文章编号:1000-5862(2014)03-0226-03

威布尔分布中尺度参数的最短区间估计

徐美萍[1],于 健[2],王 若[1]
(1.北京工商大学理学院,北京 100048;2.北京建筑大学理学院,北京 100044)

摘要:把威布尔分布中尺度参数最短置信区间的求解问题转化为非线性方程组的求解问题,并通过1个实例和数值计算对最短置信区间与常用置信区间进行长度比较,说明研究小样本情形时威布尔分布中尺度参数最短置信区间的重要性与必要性.

关键词:威布尔分布;指数分布;伽马分布;最短置信区间
中图分类号:O 212.8 **文献标志码**:A

0 引言

与未知参数的点估计相比,区间估计有着明显的优势.因为它不仅给出了参数真值所在的范围,还给出了该范围包含真值的可信程度.在给定置信水平$1-\alpha$的前提下,置信区间的长度越短越能体现估计的精度,尤其是对小样本的情形.由于当枢轴量的概率密度函数非对称时,按常规方法两侧各取$\alpha/2$所得的置信区间(下文称其为常用置信区间)一般不是最短的,在某些情形下会导致大的估计误差,因此需要考虑最短置信区间的求解问题.一些常用分布中未知参数的最短置信区间的估计已有许多文献研究,其中与尺度参数相关的文献有:文献[1]讨论了指数分布中尺度参数最短置信区间的存在唯一性及估计问题;文献[2]对于伽马分布[3](本文记为$\Gamma(\beta,\theta)$,其概率密度为$f(x;\beta,\theta)=[\theta\Gamma(\beta)]^{-1}\cdot(x/\theta)^{\beta-1}\exp(-x/\theta)$,$x>0$,参数$\beta>0$,$\theta>0$)给出了当形状参数$\beta$已知时,尺度参数$\theta$的最短区间估计,并与常用置信区间做了比较,说明在小样本情形下研究的必要性.指数分布作为伽马分布的一种特殊情形(形状参数$\beta=1$)在文献[2]也做了详尽的备注.最近文献[4-5]又给出了包含Levy分布作为特例的逆Gamma分布和Laplace分布以及广义误差分布中尺度参数的最短区间估计.

在可靠性工程中,材料或零件的寿命分布或给定寿命下的疲劳强度是经常需要关注的问题.由于威布尔分布[6-8]是根据最弱环节模型或串联模型得到的,能够充分反映材料缺陷和应力集中源对材料疲劳寿命的影响,而且具有递增的失效率,所以被广泛应用于机电类产品的磨损累计失效试验的数据处理.如本文将讨论的两参数威布尔分布就主要用于滚动轴承的寿命试验[9-10]以及高应力水平下的材料疲劳试验[11-12].

在实践中,考虑到工程的时间和成本问题,通常是小样本试验,所以本文将讨论威布尔分布中尺度参数的最短区间估计问题,给出其求解的条件,并把它们与常用置信区间做比较,说明研究小样本情形时的威布尔分布中尺度参数的最短置信区间是重要且必要的.

1 威布尔分布中尺度参数的最短区间估计

设总体X服从形状参数$c>0$,尺度参数$\theta>0$的威布尔分布,其概率密度为

$$f(x;c,\theta)=c\theta^{-c}x^{c-1}\exp\{-(x/\theta)^{c}\},x>0,\quad(1)$$

其中的形状参数c决定了威布尔分布的尾重,它的值越小,相应分布的尾越重.当$c=1$时,该威布尔分布是尺度参数为θ的指数分布.当$c>1$时,威布尔分布的尾轻于指数分布,而当$c<1$时则具有比指数分布更厚的尾.

下面考察当c已知时未知参数θ的估计问题.

假设$(X_1,X_2,\cdots,X_n)$是来自(1)式的1个简单随机样本,则有对数似然函数

$$\ln[L(\theta)]\propto-nc\ln(\theta)-t/\theta^{c},\theta>0,$$

收稿日期:2014-02-10
基金项目:北京市属高等学校人才强教计划(201106206)资助项目.
作者简介:徐美萍(1971-),女,山西太原人,副教授,博士,主要从事统计推断方面的研究.

