

基于遗传算法的维修时间分布参数非线性最小二乘估计

曾海军^{1,2} 陆中¹ 戎翔² 孙有朝¹

(1. 南京航空航天大学民航学院, 南京, 211106; 2. 中航商用航空发动机有限责任公司, 上海, 201108)

摘要: 维修时间分布参数估计是维修性验证与评定的一项重要工作。本文针对指数分布与对数正态分布两类常见的维修时间分布函数, 给出不同参数编码区间的确定方法, 以维修度函数的经验分布与含有待估计参数的维修度函数之间的误差最小为优化目标, 提出了基于遗传算法的维修时间分布参数非线性最小二乘估计方法。分别结合服从指数分布与对数正态分布的维修时间样本给出了验证实例, 结果表明本文方法精度在多数情况下要高于极大似然估计。

关键词: 遗传算法; 参数估计; 非线性最小二乘; 维修性

中图分类号: TB114.3 **文献标志码:** A **文章编号:** 1005-2615(2013)06-0859-06

GA Based Nonlinear Least-Squares Estimation for Parameter of Maintenance Time Distribution

Zeng Haijun^{1,2}, Lu Zhong¹, Rong Xiang², Sun Youchao¹

(1. College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing, 211106, China;

2. AVIC Commercial Aircraft Engine Co., Ltd., Shanghai, 201108, China)

Abstract: Parameter estimation of maintenance time distribution is an important task of maintainability verification and evaluation. The method of coding interval determination is presented for parameters from exponential and lognormal distributions that are most widely used in maintainability engineering, the minimization of error between maintainability empirical distribution and maintainability distribution with estimating parameters is chosen as objective of optimization, and nonlinear least-squares estimation based on genetic algorithm (GA) is proposed for parameter estimation of maintenance time distribution. Illustrating instances are given combined with maintenance time samples from exponential and lognormal distributions, and the results indicate that the precision of method presented herein is better than that of maximum likelihood estimation in most situations.

Key words: genetic algorithm (GA); parameter estimation; nonlinear least-squares; maintainability

维修性是产品在规定的使用条件下与规定的时间内, 按规定的程序和方法进行维修时, 保持或恢复到能完成规定功能的能力。产品的维修性是由设计赋予的使之维修简便、迅速、经济的重要特性, 随着产品的设计定型, 维修性表现为产品的固有属性^[1,2]。维修性是一种设计属性, 产品要具有良好的维修性属性, 必须在研制阶段就系统地开展

维修性设计、分析与试验、评定工作^[3]。维修时间参数是最为重要的维修性定量参数, 对该参数进行统计分析是系统维修性工程中维修性评定的重要内容。根据维修性演示试验或现场维修活动获得维修时间样本, 在基础上对维修时间进行分布拟合与参数估计, 是确定系统维修时间指标的一般性方法^[4]。

参数点估计常用方法有图估计法、矩估计法、极

基金项目: 国家自然科学基金民航联合研究基金(U1333118)资助项目; 江苏省自然科学基金(BK20130811)资助项目; 中央高校基本科研业务费(NS2011008, NZ2012118)资助项目。

收稿日期: 2013-08-15; **修订日期:** 2013-10-15

通信作者: 陆中, 男, 副教授, 1980 年出生, E-mail: luzhong@nuaa.edu.cn。

大似然估计法、最小二乘估计法等方法。在上述方法中,图估计法由于估计精度不高,导致其使用受到了很大限制;矩估计方法无法利用分布函数提供待估计参数的信息,对于复杂分布函数仍不适用;极大似然估计由于计算相对方便、精度较高而应用广泛,但是对于某些分布函数其计算过程仍然过于复杂^[5-10]。遗传算法具有极强的全局寻优能力、较好的鲁棒性,在可靠性工程中有着广泛的应用^[11,12]。本文针对指数分布与对数正态分布两种常用的维修时间分布函数,以遗传算法中的基因序列(染色体)表示待估计参数,根据上述分布各参数不同的统计特性确定编码区间,以维修度函数的经验分布与含有待估计参数的维修度函数误差最小为优化目标(误差倒数为遗传算法的适应度函数),提出了基于遗传算法的非线性最小二乘参数点估计方法。

1 维修性参数的非线性最小二乘估计

1.1 常用的维修时间分布

指数分布与对数正态分布是最为常用的维修时间分布函数^[4],指数分布的维修密度函数为

$$m(t) = \begin{cases} \mu e^{-\mu t} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (1)$$

