基于随机加权的 Bayes 方法在可靠性参数估计中的应用

白永生,温　亮
BAI Yong-sheng,WEN Liang

军械工程学院 装备指挥与管理系,石家庄 050003
Department of Equipment Command and Management,Ordnance Engineering College,Shijiazhuang 050003,China

BAI Yong-sheng,WEN Liang.Application of Bayesian method based on random weighting in reliability parameter estimation.Computer Engineering and Applications,2007,43(18):229-230.

Abstract: Under small-samples,a random weighting resampling method is implemented,which builds the prior distributions of parameters of interest,then Bayesian statistical algorithm is used to estimate the reliability parameters by combining prior information with the currently observed distribution.A Matlab program is developed to perform random weighting resampling and Bayesian calculation.A simulation example is also given to verify the effective and precision of this statistical inference.
Key words: random weighting;Bayesian method;small sample;parameter estimation

摘　要: 在小子样条件下,通过随机加权重采样技术,获得兴趣变量的分布函数,将其作为先验信息与当前样本信息相结合,然后通过贝叶斯方法对可靠性参数进行估计。仿真算例证明,这种方法在进行可靠性参数估计时较经典参数估计方法有更高的精确性。
关键词: 随机加权;贝叶斯方法;小子样;参数估计
文章编号: 1002-8331(2007)18-0229-02　**文献标识码:** A　**中图分类号:** TB114.3

1　引言

在武器装备可靠性实验的统计分析中,产品可靠性参数估计长期以来一直是一项非常重要的任务。随着装备技术的发展,武器装备的研制和开发速度越来越快,武器系统的复杂度和实验费用也越来越高,在这种情况下,只能得到较少的寿命数据,那些基于较大样本的经典参数估计方法容易产生较大偏差[2]。在小子样情况下,如何更为科学准确地进行参数估计,成为相关人员致力研究的问题。

本文提出了一种将随机加权法与 Bayes 估计相结合的新方法,即在 Bayes 评估框架下,利用随机加权重采样技术获得兴趣变量的先验分布,再与当前样本结合从而进行可靠性参数估计。

2　基于随机加权的 Bayes 方法的基本原理

贝叶斯方法是基于观测分布对总体进行参数估计的一种统计方法,其前提是任一未知参数都可以看作一个由先验分布去描述的随机变量,通过计算先验分布与观测样本的联合分布得到参数的后验分布。贝叶斯方法与经典方法的主要区别就在于是否采用先验分布,其成功的关键在于对先验信息的准确把握,如果未知变量的先验分布信息不够准确,容易导致结果的偏差[1]。因此在贝叶斯统计推断中,先验信息虽然可以来源于任何领域,可以是客观的,也可以是主观的,但必须具有足够的可靠性,否则会导致不合理的结论。

随机加权法是郑忠国教授在 Bootstrap 方法基础上提出的一种新的统计分析方法。它直接利用样本数据,借助于近代计算技术,通过对观测数据的重新抽样产生再生样本,来模拟总体分布。其主要思想如下:

设 $X=(x_1,x_2,\cdots,x_n)$ 为来自于某未知分布 F 的样本,$\theta=\theta(F)$ 为总体分布的未知参数,F_n 为抽样分布函数,$\hat{\theta}=\hat{\theta}(F_n)$ 为 θ 的估计,估计误差为 $R_n=\hat{\theta}(F_n)-\theta(F)$。通过计算机产生服从 Dirichlet 分布 $D(1,1,\cdots,1)$ 的随机变量 $(V_1,V_2,\cdots,V_n)$,得出 $\hat{\theta}_v=\hat{\theta}\left(\sum_{i=1}^{n}V_if_i(X)\right)$,其中 $f_i(X)$ 是 X 的某个 Borel 函数。记 $D_n=\hat{\theta}_v-\hat{\theta}(F_n)$,称它为随机加权统计量,以 D_n 的分布模仿 T_n 的分布,这就是所谓随机加权法[5]。

作为一种重采样技术,随机加权法非常有利于确定未知分布参数的验前分布,因此本文将其与 Bayes 方法结合,以使在估计分布函数的参数时有更高的精度[6],其基本流程如图 1 所示。

