doi:10.12068/j.issn.1005－3026.2025.20240194

基于统计最小差异原理的Weibull分布参数估计方法

谢里阳[1]，朱文慧[2]，吴宁祥[1]，杨小玉[1]
（1.东北大学 机械工程与自动化学院，辽宁 沈阳 110819；2.中国国际工程咨询有限公司 国防业务部，北京 100048）

摘 要：针对Weibull分布的参数估计，构造1个尺度参数的伪估计量，可以通过寻找有关变量的极值点的方法得到参数估计值.该参数估计方法原理是，正确的位置参数和形状参数使得根据各样本值估计出的尺度参数之间的差异最小.本质上，参数估计是基于1组具有不确定性的数据（随机变量样本）反映出的特定规律来提取（总体）信息.然而，这种规律是统计意义上的规律，而不是确定意义上的规律.其在有关函数极值点出现位置方面的表现是，估计量的准确值并非一定出现在确定性意义上的极值点.研究表明，对于上述Weibull分布参数估计问题，准确分布参数所在点与理论上的极值点之间通常存在一定的偏离，在最小值判据中引入1个偏移值（将“一阶导数等于零”修改为“一阶导数等于1个大于零的值”），能够显著提高参数估计的精度和稳健性.大量参数估计案例表明，将偏移值取为0.1，使得根据不同样本得到的真实值为1 000的Weibull分布位置参数估计值的范围从0~1 500大幅度缩小为500~1 550.

关 键 词：Weibull分布；位置参数；参数估计；极值判据；稳健性
中图分类号：TB 114.3 **文献标志码**：A **文章编号**：1005-3026（2025）07-0108-06

Weibull Distribution Parameter Estimation Method Based on Statistical Minimum Diversity Principle

XIE Li-yang[1], *ZHU Wen-hui*[2], *WU Ning-xiang*[1], *YANG Xiao-yu*[1]
（1. School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China; 2. Defense Business Department, China International Engineering Consulting Corporation, Beijing 100048, China. Corresponding author: XIE Li-yang, E-mail: lyxieneu@163.com）

Abstract: For the Weibull distribution parameter estimation, a pseudo-estimator of scale parameters is constructed, and the estimated parameter values can be obtained by finding the extreme point of relevant variables based on the principle that the right location parameter and shape parameter minimize the diversity of the scale parameter estimates associated with individual sample values. Essentially, parameter estimation extracts (overall) information based on specific patterns reflected by a set of data with uncertainty (random variable samples). However, the pattern is statistical in nature rather than deterministic. In terms of the occurrence of extreme points in the related functions, the exact value of the estimater does not necessarily occur at the extreme point in a deterministic extreme point. It is shown that there is typically a deviation between the point where the exact parameter is located and the theoretical extreme point, and the accuracy and robustness of the parameter estimation method can be greatly improved by introducing an offset value in the minimum value criterion (modifying “the first derivative being equal to zero” to “the first derivative being equal to a value greater than zero”). A large number of parameter estimation cases show that the range of the estimated value of the Weibull location parameter (true value is 1 000) is narrowed from 0~1 500 to 500~1 550 by taking an offset value of 0.1.

Key words: Weibull distribution; location parameter; parameter estimation; extreme value criterion; robustness

收稿日期：2024-10-26
基金项目：国家科技重大专项（J2019-IV-0002-0069，J2019-V-0009-0103）.
作者简介：谢里阳（1962—），男，安徽岳西人，东北大学教授，博士生导师.

科学研究和工程实际经常需要根据有限的样本数据进行随机变量概率分布的参数估计.准确性和稳健性一直是参数估计方法所面临的挑战,尤其是在样本量较小的情况下.Weibull分布是迄今为止最广泛用于描述产品寿命概率分布的模型,在小样本场合的应用更是多于其他概率分布模型[1-2].在概率统计方法中,极大似然估计法(maximum likelihood method, MLM)和最小二乘估计法(least squares method, LSM)是应用最多的参数估计方法.为了提高参数估计的精度,研究者还提出了各种各样的修正极大似然估计法[3-8]和加权最小二乘法[9-11].有关文献全面比较了这些方法的准确性、效率和稳健性[12-15].Teimouri等[16]比较了12种Weibull分布的参数估计方法,指出根据同一个样本不同方法估计出的概率分布参数差异很大.由于误差过大且很不稳定,参数估计方法在小样本场合的应用价值受到了质疑[17].

