应用概率统计　第二十三卷
第一期　2007 年 2 月

Chinese Journal of Applied Probability
and Statistics Vol. 23 No. 1 Feb. 2007

计数型二次序贯网图检验

濮晓龙 [1]　闫章更 [2]　茆诗松 [1]　张应山 [1]　李　艳 [1]
([1] 华东师范大学统计系，上海， 200062； [2] 中国华阴兵器试验中心，华阴， 714200)

摘　要

序贯概率比检验 (SPRT) 是应用非常广泛的抽样检验方法，序贯网图检验在控制最大样本量方面很好地改进了 SPRT, 但其结果还有进一步改进的余地，为此，我们建立了二次序贯网图检验，计算结果表明，它比原先的计数序贯网图检验有更好的效果．

关 键 词：序贯网图检验，二次序贯网图检验，二类风险，最大样本量.
学 科 分 类 号：O212.8.

§1.　背　　景

对成功率的检验问题，最常用到的是由瓦尔德 (A. Wald, 1947) 提出的序贯概率比检验．序贯概率比检验的优点主要有如下二点：可以保证检验犯二类错误的概率不超过设定值；可以证明在 q_0, q_1 处检验所用的平均样本量是最小的．但它也有二个不足之处：实际抽取样本量事先无法控制，可能很大，应用中不知什么时候可以结束试验，也无法对用弹量进行预算；当真正的成功率不是 q_0 或 q_1 时，检验所用的平均样本量也不再是最小的，事实上，真正的成功率通常并不等于 q_0 或 q_1. 对此，有人提出了截尾序贯概率比检验，形成了 IEC1123, 但其本身的样本量上界还是很大的，应用意义受到限制．为此，我们在文 [1] 中提出了序贯网图检验，其思路如下：用 x_i 表示每一发武器的命中情况， $x_i=1$ 表示该发武器命中功， $x_i=0$ 表示该发武器未命中，则 $x_i\sim b(1,q)$, 其中 q 为命中率，令 $S_n=\sum\limits_{i=1}^{n}x_i$ 为试验到第 n 发时的命中次数．对检验问题

$$H_0: q=q_0 \qquad \text{vs} \qquad H_1: q=q_1, \tag{1.1}$$

我们通过引入 $q_2\in(q_1,q_0)$, 取定某个 $\alpha_1=\beta_1$, 记

$$b_1=\log\Big(\frac{1-\alpha_1}{\alpha_1}\Big), \qquad a_1=\log\Big(\frac{\alpha_1}{1-\alpha_1}\Big)=-b_1, \tag{1.2}$$

$$h_1=b_1\Big/\log\Big(\frac{q_2(1-q_1)}{q_1(1-q_2)}\Big), \qquad s_1=\log\Big(\frac{1-q_1}{1-q_2}\Big)\Big/\log\Big(\frac{q_2(1-q_1)}{q_1(1-q_2)}\Big), \tag{1.3}$$

$$h_2=b_1\Big/\log\Big(\frac{q_0(1-q_2)}{q_2(1-q_0)}\Big), \qquad s_2=\log\Big(\frac{1-q_2}{1-q_0}\Big)\Big/\log\Big(\frac{q_2(1-q_0)}{q_0(1-q_2)}\Big), \tag{1.4}$$

并规定一个截尾样本量 n_t 和不合格判定数 r_t, 检验判断规则如下：

本文 2005 年 9 月 14 日收到.

从 $n=1$ 开始, 当 $n<n_t$ 时: 若 $S_n\geq s_1n+h_1$, 则停止抽样, 接受原假设 H_0; 若 $S_n\leq s_2n-h_2$, 则停止抽样, 拒绝原假设 H_0; 若 $s_2n-h_2<S_n<s_1n+h_1$, 则说明在此时尚不能做出判断, 需要进一步抽取样本.

