

相关系数对软件可靠性验证测试的影响分析

马振宇¹, 吴 伟², 张 威¹, 刘福胜¹, 韩 坤²

(1 装甲兵工程学院, 北京 100072; 2 北京特种车辆研究所, 北京 100072)

摘 要: 针对在软件可靠性验证测试过程中传统贝叶斯方法存在着不区分历史信息与试验信息的问题, 提出了基于混合贝叶斯软件可靠性验证测试方案. 引入相关系数概念, 分别通过主观定量法、拟合优度法、Kullback 信息量法、Spearman 相关系数法以及 Kendall 相关系数法确定混合贝塔分布中的相关系数. 最后通过实验分析, 证明了这 5 种计算相关系数方法的可行性, 并且同时证明了区分历史信息和试验信息, 会对软件可靠性验证测试的可接收质量水平产生影响.

关键词: 混和贝叶斯; 相关系数; 软件可靠性; 可靠性验证测试

中图分类号: TP311.5

文献标识码: A

文章编号: 1000-7180(2018)05-0069-05

Analysis of Correlation Coefficient Influence on Software Reliability Demonstration Testing

MA Zhen-Yu¹, WU Wei², ZHANG Wei¹, LIU Fu-sheng¹, HAN Kun²

(1 Academy of Armored Force Engineering, Beijing 100072, China;

2 Institute of Special Vehicle, Beijing 100072, China)

Abstract: In the process of software reliability demonstration testing there is a problem that traditional Bayesian method did not distinguish between historical information and experimental information. A hybrid Bayesian software reliability demonstration testing scheme is proposed. By introducing the concept of correlation coefficient, the correlation coefficients in hybrid Beta distribution are determined by subjective quantitative method, goodness of fit method, Kullback information method, Spearman correlation coefficient method and Kendall correlation coefficient method respectively. Finally, through experimental analysis, prove the feasibility of the 5 kinds of calculation method of correlation coefficient, and prove that differentiating the historical information and experimental information impacts acceptable quality level of software reliability demonstration testing.

Key words: hybrid Bayesian; correlation coefficient; software reliability; reliability demonstration testing

1 引言

软件可靠性验证测试^[1]是验收软件产品必须经过的步骤, 为了降低可靠性验证测试的成本, 测试员们越来越多地选用贝叶斯方法. 但是传统的贝叶斯方法默认先验信息中的历史信息和试验信息来自同一个样本, 从未实际考虑过历史信息服从样本和试验信息服从样本之间的相关程度. 本文从混和贝叶斯方法的理论出发, 提出基于混合贝叶斯方法^[2-3]的软件可靠性验证测试方案. 通过引入相关系数概念,

针对软件可靠性验证测试过程中已使用的主观定量法、拟合优度法、Kullback 信息量法, 类比地将 Spearman 相关系数法和 Kendall 相关系数法应用到可靠性验证测试中. 通过实验分析, 验证了这两种方法的可行性. 任意相关系数计算方法, 均比传统贝叶斯方法更贴近实际反映出的验证测试结果. 充分考虑历史信息和试验信息, 经过对 5 种计算方法的对比分析, 可以得到相关系数决定着软件可靠性验证测试的可接收质量水平.

2 相关工作

2.1 贝叶斯基本理论

在可靠性验证测试过程中,选用失效概率 p 作为可靠性指标.那么测试过程里出现的失效次数的条件分布密度函数为 $f(r|p)$.设 $h(p)$ 为 p 的先验分布密度函数.那么失效次数和失效概率的联合分布密度函数就为:

$$g(r, p) = f(r|p)h(p) \quad (1)$$

通过式(1)可得到关于失效次数 r 的边缘分布,即:

$$p(r) = \int g(r, p) d\lambda = \int f(r|p)h(p) dp \quad (2)$$

联立式(1)和式(2),进一步推导出失效概率的后验分布密度函数,即:

$$w(p|r) = \frac{g(r, p)}{p(r)} = \frac{f(r|p)h(p)}{\int f(r|p)h(p) dp} \quad (3)$$

2.2 混合贝叶斯软件可靠性验证测试方案

根据贝塔分布对软件的可靠性进行验证测试.

