第36卷第6期 国防科技大学学报 Vol. 36 No. 6
2014年12月 JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY Dec. 2014

doi:10.11887/j.cn.201406020 http://journal.nudt.edu.cn

寿命分布的参数 Bootstrap 拟合优度检验方法*

孙 权,周 星,冯 静,潘正强
(国防科技大学 信息系统与管理学院,湖南 长沙 410073)

摘 要: 拟合优度检验在统计和可靠性等领域具有非常重要的地位,基于参数 Bootstrap 重采样的思想,对未知参数的常用寿命分布进行拟合优度检验。数值仿真结果表明,相对于传统的经验分布函数检验,这种基于参数 Bootstrap 的拟合优度检验具有更高的功效,特别是在小样本的情况下,优势明显。

关键词: 参数 Bootstrap;拟合优度检验;寿命分布
中图分类号: TB114.3 **文献标志码:** A **文章编号:** 1001-2486(2014)06-112-05

Goodness-of-fit test for life distributions based on parametric Bootstrap

SUN Quan, ZHOU Xing, FENG Jing, PAN Zhengqiang
(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: Goodness-of-fit test plays an important role in the statistics and reliability fields. Goodness-of-fit test for the common life distribution of unknown parameters was based on the resembling idea of parametric bootstrap. Simulation results show that goodness-of-fit test based on parametric bootstrap is superior to the traditional EDF tests, such as Liefors test, CvM test and AD test, especially in small samples.

Key words: parametric Bootstrap; goodness-of-fit test; life distribution

在进行产品可靠性分析中,常常需要根据实际的样本数据对产品的寿命分布做出假设,以便进行随后的研究分析。如果假设的寿命分布根本不能反映或者不能完全反映出实际寿命数据的特性,那么基于假设分布的后续研究结果显然是不可靠的。因此,在可靠性等领域,对产品的寿命分布进行拟合优度检验是一项尤为必要的工作。

在选择合适的分布类型时,研究者需要做寿命分布的拟合优度检验。传统建立在经验分布函数上(Empirical Distribution Function, EDF)拟合优度检验方法,如 Cramer - von Mises(CvM)检验[1]、Anderson - Darling(AD)检验[2]等,在假设观测到的寿命数据服从某一特定分布函数的条件下,通过与其经验分布函数进行对比,构造一些已知分布或已知渐进分布的检验统计量,然后由观测数据计算得到统计量的数值,并与在一定显著性条件 α 下的检验临界值相比较,从而决定是否接受原先的寿命分布假设。

著名的 KS 检验的统计量表示为

$$D_n = \sqrt{n}\sup_x |F(x) - F_n(x)| \tag{1}$$

其中,$x_1, x_2, \cdots, x_n$ 为一组按升序排列的观测样本,$F_n(x)$ 为其经验分布函数,$F(x)$ 是假设的理论分布函数。在 n 取特定数值时,统计量 D_n 具有形式复杂的精确分布,当 $n\to\infty$,检验统计量 D_n 具有极限分布。

对于给定的样本数据,KS 检验统计量可以写为

$$D_n = \sqrt{n}\max_{1\leqslant i\leqslant n}\left\{F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i)\right\} \tag{2}$$

在很多情况下,Cramér 提出的 CvM 检验具有比 KS 检验更高的检验功效,CvM 型检验统计量定义为

$$\omega_n^2 = n\int_{-\infty}^{+\infty} [F(x) - F_n(x)]^2 w(x)\,\mathrm{d}F(x) \tag{3}$$

当加权函数 $w(x)=1$ 时,便得到 CvM 检验统计量

$$W_n^2 = n\int_{-\infty}^{+\infty} [F(x) - F_n(x)]^2\,\mathrm{d}F(x) \tag{4}$$

对于给定的样本数据,CvM 检验统计量可以写为

* 收稿日期:2014-06-07
基金项目:国家自然科学基金资助项目(71271212,61273041)
作者简介:孙权(1973—),男,湖北咸宁人,教授,博士,博士生导师,E-mail:sunquan@nudt.edu.cn

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(x_i) - \frac{2i-1}{2n}\right]^2 \tag{5}$$

在此基础上,Anderson 和 Darling 于 1954 年提出了 AD 统计量

$$A_n = n\int_{-\infty}^{+\infty}\frac{[F(x) - F_n(x)]^2}{F(x)[1 - F(x)]}\mathrm{d}F(x) \tag{6}$$