当 $n=n_t$ 时, 必须做出判断: 若 $S_{n_t}\geq r_t$, 则接受原假设 H_0; 若 $S_{n_t}<r_t$, 则拒绝原假设 H_0.

我们在文 [1] 中已经得到了如下二个主要结论:

(1) 理论上讲插入多个点是有可能进一步改进插入一个点的序贯网图检验方法的, 但实际计算结果表明, 在应用上可以不用考虑插入多个点的序贯网图检验方法.

(2) 序贯网图检验全面改进了国际标准 IEC1123: 或者在保持二类风险满足要求的条件下降低了截尾试验样本量, 或者在相同的截尾试验样本量下降低了二类风险.

但我们也看到在序贯网图检验中, 二类风险设定值并没有得到充分利用, 这说明对之应该还有改进的余地, 这就是本文的出发点和要讨论的问题.

§2. 随机化检验

众所周知, 在基于离散随机变量的抽样检验中, 为了足量使用显著性水平, 通常可使用随机化方法得到最优的检验.

比例 p 可看作某事件发生的概率, 即可看作二点分布 $b(1,p)$ 中的参数. 作 n 次独立试验, 以 x 记该事件发生的次数, 则 $x\sim b(n,p)$. 我们可以根据 x 检验关于 p 的一些假设. 考虑如下假设检验问题:

$$H_0: p=p_0 \qquad \text{vs} \qquad H_1: p=p_1, \tag{2.1}$$

其中 $p_0<p_1$. 直观上看, 一个显然的检验方法是取如下的拒绝域 $W=\{x\geq c\}$, 由于 x 只取整数值, 故 c 可限制在非负整数中. 然而, 一般情况下对给定的显著性水平 α, 不一定能正好取到一个 c 使

$$\mathrm{P}(x\geq c;p_0)=\sum_{i=c}^{n}\begin{pmatrix} n\\ i\end{pmatrix}p_0^i(1-p_0)^{n-i}=\alpha, \tag{2.2}$$

能恰巧使得 (2.2) 成立的 c 值是罕见的. 在这种情况下, 统计学理论告诉我们, 采用如下的随机化检验在保证第一类风险不超过 α 的条件下可降低第二类风险.

引理 2.1 在上述记号下, 检验问题 (2.1) 的最优显著性检验是如下的随机化检验, 其判断准则为:

- 若观测值 $x\geq c_0+1$, 则拒绝原假设, 认为 $p\geq p_1$;
- 若观测值 $x<c_0$, 则接收原假设, 认为 $p\leq p_0$;
- 若观测值 $x=c_0$, 则做一个成功概率为

$$\gamma=\left[\alpha-\sum_{i=c_0+1}^{n}\begin{pmatrix} n\\ i\end{pmatrix}p_0^i(1-p_0)^{n-i}\right]\Bigg/\left[\begin{pmatrix} n\\ c_0\end{pmatrix}p_0^i(1-p_0)^{n-i}\right]$$

的试验, 若试验成功, 则拒绝原假设, 试验不成功则接收原假设.

由引理 2.1 给出的随机化检验犯第一类错误的概率为 α, 而在 $p = p_1$ 处犯第二类错误的概率为

$$\begin{aligned}\beta &= \mathsf{P}_{p_1}(\text{接受原假设}) \\ &= \mathsf{P}_{p_1}(x < c_0) + \mathsf{P}_{p_1}(\text{接受原假设}|x = c_0)\mathsf{P}_{p_1}(x < c_0) \\ &= \sum_{i=1}^{c_0-1}\binom{n}{i}p_1^i(1-p_1)^{n-i} + (1-\gamma)\binom{n}{c_0}p_1^{c_0}(1-p_1)^{n-c_0}.\end{aligned}$$

