln t, t = e^y$, 于是

$$\begin{aligned} E(T) &= \eta \cdot e \cdot \int_1^{\infty} (\ln t)^{\frac{1}{m}} \cdot e^{-t} dt = \\ &= \eta \cdot e \cdot Z(m^{-1}) \end{aligned}$$

同理

$$\begin{aligned} E(T^2) &= \eta^2 \cdot e \cdot \int_1^{\infty} (\ln t)^{\frac{2}{m}} \cdot e^{-t} dt = \\ &= \eta^2 \cdot e \cdot Z(2m^{-1}) \end{aligned}$$

由于函数 $Z(\alpha)$ 非常重要,又没有解析表达式,本文在附录1中通过 matlab 计算出了 $\alpha \in (0, 10)$ 时的数表以便使用。

定义2 若 $T \sim ZZ(m, \eta)$, 那么 $X = \ln T$ 的分布函数为

$$F(x) = 1 - \exp\{1 - \exp(e^{\frac{x-\mu}{\sigma}})\}, x > 0 (\mu \in R,$$

$\sigma > 0)$

其中 $\mu = \ln \eta, \sigma = m^{-1}$, 称 X 服从参数为 μ, σ 的 EZ 分布,记 $X \sim EZ(\mu, \sigma)$ 。而 $EZ(0, 1)$ 分布称之为标准的 EZ 分布,记标准 $EZ(0, 1)$ 分布的均值和方差分别为 γ 和 ξ , 其中 $\gamma = 0.897\ 31, \xi = 1.974\ 84$ 。

性质2 若 $T \sim EZ(\mu, \sigma)$, 那么有

$$E(T) = \mu + \gamma\sigma$$

$$Var(T) = \gamma\sigma^2$$

证明: 设 $X \sim EZ(0, 1)$, 则 $T = \sigma X + \mu$

$$E(T) = E(\sigma X + \mu) = \sigma E(X) + \mu = \mu + \gamma\sigma$$

$$var(T) = var(\sigma X + \mu) = \sigma^2 var(X) = \gamma\sigma^2$$

例1中的数据在 $ZZ(m, \eta)$ 分布下的检验结果如下表3,其结果与实际比较相符,由此可见,用此分布描述这些实际问题是非常适合的,因此有必要进行下一步的研究。

表3 某产品在 ZZ 分布下检验结果

Tab.3 Test result of the product under ZZ distribution

参数	$\hat{m} = 8.9, \hat{\eta} = 18.02$
经验极小 χ^2 值	10.425
结论	不拒绝
平均寿命 $\hat{\theta}$ (经验极小方法)	$\hat{\theta} = 16.65$
平均寿命 $\hat{\theta}$ (极大似然方法)	$\hat{\theta} = 19.83$
可靠寿命 (经验极小方法)	$t_{0.8} = 14.97, t_{0.9} = 13.54, t_{0.95} = 12.26$

2 定数截尾下 ZZ(m, η) 分布的参数最佳线性无偏估计

设 $t_1 \leq t_2 \leq \dots \leq t_r$ 为来自 $ZZ(m, \eta)$ 分布的容量为 n 的定数截尾样本。

由于

$$F(t) = 1 - \exp\{1 - e^{(\frac{t}{\eta})^m}\}$$

整理得

$$\ln \ln(1 - \ln(1 - F(t_i))) = m \ln t_i - m \ln \eta \quad (i = 1, 2, \dots, r) \quad (1)$$

因为 $F(t_1) \leq F(t_2) \leq \dots \leq F(t_r)$ 是来自均匀分布 $U(0, 1)$ 的容量为 n 的前 r 个次序统计量,记

$$T = (T_1, T_2, \dots, T_r)'$$

其中 $T_i = \ln \ln(1 - \ln(1 - F(t_i)))$, ($i = 1, 2, \dots, r$)

设 $ET = a = (a_1, a_2, \dots, a_r)'$, $\Sigma = E(T - a)(T - a)'$

$-a)$, 那么 a, Σ 都是已知的。那么 (1) 式可表示为:

$$\ln t_i = \ln \eta + \frac{1}{m} a_i + \frac{1}{m} \delta_i \quad (i = 1, 2, \dots, r) \quad (2)$$

其中 $\delta = (\delta_1, \delta_2, \dots, \delta_r)'$, 并且 $E(\delta) = 0, \text{cov}(\delta) = \Sigma$

令

$$Y = (Y_1, Y_2, \dots, Y_r)' = (\ln t_1, \ln t_2, \dots, \ln t_r)'$$

$$X = \begin{pmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_r \end{pmatrix}$$

$$\beta = (\beta_0, \beta_1)'$$

其中 $\beta_0 = \ln \eta, \beta_1 = \frac{1}{m}, \varepsilon = \beta_1 \delta$ 。那么式 (2) 为式 (3)

$$\left. \begin{aligned} Y &= X\beta + \varepsilon \\ \text{cov}(\varepsilon) &= \beta_1^2 \Sigma \end{aligned} \right\} \quad (3)$$