对于给定的样本数据,AD 检验统计量可以写为

$$A_n = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\ln[F(x_i)] + (2n+1-2i)\ln[1 - F(x_i)]] \tag{7}$$

通常,在零假设的分布函数不含未知参数完全确定的情况下,可以得到这些 EDF 检验统计量的极限分布,然而特定样本容量下的准确分布却较难得到或者形式相当复杂。此外,在对不同的分布进行假设检验时,同一种检验方法的功效会有较大差异。所以,为了得到较高的统计功效,对不同的分布假设进行拟合优度检验时,需要使用不同的统计量和不同的检验方法。因此,传统的 EDF 拟合优度检验方法的推广能力不够强。

特别地,当对含有未知参数的分布函数进行复合假设检验时,EDF 检验统计量的分布不仅依赖于样本容量,也与零假设的分布类型甚至与未知参数有关。需要指出的是,在对含有未知形状参数的分布进行拟合优度检验时,检验统计量的分布与未知的形状参数有关。这将导致一定显著性水平下的准确检验临界值很难获得。

比如,KS 检验方法只能在假设的分布完全已知的情况下,进行拟合优度检验;随后 David 和 Johnson 指出在仅含有未知位置和尺度参数的情况下,可以通过极大似然法估计分布的参数,然后利用 KS 检验统计量进行分布的拟合优度检验[3]。之后,Lilliefors 用蒙特卡洛法给出了含未知均值和方差的正态分布 KS 检验临界值表[4]和含未知参数的指数分布 KS 检验临界值表[5]。Stephens 在 1974 年给出了"均值未知方差已知的正态分布""均值已知方差未知的正态分布""均值方差均未知的正态分布"以及"未知均值的指数分布"4 种未知参数检验的常用 EDF 检验统计量的临界值表[6]。随后,Woodruff 等又给出在已知形状参数而未知位置和尺度参数的情况下的 Weibull 分布和 Gamma 分布的修正检验统计量的临界值表[7-8]。

在未知分布参数的情况下,本文基于参数 Bootstrap 思想,提出了对常用寿命分布进行拟合优度检验的方法。对于不同分布类型的假设检验,最适合的检验统计量虽不相同,但由于此类方法只需简单的数值计算便可得到相应的检验临界值,无须推导检验统计量所服从的准确分布或渐进分布,所以具有很好的适用性和推广能力。

1　基于参数 Bootstrap 的未知参数寿命分布的拟合优度检验

Bootstrap 是 Stanford 大学的 Efron 教授提出的一种统计估计方法[9]。经过几十年的发展,总的来说,可以分为参数和非参数两种 Bootstrap 重采样方法。其中,参数 Bootstrap 重采样方法的思想为存在一组观测样本 $X_1, X_2, \cdots, X_n$,若总体分布类型为 $F(X,\theta)$,首先根据观测样本估计参数 $\widehat{\theta_n}$,得到估计分布函数 $F(X,\widehat{\theta_n})$,然后从中进行采样,得到采样数据 $X_1^*, X_2^*, \cdots, X_n^*$。

由于参数 Bootstrap 重采样方法是在假设观测样本的分布类型已知的情况下进行重采样,通过比较重采样数据与实际观测样本数据在统计意义上是否相近,可以判断原假设的分布类型是否可信。

基于参数 Bootstrap 方法,本文根据以下思路,对未知参数的寿命分布进行拟合优度检验。

假设在某次试验中,我们观测到 m 个产品的失效,寿命分别为 $T_1, T_2, \cdots, T_m$。产品的真实寿命分布表示为 G,仅假设其属于分布函数族 M,参数未知,则拟合优度问题可表示为:

$$H_0: G \in M$$

$$H_1: G \notin M$$

定义经验分布函数:

$$G_m(t) = \frac{1}{m}\sum_{i=1}^{m} I(t \geqslant T_i) \tag{8}$$

根据观测到的寿命数据 $T_1, T_2, \cdots, T_m$,用极大似然的方式估计出分布函数 M 的参数向量 $\hat{\beta}$,得到估计分布函数 $M(\hat{\beta})$,并将适用于此分布类型的检验统计量记为 S。

基于参数 Bootstrap 的拟合优度检验方法可以表述为如下步骤:

Step1:根据观测到的寿命数据 $T_1, T_2, \cdots, T_m$,得到分布函数的极大似然估计 $M(\hat{\beta})$,并计算相应的检验统计量 S。