例 2.1　现对某厂生产的产品不良品率 p 进行检验，一对假设是

$$H_0 : p = 0.2 \qquad \text{vs} \qquad H_1 : p = 0.4,$$

现选取 20 件进行检查，若设定第一类风险为 0.15, 应如何给出检验方案？

以 x 表示 20 件产品中的不良品件数，则 $x \sim b(20, p)$, 检验的直观拒绝域为 $W = \{x \geq c\}$, 下求 c. 由于

$$\mathsf{P}_{0.2}(x \geq 6) = 0.1958 > 0.15 > \mathsf{P}_{0.2}(x \geq 7) = 0.0867,$$

从而 $c = 7$, 拒绝域为 $W = \{x \geq 7\}$. 不难看出该拒绝域的显著水平实际上不是 0.15, 而是 0.0867, 而且，当备择假设成立时，检验犯第二类错误的概率为

$$\beta = 1 - \mathsf{P}_{0.4}(x \geq 7) = \mathsf{P}_{0.4}(x \leq 6) = 0.25,$$

也就是说该检验的第二类风险为 0.25.

接下来我们讨论随机化检验．判断准则为：

- 若 20 件产品中不良品件数达到 7 件或以上，则拒绝原假设，认为 $p \geq 0.4$;
- 若 20 件产品中不良品件数在 5 件或以下，则接收原假设，认为 $p \leq 0.2$;
- 若 20 件产品中不良品件数正好为 6 件，则做一个成功概率为

$$\gamma = \frac{0.15 - 0.0867}{0.1958 - 0.0867} = 0.5802$$

的试验，若试验成功则拒绝原假设，认为 $p \geq 0.4$, 试验不成功则接收原假设，认为 $p \leq 0.2$. 这样的检验其第一类风险恰为 0.15, 第二类风险为

$$\beta = \mathsf{P}_{0.4}(x < 6) + (1 - 0.5802)\mathsf{P}_{0.4}(x = 6) = 0.1778.$$

我们看到，由于采用了随机化检验，第二类风险从 0.25 降到了 0.1778, 两类风险均在 0.20 以下且比较接近．反过来，若要求两类风险在某个设定值内，采用一般的检验可能需要大的样本量，而采用随机化检验则可能只需要小的样本量．当然，在实际问题中，随机化检验很难被工程师接受：把接收与拒绝的权交给一个随机化试验的结果对工程师来说是不可想象的．但我们可以考虑通过追加试验的办法来部分实现这一点，这就是我们下面提出二次序贯网图检验的理论出发点．

§3. 二次序贯网图检验方法的建立

引理 2.1 告诉我们，随机化检验是可能在保持第一类风险满足要求的条件下降低第二类风险的．当然，在实际使用中，随机化检验很难被人们接受使用，因为它本身很难实施．但随机化检验的思想我们可以借用到序贯网图检验中来，这就是下面我们提出的二次序贯网图检验．

文 [1] 中的截尾序贯网图检验对传统的截尾序贯检验 IEC1123 有有效的改进，但其二类风险并没有完全达到预先的设定值，从而它所需的截尾样本量就可能不是最小的，这就是说，对文 [1] 中的序贯网图检验存在改进的可能，原因就在于实际的二类风险比设定值要小．

回顾序贯网图检验，它是这样进行的：当 $n=n_t$ 时，必须做出判断，若 $S_{n_t}\geq r_t$, 则拒绝备择假设 H_1 接受原假设 H_0; 若 $S_{n_t}<r_t$, 则拒绝原假设 H_0 接受备择假设 H_1. 但在工程应用中，人们对边界 r_t 的判断总认为有些勉强，应用中可采用无边界的截尾检验，采用如下的判断准则：在 $n=n_t$ 时，有可能会做出判断：

- 若 $S_{n_t}>r_t$, 则拒绝备择假设 H_1 接受原假设 H_0;
- 若 $S_{n_t}<r_t$, 则拒绝原假设 H_0 接受备择假设 H_1.

