

可靠性模型自助参数估计法

王智明, 杨建国

(上海交通大学 机械与动力工程学院, 上海 200240, wangzm301@sjtu.edu.cn)

摘要: 为了提高可靠性模型参数估计精度, 在分析传统可靠性模型的基础上, 指出传统的最小二乘法估计和逐次线性概似估计因受随机变量假设分布的限制存在一定的局限性. 用自助法给出了可靠性模型的参数估计, 包括点估计和区间估计. 计算结果表明: 自助法所得参数的估计区间、标准差及模型标准残差均小于传统回归法所得结果. 在可靠性模型参数估计中, 自助参数估计法通过增加模拟次数可快速逼近参数真值, 方法快速、有效, 可提高模型精度.

关键词: 可靠性模型; 自助参数估计; 最小二乘法; 逐次线性概似

中图分类号: TB114

文献标志码: A

文章编号: 0367-6234(2010)05-0820-04

Bootstrap method of parameter estimation in reliability model

WANG Zhi-ming, YANG Jian-guo

(School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai, 200240, China, wangzm301@sjtu.edu.cn)

Abstract: In order to improve the precision of parameter estimation of reliability model, based on the analyses of traditional reliability model, the limitations of least square method and gradual linear approximation because of their hypothetical distribution to random variable are pointed out, and the parameter estimators of point and interval are given by bootstrap method. The calculating results show that the estimated interval, standard error of parameter and the standard residual errors of the model are both less than those estimated by traditional method. Bootstrap method can approximate the actual value quickly than traditional way when increasing simulative times, so this method is feasible and effective and has a better accuracy than the latter.

Key words: reliability model; bootstrap parameter estimation; least square method; gradual linear approximation

可靠性模型在可靠性分析中具有十分重要的作用, 通过可靠性模型, 可在设计、试验、运行和维修等阶段对产品进行可靠性分析和优化. 参数模型是可靠性模型分析中常见的模型之一, 该模型一般具有经典的统计分布, 如指数分布、正态分布、威布尔分布等, 可用观测失效数据估计出分布中的模型参数. 其参数估计的方法有: 图形法^[1]、矩法^[2]、极大似然法^[3]、贝叶斯法^[4]等. 非线性模型参数估计的方法有^[5]: 带约束的参数估计、稳健估计、拟似然估计等. 常用方法有两种: 一是非

线性函数线性化, 用最小二乘法进行线性回归^[6-7]; 二是对于不可线性化的内在非线性函数可用泰勒级数展开, 略去高次项, 用逐次线性概似的方法解之^[8]. 上述两种方法存在的问题是: 前者有时由于模型精度过低而通不过假设检验; 后者的求解过程过于烦琐, 有时陷于局部最小而得不到全局最优解.

针对上述问题, 本文提出了可靠性模型参数自助估计法, 可在原方法的基础上给出参数的点估计和区间估计, 经实例验证, 该方法快速、有效, 模型残差小, 具有较高的精度.

1 可靠性模型参数估计的两种常见方法的局限性

可靠性模型分为参数模型和非参数模型(本

收稿日期: 2008-10-27.

基金项目: 高等学校全国优秀博士学位论文作者专项资金资助项目(200131).

作者简介: 王智明(1969—), 男, 博士研究生;

杨建国(1957—), 男, 教授, 博士生导师.

文讨论参数模型),两类模型以非线性模型居多,其中可线性化的内在线性模型可用最小二乘法进行回归分析,不可线性化的(内在)非线性模型可用逐次线性概似的方法解之.

1.1 内在线性模型参数的最小二乘估计

内在线性模型是线性化的非线性模型,即可通过变换转化成线性模型,其 p 元线性回归模型为

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon.$$
式中:随机变量 $\varepsilon \sim N(0, \sigma^2)$,且独立同分布, $\beta_0, \beta_1, \cdots, \beta_p$ 和 σ^2 是未知参数.

n 次独立观测的线性模型用矩阵形式表示为

$$Y = X\beta + \varepsilon. \tag{1}$$

式中:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$$
$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