Step2:用 step 1 中得到的估计分布 $M(\hat{\beta})$ 产生样本量为 m 的重采样样本 $T_1^*, T_2^*, \cdots, T_m^*$。

Step3:用 step 2 中产生的重采样样本 $T_1^*, T_2^*, \cdots, T_m^*$ 重新做极大似然估计,得到估计分布

$M(\hat{\beta}')$，并重新计算检验统计量 S。

Step4：重复 B 次 step 2 和 step 3 中的步骤，得到统计量序列 $\{S^k, k=1,2,\cdots,B\}$。

Step5：对序列 $\{S^k, k=1,2,\cdots,B\}$ 按升序排序，得到集合 C。在显著性水平 α 下，C 在 α 上的分位数作为此次检验的临界值，若 step1 中计算得到的统计量 S 大于此临界值，则拒绝假设 H_0，反之接受假设 H_0。

在参数估计的一致性和零假设分布的光滑性满足一定的正则条件下，Stute 等[10] 证明了参数 Bootstrap 拟合优度检验的有效性，并指出 D_n 和 W_n 等 EDF 检验统计量的实际分布可由 Bootstrap 方法近似。

事实上，上述正则条件并不严格，常用的寿命分布类型和参数估计方法都能满足文献［13］中所述的条件。所以，将参数 Bootstrap 方法用于寿命分布的拟合优度检验从理论上讲是可行的。

但与传统的 EDF 检验方法相比，特别是当应用于常用寿命分布的检验问题时，参数 Bootstrap 检验方法是否具有较高的检验功效是值得进一步探讨的。只有当参数 Bootstrap 检验方法相对于传统 EDF 检验具有较高的检验功效时，其在寿命分布的检验问题中才具有一定的应用价值。

2 寿命分布检验的功效验证

为了说明基于参数 Bootstrap 拟合优度检验方法的有效性，本文针对正态分布、指数分布等常用寿命分布，进行了拟合优度检验。通过与传统的 EDF 检验方法进行对比，验证了其对寿命分布检验的良好功效。本节的参数 Bootstrap 拟合优度检验选择 CvM 检验统计量 $W_n = \frac{1}{12n} + \sum_{i=1}^{n}\left[\hat{M}(T_i,\beta) - \frac{2i-1}{2n}\right]^2$ 进行假设检验。

在显著性为 α 的条件下，当备择假设成立而零假设不成立时，若某拟合优度检验拒绝零假设的概率为 $1-\beta$，则认为此次检验的功效为 $1-\beta$。β 为犯第二类错误的概率。所以，拟合优度检验的功效既依赖于显著性 α，又依赖于具体的备择假设。

一般来说，理论功效值较难求出。本文采取数值仿真的方式进行计算，具体的步骤如下：

Step1：根据备择假设，抽取容量为 n 的简单随机样本 $x_1, x_2, \cdots, x_n$。

Step2：对简单随机样本 $x_1, x_2, \cdots, x_n$ 进行拟合优度检验，记录零假设是否被拒绝。

Step3：重复 step1 与 step2 共 m 次，统计零假设被拒绝的次数为 k。

Step4：计算比值 $\frac{k}{m}$，当 m 足够大时，$\frac{k}{m}$ 可作为此拟合优度检验的功效近似值。

在零假设为未知参数的正态分布，备择假设分别为特定的伽马分布、指数分布、威布尔分布、对数正态分布的情况下，分别进行了传统的 EDF 拟合优度检验和基于参数 Bootstrap 的拟合优度检验，并通过上述数值仿真的步骤计算了各检验方法的功效值。

令样本数 $n=10,15,20,25,30,35,40,45,50$，均使 $m=1000$，即每种检验重复 1000 次。在基于参数 Bootstrap 的拟合优度检验中，令重采样次数 $B=500$。在以下基于传统的 EDF 检验中，本文使用 Stephens 给出的修正统计量的临界值表[9]。

在零假设为未知参数的正态分布时，可得到功效对比如表 1 所示。

表 1 未知均值和方差的正态分布拟合优度检验功效对比表

Tab. 1 The efficiency of the goodness-of-fit tests for the normal distribution with unknown means and variances

