

文章编号: 1000-5013(2010)01-0109-04

# 定时截尾缺失数据下指数分布的统计推断

田 霆, 陈祥钟, 黄春棋, 邱志平

(华侨大学 数学科学学院, 福建 泉州 362021)

**摘要:** 当寿命分布为指数分布  $\exp(m)$  时, 通过大量的 Monte-Carlo 数值模拟试验, 得出枢轴量  $\frac{m}{m}$  的近似点分布为对数正态分布, 并给出  $m$  点近似估计及近似区间估计. 结果表明, 在缺失数据数目不太大时, 参数估计的精度还是令人满意的.

**关键词:** 指数分布; 定时截尾; 数据缺失; 枢轴量; 对数正态分布

**中图分类号:** O 213.2

**文献标识码:** A

在可靠性寿命试验中, 指数分布是最常用的寿命分布之一, 其分布函数为

$$F(t; m) = 1 - \exp(-mt), \quad t > 0. \quad (1)$$

式(1)中:  $m > 0$  为失效率,  $= 1/m$  是平均寿命. 由于试验设备、观测手段等方面的困难, 使得试验中的某一些样本寿命数据丢失. 由于只知道其前后的寿命数据, 因此对不完全数据的处理是统计分析的一个重要领域. 文[1]给出了  $\frac{m}{m}$  的枢轴量的理论证明. 文[2]给出了定数截尾数据缺失场合下指数分布参数的 Bayes 估计. 文[3]给出了定时截尾缺失数据下指数分布的部分统计推断. 文[4]给出了定数截尾缺失数据下, Weibull 分布参数的点估计及区间估计, 并通过大量的 Monte-Carlo 模拟说明, 在样本比较大且缺失数较小的情况下, 点估计的精确度是令人满意的. 文[5]给出了平均寿命  $1/m$  的 Bayes 估计、近似极大似然估计和无偏估计. 本文在文[3]基础上, 给出定时截尾缺失数据指数分布参数  $m$  的近似点估计及近似区间估计.

## 1 参数近似的点估计及区间估计

设产品的寿命服从指数分布(1), 有  $n$  个产品同时参加定时截尾试验, 到 ( $t$  是预先给定的正数) 时刻停止. 设在  $t$  时刻以前有  $r$  个产品失效, 记相应的失效时间为  $t_1, t_2, \dots, t_r$ , 总试验时间为

$$T = \sum_{j=1}^r t_j + (n - r) \cdot t.$$

若由于某种原因造成数据丢失, 不妨设剩下的数据为

$$0 < t_{r_1+1} < \dots < t_{r_1+s_1} < t_{r_2+1} < \dots < t_{r_2+s_2} < \dots < t_{r_k+1} < \dots < t_{r_k+s_k},$$

而其样本为

$$\{r_1 + 1, \dots, r_1 + s_1; r_2 + 1, \dots, r_2 + s_2; \dots; r_k + 1, \dots, r_k + s_k\} \subset \{1, 2, \dots, r\}. \quad (2)$$

可记

$$\begin{aligned} S &= \sum_{i=1}^k s_i, & s(t) &= \sum_{i=1}^k \sum_{j=1}^{s_i} t_{r_i+j}, \\ m_i &= r_{i+1} - r_i - s_i, & t_{r_{k+1}} &= t, \\ r_0 &= s_0 = t_0 = 0, & r_{k+1} &= r. \end{aligned}$$

收稿日期: 2008-09-09

通信作者: 田 霆(1972-), 男, 讲师, 主要从事产品可靠性的研究. E-mail: tianting1972928@sohu.com.

基金项目: 福建省自然科学基金资助项目(Z0511027)

文[1]给出样本(2)的似然函数  $L(m)$ . 令  $\frac{\partial \ln L(m)}{\partial m} = 0$ , 则有

$$\frac{s}{m} - \sum_{i=0}^k m_i \frac{t_{r_i+s_i} \exp[-mt_{r_i+s_i}] - t_{r_{i+1}+1} \exp[-mt_{r_{i+1}+1}]}{\exp[-mt_{r_i+s_i}] - \exp[-mt_{r_{i+1}+1}]} - s(t) - (n-r) = 0, \quad (3)$$

并证明出此方程解  $\bar{m}$  的存在及唯一性. 有

$$\frac{r}{s(t) + (n-r) + \sum_{i=0}^k m_i t_{r_{i+1}+1}} < \bar{m} < \frac{r}{s(t) + (n-r) + \sum_{i=0}^k m_i t_{r_i+s_i}}. \quad (4)$$

令  $m^* = \frac{\bar{m}}{m}$ , 易知  $m^*$  的分布与参数无关.

## 2 $m^*$ 的分布

性质 1  $m^*$  的分布与参数无关<sup>[3]</sup>.