3　可靠性参数估计算法

3.1　寿命分布函数

产品的寿命分布函数是反映产品可靠性的一个重要指标,用于描述产品发生故障的统计规律,许多关于可靠性的统计推断都要在这个基础上进行。在以往样本量充分的情况下,多采用经典的参数估计方法。若样本数据较少的情况下,则可以将分布函数中的参数视为随机变量,通过随机加权重抽样获得其验前分布,然后再与当前样本结合用 Bayes 方法进行统计推断。

作者简介: 白永生(1982-),男,硕士研究生,主要研究方向为 RCM 决策建模及数据分析;温亮(1981-),讲师,主要研究方向为维修理论及建模。

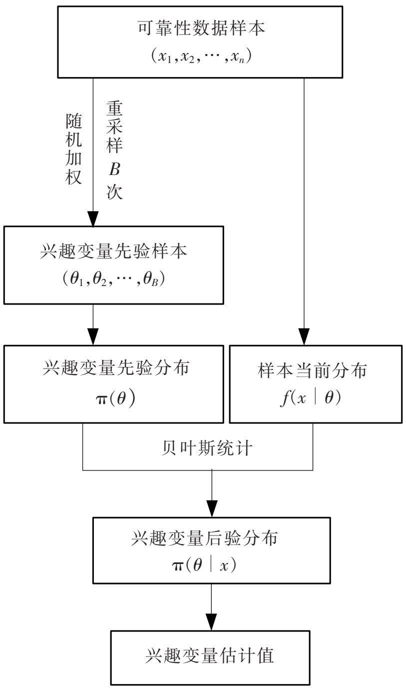

图 1 可靠性参数估计基本流程

下面以产品寿命服从正态分布为例来说明该方法的具体应用。

在正态分布条件下，产品概率密度函数为

$$f(t)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

按照贝叶斯统计推断的思想，μ、σ^2 也为随机变量，需要确定其先验分布 $\pi(\mu)$ 和 $\pi(\sigma^2)$，然后综合已知样本数据得到其后验分布 $\pi(u,\sigma^2|t)$，以确定 μ、σ^2 的值。

3.2 基于随机加权的 Bayes 可靠性参数估计

(1)通过随机加权重采样确定先验分布

随机加权法实质上是一个再抽样过程，通过对目前小样本数据 Borel 函数的加权处理，得到其再生样本，以获得其分布信息。这里，需根据小样本确定寿命分布函数中均值 μ 和 σ^2 方差的先验分布。

假设已知产品寿命数据的几个 i.i.d 样本为：$(t_1,t_2,\cdots,t_n)$，据此可估计出其均值 $\bar{t}$ 和方差 S_t^2。由于估计值和真实值之间存在偏差，记

$$R_n^{(1)}=\bar{t}-\mu,R_n^{(2)}=\frac{n}{n-1}S_t^2-\sigma^2$$

对于 $R_n^{(1)}$、$R_n^{(2)}$ 分别构造随机加权统计量：

$$D_n^{(1)}=\sum_{i=1}^{n}V_it_i-\bar{t},D_n^{(2)}=\frac{n}{n-1}\sum_{i=1}^{n}V_i(t_i-\bar{t})^2-\frac{n}{n-1}S_t^2$$

其中 $(V_1,V_2,\cdots,V_n)$ 是服从 $D(1,1,\cdots,1)$ 分布的 Dirichlet 随机变量[4]。

可证

$$E\lfloor D_n^{(1)}\rfloor=E\lfloor R_n^{(1)}\rfloor,E\lfloor D_n^{(2)}\rfloor=E\lfloor R_n^{(2)}\rfloor$$

所以按照数学期望的观点，可以用 $D_n^{(1)}$、$D_n^{(2)}$ 的分布来模仿 $R_n^{(1)}$、$R_n^{(2)}$ 的分布。

因为 $\mu=\bar{t}-R_n^{(1)},\sigma^2=\frac{n}{n-1}S_t^2-R_n^{(2)}$

由此可得 μ 和 σ^2 的估计值

$$\hat{\mu}=\bar{t}-D_n^{(1)},\hat{\sigma}^2=\frac{n}{n-1}S_t^2-D_n^{(2)}$$

取 B 组 Dirichlet $D(1,1,\cdots,1)$ 随机向量，相应地计算出 B 组随机加权子样 $D_n^{(1)}(i)$、$D_n^{(2)}(i)(i=1,2,\cdots,B)$。由于当 B 取无穷大时，$\bar{D}_n^{(1)}$ 为 $E\lfloor D_n^i\rfloor(i=1,2)$ 的无偏一致估计，所以可以分别用

$$(\hat{\mu}(1),\hat{\mu}(2),\cdots,\hat{\mu}(B)),(\hat{\sigma}^2(1),\hat{\sigma}^2(2),\cdots,\hat{\sigma}^2(B))$$