式中参数 μ 表示修复率,对指数分布而言,修复率为常数,且与平均维修时间(Mean time to repair, MTTR)互为倒数

$$\text{MTTR} = \frac{1}{\mu} \quad (2)$$

对数正态分布的维修密度函数为

$$m(t) = \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}} \quad (3)$$

式中参数为 μ 与 σ ,若随机变量 X 服从对数正态分布,则 $\ln X \sim N(\mu, \sigma^2)$,即变量的对数服从正态分布,对数正态分布下的 MTTR 为

$$\text{MTTR} = e^{\mu + \frac{\sigma^2}{2}} \quad (4)$$

1.2 维修时间参数的非线性最小二乘估计

多元线性回归方程普通最小二乘法的离差平方和为

$$Q(\theta_1, \theta_2, \dots, \theta_z) = \sum_{j=1}^n (y_j - \theta_0 - \theta_1 x_{j1} - \dots - \theta_p x_{jz})^2 \quad (5)$$

普通最小二乘估计就是寻找参数 θ_i 的估计值 $\hat{\theta}_i$ 使式(5)的离差平方和 Q 达到极小值。对于分布函数,除指数分布、威布尔分布等可以通过对数变换将累计分布函数转化为估计参数的线性表达式,大部分分布函数很难转化为线性形式,因此本

文直接应用维修时间的累计分布函数(维修度函数)计算离差平方和,通过非线性最小二乘估计求得参数的估计值。

令 $M_\theta(t)$ 表示维修度函数,即维修时间的累积分布函数, $\theta = \{\theta_1, \theta_2, \dots, \theta_z\}$ 表示维修度函数的参数, z 为参数个数,对于指数分布 $\theta = \{\mu\}$,对于对数正态分布 $\theta = \{\mu, \sigma\}$ 。令 $t_1 \leq t_2 \leq \dots \leq t_i \leq \dots \leq t_n$ 表示维修时间的样本,则维修度的经验分布函数可表示为

$$M(t) = \begin{cases} 0 & t < t_1 \\ i/n & t_i \leq t < t_{i+1} \quad i = 1, 2, \dots, n-1 \\ 1 & t \geq t_n \end{cases} \quad (6)$$

当样本数量小于 20 时,维修度的经验分布函数可由近似中位表示为

$$M(t) = \begin{cases} 0 & t < t_1 \\ (i-0.3)/(n+0.4) & t_i \leq t < t_{i+1} \\ & i = 1, 2, \dots, n-1 \end{cases} \quad (7)$$

最小二乘估计的离差平方和可以通过维修度函数的经验分布与含有待估计参数的维修度函数之间的误差来表示

$$Q(\theta_1, \theta_2, \dots, \theta_z) = \sum_{i=1}^n [M(t_i) - M(t_i, \theta_1, \theta_2, \dots, \theta_z)]^2 \quad (8)$$

式中: $M(t_i)$ 为由式(6)或式(7)获得的维修度的经验分布函数; $M(t_i, \theta_1, \theta_2, \dots, \theta_z)$ 为维修时间的累积分布函数,可通过式(1)或式(3)由式(9)求得

$$M(t_i, \theta_1, \theta_2, \dots, \theta_z) = \int_0^{t_i} m(t) dt \quad (9)$$

式中: $\theta_1, \theta_2, \dots, \theta_z$ 为相应维修时间分布函数的参数。最小二乘估计的优化目标为式(8)最小化,因此在遗传算法中,可以式(8)的倒数作为适应度函数。

2 基于遗传算法的最小二乘估计方法

基本遗传算法由初始化、选择、交叉与变异 4 个部分组成,基于遗传算法的最小二乘估计的基本思想为:确定包含参数真实值的某一区域作为参数的编码区间,该区间即为参数优化的寻优空间;随机产生一定种群数量的染色体并将每一染色体对应到编码区间内的某一实数;依据适应度函数进行选择操作,将适应度函数大的染色体保留到下一代;随机选取染色体上的基因进行交叉和变异操作;经过若干代后,适应度函数达到最大值,此时种群中最优染色体对应的实数值即为参数的点估计值。上述流程如图 1 所示。