证明 由于

$$\frac{s}{\bar{m}} - \sum_{i=0}^k m_i \frac{t_{r_i+s_i} \exp[-\bar{m} t_{r_i+s_i}] - t_{r_{i+1}+1} \exp[-\bar{m} t_{r_{i+1}+1}]}{\exp[-\bar{m} t_{r_i+s_i}] - \exp[-\bar{m} t_{r_{i+1}+1}]} - s(t) - (n-r) = 0,$$

令  $m^* = \frac{\bar{m}}{m}$ , 可得

$$\frac{s}{m^*} - \frac{1}{m} - \sum_{i=0}^k m_i \frac{t_{r_i+s_i} \exp[-m^* \cdot m t_{r_i+s_i}] - t_{r_{i+1}+1} \exp[-m^* \cdot m t_{r_{i+1}+1}]}{\exp[-m^* \cdot m t_{r_i+s_i}] - \exp[-m^* \cdot m t_{r_{i+1}+1}]} - s(t) - (n-r) = 0.$$

即

$$\frac{s}{m^*} - \sum_{i=0}^k m_i \frac{m t_{r_i+s_i} \exp[-m^* \cdot m t_{r_i+s_i}] - m t_{r_{i+1}+1} \exp[-m^* \cdot m t_{r_{i+1}+1}]}{\exp[-m^* \cdot m t_{r_i+s_i}] - \exp[-m^* \cdot m t_{r_{i+1}+1}]} - m s(t) - (n-r) m = 0.$$

由式(3)可知,  $m^*$  为母体  $m=1$  的标准指数分布的参数的估计量. 所以,  $m^*$  的分布与参数  $m$  无关,  $\frac{\bar{m}}{m}$  为一枢轴量. 由此, 用 Monte-Carlo 数值模拟的方法, 易得  $m^*$  的期望值, 记  $K_{r,n} = E(m^*)$ .

于是, 利用逆矩估计思想方法<sup>[5]</sup>, 可以得到参数  $m$  的点估计  $\hat{m}_1$ . 即有  $\frac{\bar{m}}{m} = K_{r,n}$ .

对 3 种情形进行 2 000 次 Monte-Carlo 模拟, 结果如表 1 所示. 表 1 中  $\hat{m}$ ,  $\hat{m}_1$  分别为  $\hat{m}$  的均值和均方误差. 情形 1 的参数:  $n=20, m=1.000\ 0, \alpha=1.2, k=$

6; 情形 2 的参数:  $n=30, m=0.416\ 7, \alpha=2.8, k=8$ ; 情形 3 的参数:  $n=50, m=0.250\ 0, \alpha=4.5, k=11$ . 由结果可看出, 其点估计的精度是令人满意的.

用 Monte-Carlo 模拟  $m^*$  的分位数, 结果如表 2 所示.

如果取置信水平为  $1 - \alpha$ ,  $m^*$  的分布的双侧分位数为  $\bar{m}_{1-\alpha/2}, \bar{m}_{\alpha/2}$ ,

则参数的区间估计为  $[\bar{m}_{\alpha/2}, \bar{m}_{1-\alpha/2}]$ . 而当  $m^*$  的置信水平为 90 % 时, 情形 1~3 的置信区间上限分别为 0.713 4, 0.316 9, 0.206 0, 而其置信区间下限分别为 1.425 3, 0.522 7, 0.296 5.

表 2 用 Monte-Carlo 模拟  $m^*$  的分位数

Tab.2  $m^*$  's quantiles by Monte-Carlo simulation

情形	0.01	0.05	0.10	0.25	0.75	0.90	0.95	0.99
1	0.511 9	0.639 2	0.701 6	0.836 5	1.199 8	1.401 8	1.531 4	1.859 5
2	0.653 7	0.746 2	0.797 6	0.890 4	1.156 8	1.315 7	1.404 0	1.605 3
3	0.744 7	0.809 4	0.843 2	0.907 0	1.107 2	1.213 5	1.264 2	1.391 6

类似上述的 Monte-Carlo 模拟中, 其他情形也做了大量的模拟试验. 结果表明,  $m^*$  近似服从对数

正态分布  $\ln(\mu, \sigma^2, \tau)$ . 其中:  $\mu$  为尺度参数,  $\sigma^2$  为形状参数,  $\tau$  为门限参数. 由于篇幅有限, 只给出部分模拟结果, 如图 1 所示.