通过直方图，得到 μ、σ^2 的先验分布 $\pi(\mu)$、$\pi(\sigma^2)$。重抽样的次数通常应取得较大，一般应在 1 000 次以上，本文取 B=1 000。

(2)Bayes 方法确定后验分布

在贝叶斯框架下，结合当前样本分布，得出后验分布的表示形式，从而计算 μ、σ^2 的值。

$$\pi(\mu,\sigma^2|t)=\frac{f(t|\mu,\sigma^2)\pi(\mu)\pi(\sigma^2)}{\int_0^\infty\int_0^\infty f(t|\mu,\sigma^2)\pi(\mu)\pi(\sigma^2)\mathrm{d}\mu\mathrm{d}\sigma^2}$$

其中，样本的概率密度函数为

$$f(t|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

(3)可靠性参数的计算

通过当前样本与先验分布的联合分布，利用极大似然方法进行参数估计[3]。

首先构造似然函数为：

$$L=\prod_{i=1}^{n}\pi(\mu,\sigma^2|t_i)=\prod_{i=1}^{n}\frac{f(t_i|\mu,\sigma^2)\pi(\mu)\pi(\sigma^2)}{\int_0^\infty\int_0^\infty f(t_i|\mu,\sigma^2)\pi(\mu)\pi(\sigma^2)\mathrm{d}\mu\mathrm{d}\sigma^2}$$

因为分母为常数，所以 $L\propto\prod_{i=1}^{n}f(t_i|\mu,\sigma^2)\pi(\mu)\pi(\sigma^2)$，取

$$L'=\prod_{i=1}^{n}f(t_i|\mu,\sigma^2)\pi(\mu)\pi(\sigma^2)$$

由方程组 $\begin{cases}\dfrac{\partial \mathrm{Log}L'}{\partial\mu}=0\\ \dfrac{\partial \mathrm{Log}L'}{\partial\sigma^2}=0\end{cases}$

联立求解即可得到 π、σ^2 的值。

4 仿真算例

根据以上算法，用 Matlab 编写了随机加权重采样和可靠性参数 Bayes 计算的程序对其有效性进行仿真验算[7]。

首先利用计算机产生 5 个服从正态分布 $N(\mu,\sigma^2)$ 的一组随机数，通过随机加权重抽样得到兴趣变量的分布函数，然后与当前分布结合，即可得到参数的解。为了对比明显，总体均值均取 μ=400，而方差取不同的值。分别通过经典估计方法和基于随机加权的 Bayes 方法估计参数，对比其结果(见表 1)。

表 1 经典、基于随机加权的 Bayes 方法参数估计对比表

$N(\pi,\sigma^2)$	经典法(π,σ^2) 估计值	基于随机加权的 Bayes 法 μ 的先验分布	σ^2 的先验分布	(π,σ^2)估计值
$(400,1^2)$	(401.81, 0.68)	$N(399.37,1.23^2)$	$N(0.74,1.44^2)$	(401.18, 0.89)
$(400,2^2)$	(399.12, 2.48)	$W(2.70,3.45,394.70)$	$N(1.88,1.79^2)$	(399.80, 5.09)
$(400,3^2)$	(401.20, 6.13)	$N(400.15,0.74^2)$	$N(7.15,2.01^2)$	(400.10, 7.35)
$(400,4^2)$	(397.73,11.08)	$N(397.68,1.11^2)$	$W(4.88,15.40,2.30)$	(398.32,16.29)
$(400,5^2)$	(400.48,20.56)	$N(400.55,1.68^2)$	$N(30.77,8.52^2)$	(400.37,27.09)

(下转 233 页)