β 的最小二乘估计

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \tag{2}$$

1.2 内在非线性模型参数的逐次线性概似估计

内在非线性模型的一般形式为

$$Y = f(X, \hat{\beta}) + v. \tag{3}$$

式中: v 为随机变量,设式(3)有 n 个自变量和 p 个参数, $\hat{\beta}_0$ 为参数 $\hat{\beta}$ 的初始值,则在 $\hat{\beta}_0$ 的邻域内用泰勒级数展开函数 f ,并略去高次项,有

$$Y - f(X, \hat{\beta}_0) = \sum_{i=1}^p \left[\frac{\partial f(X, \hat{\beta})}{\partial \beta_i} \right]_{\hat{\beta}_0} (\beta_i - \hat{\beta}_{i0}) + \varepsilon. \tag{4}$$

对应上式代换整理如下:

$$\Delta Y = X' \Delta \beta + \varepsilon. \tag{5}$$

式(5)形如式(1),线性回归后可得 $\Delta \beta$ 的一组估计值,因此可解出

$$\hat{\beta} = \hat{\beta}_0 + \Delta \hat{\beta}.$$

设 δ 为允许误差,如果满足

$$|(\hat{\beta}_1 - \hat{\beta}_0) / \hat{\beta}_0| \leq \delta,$$

则 $\hat{\beta}$ 为所求结果;否则以该值为新的初始值,返回式(4),重新计算,直至满足前后两次参数估计值

的相对误差小于给定的允许误差.

1.3 两种估计法存在的问题及解决的办法

从上面的分析可以看出,可线性化非线性模型和内在非线性模型,其参数估计无论是采用最小二乘估计还是逐次线性概似估计,本质都是线性回归,式(1)和式(5)的相似性也说明了这一点.前者就随机变量对模型精度的影响处理是要求残差平方和最小,但由于这类模型在对数变形代换过程中改变了因变量的形式,使得变形后模型在其残差平方和最小的情况下,原模型的残差平方和不一定最小,因此,有时会出现变形后模型精度尚可,但原模型却精度过低的情况;逐次线性概似法参数估计虽然可以使模型残差平方和达到最小而逼近最佳的待估回归系数,但其求解过程过于繁琐,且解过分依赖于初始值,因此有时会陷于局部最小而得不到全局最优解,即得不到最优值处的平方和最小.另外,两种方法均受到假设条件的限制.因为通常假定随机变量为零均值,同方差和不存在自相关,且自变量之间不存在高度多重共线性.

自助 Bootstrap 法可以解决上述问题,这是因为自助 Bootstrap 法充分利用了子样本本身的信息,只依赖于给定的观测信息,不需要其他的假设.文献[9]在时间序列多项式回归分析中,就是用自助重复抽样拟合出了参数估计值的分布.可靠性模型参数自助估计法对随机变量分布总体无须做出任何假设,利用计算机通过增加模拟次数可快速逼近参数真值,提高模型精度.

2 可靠性模型参数的自助 Bootstrap 估计

2.1 自助 Bootstrap 法

Bootstrap 法(自助法)是美国 Stanford 大学统计系教授 Efron 提出的一种统计推断方法^[10-11].该方法的思路是:设容量为 n 的独立同分布随机子样 $X = (X_1, X_2, \cdots, X_n)$ 来自未知的总体分布 F ;通过子样构造经验分布 F_n ,设 $\theta = \theta(F_n)$ 为总体分布中的未知参数, $\hat{\theta} = \hat{\theta}(F_n)$ 为 θ 的估计,估计误差

$$R_n = \hat{\theta}(F_n) - \hat{\theta}(F).$$

依 F_n 抽取一次再生样本 $X_1^* = (X_{11}^*, X_{12}^*, \cdots, X_{1n}^*)$,得到 $\theta(F)$,并作为 θ 的一个样本.设 F_n^* 是由 X^* 得出的经验分布,记:

$$R_n^* = \hat{\theta}(F_n^*) - \hat{\theta}(F_n).$$

然后用 R_n^* 的分布特性来逼近 R_n 的分布特性.因

此有

$$\theta(F) \approx \overset{\Delta}{\theta}(F_n) - R_n^*.$$

记 B 次自助抽样的再生样本 $\boldsymbol{X}_B^* = (X_{B1}^*, X_{B2}^*, \cdots, X_{Bn}^*)$, 则 B 个自助样本作为 θ 的 B 个样本构成向量:

$$\begin{aligned} \boldsymbol{X}_{\text{Bootstrap}} &= [\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_B^*]^T = \\ &\begin{bmatrix} X_{11}^* & X_{12}^* & \cdots & X_{1n}^* \\ X_{21}^* & X_{22}^* & \cdots & X_{2n}^* \\ \vdots & \vdots & \vdots & \vdots \\ X_{B1}^* & X_{B2}^* & \cdots & X_{Bn}^* \end{bmatrix}. \end{aligned} \tag{6}$$

因此可依式(6)统计求出 θ 的分布及其特征值, 如均值、方差或分布密度函数等.