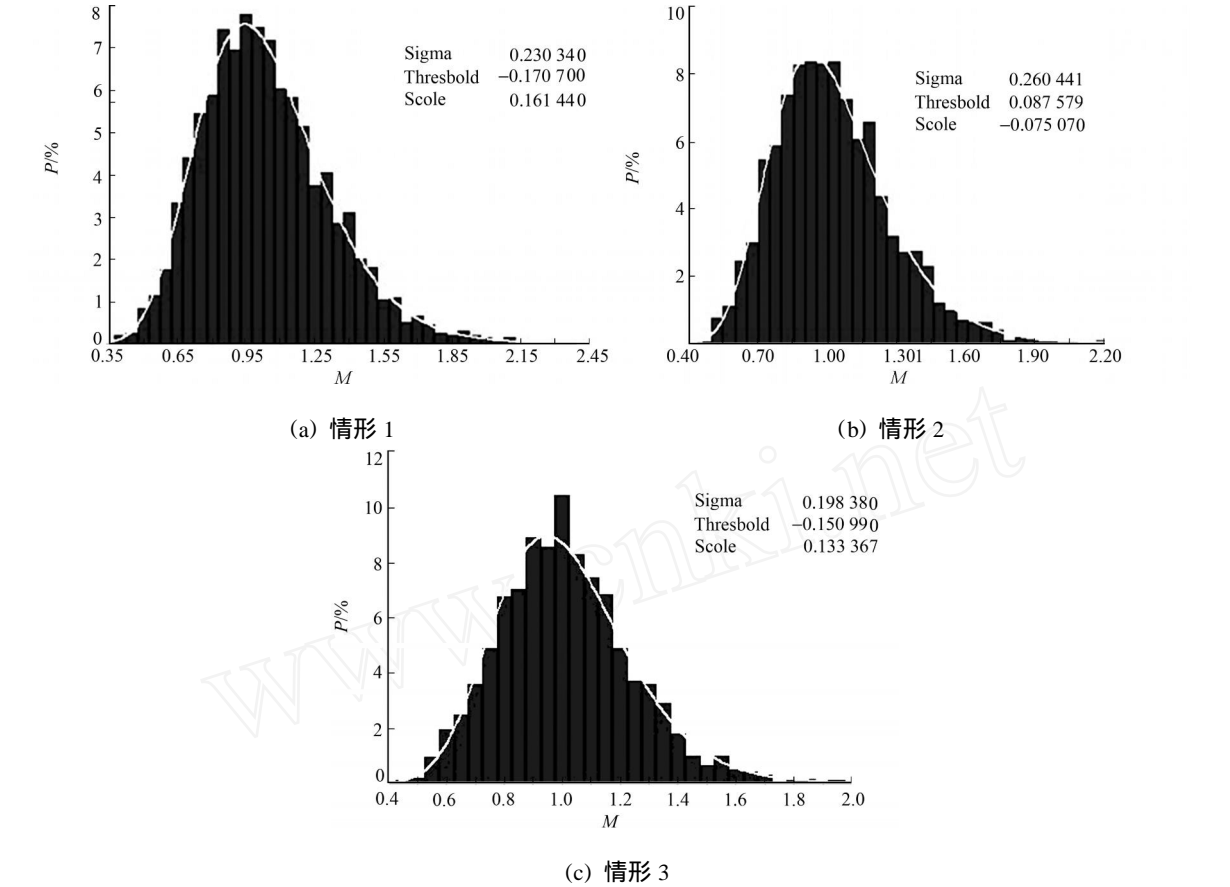


图 1  $m^*$  的部分模拟结果

Fig. 1  $m^*$  approximate lognormal distribution

3 模拟比较试验

如果随机变量  $X \simeq \ln(\mu, \sigma^2, \tau)$ , 则有

$$E(X) = \mu + \frac{\sigma^2}{2},$$

$$D(X) = \exp[2\mu + 2\sigma^2] - \exp[2\mu + \sigma^2].$$

于是, 利用逆矩估计思想方法<sup>[5]</sup>, 得参数  $m$  的点估计  $\hat{m}_2$ , 有

$$\frac{\bar{m}}{m} = \mu + \frac{\sigma^2}{2}.$$

为此, 对上述 3 种情形进行了 2 000 次 Monte-Carlo 模拟, 结果如表 3 所示. 从表 3 中可看出, 其点估计的精度是比较令人满意的. 与之前的结果比较可知, 两者的结果十分接近.

于是, 利用对数正态分布的分位表及 Monte-Carlo 模拟, 得出  $m^*$  的分位数, 如表 4 所示.

