

文章编号:1000-8055(2011)07-1531-06

基于验后概率的 Bayesian 序贯方差 检验技术(I)

刘 琦, 王 因

(国防科学技术大学 信息系统与管理学院, 长沙 410073)

摘 要: 基于验后概率和“小概率事件原理”,对复杂假设条件下,方差的 Bayesian 序贯检验,提出了基于 Bayesian 验后概率的序贯方差检验的基本假设,建立了 Bayesian 序贯方差检验(BSVT)模型,设计了临界值的求解算法,并对“强制”截尾条件下的假设检验判别准则进行了论述.最后,结合示例对 BSVT 模型的求解与应用进行了说明.

关 键 词: 航空航天装备;小概率事件原理;序贯检验;方差;Bayesian 方法;
Bayesian 序贯方差检验(BSVT)模型

中图分类号: TP114

文献标志码: A

Bayesian sequential variance test technique based on posterior probability (I)

LIU Qi, WANG Nan

(College of Information Systems and Management,
National University of Defense Technology, Changsha 410073, China)

Abstract: Based on posterior probability and “small probability event principle”, for the Bayesian sequential test on variance in complicated hypothesis conditions, the basic assumptions of Bayesian sequential variance test were presented, the Bayesian sequential variance test (BSVT) model was constructed, the algorithms of solving the critical value of sequential test were designed, and for given censored test number, the judge criteria for hypothesis testing were discussed. At last, the solution and application of BSVT model were illustrated with an example.

Key words: aerospace equipment; small probability event principle;
sequential test; variance; Bayesian method;
Bayesian sequential variance test (BSVT) model

装备试验与评价中,对于正态分布参数,包括期望 μ 和方差 σ^2 的统计检验,在简单假设条件下,经典的统计验证试验设计方法,包括固定样本抽样方法以及序贯抽样方法,其研究都已经比较成熟,已发表了大量的论文和专著.如 Desu 等^[1]对估计精度要求以及假设检验两类风险要求下,

一次抽样样本量的计算方法进行了详细的论述.郭锡福^[2]结合火炮武器系统精度分析的工程实际,对正态分布期望 μ 和方差 σ^2 的简单假设,给出了两类风险 α 和 β 要求下,试验次数 n 的计算方法. Wetherill 等^[3]对序贯方法的基本原理、应用过程进行了详细的介绍. Ghosh^[4]对正态分布

收稿日期:2010-07-04; 修订日期:2010-11-17

基金项目:国家自然科学基金(70971133);高等学校博士学科点专项科研基金(70571083)

作者简介:刘琦(1974—),男,山西运城人,副教授、硕士生导师,博士,主要从事装备试验与评价研究.

固定样本量抽样方案设计和序贯抽样方案设计的原理进行了详细的论述. 武小悦等^[5]则结合装备试验与评价的工程实际, 对固定样本和序贯检验的试验方案设计方法进行了论述, 给出了样本量的计算方法和试验决策规则.

由于试验费用、试验环境条件等的限制, 一些装备或产品的现场试验呈现出小子样的特点, 此时综合利用产品的各种验前信息, 基于 Bayes 方法进行参数的试验评价也受到工程应用领域专家的青睐, 在试验设计方面, 国内外也进行了大量的研究. 如 Bartolucci 等^[6]在环境污染分析方面, 综合考虑了 Bayes 方法、试验费用, 基于参数估计精度要求对分层抽样方法^[7]进行了拓展, 建立了 Bayes 分层抽样模型. 王国玉^[8]对 μ 的序贯验后加权检验 (sequential posterior odd test, 简称 SPOT) 方法的过程进行了推导. GUO Bo 等^[9]则以正态分布为基础, 对对数正态分布期望的截尾 SPOT 方法进行了论述.

在方差的序贯检验方面, 张堃^[10]以飞行器落点密集度的 Bayes 序贯检验分析为背景, 对给定的统计假设, 以两类风险为基础, 推导了序贯检验的相关公式. 张金槐^[11]则对方差的序贯检验与估计给出了检验的计算与评判标准, 并给出了检验中犯两类错误概率的上限的计算公式.