从文献中可以看到,学者从多个方面研究了影响参数估计结果精度的各种因素.然而,尽管极值判据(函数在其局部极值点的一阶导数等于零)在参数估计方法中起着十分重要的作用,但还从未有人探讨过极值判据在参数估计中的应用效果及有关问题.

本文结合作者提出的三参数Weibull分布参数估计原理(统计最小差异原理),提出1种通过搜寻最小极值点进行参数估计的方法.该方法不需要求解方程,不会出现极大似然估计法可能出现无解的问题,特别适合于小样本场合的三参数Weibull分布参数估计问题.针对统计意义上的最小差异点与确定性意义上的差异函数最小值点不一致的现象,以及其导致参数估计结果稳健性差的问题,提出在极值判据中采用1个合理的偏移值抵消样本不确定性对参数估计结果的影响,显著提高了参数估计的精度及稳健性.

1　Weibull分布参数估计的最小差异原理及其统计特性

1.1　Weibull分布尺度参数的伪估计量

服从Weibull分布的随机变量T的概率分布函数$F(t)$为

$$F(t)=1-\mathrm{e}^{-\left|\frac{t-\gamma}{\eta}\right|^{\beta}}. \tag{1}$$

式中:β,η和γ分别表示Weibull分布的形状参数、尺度参数和位置参数;t为样本值.

对概率分布函数进行变换,可以得到1个从样本值t到尺度参数η的映射:

$$\eta=(t-\gamma)/(-\ln(1-F(t)))^{1/\beta}. \tag{2}$$

如果位置参数γ和形状参数β已知,则尺度参数η可以根据该随机变量的1个样本值t_i及其对应的累积概率分布函数$\hat{F}(t_{(i)})$的估计值计算出来.

例如,若有1个样本量为n的随机变量T的样本,用$t_{(i)}$表示从小到大排序的n个样本值(i=1,2,…,n,$t_{(1)}\leqslant t_{(2)}\leqslant,\cdots,\leqslant t_{(n)}$),则可以通过秩估计法估计对应于$t_{(i)}$的累积概率分布函数:

$$\hat{F}(t_{(i)})=\frac{i}{i+(n+1-i)F_{2(n+1-i),2i,0.5}},\ (i=1,2,\cdots,n). \tag{3}$$

式中,$F_{2(n+1-i),2i,0.5}$是自由度为$2(n+1-i)$和$2i$的F分布的中位数.

因此,可以构建1个尺度参数的伪估计量$\hat{\eta}_i$:

$$\hat{\eta}_i=(t_{(i)}-\gamma)/(-\ln(1-\hat{F}(t_{(i)})))^{1/\beta},\ (i=1,2,\cdots,n). \tag{4}$$

将$\hat{\eta}_i$称为伪估计量,是因为它包含除样本值之外的其他参数,即位置参数γ和形状参数β.

图1是基于Weibull分布$W(2,1\,000,1\,000)$不同样本(理想样本和随机样本,样本量为7)的尺度参数估计值.理想样本的7个样本值$t_{(i)}$(i=1,2,…,7)是通过式(1)及相应的累积概率获得的,随机样本值由Monte Carlo采样生成.分别用7个理想样本值估计出的7个尺度参数值均与真实值1 000相同.使用随机样本值估计出的尺度参数值,有些接近真实参数值,而另一些则远离真实值.此外,基于含误差的形状参数和位置参数估计出的7个尺度参数值之间的差异大于基于准确形状参数和位置参数的估计结果之间的差异.

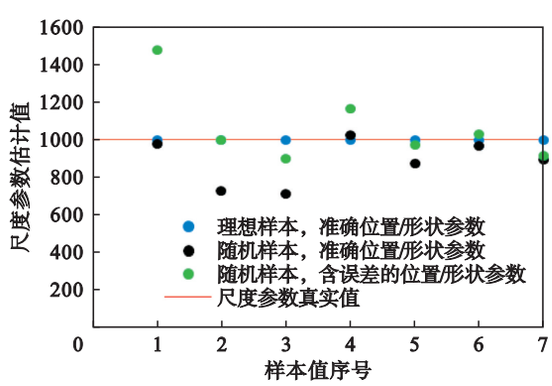

图1　采用不同样本值估计出的尺度参数

Fig. 1　Estimated scale parameters with different sample values

1.2　尺度参数伪估计量的特性

对于1个参数估计问题,所有参数都是未知的.当采用不真实的形状参数和/或位置参数时,

式(2)(从1个样本值t到尺度参数的映射)将产生1个与真实尺度参数不同的尺度参数计算值.基于1个理想的随机变量样本$t_{(1)},t_{(2)},\cdots,t_{(n)}$,在位置参数和形状参数已知的条件下,由式(2)可以准确地估计出尺度参数η.若位置参数γ和形状参数β分别含有误差$\Delta\gamma$和$\Delta\beta$,则由式(2)估计出的含误差的尺度参数为

