文章编号:1673-5196(2019)03-0034-05

基于核密度估计的性能退化数据建模分析方法

安宗文,张永明,马 强,寇海霞

(兰州理工大学 机电工程学院,甘肃 兰州 730050)

摘要:针对现有性能退化建模分析方中存在试验样本数据量少和先验分布假设过于严格的不足,提出了一种基于 核密度估计的性能退化数据建模分析方法,并详细说明了进行核密度估计时最佳窗宽的确定方法.通过实际算例对 建模分析方法进行了验证,结果表明:基于核密度估计方法能有效地克服试验样本数据量少和先验分布假设严格 的不足,从而为该情况下产品的性能退化分析提供了一种新方法. 关键词:性能退化分析;核密度估计;先验分布假设;小样本;窗宽 中图分类号:TB114.3 文献标志码:A

Analysis method of product performance degradation with data modeling based on kernel density estimation

AN Zong-wen, ZHANG Yong-ming, MA Qiang, KOU Hai-xia

(College of Mechano-Electronic Engineering, Lanzhou Univ. of Tech., Lanzhou 730050, China)

Abstract: Aimed at the problem available now in the analysis of product performance degradation with modeling approach such as shortage of sample data and excessive strictness of prior distribution as sumption, an analysis method of product performance degradation with data modeling is proposed based on kernel density estimation the determination method of the optimal window widths for kernel density estimation is explained in detail, and the analysis method proposed is verified by means of actual example. The result shows that the proposed method based on kernel density estimation can effectively deal with the above-mentioned problem, thus providing a new method for analyzing the product performance degradation under this condition.

Key words: performance degradation analysis; kernel density estimation; prior distribution hypothesis; mini-sample; window widths

随着科学技术的快速发展,高可靠、长寿命产品 被大量应用于电子工业、航空航天、能源、交通运输 等领域,该类产品的可靠性评估是可靠性工程的难 题之一.目前,基于性能退化数据的可靠性建模分析 方法是高可靠、长寿命产品可靠性评估的主要手段 之一^[1].

基于性能退化数据的可靠性建模分析方法主要 有三种:基于退化轨迹的建模分析方法、基于寿命分 布的建模分析方法和基于随机过程的建模分析方 法.基于退化轨迹的可靠性建模方法中,通常假设样 本性能退化轨迹为线性函数,模型为 $y_{ij} = \varphi_i + \theta_j t_i$, 其中,假设模型参数 φ_i 和 θ_i相互独立且均服从正 态分布.该模型由于存在过于严格的假设条件,使得 模型的应用受到很大限制^[2].基于寿命分布的建模 分析方法可分为伪失效寿命建模分析方法和退化量 建模分析方法^[3-5].伪失效寿命方法忽略了退化过程 的随机性,需要采用大量的样本以保证试验结果的 精度,并且只考虑线性化模型^[6].基于退化量分布的 可靠性评估方法存在分布函数的假设过于严格,即 在不同测量时刻的退化数据可能服从不同的分布, 并且每个测量时刻需要大量的数据,对于小样本数 据的情况可能给参数估计带来严重误差等不足^[7-8]. 基于随机过程的建模方法,通常要求存在大量的非 破坏性顺序测量的性能退化数据^[9].因此,针对高可 靠、长寿命产品进行可靠性评估存在样本数据量少

收稿日期:2017-05-03

基金项目:国家自然科学基金(51665029)

作者简介:安宗文(1968-),男,甘肃景泰人,博士,教授,博导.

等难题,建立一种基于小样本条件下克服先验分布 假设的产品可靠性评估方法显得尤为重要.

鉴于此,本文提出了一种基于核密度估计的性 能退化数据建模分析方法.首先,采用非参数 Bootstrap 方法对试验原始样本数据进行扩充.其次,采 用核密度估计法对样本数据的概率密度、累积概率 密度进行求解,并运用 ^{χ2}分布检验进行窗宽的确定. 最后,结合单侧置信区间与一定失效阈值下可靠度 的关系对产品的可靠度进行评估,从而实现对本文 提出的建模分析方法的验证.

1 分析原理

1.1 非参数 Bootstrap 方法

Bootstrap 方法是统计学中一种重要的估计统 计量方差进行区间估计的统计方法,分为参数 Bootstrap 方法与非参数 Bootstrap 方法,是现代统 计学中较为流行的一种统计方法.非参数 Bootstrap 方法与基于先验分布的蒙特卡洛方法(MC)不同.对 总体分布未知且容量为n的样本数据采用放回抽 样的方法,将其扩充至 $n_1(n_1 > n)$ 容量的新样本,新 产生的样本数据称为 Bootstrap 样本.相继独立地利 用多个 Bootstrap 样本对总体样本分布进行统计推 断,这种方法称为非参数 Bootstrap 方法.