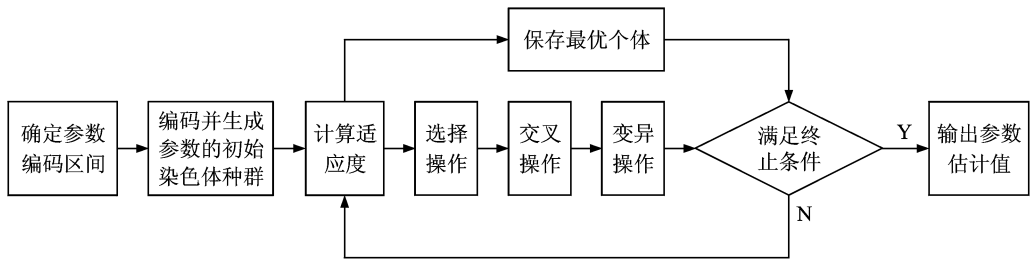


图 1 基于遗传算法的最小二乘估计流程

2.1 编码区间的确定

编码区间必须确保参数值位于该区间内,但是区间不能过大,否则染色体必须具有较长的编码长度才能保证参数的估计精度,而且还会影响算法的效率。假设维修时间样本为 $t_1 \leq t_2 \leq \dots \leq t_i \leq \dots \leq t_n$ 。

(1) 对于指数分布

由于参数 μ 为系统平均修复时间的倒数,因此以维修时间样本倒数的最大值为上限、最小值为下限组成的区间必然包含 μ 的真实值,因此指数分布参数的编码区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 的上下限可表示为

$$\hat{\mu}_L = 1/t_n, \hat{\mu}_U = 1/t_1 \quad (10)$$

若 $\hat{\theta}_L$ 与 $\hat{\theta}_U$ 相差过大,则需要进一步缩小区间间隔,由于 $M(\mu) = \int_0^{\infty} m(t) dt = 0.6231$,因此参数 μ 的值大约等于 $1/t_{[n \times 0.6231]}$,因此,进一步可按式(11)确定指数分布参数编码区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 的上下限

$$\hat{\mu}_L = 1/t_n, \hat{\mu}_U = 1/t_{[n \times 0.5]} \quad (11)$$

(2) 对于对数正态分布

由对数正态分布的性质可知,样本的对数服从正态分布,因此参数 μ 为所有样本取对数后的均值, σ 则为所有样本取对数后的标准差。

对于参数 μ ,其编码区间的上下限可表示为

$$\hat{\mu}_L = \ln t_1, \hat{\mu}_U = \ln t_n \quad (12)$$

若 $\hat{\theta}_L$ 与 $\hat{\theta}_U$ 相差过大时,可进一步按式(13)确定其上下限

$$\hat{\mu}_L = \ln t_{[n/4]}, \hat{\mu}_U = \ln t_{[3n/4]} \quad (13)$$

对于参数 σ ,可将样本取对数后分为如下两组: $\ln t_1 \leq \ln t_2 \leq \dots \leq \ln t_i \leq \ln t_j \leq \ln t_{j+1} \leq \dots \leq \ln t_n$ 与 $\ln t_{i+1} \leq \ln t_{i+2} \leq \dots \leq \ln t_{j-1}$,其中 $i = [n/4], j = [3n/4]$,前面一组样本位于原样本序列的两端,分散度很大,后面一组样本位于原样本序列的中间,分散度小,则参数 σ 的上下限可表示为

$$\begin{aligned} \hat{\sigma}_L &= \text{var}[\ln t_{[n/4]+1}, \ln t_{[n/4]+1}, \dots, \ln t_{[3n/4]-1}] \\ \hat{\sigma}_U &= \text{var}[\ln t_1, \dots, \ln t_{[n/4]}, \ln t_{[3n/4]}, \\ &\quad \ln t_{[3n/4]+1}, \dots, \ln t_n] \end{aligned} \quad (14)$$

式中 var 表示标准差。

2.2 参数编码

采用二进制编码法对维修时间分布的参数进行编码,根据染色体的编码长度和种群中染色体的个数生成 0-1 随机数矩阵,该矩阵的行向量或列向量即为初始编码,若参数的编码区间为 $[\hat{\theta}_L, \hat{\theta}_U]$,则参数编码值与实际值的关系为

$$y = \frac{(\hat{\theta}_U - \hat{\theta}_L)x}{(2^m - 1)} + \hat{\theta}_L \quad (15)$$