从现有的对 Bayes 序贯概率比检验 (Bayes sequential probability ratio test, 简称 SPRT) 或者序贯验后加权检验 (SPOT)^[12]方法的研究情况看, 还存在如下需要解决的问题: ① SPOT 方法计算复杂, 难以给出比较简易的求解方案, 从而便于试验实施人员快速制定试验计划; ② SPOT 方法基于假设检验中的两类风险, 对序贯检验过程进行推导, 给出判别准则, 但在判别完毕后, SPOT 方法并没有给出接受或者拒绝原假设的验后概率, 从而便于决策人员基于“小概率事件原理”^[13]进行决策. 鉴于此, 本文基于验后概率和“小概率事件原理”, 对正态分布方差建立了 Bayesian 序贯方差检验 (Bayesian sequential variance test, 简称 BSVT) 模型. 在给定截尾试验次数 N 的条件下, 给出了基于现场试验数据的假设检验判别准则, 最后结合一个示例, 对 BSVT 模型的求解与应用过程进行了说明.

1 方差序贯检验的基本假设

设对某装备的某项战技指标 (如命中精度) 的试验结果 X 服从正态分布 $N(\mu, \sigma^2)$, 并假设 μ 已

知, 不妨设 $\mu=0$, 仅需对 σ^2 进行假设检验. 在研制中, 用户关心的是 σ^2 是否满足给定的要求 σ_0^2 , 即要做如下的假设检验:

$$H_0: \sigma^2 \leq \sigma_0^2 \quad (1)$$

记 σ^2 的验前分布为 $\pi(\sigma^2)$, 验前分布为共轭分布时, σ^2 的验前分布^[14-16]为逆 Gamma 分布 $IG(a, b)$. 设产品现场试验结果为 $x_0, x_1, \dots, x_n$, 其中 n 为现场试验的样本量, x_i 为第 i 次试验结果 (在精度试验中为距离靶心的距离), 其中 $i = 0, 1, \dots, n$. 记 $x_0 = 0$, $D_n = \sum_{i=0}^n x_i^2$. 则由 Bayes 公式, 可得 σ^2 的验后分布 $\pi(\sigma^2 | x)$, 当验前分布为共轭分布时, σ^2 的验后分布为逆 Gamma 分布 $IG(a_1, b_1)$, 其中

$$a_1 = a + n/2, \quad b_1 = b + D_n/2 \quad (2)$$

$IG(a, b)$ 的概率密度函数 (probability density function, 简称 PDF) 为

$$f(\sigma^2 | a, b) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp\left[-\frac{b}{\sigma^2}\right] \quad (3)$$

2 BSVT 模型的建模与求解

根据第 1 节的原假设 H_0 , 可以直接给出备择假设 H_1 为

$$H_1: \sigma^2 > \sigma_0^2 \quad (4)$$

若获得的现场试验数据为 $x_0, x_1, \dots, x_n$, 则在验前分布为共轭分布条件下, 根据 $\pi(\sigma^2 | x)$ 可计算得出 H_0 和 H_1 成立的验后概率分别为

$$P(H_0 | x) = P(\sigma^2 \leq \sigma_0^2 | x) =$$

$$\int_0^{\sigma_0^2} \pi(\sigma^2 | x) d\sigma^2 = \int_0^{\sigma_0^2} f(\sigma^2 | a_1, b_1) d\sigma^2 \quad (5)$$

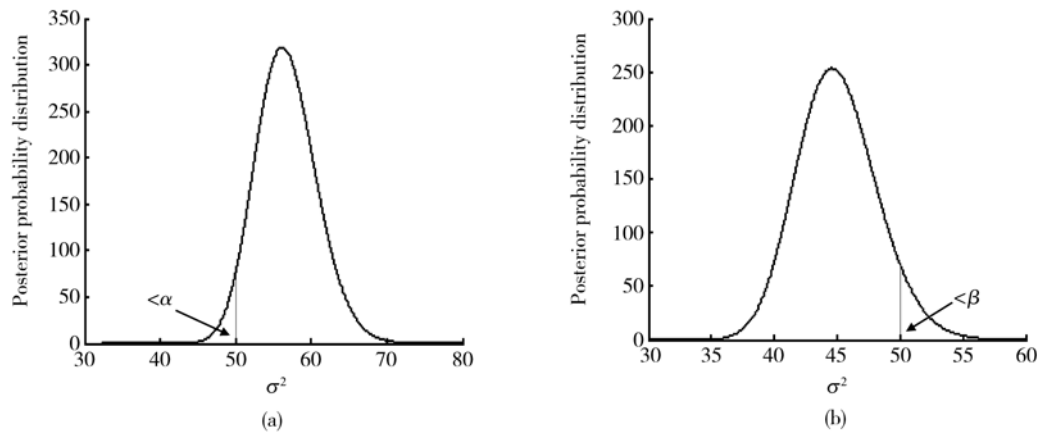
$$P(H_1 | x) = P(\sigma^2 > \sigma_0^2 | x) =$$

$$\int_{\sigma_0^2}^{\infty} \pi(\sigma^2 | x) d\sigma^2 = \int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 \quad (6)$$

图 1 为 σ^2 的验后概率分布示意图.