1.2 核密度估计

核密度估计是根据某一点处的概率密度与该点 附近包含样本个数之间的关系进行估计,利用核密 度估计得到产品的概率密度和累积概率密度曲线分 别如图1和图2所示.核密度估计在本文中主要克 服先验假设问题,利用核密度估计解决问题时,其中 最重要的是最佳窗宽的计算与选取.图3为新样本 的频数直方图,从图中可得到计算窗宽的分组信息 以及样本数据在某各区间的个数,为后续最佳窗宽 问题的解决提供了帮助.





χ²分布检验具有解决检验总体分布未知的功能,所以采用该方法对窗宽进行确定和检验.

首先假设一个窗宽,在利用非参数法绘制概率 密度曲线时可得到一自动窗宽,该自动窗宽不一定 为最佳窗宽,所以该自动窗宽为假设窗宽.然后绘制 频数直方图(见图 3),进行分组,最后根据下式进行 χ²分布检验:

$$\chi^{2} = \sum_{i=1}^{k} \frac{f_{i}^{2}}{n p_{i}} - n$$
 (1)

$$_{0}-1)$$
 (2)

式中: y_0 为总体样本数据在概率密度曲线中的柱状 图的分组个数; $f_i(i=1,2,3,\dots,k)$ 为样本观察值 x_1,x_2,\dots,x_n 落在全体样本数据分成互不相交子 集的个数; p_i 为 f_i 落在总体样本数据个数的概率; a 为显著性水平,a=0.01.

 χ^2_{a} (ν



式(1)与式(2)的结果进行对比,确定最佳窗宽:

 1) 若式(1)的 χ² 分布检验值小于等于或非常 接近根据查表获得的式(2)的值时,则认为该假设窗 宽为最佳窗宽;

2) 若式(1)的 ^{χ²}分布检验值远远小于式(2)的 值,则应调大窗宽再进行上述步骤;反之,调小窗宽 在进行上述步骤,最终确定最佳窗宽.

1.3 置信区间与可靠度关系

置信区间表现的是这个参数的真实值有一定概

率落在测量值周围的程度,它所给出的是被测量参 数的测量值的可信程度,本文选取单侧置信区间进 行研究.

首先,考虑某产品的退化量分布置信度为 0.9 的单侧置信区间:

$$Pr\{X \leqslant x_{0.9}^i\} = 0.9 \tag{3}$$

式中:x_{0.9}为第 i 次测量时退化量分布核密度估计所 得单侧置信上限.

由此可以获得不同测量时刻核密度估计获得的 退化量分布单侧置信上限 $x_{0.9}^1, x_{0.9}^2, \dots, x_{0.9}^m, 进而获$ 得 x_{0.9}(t),式(3)可表示为

产品可靠度为 0.9 时的表达式为

$$Pr\{X \leqslant x_{0.9}(t)\} = 0.9 \tag{4}$$

$$R(t) = Pr\{X \leq D_f\} = 0.9$$
 (5)
结合式(4)和式(5)可得

$$x_{0,9}(t) = D_f \tag{6}$$

cm

0.12

2.33

1.93

1.97 1.91

1.80

1.82

1.83

1.77

1.72

1.67

1.65

1.64

1.52

1.45

1.49

1.40

1.38

1.35

1.31

1.29

1.27

求解得到 T_{0.9},即可靠度或寿命分布概率为 0.9 时 所对应的时间.

最后,根据式(4)可推导出 T_{0.9}可靠度或寿命分 布概率为 0.9 时所对应的时间,进而可以推导出 $T_{0,1}, T_{0,2}, \dots, T_{1,0}$ 可靠度或寿命分布概率各时刻所 对应的时间.

算法示例 2

表1为某产品原始的疲劳裂纹扩展样本数据, 当裂纹扩展至 1.6 cm 时失效.