$$\begin{aligned} f(p|r, n, a, b) &= \frac{C_n^r p^r (1-p)^{n-r} f(p)}{\int_0^1 C_n^r p^r (1-p)^{n-r} f(p) dp} \\ &= \frac{p^{a+r-1} (1-p)^{b+n-r-1} \rho + p^r (1-p)^{n-r} (1-\rho) B(a, b)}{\rho \int_0^1 p^{a+r-1} (1-p)^{b+n-r-1} dp + (1-\rho) B(a, b) \int_0^1 p^r (1-p)^{n-r} dp} \\ &= \frac{p^{a+r-1} (1-p)^{b+n-r-1} \rho + p^r (1-p)^{n-r} (1-\rho) B(a, b)}{\rho B(a+r, b+n-r) + (1-\rho) B(a, b) B(r+1, n-r+1)} \\ \text{令 } M &= (1-\rho) B(a, b) B(r+1, n-r+1), N = \rho B(a+r, b+n-r) \\ \text{即: } f(p|r, n, a, b) &= \frac{N \text{Beta}(a+r, b+n-r) + M \text{Beta}(r+1, n-r+1)}{M+N} \end{aligned} \quad (5)$$

根据实际情况,制定软件可靠性验证测试方案.设定软件可靠性测试中执行测试用例的总数为 n ,未通过测试用例的个数为 r ,置信度为 $c = 1 - \alpha$.结合式(5),求出满足下式的可接收质量水平 p_0 .

$$\begin{aligned} &\int_0^{p_0} f(p|x=r, n, a, b) dp \\ &= \int_0^{p_0} \frac{N \text{Beta}(a+r, b+n-r)}{M+N} + \frac{M \text{Beta}(r+1, n-r+1)}{M+N} dp \\ &= c \end{aligned} \quad (6)$$

2.3 相关系数

无论是两个变量,还是两个数组,或是两个分布之间都会存在着某种非确定性的关系,为了量化两者之间的相关程度,提出了相关系数这个概念.相关系数^[5]可以真实客观地反映变量、数组、分布各自之

通常情况下,测试员认为更新后的软件和原软件在可靠性上是没有区别的,即认为历史信息 and 试验信息来自同一总体.但是在更新软件产品中,新旧产品之间的可靠性一定存在不一样的地方,那么历史信息和试验信息有可能来自不一样的总体.所以运用混合贝叶斯方法,引入继承因子和更新因子,那么混合贝塔分布^[4]的混合先验分布为:

$$\begin{aligned} f(p) &= \rho \text{Beta}(a, b) + (1-\rho) \\ &= \rho \frac{p^{a-1} (1-p)^{b-1}}{B(a, b)} + (1-\rho) \end{aligned} \quad (4)$$

继承因子 ρ 体现了历史信息符合总体与试验信息符合总体的相关程度,因此本文后面都用相关系数这个概念代替继承因子这个概念,即 ρ 为相关系数.

在软件可靠性测试过程中,用 n 个测试用例进行遍历测试,共出现了 r 次未通过测试.再根据混合贝叶斯的先验分布,可以推导出失效概率的后验分布,即:

间的关联程度.本文给出常用的三种相关系数.

(1)Person 相关系数适用于有线性关系的数据,且整体上满足正态分布的连续变量.

(2)Spearman 相关系数对初始数据没有强制性要求,即数据之间不需要符合任意分布规律,就能够通过相关系数量化彼此间的相关程度.

(3)Kullback 相关系数是用来度量两个分类变量相关程度的指标,它能够针对两个有序变量进行分类,并且可以进行非参数相关度量.

3 相关系数计算方法

3.1 主观定量法

把相关系数 ρ 看为随机变量,同时构造关于 ρ 的概率密度函数 $\pi(\rho)$.本文选取 $\pi(p)$ 为均匀分布,进行相关系数对于软件可靠性验证测试结果的影响

分析^[6].

$$\text{设: } \pi(\rho) = \begin{cases} \frac{1}{\rho_2 - \rho_1}, & \rho_1 < \rho < \rho_2 \\ 0, & \text{其他} \end{cases} \quad (7)$$

式中 ρ_1 和 ρ_2 都是根据专家们的经验确定的取值.