$$\eta=(t-(\gamma+\Delta\gamma))/(-\ln(1-F(t)))^{1/(\beta+\Delta\beta)}.$$

根据式(1)可知,

$$(-\ln(1-F(t)))^{1/(\beta+\Delta\beta)}=\left(-\ln\left(1-\left(1-\mathrm{e}^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}\right)\right)\right)^{1/(\beta+\Delta\beta)}=$$

$$\left(-\ln\left(\mathrm{e}^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}\right)\right)^{1/(\beta+\Delta\beta)}=\left(\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right)^{1/(\beta+\Delta\beta)}=\left(\frac{t-\gamma}{\eta}\right)^{\beta/(\beta+\Delta\beta)}.$$

由此可知,尺度参数计算值的误差与位置参数误差和形状参数误差之间的关系为

$$\Delta\eta(t)=\frac{t-(\gamma+\Delta\gamma)}{((t-\gamma)/\eta)^{\frac{\beta}{\beta+\Delta\beta}}}-\eta. \tag{5}$$

式中:$\Delta\eta$,$\Delta\gamma$和$\Delta\beta$分别表示尺度参数误差、位置参数误差和形状参数误差.

由式(5)可知,一方面,较大的位置参数误差和形状参数误差会导致较大的尺度参数误差;另一方面,在位置参数误差和形状参数误差一定的条件下,尺度参数的误差大小与样本值t(由$(t,F(t))$定义)的取值有关.也就是说,对于相同的形状参数误差和位置参数误差,不同的$(t,F(t))$使得式(5)产生不同的尺度参数误差.

图2为在不同位置参数误差/形状参数误差条件下,由式(2)计算产生的尺度参数误差,即对于1个Weibull分布W(2.0,1 000,1 000)(表示形状参数为2.0,尺度参数为1 000,位置参数为1 000的Weibull分布,下同),对应于不同的位置参数误差和样本值数值,尺度参数计算值误差与形状参数误差之间的关系.可以看到,位置参数误差和/或形状参数误差增大,不仅会导致尺度参数误差增大,同时还导致基于不同样本值计算出的尺度参数值之间的差异增大.显然,式(2)的特性也是伪估计量(式(4))的特性.

2 Weibull分布参数估计计算

基于伪估计量的上述特性,可以建立相应的判据,并通过"搜索-判断"估计出正确的位置参数γ和形状参数β.也就是说,当有1个样本时,可以基于式(4)估计Weibull分布参数.具体做法如下:尝试一系列可能的位置参数γ和形状参数β值,基于每1对(γ,β)和1个样本值,都能够由式(4)估计出1个尺度参数η;根据n个样本值,可以得到n个尺度参数的估计值$\eta_i(\gamma,\beta)$(i=1,2,…,n);分别由n个样本值估计得到的n个尺度参数值之间的差异最小的位置参数γ和形状参数β即为形状参数和尺度参数的估计值.

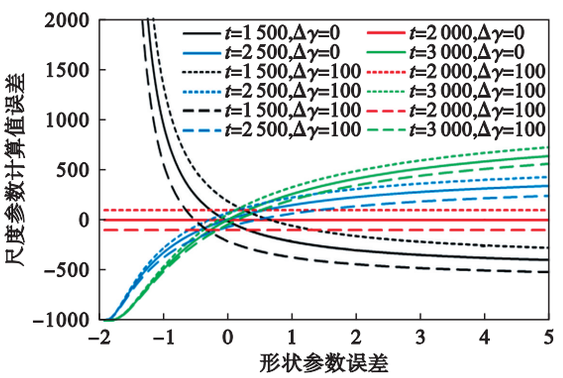

图2 尺度参数计算值误差

Fig. 2 Errors of the calculated scale parameter

基于上述参数估计原理(由不同样本值估计出的尺度参数之间的差异最小——最小差异原理),可以通过一个搜索过程尝试一系列的可能参数值,并最终找到待估计的参数.