表1 疲劳裂纹长度 Tab.1 Fatigue crack length

样本		载荷循环次数/(×10 ⁶ 次)										
编号	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11
1	0.90	0.95	1.00	1.05	1.12	1.19	1.27	1.35	1.48	1.64	1.80	2.03
2	0.90	0.94	0.98	1.03	1.08	1.14	1.21	1.28	1.37	1.47	1.60	1.74
3	0.90	0.94	0.98	1.03	1.08	1.13	1.19	1.26	1.35	1.46	1.58	1.77
4	0.90	0.94	0.98	1.03	1.07	1.12	1.19	1.25	1.34	1.43	1.55	1.73
5	0.90	0.94	0.98	1.03	1.07	1.12	1.19	1.24	1.34	1.43	1.55	1.71
6	0.90	0.94	0.98	1.03	1.07	1.12	1.18	1.23	1.33	1.41	1.51	1.68
7	0.90	0.94	0.98	1.02	1.07	1.11	1.17	1.23	1.32	1.41	1.52	1.60
8	0.90	0.93	0.97	1.00	1.06	1.11	1.17	1.23	1.30	1.39	1.49	1.62
9	0.90	0.92	0.97	1.01	1.05	1.09	1.15	1.21	1.28	1.36	1.44	1.55
10	0.90	0.92	0.96	1.00	1.04	1.08	1.13	1.19	1.26	1.34	1.42	1.52
11	0.90	0.93	0.96	1.00	1.04	1.08	1.13	1.18	1.24	1.31	1.39	1.49
12	0.90	0.93	0.97	1.00	1.03	1.07	1.10	1.16	1.22	1.29	1.37	1.48
13	0.90	0.92	0.97	0.99	1.03	1.06	1.10	1.14	1.20	1.26	1.31	1.40
14	0.90	0.93	0.96	1.00	1.03	1.07	1.12	1.16	1.20	1.26	1.30	1.37
15	0.90	0.92	0.96	0.99	1.03	1.06	1.10	1.16	1.21	1.27	1.33	1.40

1.03

1.05

1.04

1.02

1.02

1.02

1.07

1.08

1.07

1.05

1.05

1.04

1.11

1.11

1.09

1.08

1.08

1.07

1.16

1.16

1.14

1.12

1.12

1.11

1.22

1.20

1.19

1.16

1.16

1.14

1.26

1.24

1.23

1.20

1.19

1.18

1.33

1.32

1.28

1.25

1.24

1.22

1)利用非参数法中 Bootstrap 方法进行样本扩 充.

0.95

0.96

0.94

0.94

0.94

0.94

0.97

0.97

0.97

0.97

0.97

0.97

1.00

1.00

1.01

0.99

0.99

0.99

文献[2]给出 12 组样本数据,进行原始样本扩 充,扩充后的样本容量为10000个,结果见表2.

2) 根据扩充后的样本数据,利用 χ^2 分布确定 窗宽,并采用核密度估计进行绘图.

以扩充后的第一组数据为基础,对窗宽的计算 与检验进行演示.首先假设窗宽为 0.003.然后根据 表 2^[2]中的第一组数据,利用图 3 进行样本数据分 组,分组情况见表 3.

根据式(1)进行计算,结果为 0.304 54,利用

表 2 原始裂纹长度的扩充数据

Tab.2 Augmented data of original fatigue crack length cm

式(2) 查表得出在 0.01 所对应的值为 13.277. 两式 结果对比可以看出式(1)计算的结果远远小于式(2)

样本			样	羊本个梦	銰		
编号	1	2	3		9 998	9 999	10 000
1	0.92	0.94	0.94	•••	0.92	0.93	0.94
2	0.96	0.94	0.98		0.96	0.98	0.96
				•••			
10	1.58	1.19	1.30		1.42	1.55	1.18
11	1.25	1.49	1.55		1.33	1.52	2.03
12	1.83	1.72	1.40	•••	1.93	1.35	1.93

16

17

18

19

20

21

0.90

0.90

0.90

0.90

0.90

0.90

0.92

0.93

0.92

0.92

0.92

0.92

的结果,所以应调大窗宽,令窗宽为 0.004 5.经上述 计算步骤得到式(1)值为 11.364 与式(2)值接近,即 选定最佳窗宽为 0.004 5.

表 3 样本数据分组情况

Tab.3 Grouping list of sample data

样本数据划分区间	概率	样本数据个数
$x \le 0.915$	0.020 48	0
$0.915 < x \le 0.925$	0.401 22	4 234
0 . 925≪ <i>x</i> ≪0 . 935	0.292 60	2 901
0 . 935< <i>x</i> ≤0 . 945	0.272 00	2 865
x>0.945	0.013 70	0

根据以上过程,利用其他几组数据进行最佳窗 宽的计算与检验,见表 4.

表 4 每组数据对应的最佳窗宽

Tab.4 Optimal window width of each data set

样本编号	1	2	3	4	5	6
最佳窗宽	0.004 5	0.002 5	0.005 0	0.007 5	0.010 0	0.012 0
样本编号	7	8	9	10	11	12
最佳窗宽	0.015 0	0.020 0	0.025 0	0.032 0	0.042 0	0.050 0

根据第一组扩充后的数据进行核密度估计,即 绘制概率密度曲线和累积概率密度曲线,分别如 图 4和图 5 所示.



Fig.4 Probability density curve





3) 根据累积概率密度曲线进行数据取点.

以扩充后的第一组数据为例,根据第一组数据 绘制的累积概率密度曲线进行取点,对应的累积概 率密度取值点如图 6 所示.





得到扩充后第一组数据的数据信息,命名为 x_j ($j=1,2,\dots,9$),分别对应累积概率密度为 0.9,0.8, 0.7,…,0.1,即扩充后第一组样本数据的累积概率 密度对应值见表 5.