2.2 Bootstrap 法在可靠性模型参数估计中的具体应用

对于内在线性模型或内在非线性模型, 可分别依式(2)或经验初值得到模型参数的估计 $\overset{\Delta}{\boldsymbol{\beta}}$, 则由式(1)得

$$\overset{\Delta}{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\overset{\Delta}{\boldsymbol{\beta}}.$$

因此可计算 $\overset{\Delta}{\boldsymbol{\varepsilon}}$, 构成子样 $\overset{\Delta}{\boldsymbol{\varepsilon}} = (\overset{\Delta}{\varepsilon}_1, \overset{\Delta}{\varepsilon}_2, \cdots, \overset{\Delta}{\varepsilon}_n)$, 随后抽去一个自助样本 $\boldsymbol{\varepsilon}_1^* = (\varepsilon_{11}^*, \varepsilon_{12}^*, \cdots, \varepsilon_{1n}^*)$, 由式(1)重新计算 $\boldsymbol{Y}^*$, 得到自助观测值 $(\boldsymbol{X}, \boldsymbol{Y}^*)$, 最后用该值和公式

$$\boldsymbol{Y}^* = \boldsymbol{X}\overset{\Delta}{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}^*,$$

得到 $\boldsymbol{\beta}^*$ 的第一个估计 $\boldsymbol{\beta}_1^*$, 重复上述过程至第 B 次, 得 $\boldsymbol{\beta}^*$ 的第 B 个估计 $\boldsymbol{\beta}_B^*$. 至此, 可由 B 个自助抽样估计值 $\boldsymbol{\beta}^*$ 统计得出其均值, 作为参数的最后估计.

作为验证, 这里以文献[12]提供的数据为例, 可靠性模型选取威布尔分布.

具有经典统计分布的威布尔分布是典型的可线性化的非线性模型, 两参数威布尔分布可靠性模型为

$$R(t) = \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right].$$

这里 β 为形状参数, η 为尺度参数, t 为时间, $R(t)$ 为可靠度, 设

$$y = \ln[-\ln(R(t))], x = \ln(t), \tag{7}$$

变换后的线性模型为

$$y = \beta(x - \ln(\eta)). \tag{8}$$

对应文献[12]的一组失效数据: $t_1, t_2, \cdots, t_n$, 首先用中位秩估计可得模型的可靠性估计. 公式如下:

$$\begin{aligned} R(t_i) &= 1 - (i - 0.3)/(n + 0.4), \\ i &= 1, 2, \cdots, n. \end{aligned}$$

因此, 得到数据对 $(t_i, R(t_i))$, 由式(7), 有数据对 (x_i, y_i) , 随后对应式(8)作回归分析, 最后应用自助法得到各参数的最后估计值, 其中, 取 $B = 1\,000$, 同时计算出其置信区间和标准差及各模型的标准残差^[13], 计算结果如表 1.

表 1 不同方法结果比较

方法	β			η			模型标准残差	
	估计值	置信区间	标准差	估计值	置信区间	标准差	最小二乘估计	逐次线性概率估计
传统回归法	1.977 298	[1.710 485, 2.244 111]	10.560	553.071 393	[537.250 246, 574.600 564]	11.200	6.199	3.676
Bootstrap 法	1.976 424	[1.968 974, 1.983 875]	7.373	553.406 942	[552.895 520, 553.920 960]	7.824	5.232	2.567

自助法所得可靠性曲线与原始曲线如图 1 所示, 其可靠度在图中分别用 $R_b(t)$ 、 $R(t)$ 表示. 图 2 为自助法残差随机变量有放回的重复抽样 100、

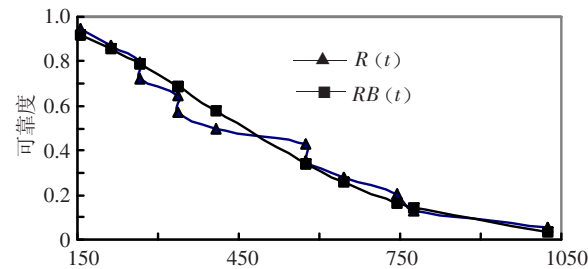


图 1 自助法可靠性曲线与原始曲线比较图

1 000 次后的直方图、密度估计曲线和正态分布概率密度曲线图.