由“小概率事件原理”, 可给出如下的基本假设和决策规则:

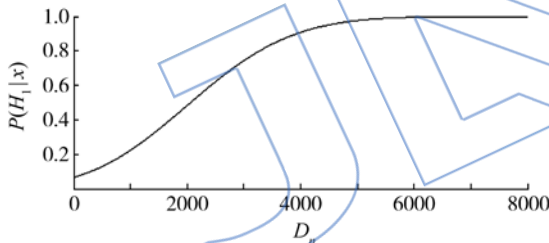
1) 当 H_0 成立的验后概率 $P(H_0 | x)$ 非常小时, 说明获得的现场试验数据不能证明 H_0 是正确的, 所以可以作出拒绝 H_0 的决策. 由于拒绝 H_0 可能导致使用方“弃真”, 所以可引入“弃真”风险 α (α 也称为犯第 I 类错误的概率), 要求 $P(H_0 | x) < \alpha$ (等价于 $P(H_1 | x) \geq 1 - \alpha$) 时, 拒绝 H_0 . 此时“弃真”的概率小于 α , 如图 1(a) 所示.

图1 σ^2 的验后概率分布示意图Fig. 1 Schematic diagram of posterior probability distribution of σ^2

根据上述假设,当试验进行到第 n 次,作出拒绝 H_0 的决策时, D_n 应满足如下条件:

$$P(H_1 | x) = \int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 \geq 1 - \alpha \quad (7)$$

式(7)中,对于给定的 a, b, n, σ_0^2 ,随着 D_n 的增加, $P(H_1 | x)$ 不断增加.如 $a=200, b=9000, n=2, \sigma_0^2=50$ 时, $P(H_1 | x)$ 随 D_n 的变化曲线如图2所示.

图2 $P(H_1 | x)$ 随 D_n 的变化曲线Fig. 2 Variation curve of $P(H_1 | x)$ changing with D_n

结合图2,记 $D_{u,n}$ 为满足式(7)的最小 D_n , 则有

$$D_{u,n} = \max \left\{ -1, \min \left\{ D_n \left| \int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 \geq 1 - \alpha, D_n \geq 0 \right\} \right\} \right\} \quad (8)$$

式(8)中,当取 $D_{u,n} = -1$ 时,表示对所有的 $D_n \geq 0$, 都不能找到满足式(7)要求的最小 D_n , 即使 $D_n = 0$, 仍有

$$\int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 \geq 1 - \alpha + \epsilon_a > 1 - \alpha \quad (9)$$

其中 $\epsilon_a > 0$.

2) 当 H_1 成立的验后概率非常小时,说明获

得的现场试验数据不能证明 H_1 是错误的假设, 所以可作出采纳 H_0 的决策. 由于接受 H_0 可能导致使用方“采伪”, 所以引入“采伪”风险 β (β 也称为犯第II类错误的概率), 要求 $P(H_1 | x) < \beta$ (等价于 $P(H_0 | x) \geq 1 - \beta$) 成立时, 接受 H_0 , 此时“采伪”的概率小于 β , 如图1(b)所示.

根据上述假设,当试验到第 n 次,作出接受 H_0 的决策时, D_n 应满足如下条件:

$$P(H_1 | x) = \int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 < \beta \quad (10)$$

结合图2,记 $D_{l,n}$ 为满足式(10)的最大 D_n , 则有

$$D_{l,n} = \max \left\{ -2, \max \left\{ D_n \left| \int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 < \beta, D_n \geq 0 \right\} \right\} \right\} \quad (11)$$

式(11)中,当取 $D_{l,n} = -2$ 时,表示对所有的 $D_n \geq 0$, 都不能找到满足式(10)要求的最大 D_n , 即使 $D_n = 0$, 仍有

$$\int_{\sigma_0^2}^{\infty} f(\sigma^2 | a_1, b_1) d\sigma^2 \geq \beta + \epsilon_\beta > \beta \quad (12)$$

其中 $\epsilon_\beta > 0$.