备择假设	检验方法	样本大小 n								
		10	15	20	25	30	35	40	45	50
伽马分布（1,2）	KS	0.317	0.465	0.582	0.69	0.799	0.839	0.899	0.933	0.956
	CvM	0.394	0.554	0.733	0.813	0.874	0.950	0.967	0.982	0.99
	AD	0.004	0.221	0.549	0.706	0.83	0.931	0.964	0.981	0.995
	Bootstrap	0.401	0.569	0.737	0.829	0.897	0.936	0.96	0.981	0.991

（续表）

备择假设	检验方法	样本大小 n								
		10	15	20	25	30	35	40	45	50
伽马分布（2,2）	KS	0.171	0.243	0.317	0.399	0.464	0.543	0.594	0.634	0.688
	CvM	0.207	0.307	0.421	0.508	0.580	0.695	0.742	0.803	0.83
	AD	0.030	0.072	0.206	0.372	0.469	0.622	0.696	0.781	0.824
	Bootstrap	0.319	0.434	0.536	0.604	0.724	0.747	0.824	0.833	0.895
伽马分布（3,2）	KS	0.144	0.162	0.209	0.29	0.335	0.377	0.406	0.498	0.534
	CvM	0.133	0.212	0.281	0.365	0.389	0.503	0.531	0.615	0.637
	AD	0.010	0.028	0.105	0.218	0.279	0.412	0.467	0.573	0.612
	Bootstrap	0.230	0.325	0.435	0.522	0.564	0.657	0.685	0.721	0.811
指数分布（1）	KS	0.282	0.46	0.566	0.7	0.787	0.859	0.909	0.926	0.967
	CvM	0.362	0.551	0.749	0.838	0.912	0.926	0.965	0.974	0.99
	AD	0.102	0.214	0.525	0.736	0.867	0.912	0.95	0.976	0.992
	Bootstrap	0.403	0.569	0.714	0.805	0.906	0.945	0.957	0.983	0.985
威布尔分布（1,2）	KS	0.068	0.09	0.112	0.122	0.128	0.159	0.187	0.187	0.244
	CvM	0.061	0.097	0.102	0.147	0.181	0.200	0.22	0.222	0.278
	AD	0	0.003	0.027	0.063	0.112	0.106	0.165	0.161	0.214
	Bootstrap	0.137	0.172	0.171	0.244	0.25	0.32	0.324	0.375	0.395
对数正态（1,1）	KS	0.470	0.670	0.758	0.876	0.929	0.963	0.9890	0.981	0.995
	CvM	0.560	0.742	0.89	0.943	0.972	0.991	0.998	0.994	1
	AD	0.026	0.46	0.756	0.909	0.949	0.987	0.999	0.995	1
	Bootstrap	0.552	0.757	0.862	0.955	0.976	0.986	0.998	0.996	1

当零假设为未知均值的指数分布时，可以得到功效对比如表 2 所示。

表 2　均值未知的指数分布拟合优度检验功效对比表

Tab. 2　The efficiency of the goodness-of-fit tests for the exponential distribution with unknown means

备择假设	检验方法	样本大小 n								
		10	15	20	25	30	35	40	45	50
伽马分布（2,2）	KS	0.221	0.285	0.392	0.4640	0.584	0.677	0.744	0.788	0.850
	CvM	0.246	0.362	0.456	0.596	0.689	0.751	0.805	0.872	0.876
	AD	0.195	0.316	0.437	0.583	0.688	0.756	0.813	0.892	0.893
	Bootstrap	0.249	0.336	0.478	0.568	0.677	0.747	0.831	0.857	0.903
伽马分布（3,2）	KS	0.469	0.672	0.814	0.903	0.961	0.979	0.986	0.996	0.998
	CvM	0.524	0.736	0.890	0.942	0.982	0.991	0.997	0.999	0.999
	AD	0.470	0.713	0.882	0.944	0.981	0.997	0.999	0.999	0.999
	Bootstrap	0.559	0.782	0.882	0.954	0.983	0.995	0.996	1	1
正态分布（1,1）	KS	0.887	0.944	0.980	0.996	0.998	1	1	1	1
	CvM	0.889	0.955	0.983	0.997	0.999	0.999	1	1	1
	AD	0.877	0.950	0.982	0.997	0.999	1	1	1	1
	Bootstrap	0.885	0.951	0.984	0.997	0.998	1	1	1	1
威布尔分布（1,2）	KS	0.494	0.738	0.858	0.919	0.967	0.988	0.992	1	1
	CvM	0.548	0.810	0.913	0.968	0.981	0.987	0.995	0.999	1
	AD	0.516	0.784	0.916	0.967	0.989	0.999	1	1	1
	Bootstrap	0.593	0.853	0.928	0.970	0.990	0.997	0.998	1	1
对数正态（1,1）	KS	0.072	0.102	0.113	0.150	0.164	0.200	0.214	0.251	0.265
	CvM	0.092	0.123	0.152	0.176	0.196	0.185	0.250	0.268	0.286
	AD	0.092	0.123	0.152	0.176	0.196	0.185	0.250	0.268	0.286
	Bootstrap	0.093	0.133	0.133	0.174	0.187	0.231	0.244	0.289	0.299