这种无边界的截尾检验与前述一般的截尾检验的区别就在于当 $n=n_t$ 时，前者必须做出判断，而这里并非一定要做出判断．直观上看，这种无边界的截尾与普通的截尾，不应有本质上的差别，但实际计算表明这种差别是相当大的，用这种无边界的截尾所得到的最优方案，要比普通的截尾所得到的最优方案相应的实验次数少得多．

事实上，如果在 $n=n_t$ 时 $S_{n_t}=r_t$, 该如何继续下去是一个很重要的问题．依照前面的分析，我们可以知道使用一个随机化检验是可以降低风险的，也就是说，如果 $S_{n_t}=r_t$, 则我们可以做一个成功概率为 γ 的试验，若试验成功则拒绝原假设，若试验不成功则接收原假设，这样的检验在理论上是比一般的检验更优的．

然而这种 "理论上更优的" 检验没有实用价值，因为没有任何一个人愿意采用这样的检验，也无法使用这种检验去作判断．但基于这样的思想，我们提出如下的检验方法：在 $S_{n_t}=r_t$ 时通过附加试验来作判断，这就是我们下面提出的二次检验．

对贝努里试验，考虑检验问题 (1.1), 我们通过引入 $q_2\in(q_1,q_0)$ 将之拆分为如下二对假设检验问题

$$
\begin{aligned}
&H_{01}: q=q_2 \quad \text{vs} \quad H_{11}: q=q_1,\\
&H_{02}: q=q_0 \quad \text{vs} \quad H_{12}: q=q_2.
\end{aligned}
$$

采用 (1.2), (1.3), (1.4) 的记号，在序贯网图检验中取一个截尾样本量 n_t 以及一个待判断合格数 r_t, 采用如下判断规则：从 $n=1$ 开始，

当 $n<n_t$ 时：

- 若 $S_n\geq s_1n+h_1$, 则停止抽样，拒绝假设 H_{12}, 即接受原假设 H_0;
- 若 $S_n\leq s_2n-h_2$, 则停止抽样，拒绝假设 H_{01}, 即拒绝原假设 H_0;
- 若 $s_2n-h_2<S_n<s_1n+h_1$, 则说明在此时尚不能做出判断，需要进一步抽取样本．

当 $n=n_t$ 时：

- 若 $S_{n_t}>r_t$, 则接受原假设 H_0;

- 若 $S_{n_t} < r_t$, 则拒绝原假设 H_0;
- 若 $S_{n_t} = r_t$, 则再增加检验 n' 个样品，以 S' 记这 n' 次试验中的成功次数，取一个追加检验合格判断数 r', 采用如下判断准则：
  - 若 $S' \ge r'$, 则拒绝备择假设 H_1 接受原假设 H_0;
  - 若 $S' < r'$, 则拒绝原假设 H_0 接受备择假设 H_1.

这样形成的检验我们称之为二次序贯网图检验．

二次序贯网图检验中有四个值需要确定，它们是：

- 第一次序贯网图检验截尾样本量 n_t;
- 第一次截尾待判断合格数 r_t 或其相应的待判断失败数 $f_t = n_t - r_t$;
- 追加检验样本量 n';
- 追加检验合格判断数 r' 或其相应的追加检验失败判断数 $f' = n' - r' + 1$.

这四个值的确定应由二类实际风险进行：由这四个值制定的检验方案满足二类风险的设定要求，然后希望总的样本量 $n_t + n'$ 尽可能小．这里要说明一点：在第一次序贯网图检验截尾时，由于是待判断，因此，合格数与失败数加起来就等于截尾样本量，而在追加检验时，由于必须做出判断，因此，合格判断数与失败判断数加起来等于追加样本量 +1.