表 5 扩充后第一组样本数据对应的累积概率密度 Tab.5 Accumulative probability density values corresponding to first set of augmented sample data

		0	-		
x_j	0.942 0	0.939 8	0.936 7	0.931 4	0.928 7
у	0.900 0	0.800 0	0.700 0	0.600 0	0.500 0
x_j	0.923 7	0.921 2	0.919 7	0.912 0	
у	0.400 0	0.300 00	0.200 00	0.100 0	

 4)根据累积概率密度曲线所采集的累积概率 密度进行仿真模拟.

根据文献[2]对产品进行试验的试验循环次数 与表 6 对应的 *x* 值进行拟合.

表 6 载荷循环次数对应的 0.9 时刻所对应的累积概率密度 Tab.6 Accorresponding to the Accumulative probability density corresponding to load cycle times at 0.9 instant

载荷循环次 数/(×10 ⁶ 次)	0.01	0.02	0.03	0.04	0.05	0.06
x	0.942	0.982	1.034	1.080	1.136	1.204
载荷循环次 数/(×10 ⁶ 次)	0.07	0.08	0.09	0.10	0.11	0.12
r	1.270	1.362	1.468	1.660	1.855	2.046

通过拟合发现,疲劳失效次数与载荷循环次数 符合以下关系:

$$g(t) = p_1 t^2 + p_2 t + p_3 \tag{7}$$

式中:g(t)为t时刻循环次数下疲劳破坏的次数; p_1 、 p_2 、 p_3 为模型参数.

将 0.9,0.8,…,0.1 时刻所对应的累积概率密度 利用式(7)进行拟合,得到模型参数 p_1 、 p_2 、 p_3 ,同时 得到拟合精度 R^2 ,见表 7.

5)利用 Matlab 进行数据仿真,得到载荷循环 次数与失效概率之间的关系,如图 7 所示.

表 7 不同时刻对应的模型参数和拟合精度 Tab.7 Model parameters and fitting accuracy at different instants

参数		时刻									
	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1		
p _1	88.94	69.05	56.48	47.93	41.94	32.70	24.84	15.47	2.18		
₱ 2	-2.009	-0.7108	0.132 5	0.593 1	0.551 8	1.063 0	1.357 0	1.818 0	2.692 0		
<i>₱</i> ₃	0.991 3	0.961 7	0.951 2	0.936 7	0.936 7	0.924 8	0.918 6	0.925 8	0.885 1		
$R^{_2}$	0.993	0.997	0.997	0.998	0.997	0.997	0.997	0.998	0.993		



Fig.7 Relationship of failure probability to load cycles times

根据图7得到如下结果:

 1) 在载荷循环次数达到 0.11 万次时,本文方 法得到失效概率为 0.39,文献[2]方法的失效概率为 0.35,而根据原始数据得到的失效概率为 0.38,因此 本文分析方法较为准确.

2)在载荷循环次数达到 0.12 万次之后,由于 试验数据不足,本文分析方法所得结果偏差较大.

3 结论

 1)本文提出的基于核密度估计的性能退化数据建模新方法能有效地解决小样本及先验分布假设 过于严格的问题.

2)本文分析方法对于小样本试验数据和克服
先验假设分布的可靠性评估较为准确.

 3)本文分析方法对于样本数据缺失的情况下 预测能力不足.

参考文献:

- [1] 邓爱民.高可靠长寿命产品可靠性技术研究 [D].长沙:国防科 学技术大学,2006.
- [2] LU C J, MEEKER W O. Using degradation measures to estimate a time-to-failure distribution [J]. Technometrics, 1993, 35 (2):161-174.
- [3] 邓爱民,陈 循,张春华,等.基于性能退化数据的可靠性评估 [J].宇航学报,2006,27(3):546-552.
- [4] 邓爱民,陈 循,张春华,等.基于加速退化数据的可靠性评估 [J].弹箭与制导学报,2006,26(2):808-812.
- [5] 安宗文,刘小刚,高建雄.基于裂纹扩展理论的结构强度退化模型[J].兰州理工大学学报,2014,40(6):45-48.
- [6] 林逢春,王前程,陈云霞,等.基于伪寿命的加速退化机理一致 性边界检验[J].北京航空航天大学学报,2012,38(2):233-238.
- [7] 赵建印.基于性能退化数据的可靠性建模与应用研究 [D].长 沙:国防科学技术大学,2005.
- [8] ZEHUA C, SHURONG Z.Lifetime distribution based degradation analysis [J]. IEEE Transactions on Reliability, 2005, 54 (1):3-10.
- [9] 彭宝华.基于 Wiener 过程的可靠性建模方法研究 [D].长沙:国 防科学技术大学,2010.