其中 $t = \sum_{i=1}^{n} X_i^{\,c}$. 由此可得 θ 的最大似然估计为 $\hat{\theta} = (t/n)^{1/c}$. 通过计算可得 $X_i^{\,c} \overset{i.i.d.}{\sim} \Gamma(1,\theta^c)$，$i = 1,2,\cdots,n$，从而 $t \sim \Gamma(n,\theta^c)$ [13]，由此可得

$$E(\hat{\theta}) = \theta(1/n)^{1/c}\Gamma(n + c^{-1})/\Gamma(n).$$

接下来，考虑 $\hat{\theta}$ 和 θ 的函数 $Z = 2n\hat{\theta}^c/\theta^c = 2t/\theta^c$. 由 $t \sim \Gamma(n,\theta^c)$ 可得 $Z \sim \Gamma(n,2)$，所以可用 Z 作为枢轴量来构造 θ 的置信区间. 事实上，对于给定的置信水平 $1 - \alpha$，令

$$P(a < Z < b) = P(\hat{\theta}(2n/b)^{1/c} < \theta < \hat{\theta}(2n/a)^{1/c}) = 1 - \alpha.$$

把 $\hat{\theta} = (t/n)^{1/c}$ 和 $t = \sum_{i=1}^{n} X_i^{\,c}$ 代入上式，得 θ 的置信水平为 $1 - \alpha$ 的置信区间的广义形式如下:

$$(\hat{\theta}(2n/b)^{1/c},\hat{\theta}(2n/a)^{1/c}) = \left(\left(2\sum_{i=1}^{n} X_i^{\,c}/b\right)^{1/c},\left(2\sum_{i=1}^{n} X_i^{\,c}/a\right)^{1/c}\right),$$

其平均长度为

$$L(a,b) = (2n)^{1/c}(a^{-1/c} - b^{-1/c})E(\hat{\theta}) = 2^{1/c}(a^{-1/c} - b^{-1/c})\theta\Gamma(n + c^{-1})/\Gamma(n).$$

从而求解 θ 的最短置信区间的问题转化为条件极值问题: 求解 a^*，b^* 使得带约束条件

$$\int_a^b f_Z(z)\,\mathrm{d}z = 1 - \alpha \tag{2}$$

的 2 元函数 $a^{-1/c} - b^{-1/c}$ 取最小值，其中 $f_Z(z)$ 是 Z 的概率密度函数.

类似于文献 [2] 中定理 2 的讨论，求解 θ 的最短置信区间的问题可以进一步转化为方程组(3) 的求解问题.

定理 1　θ 的最短区间估计等价于问题: 求解 a^*，b^*，使方程组

$$\begin{cases} a^{n+1/c}\mathrm{e}^{-a/2} = b^{n+1/c}\mathrm{e}^{-b/2} \\ \int_a^b f_Z(z)\,\mathrm{d}z = 1 - \alpha \end{cases} \tag{3}$$

成立，其中 $f_Z(z)$ 是 Z 的概率密度函数.

2　与常用置信区间的比较

接下来将对给定的置信水平 $1 - \alpha = 0.95$ 和样本容量 n 比较尺度参数的最短置信区间与常用置信区间来说明研究威布尔分布中尺度参数最短置信区间的必要性.

首先，对给定的形状参数 c，其尺度参数 θ 的常用置信区间的长度为 $(2\sum_{i=1}^{n} X_i^{\,c})^{1/c} \cdot [\hat{a}(n)^{-1/c} - \hat{b}(n)^{-1/c}]$，其中 $\hat{a}(n) = \Gamma_{1-\alpha/2}(n,2)$，$\hat{b}(n) = \Gamma_{\alpha/2}(n,2)$ 分别为分布 $\Gamma(n,2)$ 的下、上 $\alpha/2$ 分位点. 记 $L_m(n) = n[a^*(n)^{-1/c} - b^*(n)^{-1/c}]$，$L(n) = n[\hat{a}(n)^{-1/c} - \hat{b}(n)^{-1/c}]$，两者的相对差异记为 $d(n) = [L(n) - L_m(n)]/L_m(n)$，其中 $a^*(n)$，$b^*(n)$ 使用 Matlab 软件中的非线性规划命令 fmincon 对(2) 式求解得到. 当然也可以取 $\hat{a}(n)$，$\hat{b}(n)$ 为初值，通过对(3) 式迭代求解得到 [14].

表 1 中列出了当 $c = 1/2$ 时威布尔分布中尺度参数 θ 的最短置信区间与常用置信区间的比较. 为方便对比，在表 2 中列出了指数分布(当 $c = 1$ 时的威布尔分布) 情形的相应值.

表 1　当 $c = 1/2$ 时，θ 的最短置信区间与常用置信区间的比较

n	3	5	7	9	10	11	13	15	17
$L_m(n)$	1.119 1	0.318 4	0.158 1	0.097 9	0.080 7	0.068 0	0.050 8	0.039 8	0.032 3
$L(n)$	1.945 1	0.462 3	0.210 7	0.123 8	0.100 2	0.083 1	0.060 4	0.046 4	0.037 0
$d(n)$/%	73.81	45.19	33.27	26.46	24.16	22.21	18.90	16.58	14.55