由此可得到参数最小二乘估计为

$$\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Y$$

当 n, r 给定时, 矩阵 X 与 Σ 都是已知的, 所以为了便于使用, 将上式简化为

$$\left\{ \begin{aligned} \hat{\beta}_0 &= \sum_{i=1}^r A_0(n, r, i) \ln t_i \\ \hat{\beta}_1 &= \sum_{i=1}^r A_1(n, r, i) \ln t_i \end{aligned} \right.$$

其中 $A_0(n, r, i), A_1(n, r, i)$ 的值在 $n \leq 20$

由此得到 $ZZ(m, \eta)$ 分布两个参数的估计为 $\hat{\eta} = e^{\hat{\beta}_0}, \hat{m} = \hat{\beta}_1^{-1}$ 。

3 定数截尾下 $ZZ(m, \eta)$ 分布的参数简单线性无偏估计

设 $t_1 \leq t_2 \leq \dots \leq t_r$ 为来自 $ZZ(m, \eta)$ 分布的容量为 n 的定数截尾样本。那么 $x_1 = \ln t_1 \leq x_2 = \ln t_2 \leq \dots \leq x_r = \ln t_r$ 便是来自 $EZ(\mu, \sigma)$ 分布的容量为 n 的定数截尾样本, 其中 $\mu = \ln \eta, \sigma = m^{-1}$ 。

按照 Max Engelhardt, Lee J. Bain^[2] 茆诗松, 王玲玲^[3] 的方法, 首先选定一个 $s (1 \leq s \leq r)$, 然后构造 σ 形式如下的无偏估计

$$\hat{\sigma} = \frac{1}{nk} \sum_{i=1}^r |x_i - x_s| = \frac{1}{nk} \left[(2s - r - 1) x_s - \sum_{i=1}^{s-1} x_i + \sum_{i=s+1}^r x_i \right] \quad (4)$$

为了便于计算 $\hat{\sigma}$ 的均值与方差, 把式 (4) 改

写为

$$\begin{aligned} \hat{\sigma} &= \frac{\sigma}{nk} \sum_{i=1}^r \left| \frac{x_i - \mu}{\sigma} - \frac{x_s - \mu}{\sigma} \right| = \\ &= \frac{\sigma}{nk} \sum_{i=1}^r |y_i - y_s| = \\ &= \frac{\sigma}{nk} \left[(2s - r - 1) y_s - \sum_{i=1}^{s-1} y_i + \sum_{i=s+1}^r y_i \right] \end{aligned} \quad (5)$$

这里 $y_i = (x_i - \mu) / \sigma (i = 1, 2, \dots, r)$ 就是来自标准 $EZ(0, 1)$ 分布容量为 n 的定数截尾样本。

记 $Y = (y_1, y_2, \dots, y_r)'$, $EY = a = (a_1, a_2, \dots, a_r)'$, $\text{cov}(Y) = \Sigma = (\sigma_{ij})_{r \times r}$ 。因 $y_i (i = 1, 2, \dots, r)$ 的分布完全已知, 所以 a, Σ 都是可确定的, 令

$$b_s = (-1, -1, \dots, -1, 2s - r - 1, 1, 1, \dots, 1)'$$

由于 $E(\hat{\sigma}) = \sigma$, 所以容易得到式 (4) 中的 k 满足

$$k = \frac{1}{n} \left[(2s - r - 1) a_s - \sum_{i=1}^{s-1} a_i + \sum_{i=s+1}^r a_i \right] = \frac{1}{n} b_s' a$$

式 (5) 也可表示为 $\hat{\sigma} = \frac{\sigma}{nk} \left[(2s - r - 1) y_s - \sum_{i=1}^{s-1} y_i + \sum_{i=s+1}^r y_i \right] = \frac{\sigma}{nk} b_s' Y$, 所以有

$$\text{var}(\hat{\sigma}) = \frac{\sigma^2}{n^2 k^2} \text{var}(b_s' Y) = \frac{b_s' \Sigma b_s}{(b_s' a)^2} \sigma^2$$

s 的确定应使得 $\frac{b_s' \Sigma b_s}{(b_s' a)^2}$ 达到最小。此时 k 只与 n, r 有关, 不妨记为 $k = k_{n,r}$ 。这样就得到了 σ 的简单线性无偏估计, 且方差最小。 μ 的无偏估计自然取为

$$\hat{\mu} = x_s - \hat{\sigma} E(y_s) = x_s - a_s \hat{\sigma}$$

所以 $ZZ(m, \eta)$ 分布的参数估计为 $\hat{\eta} = e^{\hat{\mu}}, \hat{m} = \hat{\sigma}^{-1}$ 。