整体上来看,在对均值和方差未知的正态分布进行检验时,CvM 检验和参数 Bootstrap 检验方法有较高的检验功效,并且参数 Bootstrap 检验方法在小样本的情况下,更具优势。特别当备择假设是伽马分布(2,2)、伽马分布(3,2)、威布尔分布(1,2)时,参数 Bootstrap 检验方法优势明显。此外,当备择假设为指数分布(1)和对数正态分布(1,1)时,参数 Bootstrap 方法的检验功效与 CvM 检验相当。

在对未知均值的指数分布进行检验时,4 种检验方法的功效大致相当,CvM、AD、参数 Bootstrap 方法略好于 KS 方法。

在进行 KS 检验、CvM 检验和 AD 检验时,本文使用的是由 Stephens 通过蒙特卡洛法给出临界值表。从以上功效对比表也可看出,参数 Bootstrap 方法要稍好于蒙特卡洛法。参数 Bootstrap 的重采样样本来自于由原始样本数据估计出的分布函数,生成的统计量的近似分布不仅与零假设有关还与原始的样本数据有关。而蒙特卡洛仿真是由预先设定的固定参数的分布函数产生样本,生成的统计量的近似分布只与零假设有关,这可能导致蒙特卡洛法的功效相对较低。此外,对比 KS 检验、CvM 检验和 AD 检验的检验功效可发现,在对正态分布的零假设进行检验时,CvM 检验统计量具有明显优势;在对指数分布的零假设进行检验时,三者功效相当。然而,对于不同的检验问题,参数 Bootstrap 检验方法可以灵活地选择合适的检验统计量,这也是其很好的应用优势。

3 结论

从理论和数值实验均可看出,参数 Bootstrap 方法可以较好地近似检验统计量的实际分布,避免了复杂的分布推导,从而方便地获取检验统计量的临界值,完成拟合优度检验。基于 Bootstrap 的重采样思想为分布检验乃至模型选择都提供了新的选择,这也是有待进一步研究的领域。

参考文献(References)

[1] Cramér H. On the composition of elementary errors [J]. Scandinavian Actuarial Journal, 1928(1): 141 – 180.

[2] Anderson T W, and Darling D A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes[J]. The Annals of Mathematical Statistics, 1952, 23(2): 193 – 212.

[3] David F N, Johnson N L. The probability integral transformation when parameters are estimated from the sample[J]. Biometrika, 1948, 35(1/2): 182 – 190.

[4] Lilliefors H M. On the kolmogorov-smirnov tests for normality with mean and variance unknown[J]. Journal of the American Statistical Association, 1967, 62(318): 399 – 402.

[5] Lilliefors H M. On the kolmogorov-smirnov tests for exponential distribution with unknown mean[J]. Journal of the American Statistical Association, 1969, 64(325): 387 – 389.

[6] Stephens M A. EDF statistics for goodness of fit and some comparisons[J]. Journal of the American Statistical Association, 1974, 69(347): 730 – 737.

[7] Woodruff B W, Moore A H, Dunne E J, et al. A modified kolmogorov-smirnov test for weibull distribution with unknown location and scale parameters [J]. IEEE Transaction on Reliability, 1983, R – 32(2): 209 – 213.

[8] Woodruff B W, Viviano P J, Moore A H, et al. Modified goodness-of-fit tests for gamma distribution with unknown location and scale parameters [J]. IEEE Transaction on Reliability, 1984, R – 33(3): 241 – 245.

[9] Efron B. Bootstrap methods: another look at the jackknife[J]. Annals of Statistics, 1979, 7(1): 1 – 26.

[10] Stute W, González-Manteiga W, Presedo-Quindimil M. Bootstrap based goodness-of-fit tests [J]. Metrika, 1993, 40(1): 243 – 256.