威布尔分布的极大似然估计的精度分析

李 进 黄 敏 赵 宇

(北京航空航天大学 工程系统工程系, 北京 100083)

摘 要: 威布尔分布参数估计的极大似然方法是一种常用的方法,在故障数不小于10的情况下推荐使用.但在工程中,由于产品的可靠性高,或者是样本量少,试验的故障数常常是小于10,在这种情况下需要明确所得的评估结果的精度是否满足要求.运用区间估计的思想,提出了一种解决上述问题的评价和判断的方法,并应用此方法对完全样本情况下,形状参数的极大似然估计量的精度进行了讨论.工程上,可以依据文中提供的结论定量分析威布尔分布形状参数极大似然估计量的精度.

关 键 词: 威布尔分布; 极大似然估计; 分析

中图分类号: V 241.01

文献标识码: A **文 章 编 号**: 1001-5965(2006)08-0930-03

Analysis of precision for maximum likelihood estimation in the Weibull distribution

Li Jin Huang Min ZhaoYu

(Dept. of System Engineering of Engineering Technology, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Maximum likelihood estimation(MLE) in the Weibull distribution of parameters is considered an effective method, which is recommended in the case that the number of failure is more than 10 or equal 10. In practice, the number of failure is often less than 10 because of the high reliability or the fewness of sample. It is wanted to be known whether the MLE can be used in the case above; the number of failure, which makes the estimator have a given precision, will be known, sometimes, before the experimentation applying. So a method of analysis and estimation, which resolves that problems, was given with the idea of interval estimation. Then the precision of the estimator of shape parameter was discussed by this way under complete sample. In the practice projects, the precision of maximum likelihood estimation of Weibull shape parameter is quantitatively provided by the conclusion.

Key words: Weibull distribution; maximum likelihood estimation; analysis

1 问题的提出

威布尔分布是一种常用的寿命分布.极大似然估计法(MLE, Maximum Likelihood Estimation)是一种对威布尔分布参数估计普遍采用的方法,一般在 $n \geqslant r \geqslant 10$($n$ 为样本量,r 为失效数)的情况下推荐使用.但是在许多的工程领域中,由于产品的可靠性比较高,或者样本量比较少,要达到故障数大于或等于10这样的条件是比较困难的.因此,工程中常常在 $r \leqslant 10$ 情况下,使用极大似然估计.在通常情况下,不清楚估计出参数的精度,也同样不清楚估计出的参数是否满足要求.

实际上,无论是故障数大于或者小于10,参数的估计值都不能肯定等于参数的真实值,通常的做法是进行区间估计,这样在给定置信度的情况下,就可以得到置信区间.但是,要是想要了解估计出的参数在给定真值区间的概率,或者在试验还没有进行时想要了解使参数的估计值具有一定的精度的条件下所需要进行的试验的故障数,

收稿日期: 2005-11-29

作者简介: 李 进(1980-), 男, 北京人,硕士生, jinlee@sohu.com.

区间估计显然不能解决这样的问题.

本文为解决此类的问题,提出一种方法. 同时使用此方法对于完全样本的情况下,威布尔分布的形状参数 β 的极大似然估计量 $\hat{\beta}$,进行了估计精度的讨论,同时得到了满足相应条件的最小样本失效数.

2　极大似然估计量的评价方法

两参数威布尔分布,其密度函数为

$$f(t)=\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]\qquad t\geqslant 0 \tag{1}$$

其中 $\beta>0$ 和 $\alpha>0$ 是参数,分别称为威布尔分布的形状参数和尺度参数.

对于威布尔分布的两个参数的极大似然估计量是由下面的方程组解出的[1].

$$\left.\begin{aligned}&\frac{\sum_{i=1}^{r}t_i^{\beta}\ln t_i+(n-r)\cdot t_Z^{\beta}\ln t_Z}{\sum_{i=1}^{r}t_i^{\beta}+(n-r)\cdot t_Z^{\beta}}-\\&\frac{1}{r}\sum_{i=1}^{r}\ln t_i-\frac{1}{\beta}=0\\&\alpha=\left(\frac{\sum_{i=1}^{r}t_i^{\beta}+(n-r)\cdot t_Z^{\beta}}{r}\right)^{\frac{1}{\beta}}\end{aligned}\right\} \tag{2}$$

其中,t_i 为第 i 个失效时间;t_Z 为截尾时间.