表 2　当 $c = 1$ 时，θ 的最短置信区间与常用置信区间的比较

n	2	3	4	5	6	7	8	9	10
$L_m(n)$	2.728 9	1.716 2	1.319 4	1.103 4	0.965 0	0.867 4	0.794 0	0.736 3	0.689 6
$L(n)$	3.949 2	2.216 9	1.607 0	1.295 8	1.105 4	0.975 6	0.880 8	0.808 0	0.750 0
$d(n)$/%	44.72	29.17	21.80	17.44	14.55	12.47	10.93	9.74	8.76

从表 1 和表 2 可以看出，随着样本容量 n 的增大，威布尔分布中尺度参数的常用置信区间与最短置信区间的长度差异逐渐减小. 但是当样本容量 n 不够大时，如当 $n \leqslant 10$ 时两者的相对差异还是比较大的，与 $c = 1$ 相比，当 $c = 1/2$ 时这种差异更为显著，即使样本容量 n 取到 17，两者的相对差异仍为 14.55%，这是相当大的.

接下来，看 1 个实例，数据来自国内某核电站 1998 年 5 月 5 日至 1999 年 12 月 31 日的设备维修记录，共采集到 RRI 泵泄漏失效记录的 7 个截尾样

本(单位:h): 2 185.37, 3 421.75, 5 586.10, 8 109.71,8 289.89,12 153.77,14 363.27,采用3参数威布尔分布拟合RRI泵无故障运行时间,计算得到形状参数 c、尺度参数 θ 和位置参数的最大似然估计分别为1.4,1 228.5,115.0[15].从而,上述数据与115.0的差值可以看作是来自2参数威布尔分布(本文所考虑)的1个样本,经过数值计算可得尺度参数 θ 的最短置信区间与常用置信区间分别为[559.1,2 021.4]和[737.6,2 207.5],两者的相对长度差异为7.91%.由此可见这种误差是不容忽视的.

通过上述分析可以看出,研究威布尔分布中尺度参数的最短置信区间,尤其是小样本情形,是非常必要且有意义的.

3 参考文献

[1] 蒋福坤,刘玉春.指数分布参数的最短区间估计[J].数理统计与管理,2004,23(3):43-45.

[2] 袁长迎,徐明民.伽马分布参数的最短置信区间[J].数理统计与管理,2006,25(4):435-437.

[3] 丁新月,徐美萍.正态和对数正态分布中参数的损失和风险函数的Bayes推断[J].江西师范大学学报:自然科学版,2014,38(1):70-73.

[4] 徐美萍,于健,马玉兰.几种厚尾分布尺度参数的最短区间估计[J].统计与决策,2014(10):74-77.

[5] 徐美萍,于健.广义误差分布中尺度参数的最短区间估计[J].统计与决策,2014(12):76-80.

[6] 姜培华,范国良.威布尔分布尺度参数的最短区间估计[J].统计与决策,2013(17):81-83.

[7] 张晓勤,王煜,卢殿军.混合指数威布尔分布的参数估计[J].河南大学学报:自然科学版,2012,42(3):230-233.

[8] 吴启光,李国英,顾岚,等.双参数威布尔分布下可靠性抽样检验[J].应用概率统计,2004,20(3):270-286.

[9] 毕然,武东.Weibull分布恒定应力加速寿命试验的Bayes估计[J].纯粹数学与应用数学,2014,30(1):93-99.

[10] 游达章,钟毓宁,唐小琦.基于双截尾数据消除法的自适应加速寿命试验数据统计分析[J].中国机械工程,2013,24(5):614-616.

[11] 徐微,胡伟明,孙鹏.基于两参数威布尔分布的设备可靠性预测研究[J].中国工程机械学报,2013,11(2):112-116.

[12] 石立华,张祥,马如坡,等.基于Weibull分布的小子样电磁脉冲效应试验与评估[J].解放军理工大学学报:自然科学版,2013,14(4):441-447.

[13] 范金城,吴可法.统计推断导引[M].北京:科学出版社,2001.

[14] 章栋恩,马玉兰,徐美萍,等.MATLAB数学实验[M].北京:电子工业出版社,2008.

[15] 陈豫龙,陈晓明,赵炳全.威布尔分布在核电站可靠性数据库中的应用[J].原子能科学技术,2003,37(4):349-352.

The Estimation of Shortest Confidence Interval for Scale Parameter of Weibull Distribution

XU Mei-ping[1], YU Jian[2], WANG Ruo[1]

(1. School of Science, Beijing Technology and Business University, Beijing 100048, China;

2. School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)

Abstract: The solving problem of the shortest confidence interval of scale parameter and its equivalent proposition for Weibull distribution are transferred to a system of nonlinear equations. Then the importance and necessary of this research are showed by a real example and numerical calculations through comparing lengths of these shortest confidence intervals with the corresponding ones of usually used confidence intervals.

Key words: Weibull distribution; exponential distribution; Gamma distribution; the shortest confidence interval

(责任编辑:曾剑锋)