由 $\pi(\sigma^2 | x)$ 为连续分布的性质可知,当 $\alpha + \beta < 1$ 时,总会存在试验数据 $x_0, x_1, \dots, x_n$, 使得

$$\beta \leq P(H_1 | x) < 1 - \alpha \quad (13)$$

或

$$\alpha \leq P(H_0 | x) < 1 - \beta \quad (14)$$

成立. 所以,当上述1)和2)两种决策规则都不满足时,可进行如下决策:

3) 当式(13)或式(14)成立时,继续试验.

从上述计算过程可见,当 $\alpha + \beta < 1$ 时,继续试验区在序贯试验中总是存在的,也就是说不能找一个最大的 N ,使得最多试验到第 N 次时就会作出最终的判断。

综上所述,在给定验前分布 $IG(\alpha, b)$ 的条件下,进行序贯试验时,试验从 $n = 0$ 开始,进行到第 n 次试验,记前 n 次试验的结果为 D_n 。序贯试验的判别规则为

- 1) 若 $D_n \leq D_{l,n}$, 则停止试验,采纳 H_0 。
- 2) 若 $D_n \geq D_{u,n}$, 则停止试验,拒绝 H_0 。
- 3) 若 $D_{l,n} < D_n < D_{u,n}$, 则继续试验。

3 截尾 BSVT 模型

从第2节的分析可见,当 $\alpha + \beta < 1$ 时,继续试验区在序贯试验中总是存在的,也就是说,当现场试验数据 $x_0, x_1, \dots, x_n$ 不满足试验停止决策规则1)或者2)时,试验可以无休止地进行下去。这在工程实际中是不可行的,因为一方面,厂家生产的产品总是有限的,不可能让试验无休止地进行下去;另一方面,通常在试验前,生产方、使用方会通过协商给定一个最大的投试样本量 N ,当现场试验的样品数达到 N 时必须终止试验。

当给定最大试验样本量 N 时,可以考虑“强制”截尾,可考虑如下三种截尾决策规则。

1) 控制 α , 降低使用方风险。当 $n < N$ 时,按照 BSVT 模型的序贯试验判别规则进行决策,当 $n = N$ 时,按如下规则判断:

- (a) 若 $D_n < D_{u,n}$, 采纳 H_0 。
- (b) 若 $D_n \geq D_{u,n}$, 拒绝 H_0 。

2) 控制 β , 降低生产方风险。当 $n < N$ 时,按照 BSVT 模型的序贯试验判别规则进行决策,当 $n = N$ 时,按如下规则判断:

- (a) 若 $D_n \leq D_{l,n}$, 采纳 H_0 。
- (b) 若 $D_n > D_{l,n}$, 拒绝 H_0 。

3) 折中原则。当 $n < N$ 时,按照 BSVT 模型的序贯试验判别规则进行决策,当 $n = N$ 时,取决策的临界值为 $D_m = (D_{l,n} + D_{u,n})/2$, 按如下规则判断:

- (a) 若 $D_n < D_m$, 采纳 H_0 。
- (b) 若 $D_n \geq D_m$, 拒绝 H_0 。

4 BSVT 模型示例分析

设在某装备的命中精度试验中,根据类似型号装备和该装备的研制试验等数据,计算得出 σ^2 的验前分布为 $IG(200, 9\ 000)$, 若要求 $\alpha_0^* = 50$,

$\alpha = \beta = 0.05$ 。则根据上述的计算求解过程,计算可得 $D_{u,n}$ 和 $D_{l,n}$ 随试验次数 n 的变化情况如图3所示。从而可得到图3所示的样本空间划分方法。在 $n \leq 29$ 的条件下, $D_{u,n}$ 和 $D_{l,n}$ 的具体数值如表1所示。

表1 给定不同 n 值下 BSVT 模型的 $D_{l,n}$ 和 $D_{u,n}$ 值

Table 1 Numerical values of $D_{l,n}$ and $D_{u,n}$ for different n of BSVT model

n	$D_{l,n}$	$D_{u,n}$
0	-2.00	4381.62
1	-2.00	4434.53
2	-2.00	4487.43
3	-2.00	4540.34
4	-2.00	4593.23
5	-2.00	4646.13
6	14.66	4699.02
7	61.77	4751.90
8	108.89	4804.78
9	156.01	4857.66
10	203.13	4910.54
11	250.26	4963.41
12	297.39	5016.28
13	344.53	5069.14
14	391.67	5122.01
15	438.81	5174.86
16	485.95	5227.72
17	533.10	5280.57
18	580.26	5333.42
19	627.41	5386.26
20	674.57	5439.10
21	721.73	5491.94
22	768.90	5544.77
23	816.07	5597.60
24	863.24	5650.43
25	910.42	5703.26
26	957.60	5756.08
27	1004.78	5808.90
28	1051.96	5861.71
29	1099.15	5914.52