首先,设置一系列可能的位置参数γ_j(从最小样本值$t_{(1)}$开始逐步递减),并对每个位置参数值测试一系列可能的形状参数值β_k(例如,0.10,0.11,0.12,直到可能的最大值,例如20).基于每对(γ_j,β_k),根据式(4)可获得与n个样本值对应的n个尺度参数估计值$\hat{\eta}_i(\gamma_j,\beta_k)$($i$=1,2,…,$n$).由上述最小差异原理可知,使得这$n$个尺度参数估计值之间的差异达到最小的参数对$(\hat{\beta},\hat{\gamma})$即为真实的形状参数和位置参数.本文采用$n$个尺度参数估计值的标准差$\sigma_\eta$表征$n$个尺度参数估计值的差异,根据该标准差随采用的位置参数值变化的梯度(标准差对位置参数的一阶导数)判断极值点的位置.

根据1个样本量为n的随机变量样本,进行Weibull分布参数估计的具体流程如下:

首先,给定1个位置参数$\gamma_1=t_{(1)}$,它与每个不同的形状参数β_k(k=1,2,…)共同使用,都可由式(4)产生n个不同的尺度参数估计值$\widehat{\eta_i}(\gamma_1,\beta_k)$($i$=1,2,…,$n$),这$n$个尺度参数估计值的标准差为$\sigma_\eta(\gamma_1,\beta_k)$.图3为基于来自同1个Weibull分布的样本量为7的3个不同样本进行参数估计过程中计算所得的$\sigma_\eta(\gamma_j,\beta_k)$-$\beta_k$关系.文中样本A和样本

B 是采用 Monte Carlo 抽样方法从 Weibull 分布 $W(2.0,1\,000,1\,000)$ 中随机抽取的,理想样本是根据 Weibull 分布的累积分布函数反算出来的 7 个样本值,分别对应于累积分布概率 0.094, 0.228, 0.364, 0.500, 0.636, 0.772 和 0.906(这 7 个概率值是应用式(3)在 n=7 的条件下计算出来的 $\hat{F}(t_{(1)}),\hat{F}(t_{(2)}),\cdots,\hat{F}(t_{(7)})$. 显然,对应于不同的 γ_j 值,不同的形状参数 β_k 将产生不同的 7 个尺度参数估计值的标准差 $\sigma_\eta(\gamma_j,\beta_k)$,并有 1 个相应的条件最小标准差 $\sigma_{\eta,\min}(\gamma_j)$(以位置参数 γ_j 为条件,对应于 1 个特定的形状参数),如图 3 所示(其中,L 表示位置参数 γ_j 的取值). 不同的位置参数 γ_j 对应不同的 $\sigma_{\eta,\min}(\gamma_j)$,使得 $\sigma_{\eta,\min}(\gamma_j)$ 达到最小(用 $\sigma_{\eta,\min}$ 表示)的位置参数 $\hat{\gamma}$ 和与之一起实现该标准差的最小值的形状参数 $\hat{\beta}$ 即为估计出的位置参数和形状参数. 在此基础上,可以根据式(4),由任意 1 个样本值 $t_{(i)}$ 估计出 1 个尺度参数.

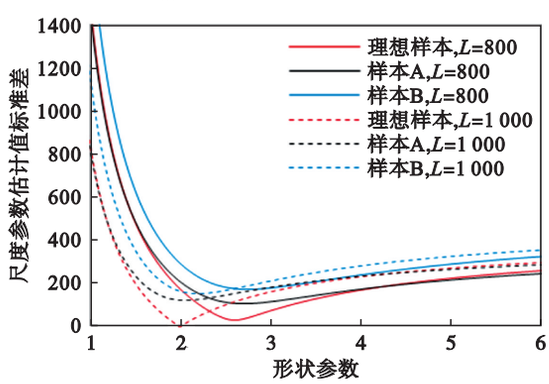

图 3　$\sigma_\eta(\gamma_j,\beta_k)$-$\beta_k$ 关系曲线

Fig. 3　$\sigma_\eta(\gamma_j,\beta_k)$-$\beta_k$ plots

与样本值 $t_{(i)}$ 对应的累积概率分布 $\hat{F}(t_{(i)})$ 可以由秩估计得到. 然而,由于样本具有不确定性,与累积分布概率 $\hat{F}(t_{(i)})$ 精确对应的样本值未必等于 $t_{(i)}$,甚至有很大的差异. 因此,将不同的样本值代入式(4)可以得到不同的尺度参数估计值(如图 1 所示). 为了减小样本不确定性的影响,本文将由 n 个样本值估计出的 n 个尺度参数的平均值作为最终估计出的 Weibull 分布的尺度参数:

$$\hat{\eta}=\sum_{i=1}^{n}\left\{\frac{t_{(i)}-\hat{\gamma}}{(-\ln(1-F(t_{(i)})))^{1/\hat{\beta}}}\right\}/n. \tag{6}$$