表 3 近似结果

Tab. 3 Approximate result

情形	$(\mu, \sigma^2, \tau)$	$\hat{m}_2$	$\hat{m}_1$
1	(0.161 4, 0.026 1, -0.170 8)	1.036 0	0.281 7
2	(-0.118 61, 0.049 1, 0.126 5)	0.432 0	0.117 3
3	(-0.410 4, 0.210 6, 0.337 1)	0.253 8	0.063 5

表 4 用对数正态分布及 Monte-Carlo 模拟  $m^*$  的分位数

Tab. 4  $m^*$ 's quantiles by Monte-Carlo simulation and approximate lognormal distribution

情形	0.01	0.05	0.95	0.99
1	0.516 9	0.633 8	1.545 8	1.837 6
2	0.656 8	0.743 3	1.405 4	1.614 0
3	0.743 5	0.806 2	1.275 1	1.419 9

4 Monte-Carlo 模拟实验

关于参数估计的精度,仅利用式(3),对真值  $\theta = 1.0, 2.4, 3.0, 4.0 (m = 1/ \theta)$ ,  $n = 20$  进行了 2 000 次 Monte-Carlo 模拟实验. 部分模拟结果,如表 5 所示. 表 5 中:  $m$  为包含  $m$  的比例.

从表 5 可知,当  $n$  固定时,随  $k$  的增大,精度愈高;而当  $k$  很小时,即缺失数太大,参数估计误差偏大. 因此,应尽量避免数据缺失. 总的来说,在缺失数据数目不太大时,参数估计的精度还是令人满意的.

表 5 模拟结果

Tab. 5 Parameter estimation

序号	模拟情形	$m$	$m/ \%$	序号	模拟情形	$m$	$m/ \%$
1	$\theta = 1.0, \theta = 1.2, k = 5$	1.000 0	42	11	$\theta = 3.0, \theta = 3.5, k = 5$	0.333 3	43
2	$\theta = 1.0, \theta = 1.2, k = 6$	1.000 0	63	12	$\theta = 3.0, \theta = 3.5, k = 6$	0.333 3	64
3	$\theta = 1.0, \theta = 1.2, k = 7$	1.000 0	70	13	$\theta = 3.0, \theta = 3.5, k = 7$	0.333 3	71
4	$\theta = 1.0, \theta = 1.2, k = 8$	1.000 0	75	14	$\theta = 3.0, \theta = 3.5, k = 8$	0.333 3	74
5	$\theta = 1.0, \theta = 1.2, k = 9$	1.000 0	82	15	$\theta = 3.0, \theta = 3.5, k = 9$	0.333 3	81
6	$\theta = 2.4, \theta = 2.8, k = 5$	0.416 7	43	16	$\theta = 4.0, \theta = 4.5, k = 5$	0.250 0	51
7	$\theta = 2.4, \theta = 2.8, k = 6$	0.416 7	68	17	$\theta = 4.0, \theta = 4.5, k = 6$	0.250 0	60
8	$\theta = 2.4, \theta = 2.8, k = 7$	0.416 7	74	18	$\theta = 4.0, \theta = 4.5, k = 7$	0.250 0	72
9	$\theta = 2.4, \theta = 2.8, k = 8$	0.416 7	81	19	$\theta = 4.0, \theta = 4.5, k = 8$	0.250 0	81
10	$\theta = 2.4, \theta = 2.8, k = 9$	0.416 7	87	20	$\theta = 4.0, \theta = 4.5, k = 9$	0.250 0	87

参考文献:

[1] BALASUBRANMANIAN K, BLAKRISHNA N. Estimation for one- and two-parameter exponential distributions under multiple type- censoring[J]. Statistical Papers, 1992, 33: 203-216.  
[2] 王乃生, 王玲玲. 定数截尾数据缺失场合下指数分布参数的 Bayes 估计[J]. 应用概率统计, 2001, 8(3): 229-235.  
[3] 田 霖, 刘次华. 定时截尾缺失数据下指数分布的统计推断[J]. 华侨大学学报: 自然科学版, 2006, 27(1): 20-23.  
[4] 徐晓岭. 定数截尾缺失数据下 Weibull 分布的统计推断[J]. 概率论与数理统计, 1997, 12(2): 363-370.  
[5] 王炳兴. Weibull 分布的统计推断[J]. 应用概率统计, 1992, 8(4): 357-364.

Statistical Inference for the Exponential Distribution  
under Multiply Type- Censoring

TIAN Ting, CHEN Xiang-zhong,  
HUANG Chun-qi, QIU Zhi-pin

(School of Mathematical Sciences, Huaqiao University, Quanzhou 362021, China)

**Abstract:** It is prove that the approximate distribution of pivot  $\frac{\bar{m}}{m}$  is lognormal distribution when product life time follows exponential distribution by means of Monte-Carlo numeral simulation. And the approximate point estimation and the approximate interval estimation of  $m$  are given. It is shown by result in the paper that precision of the parameter estimation is satisfied when the missing number is not too big.

**Keywords:** exponential distribution; multiply type- censoring; pivot; lognormal distribution

(责任编辑: 陈志贤 英文审校: 张金顺, 黄心中)