进而推导出关于失效概率 p 的后验分布密度函数,即:

$$\begin{aligned} & w(p | n, r, \rho_1, \rho_2) \\ &= \frac{1}{\rho_2 - \rho_1} (\varphi(\rho_2) - \varphi(\rho_1)) \end{aligned} \quad (8)$$

其中:

$$\begin{aligned} \varphi(\rho) &= \frac{AD - BC}{(C - D)^2} \ln(C - (C - D)\rho) + \frac{A - B}{C - D} \rho \\ A &= p^r (1 - p)^{n-r} \\ B &= \frac{p^{a+r-1} (1 - p)^{b+n-r-1}}{B(a, b)} \end{aligned} \quad (9)$$

$$C = B(1 + r, 1 + n - r)$$

$$D = \frac{B(a + r, b + n - r)}{B(a, b)}$$

3.2 拟合优度检验法

不妨假设历史信息 (N, R) , 其中 N 为历史执行测试用例的测试用例数, R 为历史通过测试的测试用例数, 那么历史未通过测试的测试用例数就应该为 $N - R$; 试验信息 (n, r) , 其中 n 为试验执行测试用例的测试用例数, r 为试验通过测试的测试用例数, 那么试验未通过测试的测试用例数就应该为 $n - r$.

$$\text{设: } K = \frac{((n - r)R - (N - R)r)^2 (n + N)}{n N (n - r + N - R) (r + R)} \quad (10)$$

式中 K 为 Person 统计量. 在确定检验水平为 α 的条件下, 此时 K 作为检验历史信息与试验信息是否都来自同一个总体的度量指标.

只有当上式中的 n, r, N, R 均不小于 6 时, 此时 K 近似符合自由度为 1 的卡方分布. 那么当上式中的条件不满足时, 必须选用其他的 K 统计量. Yates 在原有的基础上对 K 做出了修正^[7], 即:

$$\begin{aligned} K &= \\ & \frac{(|(n - r)R - (N - R)r| - 0.5(n + N))^2 (n + N)}{n N (n - r + N - R) (r + R)} \end{aligned} \quad (11)$$

$$\text{设: } Q(K) = P(\chi_\alpha^2(1) > K) \quad (12)$$

式中 $Q(K)$ 为确定历史信息和试验信息的拟合优度.

下面找出拟合优度 $Q(K)$ 和相关系数 ρ 之间的关系式, 即:

$$\rho = f(Q) = Q(K)^i (i > 0) \quad (13)$$

式中 i 的取值由历史信息 and 试验信息共同决定, 通常情况下, 为了便于计算, 取 $i = 0.5$.

当有 n 个试验信息

$$\rho_i = f(Q_i) = \frac{Q_i}{\sum_{i=1}^n Q_i} (0 < i < n) \quad (14)$$

3.3 Kullback 信息量法

在空间 Ω 里, 确定历史信息所服从的分布 P , 以及试验信息所服从的分布 Q . 这两个分布之间的 Kullback 信息定义是:

$$I(Q; P) = \begin{cases} E_Q(\ln(\frac{dQ}{dP})), & Q \ll P \\ \infty, & \text{其他} \end{cases} \quad (15)$$

式中, E_Q 为 Q 的期望, $\frac{dQ}{dP}$ 是 Q 对 P 的 Radon-Nikodym 导数.

假如 Ω 为实数空间, 就可以得到以下关系式:

$$P(dx) = p(x)dx, Q(dx) = q(x)dx \quad (16)$$

那么与此同时, Kullback 的信息定义等同于

$$I(Q; P) = I(q; p) = \int q(x) \ln(\frac{q(x)}{p(x)}) dx \quad (17)$$

下面给出 Kullback 信息量和相关系数 ρ 之间的转换关系.

(1) 当存在有且只有一个试验信息

Kullback 信息量是测量两个分布之间的距离, 如果两个分布之间的相关程度越高, 那么 Kullback 的信息量就会越少. 同时存在 Kullback 信息量的取值不在区间 $[0, 1]$ 的问题, 相关系数是在区间 $[0, 1]$ 里的一个数. 根据函数 $f(x) = x/(x + 1)$ 的递增思想, 给出二者之间的关系式, 即:

$$\rho = 1 - \frac{|I(Q; P)|}{|I(Q; P)| + 1} \quad (18)$$

(2) 当有 n 个试验信息

$$\rho_i = \frac{\rho_i}{\sum_{i=1}^n \rho_i} (0 < i < n) \quad (19)$$

$$\text{其中: } \rho_i = 1 - \frac{|I(Q_i; P_i)|}{|I(Q_i; P_i)| + 1}$$

3.4 Spearman 相关分析法

通过 Spearman 等级相关系数, 量化历史性与试验信息之间的相关性. 把历史信息与试验信息的数据看成两个 N 维向量. 选取历史信息这组数据作为向量 $\mathbf{X}$, 并按照升序顺序依次排列, 其中 X_i 表示第 i 个值. 然后再对试验信息数据构成的向量 $\mathbf{Y}$, 通过第 i 个值所对应的 Y_i , 寻找向量 $\mathbf{Y}$ 里 Y_i 值所对应的升序排列位置 j , 与向量 $\mathbf{X}$ 第 i 位置相互对应. 然