3　误差分析与参数估计方法验证

对于 1 个连续函数 $y=f(x)$,其一阶导数 $\mathrm{d}y/\mathrm{d}x$ 在极值点的值为零,这是常用的极值判据. 图 4 是在参数估计过程中产生的、n 个尺度参数估计值标准差的条件最小值 $\sigma_{\eta,\min}(\gamma_j)$ 对位置参数 γ_j 的导数(尺度参数估计值标准差的条件最小值 $\sigma_{\eta,\min}(\gamma_j)$ 的梯度)与所采用的位置参数之间的关系曲线. 尺度参数估计值标准差的条件最小值 $\sigma_{\eta,\min}(\gamma_j)$ 的梯度 $\nabla\gamma$ 以离散的形式定义为

$$\nabla\gamma=\frac{\sigma_{\eta,\min}(\gamma+\nabla\gamma)-\sigma_{\eta,\min}(\gamma)}{\nabla\gamma}. \tag{7}$$

理论上(根据最小差异原理),尺度参数估计值标准差的条件最小值 $\sigma_{\eta,\min}(\gamma_j)$ 的梯度在位置参数的准确值处等于零. 然而,这些曲线表明,由于样本不确定性效应,该梯度在位置参数的真实值处未必等于零(理想样本除外). 基于某一样本得到的尺度参数估计值标准差的条件最小值 $\sigma_{\eta,\min}(\gamma_j)$ 的梯度在真实位置参数点的值可能大于零,也可能小于零. 由图 4 可见,如果简单地应用极值判据,则对于某些样本,位置参数估计值的误差非常大.

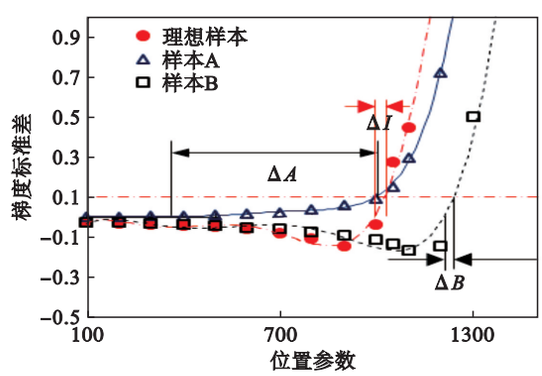

图 4　$\nabla(\gamma)$-γ 关系曲线及最小值判据补偿效果

Fig. 4　$\nabla(\gamma)$-γ plots and effect of minimum value criterion offset

由图 4 可知,这种基于最小值判据的参数估计方法,只有理想样本才能得到准确的位置参数估计值,而随机样本得到的位置参数估计值结果误差很大;如果使用 1 个大于零的数,例如 0.1 替换最小值判据中的零,即用梯度 $\nabla\gamma$=0.1 的点定位位置参数的估计值,则在整体上可以显著提高参数估计结果的精度和稳健性. 这样,由理想样本、样本 A 和样本 B 得到的位置参数估计值分别为 1 025, 1 000 和 1 240. 当直接应用最小值判据时,3 个样本位置参数的估计值分别为 990, 350 和 1 210. 与样本 A 估计的位置参数误差从−65.0% 降低到 0.0% 的有益效果(参数估计值调整幅度 ΔA)相比,由样本 B 估计出的位置参数的误差从 21.0% 增加到 24.0%(参数估计值调整幅度 ΔB),由理想样本估计出的位置参数误差从−1.0% 增加到 2.5%(参数估计值调整幅度 ΔI)的不利影响显

然小得多.

从总体上看,在最小值判据中使用1个大于零(例如0.1)的偏移值,对于提高参数估计方法(即本文采用的最小差异法)的精度和稳健性效果显著.本文偏移值0.1是根据120个参数估计案例得到的经验值,更适当的偏移值的取法需要进一步研究.由于样本存在不确定性,需要对极值判据进行偏置,而上文中所述参数估计方法依据的"最小差异原理"也具有统计不确定性,因而应称之为"统计最小差异原理".下面采用更多的随机样本分析该参数估计方法的误差特性,研究减小误差的途径.

