文章编号: 1000-5013(2012)02-0232-03

定数截尾缺失数据下 Weibull 分布的形状参数近似估计

田霆[1], 刘次华[2]

(1. 华侨大学 数学科学学院, 福建 泉州 362021;
2. 华中科技大学 数学与统计学院, 湖北 武汉 430074)

摘要: 研究定数截尾寿命试验数据缺失场合下,Weibull 分布中形状参数的点估计问题. 利用次序统计量的方法给出形状参数的近似估计,并通过大量的 Monte-Carlo 数值模拟试验,验证方法的可行性.

关键词: Weibull 分布; 定数截尾; 数据缺失; 形状参数; 极小值分布; 次序统计量

中图分类号: O 213.2 **文献标志码:** A

1 预备知识

在用统计方法处理实际问题时,常会遇到数据缺失问题,如产品寿命试验中由于试验设备、观测手段或有其他方面的困难,造成某些试验数据丢失或未观测到等现象. 因此,对不完全数据的处理是统计分析的一个重要领域. 在可靠性寿命试验中,Weibull 分布是最常用的寿命分布之一,其分布函数为

$$F(t) = 1 - \exp\{-(\frac{t}{\eta})^m\}, \qquad t \geqslant 0. \tag{1}$$

式(1)中: m 为形状参数, η 为刻度参数.

由于寿命 T 服从分布式(1),易知 $W = (\frac{T}{\eta})^m$ 服从标准指数分布,而 $\ln W$ 服从标准极小值分布. 现假定有 n 个产品同时参加定数截尾试验,试验进行到 r(r 是预先给定的正数)个产品失效时停止. 设相应的失效时间为

$$t_1 \leqslant t_2 \leqslant \cdots \leqslant t_r, \tag{2}$$

若由于某种原因造成数据丢失,不妨设剩下的数据为

$$0 < t_{r_1} \leqslant t_{r_2} \leqslant \cdots \leqslant t_{r_k}, \tag{3}$$

其中: $\{r_1, r_2, \cdots r_k\} \subset \{1, 2, \cdots, r\}$, 则

$$\ln(\frac{t_{r_1}}{\eta})^m \leqslant \ln(\frac{t_{r_2}}{\eta})^m \leqslant \cdots \leqslant \ln(\frac{t_{r_k}}{\eta})^m \tag{4}$$

服从标准极小值分布.

对于试验数据(3)的统计分析,已有一些文献作了相关的研究[1-3]. 文献[1]给出了 Weibull 分布及极值分布的参数的点估计及区间估计;文献[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;文献[3]给出了定时截尾缺失数据下指数分布的统计推断. 文中试图利用次序统计量的方法[4]给出形状参数的一种近似点估计,通过大量的 Nonte-Carlo 模拟试验说明所给方法的可行性.

2 形状参数 m 的一种点估计

Murthy 等[4]从 n 个失效样本中两个次序统计量出发,提出了 $m^* = 1/m$ 的无偏估计. 根据次序统

收稿日期: 2011-09-19

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)

计量的选择准则，在估计量的方差达到最小的条件下，渐近率为 70%左右，在 Weibull 分布的刻度参数未知时，进行了 m 的假设检验.

若次序统计量观察值为 t_m 与 t_l，$1\leqslant l<k<n$，$m^*=\dfrac{1}{m}$，记

$$Y=\frac{\ln t_k-\ln t_l}{m^*}=\ln(\frac{t_k}{\eta})^m-\ln(\frac{t_l}{\eta})^m. \tag{5}$$

易知 Y 的分布与未知参数 m 无关，其为一枢轴量.

3 Y 的近似分布

引理 1 若 $x_1\leqslant x_2\leqslant\cdots\leqslant x_r$ 为标准极小值分布的前 r 个次序统计量，$x_{i+1}-x_i$ 近似服从参数为 i 的指数分布，$i=1,2,\cdots,r-1$，即 $\lim\limits_{n\to\infty}P\{x_{i+1}-x_i<x\}=1-\exp(\mathrm{i}x)$[5]，而

$$Y=\ln(\frac{t_k}{\eta})^m-\ln(\frac{t_l}{\eta})^m=\ln(\frac{t_k}{\eta})^m-\ln(\frac{t_{k-1}}{\eta})^m+\ln(\frac{t_{k-1}}{\eta})^m-\ln(\frac{t_{k-2}}{\eta})^m+\cdots+\ln(\frac{t_{l+1}}{\eta})^m-\ln(\frac{t_l}{\eta})^m. \tag{6}$$

故由式(6)及引理 1 的结果可知，Y 的近似分布是由 $k-l$ 个参数不同的指数分布的和的分布.

4 m^* 的无偏估计的构造

令 $\hat{m}^*=\dfrac{\ln t_k-\ln t_l}{E(Y)}$，显然 $\hat{m}^*$ 是 m^* 的无偏估计，由 Y 的构造不难推导出

$$E(Y)=2A\sum_{i=0}^{l-1}\sum_{j=0}^{k-l-1}C_{i,j}\ \frac{1}{4\rho b}\ln\frac{\rho+1}{\rho},$$

$$E(Y^2)=2A\sum_{i=0}^{l-1}\sum_{j=0}^{k-l-1}C_{i,j}\ \frac{1}{4\rho b^2}g(\rho).$$

其中：$A=\dfrac{n!}{(l-1)!(k-l-1)!(n-m)!}$；$g(x)=\sum\limits_{t=1}^{\infty}(-\dfrac{1}{t^2})(-\dfrac{1}{x})^t$；$\rho=\dfrac{n+j-k+1}{b}$；$b=k-l+i-j$；$C_{i,j}=(-1)^{i+j}\dbinom{l-1}{i}\dbinom{k-j-1}{j}$.