对于方程(2)解出的极大似然的估计值 $\hat{\beta}$ 和 $\hat{\alpha}$,理论上符合某种分布. 下面就以 β 为例来说明这个方法. 首先,构造枢轴量:$Z=\hat{\beta}/\beta$(证明见文献[2]),所以 Z 的分布和 β 的取值是无关的. 这样,就可以得到 Z 的分布 F_Z(得到 Z 的分布 F_Z 的方法有很多,例如用公式推导或是用模拟的方法). 然后,利用这个分布,计算概率 p. 计算的步骤是:①取区间 $[z_1,z_2]$;②计算 $p=F_Z(z_2)-F_Z(z_1)$. 这样,得到的概率 p 可以描述为:在给定 n 和 r 的情况下,$\hat{\beta}$ 在真值 β 的区间 $[z_1\beta,z_2\beta]$ 的概率为 p. 由于 Z 为枢轴量,所以这个概率 p 只与 n 和 r 有关. 这样,就可以用这个概率 p 来评价和判断 n 和 r 在各种情况下 $\hat{\beta}$ 的优劣了.

3　完全样本情况下的 MLE 的评价

下面就利用上面的方法对完全样本情况下的威布尔分布形状参数的 $\hat{\beta}$ 进行评价.

3.1　枢轴量 $Z=\hat{\beta}/\beta$ 的分布的研究

极值分布是与威布尔分布等价的,可以代替其的一种更为方便的形式,其密度函数为

$$f(x,u,b)=\frac{1}{b}\exp[(x-u)/b]\exp\{-\exp[(x-u)/b]\}$$
$$-\infty<x<\infty \tag{3}$$

其中 $u(-\infty<u<\infty)$ 和 $b(b>0)$ 是参数. 若 T 服从威布尔分布,则随机变量 $X=\log T$ 就服从极值分布,其中 $u=\log\alpha$,$b=\beta^{-1}$.

对于极值分布函数的参数 b,$Z'=\hat{b}/b$ 同样为枢轴量. 对于 Z' 的分布,Lawless(1972)已经给出了其分布密度,但是同时也指出对于 Z' 的分布密度的积分不可能有解析的表达式. 所以只能找出一个近似的分布,对于 Z' 的分布,从经验中发展起来有如下形式的近似:

$$gZ'\sim\chi^2_{(h)} \tag{4}$$

其中 g 和 h 只与 n 和 r 有关. Harter 和 Moore(1968),McCool(1975)和 Lawless 和 Mann(1976)分别得到结果组成了表,给出了各种 (r,n) 组合的 h 的值,g 由 $g=h+2$ 给出. 表 1 只列出了对于本文有用的值[2].

表 1　近似所用的 h 值

r/n	n				
	5	10	20	40	60
1.0	12.9	29.3	62.4	128.2	194.8

对于表 1 中没有给出的 h 值,可以通过线性插值获得(本段论述的各种近似和取值以及具体的证明和解释见文献[2]).

这样就可以得到 F_Z 的近似的计算公式:

$$F_Z(x)=P(Z\leqslant x)=P\left(\chi^2_{(h)}\geqslant\frac{g}{x}\right)=$$
$$1-P\left(\chi^2_{(h)}\leqslant\frac{g}{x}\right) \tag{5}$$

式(5)是非常便于计算的.

3.2　对于 r 的各种情况的计算

关于下面的讨论,假定讨论的 r 的范围是 $[3,40]$. 这是由于,对于 $r<3$ 的情况,认为不适合用极大似然法进行估计;同时,对于 $r>40$ 的情况,视为其估计的精度已经足够精确.

利用第 2 节中所给的方法,同时结合第 3.1 节中所推倒出的 F_Z 就可以很容易地求出概率 p. 具体的计算步骤如下:

1)给定 $n=r=n_0$;

2)给出 $Z=\dfrac{\hat{\beta}}{\beta}$ 的区间 $[z_1,z_2]$;

3)利用线性内插值,对于给定的 n 求出相应的 h 和 g;

4）分别计算概率 $P\left(\chi^2_{(h)} \leqslant \frac{g}{z_1}\right)$ 和 $P\left(\chi^2_{(h)} \leqslant \frac{g}{z_2}\right)$；

5）则 $p = F_Z(z_2) - F_Z(z_1) = P\left(\chi^2_{(h)} \leqslant \frac{g}{z_1}\right) - P\left(\chi^2_{(h)} \leqslant \frac{g}{z_2}\right)$.