图5为基于Weibull分布$W(2.0,1\,000,1\,000)$的30个随机样本(样本量为7)分别进行参数估计的过程中产生的30条尺度参数估计值标准差梯度曲线,即$\nabla\gamma'$–γ'曲线.通过这些曲线可以清楚地看到,直接应用最小值判据,基于不同样本估计出的位置参数差异很大(这是目前所有参数估计方法共同存在的问题),范围在0~1 475之间(位置参数的真实值为1 000),尽管这些样本都是从同一个Weibull分布母体中随机选出的.若适当地采用1个偏移值来修正最小值判据,例如使用1个大小等于0.1的偏移值,那么不同样本估计出的位置参数之间的差异则显著减小,范围仅为463~1 506.对于这30个样本,传统的极大似然估计法得到的位置参数范围为0~1 691,最小二乘估计法的位置参数范围为0~1 686,与本文方法直接用于最小值判据的结果0~1 475相当.

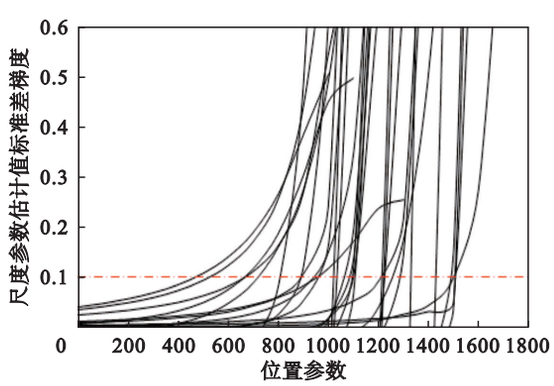

图5 对应于30个样本的$\nabla\gamma'$–γ'曲线

Fig. 5 $\nabla\gamma'$–γ' plots of thirty samples

显然,偏移值的大小直接影响参数估计结果的精度及稳健性.最优偏移值与所估计随机变量的概率分布的特征(例如分散性、对称性等)和样本大小有关.尽管目前尚未进行更全面、更细致的分析、归纳与总结,但已尝试过来自不同Weibull分布的近千个样本(其总体概率分布的形状参数值在2~5之间),表明补偿值取0.1对提高参数估计的精度和稳健性的效果很好.

4 结 语

采用尺度参数的1个伪估计量和1个参数估计的所提原理(正确的位置参数和形状参数使得根据1个Weibull分布样本中各样本值估计出的尺度参数之间的差异最小——在统计学意义上),通过对位置参数和形状参数的二维搜索可估计出Weibull分布的3个参数.这种参数估计方法简单易行,但存在与极大似然估计法和最小二乘估计法等传统参数估计方法相同的缺点,即参数估计结果对样本不确定性敏感.与其他方法不同,这种基于搜索和最小值判断的参数估计方法的过程——位置参数逼近其真实值的过程是可见的.而且,大量案例表明,位置参数逼近其真实值的过程中位置参数估计值的误差具有非对称性.基于位置参数估计值误差的非对称性,1个样本中不同样本值估计出的各尺度参数的标准差的条件最小值(该值是估计尺度参数时采用的位置参数值的函数)对位置参数的一阶导数等于1个大于零的值(例如0.1),参数估计结果的精度和稳健性大幅度提升.

参考文献:

[1] Park J P, Ham J, Mohanty S, et al. Statistical analysis of SN type environmental fatigue data of Ni–base alloy welds using Weibull distribution [J]. *Nuclear Engineering and Technology*, 2023, 55(5): 1924–1934.

[2] Klemenc J. Influence of fatigue-life data modelling on the estimated reliability of a structure subjected to a constant-amplitude loading [J]. *Reliability Engineering & System Safety*, 2015, 42:238–247.

[3] Çankaya M N, Vila R. Maximum log*q* likelihood estimation for parameters of Weibull distribution and properties: Monte Carlo simulation [J]. *Soft Computing*, 2023, 27(11):6903–6926.

[4] Jokiel-Rokita A, Piątek S. Estimation of parameters and quantiles of the Weibull distribution [J]. *Statistical Papers*, 2024, 65(1): 1–18.

[5] Sazak H S, Zeybek M. The modified maximum likelihood estimators for the parameters of the regression model under bivariate median ranked set sampling [J]. *Computational Statistics*, 2022, 37(3):1069–1109.

[6] Acitas S, Aladag C H, Senoglu B. A new approach for estimating the parameters of Weibull distribution via particle swarm optimization: an application to the strengths of glass fibre data [J]. *Reliability Engineering & System Safety*, 2019, 183: 116–127.

(下转第130页)