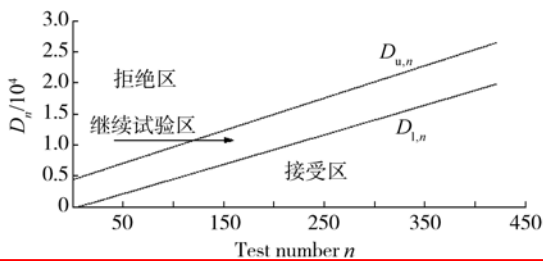


图3 BSVT模型样本空间划分结果

Fig.3 Dividing result of sample size of BSVT model

在具体的试验实施中,序贯试验从 $n=0$ 开始,设试验进行到第10次,即 $n=10$,那么此时就需要根据所获得的样本观测值 $x_1, x_2, \dots, x_{10}$ 计算 D_{10} ,然后根据第2节的判别准则进行判断,如 $203.13 < D_{10} < 4910.54$,继续试验,否则停止试验,根据 D_{10} 的大小作出拒绝或采纳 H_0 的决策。

若采用传统的 SPOT 方法^[5],取 A 和 B 分别为

$$A = \frac{\beta}{\pi_0 - \alpha}, \quad B = \frac{\pi_1 - \beta}{\alpha} \quad (15)$$

其中

$$\pi_0 = P(H_0), \quad \pi_1 = 1 - P(H_0)$$

根据本例的数据,计算可得

$$\pi_0 = P(H_0) = P(\sigma^2 \leq 50) = \int_0^{50} f(\sigma^2 | 200, 9000) d\sigma^2 = 0.9251$$

所以 $A=0.0571, B=0.4980$. 由于 $B < 1$, 与 SPOT 方法要求的 $B \geq 1$ 矛盾,所以对于本例,难以使用传统的 SPOT 方法进行分析。

若采用传统的 Bayes 假设检验方法,当采用 0-1 损失函数时,其检验的判别公式为^[5]

$$\frac{P(H_1 | x)}{P(H_0 | x)} \underset{\text{accept } H_1}{\overset{\text{accept } H_0}} < \underset{\text{accept } H_1}{\overset{\text{accept } H_0}} 1 \quad (16)$$

该方法虽然简单易懂,但仅考虑了验后的损失,而未考虑统计推断中应用最为广泛的“小概率事件原理”。

如现场试验次数为 8 次,得出

$$D_8 = \sum_{i=1}^8 x_i^2 = 2000$$

则计算可得

$$a_1 = 204, \quad b_1 = 10000$$

从而

$$P(H_0 | x) =$$

$$\int_0^{50} f(\sigma^2 | 204, 10000) d\sigma^2 = 0.6020$$

$$P(H_1 | x) = 1 - 0.6020 = 0.3980$$

由式(16)可作出采纳 H_0 的决策,但从 Bayes 统计分析的角度,因 H_1 成功的概率为 0.3980,所以若采纳 H_0 ,则“采伪”的概率为 0.3980,这是不符合“小概率事件原理”的,从而需要继续试验,找出充分的证据验证 H_0 (或者 H_1) 的正确性。

从上述比较分析可见:①相比传统的 SPOT 方法,BSVT 模型解决了 SPOT 方法有可能无法解决的问题;②相比传统的 Bayes 模型假设检验方法,BSVT 模型能从统计上(利用“小概率事件原理”)充分地解释采纳 H_0 (或者 H_1) 决策的统计现象,而传统的 Bayes 模型假设检验方法则不能做到这一点,所以相比而言,BSVT 模型具有较大的优越性。

5 结束语

论文以装备试验评价中正态分布方差的假设检验需求为背景,对小子样条件下,基于验后概率和“小概率事件原理”,建立了方差序贯假设检验的 BSVT 模型,并通过示例分析说明了该模型的计算与应用过程,通过比较分析,论述了该模型的优越性.在 BSVT 模型的应用中需注意以下几个方面。