为了工程应用的方便，在构造表2的过程中给定 $p = p_0$，然后对于 $Z = \frac{\hat{\beta}}{\beta}$ 不同的区间，求出满足 $p(r) \geqslant p_0$ 最小的 r 值.

表2 满足条件最小 r 值表

Z	p/%				
	50	60	70	80	90
[0.85,1.15]	15	22	32	–	–
[0.8,1.2]	9	13	19	29	–
[0.75,1.25]	7	10	14	20	32
[0.7,1.3]	6	7	10	15	24
[0.6,1.4]	4	5	7	10	16
[0.5,1.5]	3	4	6	8	12
[0.4,1.6]	3	4	5	7	10
[0.3,1.7]	3	4	4	6	8
[0.2,1.8]	3	3	4	5	7

注：①表中“–”代表 r 的值大于40；②由于 r 的值最小为3，所以表的左下方都是3；③对于表中没有给出的数据，用二次插值可以近似地得到.

表2中数据的解释和使用的方法：以第2行中的数13为例，13代表故障数为13的完全样本情况下，$\hat{\beta}$ 在真值区间 $[0.8\beta, 1.2\beta]$ 内的概率为60%. 这样在工程中，如果给出形状参数的估计值可以接受的区间，同时给出可以接受的概率，就可以得到要进行的试验在完全样本情况下最少的故障数.

这里还要说明的一点就是这种评价的方法和区间估计的区别：

1）这里只是一种评价和判断的方法，得到的结论只是可以用和不可以用，可以不用知道参数 β 的真值，所以这里考虑的是估计的参数在真值区间的概率；而区间估计所要求的是区间的数值，所以求的是在一定概率下的真值在估计参数的区间；

2）这里所用的方法是给定区间求概率；而区间估计是给定概率求区间.

所以，本文的方法和区间估计可以说是一种理念的两种不同角度的理解和看法.

4 结束语

本文对于威布尔分布的极大似然法的估计量，在样本量和故障数的各种不同情况的精度，给了一种评价和判断的方法. 同时对于完全样本情况下，形状参数的极大似然估计进行了讨论，文中并没有给出好坏的结论. 这是由于在不同的情况下，要求的精度是不同的，所以可以根据不同的情况，依照表2进行选择.

表2中的数据定量地给出了满足一定精度要求的威布尔分布形状参数的极大似然估计值所要求的最少样本量，同时使得在工程中应用极大似然估计威布尔形状参数时对估计值精度有了定量的了解.

对于文中没有讨论的尺度参数和不是完全样本的情况，同样可以按照第2节中所述的方法，根据不同的性质，分别进行分析和处理.

参考文献（References）

[1] 杨华波，张士峰，蔡洪. Weibull分布可靠性分析[J]. 飞行器测控学报，2003，22(4)：37-42
Yang Huabo, Zhang Shifeng, Cai Hong. Reliability analysis of Weibull distribution[J]. Journal of Spacecraft TT&C Technology, 2003, 22(4):37-42(in Chinese)

[2] Lawlee J F. 寿命数据中的统计模型与方法[M]. 北京：中国统计出版社，1998：141-199
Lawlee J F. Statistical models and methods for lifetime data [M]. Beijing: China Statistical Press, 1998: 141-199 (in Chinese)

[3] 叶慈南. 完全样本情形下威布尔分布参数的估计[J]. 应用概率统计，2003，19(3)：259-266
Ye Cinan. Estimator for the parameters of the Weibull distribution based on complete sample [J]. Chinese Journal of Applied Probability and Statistics, 2003, 19(3):259-266(in Chinese)

[4] 肖良华，赵宇，黄敏. 引入位置参数的三参数Weibull过程以及其点估计方法[J]. 北京航空航天大学学报，2004，30(9)：897-900
Xiao Lianghua, Zhao Yu, Huang Min. Three-parameter Weibull process of using position element and point estimation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2004, 30(9):897-900(in Chinese)

（责任编辑：赵海容）