二次序贯网图检验的实际风险的是可以计算的．对如上定义的二次序贯网图检验，检验犯第一类错误的概率为

$$\begin{aligned}
\alpha' \quad = \quad & \mathsf{P}_{q_0}(S_1 \le s_2 - h_2) \\
& + \sum_{i=2}^{n_t-1} \mathsf{P}_{q_0}(S_i \le s_2 i - h_2, s_2 j - h_2 < S_j < s_1 j + h_1, j = 1, 2, \cdots, i-1) \\
& + \mathsf{P}_{q_0}(S_{n_t} < r_t, s_2 j - h_2 < S_j < s_1 j + h_1, j = 1, 2, \cdots, n_t - 1) \\
& + \mathsf{P}_{q_0}(S' < r')\mathsf{P}_{q_0}(S_{n_t} = r_t, s_2 j - h_2 < S_j < s_1 j + h_1, j = 1, 2, \cdots, n_t - 1),
\end{aligned}$$

而犯第二类错误的概率为

$$\begin{aligned}
\beta' \quad = \quad & \mathsf{P}_{q_1}(S_1 \ge s_1 + h_1) \\
& + \sum_{i=2}^{n_t-1} \mathsf{P}_{q_1}(S_i \ge s_1 i + h_1, s_2 j - h_2 < S_j < s_1 j + h_1, j = 1, 2, \cdots, i-1) \\
& + \mathsf{P}_{q_1}(S_{n_t} > r_t, s_2 j - h_2 < S_j < s_1 j + h_1, j = 1, 2, \cdots, n_t - 1) \\
& + \mathsf{P}_{q_1}(S' \ge r')\mathsf{P}_{q_1}(S_{n_t} = r_t, s_2 j - h_2 < S_j < s_1 j + h_1, j = 1, 2, \cdots, n_t - 1).
\end{aligned}$$

例 3.1 考虑 $q_0 = 0.90$, $q_1 = 0.7$, $\alpha = \beta = 0.20$ 的情况，根据 IEC1123, 截尾数和判别数为 (49,7), 它的二类风险分别为 0.1977 和 0.1991; 若选用序贯网图方法，则可选出最优方案为 (38,6), 它的二类风险分别为 0.1748 和 0.2004 (见文 [1]). 这里的第一类风险 0.1748 与设定值 0.20 还有较大的差距，因此有改进的可能．我们采用二次序贯网图方法，采用搜索的方法可以找出当 $(n_t, f_t) = (30, 5)$, $(n', f') = (7, 1)$ 时检验的实际二类风险分别为 0.2020 和 0.1991.

关于该例我们说明几点：

(1) 该检验是分二步进行的，第一步是无边界的序贯网图检验，它的理论、方法与实施类似于我们前面讲述的序贯网图检验，只是在最后一步有差别．如果在 $n = n_t$ (30) 时 $S_{n_t} = r_t = n_t - f_t$ (25), 我们有必要作一个补充检验，补充的则是一个定数检验．

(2) 该检验的判断准则在前面一部分就是采用序贯网图判别准则，若 $n = n_t$ (30) 时 $S_{n_t} = r_t = n_t - f_t$ (25), 则再一次性作 7 次试验，在这 7 次试验中若 7 次试验全部成功，则接收原假设，只要有 1 次或 1 次以上失败，则拒绝原假设.

(3) 该检验的二类风险总的来说略大于前面二种检验的二类风险，但没有超出所设定的要求，或者说更接近二类风险的设定值. 天下没有免费的午餐，正因为该检验的实际风险更接近设定值，才使得它所用的检验样本量更小.

(4) 该检验实际上就是借鉴了随机化检验的思想，但没有使用随机化的方法. 随机化在理论上有其优势，但无法在实际中得以实现，我们这里的检验，当出现应该随机化的时候，我们采用了追加试验的方法，这就是我们提出二次序贯网图检验的理论依据.

(5) 这里的二次序贯网图检验在最大样本量上是明显优的：即使要追加试验，总试验量 $(30 + 7 = 37)$ 也比一次序贯网图的截尾数 (38) 要略小一点，更加小于 IEC1123 的截尾数 (49).