式中: x 为由二进制编码的染色体转化而来的十进制值; y 为真实值; m 为编码长度。若种群染色体个数(种群数)为 p ,则初始种群的参数编码可以由 $p \times m$ 的 0-1 随机数矩阵的行向量表示。

编码长度 m 与编码区间长度以及参数估计的精度要求有关,若以编码区间下限 $\hat{\theta}_L/10^l$ (l 为自然数)表示参数估计的精度要求,则编码长度必须满足

$$10^l(\hat{\theta}_U - \hat{\theta}_L)/\hat{\theta}_L \leq 2^m \quad (16)$$

2.3 选择

选择算子是实现优胜劣汰的关键操作,而适应度函数的确定是选择操作的关键,本文选取维修度函数的经验分布与含有待估计参数的维修度函数的误差的倒数为适应度函数,如式(17)所示

$$f(\theta_1, \theta_2, \dots, \theta_z) = 1/Q(\theta_1, \theta_2, \dots, \theta_z) \quad (17)$$

式中: $Q(\theta_1, \theta_2, \dots, \theta_z)$ 可由式(8)求得。染色体 i 被选择的概率为

$$p_j = f_j / \sum_{j=1}^p f_j \quad (18)$$

式中: f_j 为第 j 条染色体的适应度函数,选择的具体操作按照赌轮方法进行。

2.4 交叉与变异

取 $[0, 1]$ 上的一个随机数,当该数小于交叉概率或变异概率时,则进行交叉或变异操作。进行交叉操作时随机取一个 1 到 m (m 为编码长度)上的数 r ,对相邻两个染色体的 r 至 m 位进行交换,即可完成交叉作业。同样,进行变异操作时,仍随机

选取一个1到 m 上的数 r ,对染色体上的第 r 位进行变异(0与1互换)。

为了提高寻优速度,将每一代染色体中适应度函数最大的一个,不进行交叉与变异操作,直接保存到下一代。

2.5 迭代终止条件

当迭代终止时,所保存的染色体绝大部分必须为最优染色体,此时相邻两代的所有染色体适应度函数值之和应该相差很小,但是仅依据此条件,可能会使得求解过程陷入局部最优点,为防止出现此类情况,相邻 n 代的所有染色体的适应度函数之差也应当很小,因此迭代终止条件必须同时满足式(19)的两个不等式。

$$\left| \sum_{j=1}^p f_q(\theta_1, \theta_2, \dots, \theta_z) - \sum_{j=1}^p f_{q-1}(\theta_1, \theta_2, \dots, \theta_z) \right| < \varepsilon$$

$$\left| \sum_{j=1}^p f_q(\theta_1, \theta_2, \dots, \theta_z) - \sum_{j=1}^p f_{q-n}(\theta_1, \theta_2, \dots, \theta_z) \right| < \varepsilon$$

(19)

式中: q 为迭代终止时的代数; p 表示染色体个数(种群数); j 为染色体的序号; n 的取值一般建议在10~100之间; ε 为一小正实数,可根据待估计参数的精度要求取值($\hat{\theta}_1/10^4$)。

3 实例分析

本节以服从指数分布与对数正态分布的维修时间序列为样本给出了分析实例。

3.1 指数分布实例

随机生成参数 $\mu=0.001$ 指数分布样本。

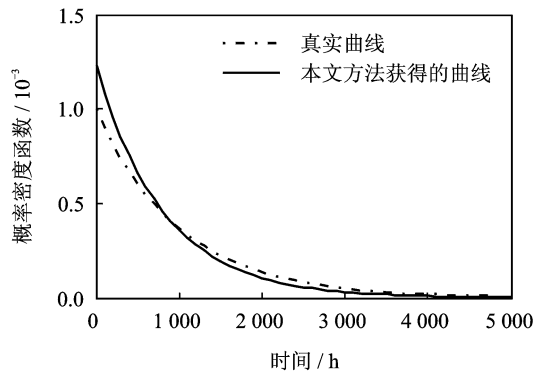
当样本量为15时,样本序列为:28.78,29.68,226.84,230.97,340.88,359.99,490.94,514.05,754.48,893.62,948.63,963.11,1288.33,1613.42,2467.75;按照式(11)取编码空间为 $[1/514.05, 1/2467.75]$,染色体长度为20,种群数为200,迭代终止时求得参数 μ 的估计值为0.0012,由极大似然估计可求得参数估计值为0.0014。