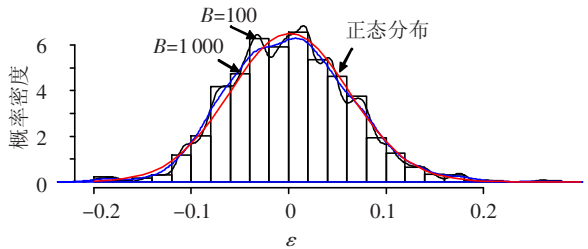


图 2 自助法随机变量直方图、密度估计曲线和正态分布概率密度曲线图

3 讨 论

1)从表1可以看出:自助法较之传统的回归法,其各项指标均明显优于后者.特别是对参数置信区间的估计,在同等置信水平下,前者显著好于后者.

2)从模型标准残差看,若分别采用传统回归法中的最小二乘估计和逐次线性概似估计,其值分别为6.199、3.676,后者小于前者,约为前者的一半左右.后者较高的模型精度其实是以其增加迭代次数为代价的,其中,迭代次数与允许误差有关;若在两种传统方法的基础上均采用自助法,模型标准误差均会有不同程度的减小.前者减小0.967,后者减小1.109.且后者仍为前者的一半左右,自助概似估计仍优于自助平方估计.

3)前文已述及原模型的精度均低于变换后模型精度,这是因为等量变换改变了因变量形式所致.

4)从图2可以看出:只有在大样本($B = 1\ 000$)情况下,残差分布才接近正态分布.因此传统回归视随机变量 $\varepsilon \sim N(0, \sigma^2)$ 的假定必然会影响模型的精度,而自助参数估计只依赖于给定的观测信息,不需要其他假设的这一特点正好弥补了其不足,因此可获得具有较高精度的模型.

4 结 论

1)传统参数估计中,视随机变量为零均值,同方差和不存在自相关,且自变量之间不存在高度多重共线性的假定是影响模型精度的原因之一.

2)传统参数估计中的最小二乘估计和逐次线性概似估计,本质都是线性回归.

3)自助参数估计法对随机变量分布总体不做任何假设,利用计算机通过增加模拟次数可快速逼近参数真值,快速、有效,所得模型残差小,与传统的参数估计相比,可靠性模型具有较高的精度.

参考文献:

- [1] ALLAN R N, ABU-SHEIKHAH N M. Analysis of reliability data using graphic-based interactive techniques[J]. International Journal of Quality & Reliability Management, 1987, 4(3): 57–70.
- [2] ZHAO Y G, ONO T, KATO M. Second-order third-moment reliability method[J]. Journal of structural engineering – ASCE, 2002, 128(8): 1087–1090.
- [3] HUANG W, DIETRICH D L. An alternative degradation reliability modeling approach using maximum likelihood estimation[J]. IEEE transactions on reliability, 2005, 54(2): 310–317.
- [4] BAI Yongsheng, WEN Liang. Application of Bayesian method based on random weighting in reliability parameter estimation[J]. Computer Engineering and Applications, 2007(6): 229–233.
- [5] 王新洲. 非线性模型参数估计理论及应用[M]. 武汉: 武汉大学出版社, 2002.
- [6] 张英芝, 申桂香, 贾亚洲, 等. 数控车床故障分布规律及可靠性[J]. 农业机械学报, 2006, 37(1): 156–159.
- [7] 高翔, 王若平, 夏长高, 等. 随机变量多重 Weibull 统计模型及参数最优估计[J]. 农业机械学报, 2006, 37(11): 41–44.
- [8] 梅文华. 可靠性增长试验[M]. 北京: 国防工业出版社, 2003: 90–201.
- [9] PARRELLA L M, VITALE C. Bootstrap inference in local polynomial regression of time series[J]. Stat Meth & Appl, 2007, 16: 117–139.
- [10] EFRON B. Bootstrap methods: another look at the jackknife[J]. The Annals of Statistics, 1979, 7(1): 1–26.
- [11] EFRON B. Better bootstrap confidence interval[J]. Amer Statist Assor, 1987, 82: 171–200.
- [12] LUXHOY, JAMES T, HUAN-JYH S. Reliability curve fitting for aging helicopter components[J]. Reliability Engineering and System Safety, 1995, 48: 229–234.
- [13] EFRON B, TIBSHIRANI R J. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy[J]. Statistical Science, 1986, 1(1): 54–77.

(编辑 杨波)