1) 论文仅对正态分布的方差建立了 BSVT 模型,在具体的应用中,针对装备试验评价的不同参数(如可靠性分析中的 MTBF(平均故障间隔时间)、维修性分析中的 MTTR(平均恢复时间)等),读者可结合本文的基本理论,推导相关公式,给出相关参数的 Bayesian 序贯检验模型和计算用表。

2) 从第2节的论述可见,当 $\alpha + \beta < 1$ 时,继续试验区在序贯试验中总是存在的,而且例1的结果图3也说明了这一点.而从产品或装备试验的具体情况看:在试验中,一方面无论是生产方还是使用方都希望在试验前能给出一个试验次数,从而便于试验的准备;另一方面,厂家生产的产品也是有限的,能供应给试验的样品数也是有限的,所以难以允许试验无休止地进行下去.为了解决这种情况,须对 BSVT 模型进行修正,具体请参见《基于验后概率的 Bayesian 序贯方差检验技术(II)》一文。

参考文献:

- [1] Desu M M, Raghavarao D. Sample size methodology[M]. Boston: Academic Press, 1990.
- [2] 郭锡福. 远程火炮武器系统射击精度分析[M]. 北京: 国防工业出版社, 2004.
- GUO Xifu. Firing accuracy analysis for long range gun weapon systems[M]. Beijing: National Defense Industry Press, 2004. (in Chinese)
- [3] Wetherill G B, Glazebrook K D. Sequential methods in statistics[M]. 3rd ed. London: Chapman and Hall, 1986.
- [4] Ghosh B K. Sequential tests of statistical hypotheses[M]. Reading, Mass.: Addison-Wesley, 1970.
- [5] 武小悦, 刘琦. 装备试验与评价[M]. 北京: 国防工业出版社, 2008.
- WU Xiaoyue, LIU Qi. Military equipment test and evaluation[M]. Beijing: National Defense Industry Press, 2008. (in Chinese)
- [6] Bartolucci A A, Bartojucci A D, Bae S, et al. A Bayesian method for computing sample size and cost requirements for stratified random sampling of pond water[J]. Environmental Modelling & Software, 2006, 21(9): 1319-1323.
- [7] 武小悦, 刘琦. 应用统计学[M]. 长沙: 国防科技大学出版社, 2009.
- WU Xiaoyue, LIU Qi. Applied statistics[M]. Changsha: National University of Defense Technology Press, 2009. (in Chinese)
- [8] 王国玉. 电子系统小子样试验理论方法[M]. 北京: 国防工业出版社, 2003.
- WANG Guoyu. Small sample test theory and method of electronic system[M]. Beijing: National Defense Industry Press, 2003. (in Chinese)
- [9] GUO Bo, JIANG Ping, XING Yunyan. A censored sequential posterior odd test (SPOT) method for verification of the mean time to repair[J]. IEEE Transactions on Reliability, 2008, 57(2): 243-247.
- [10] 张堃. Bayes 序贯检验方法[J]. 装备指挥技术学院学报, 1998, 9(4): 89-95.
- ZHANG Kun. Bayesian sequential testing method[J]. Journal of the Academy of Equipment Command & Technology, 1998, 9(4): 89-95. (in Chinese)
- [11] 张金槐. 正态总体下分布参数的 Bayes 序贯估计[J]. 国防科技大学学报, 2002, 24(2): 95-100.
- ZHANG Jinhuai. Bayes sequential estimation of the parameter for normal sample[J]. Journal of National University of Defense Technology, 2002, 24(2): 95-100. (in Chinese)
- [12] 张金槐, 刘琦, 冯静. Bayes 试验分析方法[M]. 长沙: 国防科技大学出版社, 2007.
- ZHANG Jinhuai, LIU Qi, FENG Jing. Bayes test analysis method[M]. Changsha: National University of Defense Technology Press, 2007. (in Chinese)
- [13] Li X R, Li X B. Common fallacies in applying hypothesis testing[C]//Proceedings of the 11th International Conference on Information Fusion. Cologne: IEEE, 2008: 1-8.
- [14] Hoff P D. A first course in Bayesian statistical methods[M]. New York: Springer-Verlag, 2009.
- [15] Hamada M S, Wilson A G, Reese C S, et al. Bayesian reliability[M]. New York: Springer Science and Business Media, LLC, 2008.
- [16] Koch K R. Introduction to Bayesian statistics[M]. 2nd ed. Heidelberg, Berlin: Springer-Verlag, 2007.