后依次用 $X_i \sim Y_i$ 得到差值 d_i ,最后两个向量之间的 Spearman 等级相关系数可以定义为:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{N^3 - N} \tag{20}$$

3.5 Kendall 相关分析法

将历史信息 and 实验信息的数据分别看成数组 X 和数组 Y ,每组的元素个数都是 N 个. X_i, Y_i 用来表示第 i 个元素的值,同时构成以 (X_i, Y_i) 为元素形式的集合 XY . 对集合 XY 中任意两对元素进行比较时,如果出现 $X_i > X_j, Y_i > Y_j$ 或者 $X_i < X_j, Y_i < Y_j$,那么认为两对元素符合一致性;如果出现 $X_i > X_j, Y_i < Y_j$ 或者 $X_i < X_j, Y_i > Y_j$,那么认为两对元素满足非一致性;如果出现 $X_i = X_j, Y_i = Y_j$,那么这两对元素既不满足一致性也不满足非一致性. 由此可以得到 Kendall 相关系数,即:

$$\rho = \frac{A - B}{\sqrt{(a - b)(a - c)}} \tag{21}$$

式中, A, B 分别为集合 XY 中满足一致性元素以及非一致性元素的数量. $a = \frac{1}{2}N(N - 1), b = \sum_{i=1}^m \frac{1}{2}m_i(m_i - 1), c = \sum_{i=1}^n \frac{1}{2}n_i(n_i - 1)$. m, n 分别表示数组 X 以及数组 Y 中出现相同数值构成小集合的个数. m_i, n_i 表示第 i 个小集合中元素的个数.

4 实验分析

实验数据来自特种车辆软件测评中心. 根据测试人员收集、整理、统计的某软件的测试数据,通过当前的现场试验,得到实验分析所需数据,如表 1 所示.

表 1 数据表

	未通过测试用例数	通过测试用例数	执行测试用例总数
历史信息	2	998	1 000
试验信息	1	49	50

首先依据文献[8]中给出的 a, b 估值方法,计算出 $a = 1, b = 49$. 然后根据表 1,计算出传统贝叶斯方法($\rho = 1$,不对历史信息 and 试验信息加以区分)求解得到的可接收质量水平 $p_0 = 0.038\ 73$. 接着结合不同的相关系数计算方法,最终得到以下结果.

(1) 采用主观定量法. 计算出 $A = 0.007\ 43, B = 0.138\ 09, C = 0.000\ 04, D = 0.005\ 05$. 再根据专家的经验,给出 $\rho_1 = 0.3, \rho_2 = 0.8$. 最终得到可接收

质量水平 $p_0 = 0.033\ 62$.

(2) 采用拟合优度检验法. 由于未通过测试用例数均小于 5,所以选用 Yates 修正法. 计算得到 $K = 0.940\ 19$,进而求得 $Q(K) = 0.332\ 24$. 分别选取 $i = \frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \frac{1}{10}$,计算出与其对应的相关系数 $\rho = 0.576\ 4, 0.110\ 75, 0.066\ 45, 0.033\ 22$. 最终计算得到可接收质量水平为 $p_0 = 0.040\ 83, 0.055\ 11, 0.060\ 51, 0.066\ 22$.

(3) 采用 Kullback 信息量法. 首先求得 $I(Q; P) = 1.1625$. 进而得到相关系数 $\rho = 0.462\ 43$. 最终得到可接收质量水平 $p_0 = 0.042\ 01$.

(4) 采用 Spearman 相关系数法. 求得 $\sum_{i=1}^3 d_i = 1 + 1 + 0.5 = 2.5$,进一步得到相关系数 $\rho = 0.437\ 5$. 那么可接收质量水平 $p_0 = 0.042\ 35$.

(5) 采用 Kendall 相关系数法. 计算得到 $A = 2, B = 0, a = 3, b = 0, c = 0$,则相关系数 $\rho = \frac{2}{3}$. 那么可接收质量水平 $p_0 = 0.040\ 17$.

根据以上实验结果,汇总得到实验结果,如表 2 所示.