4 结 论

本文针对可靠性统计中常见的四种寿命分布使用时具有局限性的问题, 提出了新型 ZZ 分布刻画一类元件存储寿命的理论和方法, 得到了预期的效果, 在一定程度上丰富了寿命分布的研究内容。在此基础上给出了 ZZ 分布参数的简单线性无偏估计与最佳线性无偏估计。

参 考 文 献:

- [1] LEE J. BAIN. Inferences Based on Censored Sampling From the Weibull or Extreme-Value distribution [J]. Technometrics, 1972, 14(3): 693.
- [2] MAX ENGELHARDT, LEE J. BAIN. Some Complete and Cen-

- sored Sampling Results for the Weibull or Extreme-Value Distribution [J]. *Technometrics*, 1973, 15(3):541.
- [5] SZAJNOWSKI W J. Estimators of Log-Normal Distribution Parameters [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1977, AES-13(5):533.
- [6] CLIFFORD COHEN A, BETTY JONES WHITTEN. Estimation in the Three-Parameter Log-normal Distribution [J]. *Journal of the American Statistical Association*, 1980, 75(370):399.
- [7] FARNUM N R, BOOTH P. Uniqueness of Maximum Likelihood Estimators for the 2-parameter Weibull Distribution [J]. *IEEE Transactions on Reliability*, 1997, 46(4):523.
- [8] ZHENG R. Parameter Estimation of Three-parameter Weibull Distribution and Its Application in Reliability Analysis [J]. *Zhendong Yu Chongji/Journal of Vibration & Shock*, 2015, 34(5):78.
- [9] DAVIES I J. Unbiased Estimation of the Weibull Scale Parameter Using Linear Least Squares Analysis [J]. *Journal of the European Ceramic Society*, 2017, 37(8):2973.
- [10] WILLIAM SHELNUTT J, ALBERT H. MOORE, LEON HARTER H. Linear Estimation of the Scale Parameter of the First Asymptotic Distribution of Extreme Values [J]. *IEEE Transactions on Reliability*, 1973, 22(5):259.
- [11] K. MURALIDHARAN, P. LATHIKA. Analysis of Instantaneous and Early Failures in Weibull Distribution [J]. *Metrika*, 2006, 64(3):305.
- [12] WAYNE NELSON. Accelerated Life Testing-Step-Stress Models and Data Analyses [J]. *IEEE Transactions on Reliability*, 1980, 29(2):103.
- [13] BALAKRISHNAN N, XIE QIHAO, KUNDU D. Exact Inference for a Simple Step-stress-model from the Exponential Distribution Under Time Constraint [J]. *Annals of the Institute of Statistical Mathematics*, 2009, 61(1):251.
- [14] XIONG C. Inferences on a Simple Step-stress Model with Type-II Censored Exponential Data [J]. *IEEE Transactions on Reliability*, 1998, 47(2):142.
- [15] THOMAN D R, BAIN L J, ANTLE C E. Inferences on the Parameters of the Weibull Distribution [J]. *Technometrics*, 1969, 11:445.
- [16] MANN N R, FERTIG K W. Tables for Obtaining Confidence Bounds and Tolerance Bounds Based on Best Linear Invariant Estimates of Parameters of the Extreme-value Distribution [J]. *Technometrics*, 1973, 15:87.
- [17] BAIN L J, MAX ENGELHARDT. Simple Approximate Distribution Results for Confidence and Tolerance Limits for the Weibull Distribution Based on Maximum Likelihood Estimators [J]. *Technometrics*, 1981, 23(1):15.
- [18] LAWLESS J F. Conditional Versus Unconditional Confidence Intervals for the Parameters of the Weibull Distribution [J]. *Technometrics*, 1975, 68:665.
- [19] MANN N R, FERTIG K W. Simplified Efficient Point and Interval Estimators for Weibull Parameters [J]. *Technometrics*, 1975, 17:361.
- [20] 张尧庭, 刘鸿翔. 可靠性统计分析中“倒挂”数据的处理方法 [J]. *应用数学*, 1991, 4(4):23.
- ZHANG Yaoting, LIU Hongxiang. Disordered Data Analysis in the Reliability Statistics [J]. *Mathematica Applicata*, 1991, 4(4):23.
- [21] 杨振海, 安保社. 基于成败型不完全数据的参数估计 [J]. *应用概率统计*, 1994, 10(2):142.
- YANG Zhenhai, AN Baoshe. Parameter Estimation Based On Binary Data [J]. *Chinese Journal of Applied Probability and Statistics*, 1994, 10(2):142.
- [22] 茆诗松, 王玲玲. 可靠性统计 [M]. 上海: 华东师范大学出版社, 1989:141.

(编辑: 王 萍)