当样本量为30时,样本序列为:6.36,14.14,35.74,41.70,47.34,50.18,65.48,134.53,140.41,184.60,546.67,569.76,635.86,649.54,802.05,879.86,907.44,944.77,1022.10,1033.58,1083.64,1182.13,1311.13,1320.56,1345.27,1530.07,2309.60,2533.68,2651.34,2668.86;按照式(11)取编码空间为 $[1/802.05, 1/2467.75]$,染色体长度为20,种群数为200,迭代终止时求得参数 μ 的估计

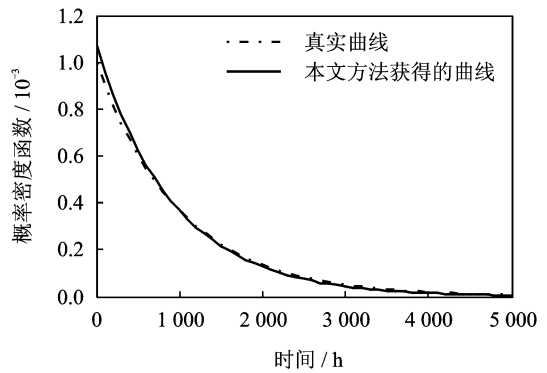
值为0.00106,由极大似然估计可求得参数估计值为0.00113。

此例中,本文方法的精度要略高于极大似然估计;并且多次随机生成参数 $\mu=0.001$ 指数分布样本,分别用本文方法与极大似然估计方法进行参数估计,本文方法的精度均要略高于极大似然估计,这就排除了由于生成随机数而导致的随机误差,证明对于指数分布,本文方法的精度是可信的。

图2给出了由本文方法确定的维修概率密度函数曲线与真实曲线的比较,由此可见,样本量为30时的估计精度要高于样本量为15时的估计精度。



(a) 样本量为15



(b) 样本量为30

图2 指数分布下的维修概率密度函数

3.2 对数正态分布实例

随机生成参数 $\mu=6.5, \sigma=1$ 的对数正态分布样本。

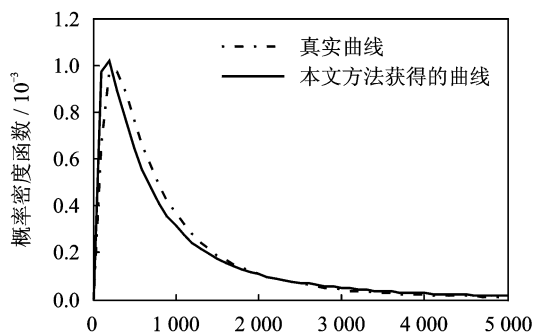
当样本量为15时,样本序列为:172.56,196.90,203.18,215.22,231.39,243.73,257.98,457.41,512.30,638.28,679.57,703.27,840.55,1725.83,2900.04;由式(13,14)分别取 μ 的编码空间为 $[\ln 215.22, \ln 703.27]$, σ 的编码空间为 $[0.2297, 1.4016]$,染色体长度为20,种群数为200,迭代终止时求得参数估计值 $\mu=6.49, \sigma=$

1.20,由极大似然估计可求得参数估计值 $\mu = 6.11, \sigma = 0.84$ 。

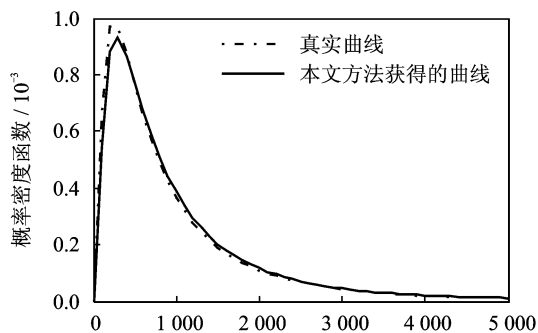
当样本量为 30 时,样本序列为:145.83, 162.47, 189.31, 205.50, 217.23, 237.70, 239.57, 259.47, 372.31, 384.59, 445.12, 470.04, 514.08, 520.41, 592.22, 656.43, 671.65, 714.35, 791.29, 848.18, 863.34, 912.82, 921.08, 1 096.44, 1 236.88, 1 927.77, 2 367.95, 2 387.70, 3 905.34, 5 632.450。由式(13,14)分别取 μ 的编码空间为 $[\ln 259.47, \ln 921.08]$, σ 的编码空间为 $[0.1308, 1.6718]$, 染色体长度为 20, 种群数为 200, 迭代终止时求得参数估计值 $\mu = 6.51, \sigma = 0.97$, 由极大似然估计可求得参数估计值 $\mu = 6.43, \sigma = 0.93$ 。