表 2 实验结果分析表

	相关系数	可接收质量水平
传统贝叶斯法	1	0.038 73
主观定量法	—	0.033 62
拟合优度检验法(1/2)	0.576 40	0.040 83
拟合优度检验法(1/3)	0.110 75	0.055 11
拟合优度检验法(1/5)	0.066 45	0.060 51
拟合优度检验法(1/10)	0.033 22	0.066 22
Kullback 信息量法	0.462 43	0.042 01
Spearman 相关系数法	0.437 5	0.042 35
Kendall 相关系数法	0.666 67	0.040 17

从表 2 可以得到以下结论:

(1) 可以很明显地看出随着相关系数的不断增加,软件可靠性验证测试的可接收质量水平一直在降低. 也就是说在验收软件产品时,允许失效的概率越来越低,对于验收的条件越来越严苛. 但同时也不能使可接收质量水平过高,以免以后在使用软件产品时,出现的失效次数偏多,影响使用. 通过分析可以得到,相关系数 ρ 在 $[0.4, 0.7]$ 这个区间里,可接收质量水平的取值较为合适.

(2) 对 i 取不同的值分析,通过实际的实验分析. 可以看出专家们对于 i 取值的估计还是比较准

确的,当 $i=0.5$ 时,软件产品的可接收质量水平较为恰当。

(3)选用主观定量法,主要因为人为因素过大,主观性太强,导致软件产品的可接收质量水平较低。可以人为调节相关系数,使得可接收质量水平较为合理。但是这样就让结果变得随机性太强,不建议在软件可靠性验证测试过程中,过多使用主观定量法。

(4)在相关系数概念的基础上,类比现有的算法,把 Spearman 相关系数法和 Kendall 相关系数法应用到软件可靠性验证测试方案中。通过实验,这两种方法的结果都较为适宜,就证明了这两种方法在软件可靠性验证测试方案中的可行性。

(5)实验结果表明,在进行软件可靠性验证测试时,需要把历史信息 and 试验信息加以区分,否则就会使得可接收质量水平过低,造成对生产方的要求过于严苛,不利于软件产品的检验验收。

5 结束语

本文提出了基于混合贝叶斯方法的软件可靠性验证测试方案,并引入了相关系数的相关概念。在已有的相关系数的计算方法之上,提出了基于相关系数概念的两种计算方法。通过实验分析,论证了这两种方法的可行性,同时证明了对历史信息 and 试验信息进行区分,使得软件可靠性验证测试的结果更加合理化。但是这 5 种方法都未有明确的基础理论证明,该问题仍需要进行深入的探讨研究。

参考文献:

[1] 李海峰,刘畅,郑军. 安全关键软件可靠性验证测试

研究[J]. 航空标准化与质量, 2013, 32(3): 46-50, 58.

[2] 明志茂,陶俊勇,陈循,等. 基于混合 Beta 分布的成败型产品 Bayes 可靠性鉴定试验方案研究[J]. 兵工学报, 2008, 29(2): 204-207.

[3] 张文杰,杨华波,张士峰. 基于 Bayes 混合验前分布的成败型产品可靠性评估[J]. 兵工学报, 2016, 37(3): 505-511.

[4] 赵勇,刘建新,牛青坡. 基于混合 Beta 先验分布的成败型产品的可靠性评估[J]. 航空兵器, 2014, 51(6): 59-61.

[5] 马振宇,张威,卜新旺. 软件缺陷与软件测试之间的关联性分析[J]. 信息通信, 2015, 29(2): 124-125.

[6] 严惠云. Bayes 理论在二项分布可靠性分析中的应用[D]. 西安:西北工业大学, 2007.

[7] 刘解放,刘思峰,方志耕. 成败型产品可靠性评价的加权 Bayes 方法[J]. 中国机械工程, 2013, 24(24): 3371-3374.

[8] 刘晗,郭波. 融合多源验前信息的成败型产品可靠性 Bayes 评估[J]. 弹箭与制导学报, 2007, 27(2): 338-341.

作者简介:

马振宇 男, (1991-), 博士研究生. 研究方向为软件可靠性、可靠性验证测试、软件测试. E-mail: 625181316@qq.com.

吴 伟 男, (1963-), 博士, 高级工程师. 研究方向为装备全寿命、可靠性增长.

张 威 男, (1968-), 博士, 教授. 研究方向为软件工程、软件测试、军用软件保障.

刘福胜 男, (1975-), 博士, 副教授. 研究方向为装备保障.

韩 坤 男, (1985-), 博士, 工程师. 研究方向为软件可靠性、可靠性测试.