为了充分利用缺失后信息及以上结果，从数据(3)出发，导出 m 的估计. 对于数据(3)，记

$$Y_1=\frac{\ln t_{r_k}-\ln t_{r_{k-1}}}{m^*},\ Y_2=\frac{\ln t_{r_{k-1}}-\ln t_{r_{k-2}}}{m^*},\cdots,Y_{k-1}=\frac{\ln t_{r_2}-\ln t_{r_1}}{m^*}.$$

由式(6)可知，Y_1，Y_2，Y_{k-1} 的联合分布是完全确定的. 令

$$\hat{m}^*=\frac{\ln t_{r_k}-\ln t_{r_{k-1}}}{E(Y_1)}\cdot\frac{r_k-r_{k-1}}{r_k-r_1}+\frac{\ln t_{r_{k-1}}-\ln t_{r_{k-2}}}{E(Y_2)}\cdot\frac{r_{k-1}-r_{k-2}}{r_k-r_1}+\cdots+\frac{\ln t_{r_2}-\ln t_{r_1}}{E(Y_{k-1})}\cdot\frac{r_2-r_1}{r_k-r_1},$$

记 $\alpha_i=\dfrac{r_{k-i+1}-r_{k-i}}{E(Y_i)(r_k-r_1)}$，$i=1,2,\cdots,k-1$；$\beta_1=\alpha_1$，；$\beta_j=\alpha_j-\alpha_{j-1}$，$j=2,3,\cdots,k-1$；$\beta_k=-\alpha_{k-1}$，则有

$$\hat{m}^*=\sum_{j=1}^{k}\beta_j\ln t_{r_j}. \tag{7}$$

从而 $\hat{m}_1=\dfrac{1}{\sum\limits_{j=1}^{k}\beta_j\ln t_{r_j}}$. 显然，$E(\hat{m}^*)=m^*=\dfrac{1}{m}$，$\hat{m}^*$ 为 $m^*=\dfrac{1}{m}$的线性无偏估计.

5 Monte-Carlo 模拟实验

为确定参数估计的精度，设 $\eta=1.0$，利用式(6)对 m 的真值 $m=1.0,0.5,0.25$ 进行 2 000 次的 Monte-Carlo 模拟实验，部分模拟结果如表 1 所示. 从表 1 中可以看到：当 n 固定时，随着 k 的增大，精度愈高；而当 k 很小即缺失数太大时，参数估计误差偏大，故应尽量避免数据缺失 . 总的来说，在缺失数据

数目不太大的情况下,参数估计的精度还是令人满意的.

表1 Monte-Carlo模拟实验部分结果

Tab. 1 Part of the Monte-Carlo simulation results

m,n,r	缺失情况	$\hat{m}$	偏差
$m=1.00,n=20,r=15$	$k=15$,无缺失	0.978	−0.022
	$k=13,t_1,t_2$ 缺失	0.950	−0.050
	$k=13,t_1,t_7$ 缺失	0.946	−0.054
	$k=10,t_1,t_2,t_6,t_8,t_{10}$ 缺失	0.905	−0.095
$m=0.50,n=25,r=20$	$k=20$,无缺失	0.490	−0.010
	$k=17,t_1,t_2,t_9$ 缺失	0.512	0.012
	$k=17,t_3,t_5,t_{12}$ 缺失	0.482	−0.018
	$k=15,t_2,t_4,t_5,t_8,t_{14}$ 缺失	0.436	−0.064
$m=0.25,n=30,r=24$	$k=24$,无缺失	0.247	−0.003
	$k=22,t_4,t_{20}$ 缺失	0.254	0.004
	$k=22,t_3,t_{18}$ 缺失	0.245	−0.005
	$k=20,t_1,t_5,t_7,t_{17},t_{22}$ 缺失	0.239	−0.011

参考文献:

[1] FEI H,KONG F,TANG Y. Estimation for two-parameter Weibull distribution and extreme-value distribution under multiple type-Ⅱ censoring[J]. Comm Statist Theory and Meth,1995,24:2087-2104.

[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] 田霆,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] MURTHY V K,SWARTZ G B. Estimation of Weibull parameters from two-order statisticals[R]. Ohio:Wright-Patterson Air Force Base,1974.

[5] 曹晋华,程侃.可靠性数字引论[M].北京:科学出版社,1986.

Approximate Estimation for the Shape Parameter of Weibull Distribution under Multiply Type-Ⅱ Censoring

TIAN Ting[1], LIU Ci-hua[2]

(1. School of Mathematical Sciences, HuaqiaoUniversity, Quanzhou 362021, China;
2. School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: In this paper, we study the point estimation for the shape parameter of Weibull distribution under multiply type-Ⅱ censoring. The approximate point estimation of unknown parameter is obtained by using the method of order statistics. By the Monte-Carlo simulation, the precision for the parameter estimation under a large amount of samples is satisfied.

Keywords: Weibull distribution; multiply type-Ⅱ censoring; missing date; shape parameter; minimum value distribution; order statistics

(**责任编辑**:陈志贤 **英文审校**:张金顺,黄心中)