此例中,本文方法的精度同样要略高于极大似然估计;同样,多次随机生成参数 $\mu = 6.5, \sigma = 1$ 的对数正态分布样本,分别用本文方法与极大似然估计方法进行参数估计,本文方法的精度仍然均要略高于极大似然估计,这也排除了由于生成随机数而导致的随机误差,证明对于对数正态分布,本文方法的精度是可信的。

图 3 给出了由本文方法确定的维修概率密度



(a) 样本量为15



(b) 样本量为30

图 3 对数正态分布下的维修概率密度函数

函数曲线与真实曲线的比较,同样可见,样本量为 30 时的估计精度要高于样本量为 15 时的估计精度。

4 结束语

本文给出了常用维修时间分布参数编码区间的确定方法,以维修度函数的经验分布与含有待估计参数的维修度函数之间的误差最小为优化目标,利用遗传算法极强的全局优化能力,提出了维修时间分布参数的非线性最小二乘估计方法,本文方法避免了将维修度函数转化为线性表达式的繁琐过程,在已知参数编码区间的条件下,适用于任何复杂的分布类型。在大多数情况下,本文方法的精度要高于极大似然估计,并且非常接近参数的真实值,随着维修时间样本量的增加,本文方法的精度也会提高。

参考文献:

- [1] Lu Zhong, Sun Youchao. Maintainability virtual evaluation method based on fuzzy multiple attribute decision making theory for civil aircraft system[C]// The 8th International Conference on Reliability, Maintainability and Safety. New York: IEEE, 2009: 684-689.
- [2] Lu Zhong, Sun Youchao, Okafor E G, et al. Disassembly sequence planning for maintenance based on metaheuristic method[J]. Aircraft Engineering and Aerospace Technology, 2011,83(3):138-145.
- [3] United States Department of Defense. DOD-HDBK-791 (AM) Maintainability Design Techniques [S]. Washington D C: United States Department of Defense, 1988.
- [4] Rudolph Frederick Stapelberg. Handbook of reliability, availability, maintainability and safety in engineering design[M]. London: Springer-Verlag London Limited, 2009:299-300.
- [5] 杨谋存, 聂宏. 三参数 Weibull 分布参数的极大似然估计数值解法[J]. 南京航空航天大学学报, 2007, 39(1): 22-25.
Yang Moucun, Nie Hong. Advanced algorithm for maximum likelihood estimation of three parameter Weibull distribution[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2007, 39(1): 22-25.
- [6] 赵继俊, 邹经湘. 可靠性实验数据混合威布尔分布的参数优化估计[J]. 农业机械学报, 2000, 31(5): 106-109.

- Zhao Jijun, Zou Jinxiang. Estimation of parameter optimization for mixed Weibull distributions from the experimental data of reliability[J]. Transactions of the Chinese Society for Agricultural Machinery, 2000,31(5):106-109.
- [7] Lu Hailin, Chen Chonghong, Wu Jong-wuu. A note on weighted least-squares estimation of the shape parameter of the Weibull distribution[J]. Quality and Reliability Engineering International, 2004, 20(6): 579-586.
- [8] Nagode M, Fajdiga M. An improved algorithm for parameter estimation suitable for mixed Weibull distribution[J]. International Journal of Fatigue, 2000, 22(1):75-80.
- [9] Mourad Krifa. A mixed Weibull model for size reduction of particulate and fibrous materials[J]. Powder Technology, 2009,194(3):233-238.
- [10] Lu Zhong, Sun Youchao, Wu Haiqiao, et al. Maintainability parameter estimation model using particle swarm optimization based nonlinear least-squares method[J]. Journal of Computational Information Systems, 2010, V6(14):4665-4672.
- [11] Levitin G. Genetic algorithms in reliability engineering[J]. Reliability Engineering & System Safety, 2006,91(9):975-976.
- [12] 朱剑英. 智能系统非经典数学方法[M]. 武汉:华中科技大学出版社,2001:274-275.
- Zhu Jianying. Non-classical mathematics for intelligent systems[M]. Wuhan: Press of Huazhong University of Science and Technology, 2001:274-275.