§4. 二次序贯网图检验与一次序贯网图检验的结果比较

通过上面的叙述，我们看到当序贯网图检验的实际的二类风险值与设定值有一定的差距时，二次序贯网图检验是有可能对序贯网图检验作进一步改进的.

为了说明二次序贯网图检验的效果，我们将 IEC1123 中所有截尾数不超过 20 的检验方案单列出来，给出了二次序贯网图检验的样本量和二类风险，这样的检验方案一共有 21 组，结果罗列成表 4.1. 由于这些方案序贯网图检验与 IEC1123 的比较已在文 [1] 中给出，故这里不再列出 IEC1123 的结果，而只将二次序贯网图检验与一次序贯网图检验的结果进行比较.

从表 4.1 中我们可以看到，所有的参数组合下，二次序贯网图检验二步所需总样本量均不超过一次序贯网图检验的截尾数，这说明在降低试验样本量方面，二次序贯网图检验确实起到了比较明显的作用.

二次序贯网图检验也有它自身的困难：要找出这样一个最佳的二次序贯网图检验方案不是一件容易的事，需要对所有可能的方案进行搜索，计算量是十分巨大的，因此，在应用上需要有专家的帮助才能设计一个好的检验方案.

为了能将二次序贯网图检验用于实践，我们已经给出了 100 组常见检验问题的二次序贯网图检验方案，也给出了计算程序，若有需要可与作者联系.

参 考 文 献

[1] 濮晓龙，闫章更，茆诗松，张应山，李艳，计数型序贯网图检验，华东师范大学学报，**125**(2006), 67–71.

[2] IEC (International Electrotenical Commission) 1123, Reliability testing–compliance test plans for success ratio.

[3] 陈家鼎， 序贯分析, 北京大学出版社， 1995.

[4] Anderson, T.W., A modification of the sequential probability ratio test to reduce the sample size, *Ann. Math. Statist.*, **31**(1960), 165–197.

[5] Lorden, G., 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size, *Ann. Math. Statist.*, **47**(1976), 281-291.

[6] Wald, A., *Sequential Analysis*, Wiley, New York, 1947.

表 4.1　IEC1123 小样本结果比较 (21 组)

序号	q_0	D	$\alpha=\beta$	一次序贯网图检验				二次序贯网图检验					
				n_t	f_t	α'	β'	n_t	f_t	n'	f'	α'	β'
1	0.85	3.00	0.30	2	1	0.2775	0.3025	1	1	1	0	0.1808	0.1792
2	0.80	3.00	0.30	4	2	0.1808	0.1792	2	2	1	0	0.2320	0.2560
3	0.80	3.00	0.20	4	2	0.1808	0.1792	3	2	1	0	0.1808	0.1792
4	0.90	3.00	0.30	8	2	0.1704	0.2916	3	1	3	1	0.2914	0.2689
5	0.85	3.00	0.20	9	3	0.1409	0.1495	5	2	3	1	0.1838	0.1862
6	0.91	3.00	0.30	9	2	0.1782	0.2820	3	1	4	1	0.2788	0.2739
7	0.80	2.00	0.30	9	3	0.2549	0.2714	5	2	2	1	0.2758	0.3038
8	0.92	3.00	0.30	10	2	0.1777	0.2886	4	1	3	1	0.2966	0.2852
9	0.80	3.00	0.10	9	4	0.0856	0.0994	8	4	1	0	0.0856	0.0994
10	0.93	3.00	0.30	11	2	0.1747	0.2988	5	1	3	2	0.3046	0.3048
11	0.85	2.00	0.30	12	3	0.2601	0.2637	7	2	3	1	0.3040	0.2837
12	0.80	1.75	0.30	10	3	0.3015	0.2955	8	3	2	0	0.3015	0.2955
13	0.90	3.00	0.20	13	3	0.1335	0.2032	8	2	4	1	0.2044	0.1951
14	0.94	3.00	0.30	13	2	0.1758	0.3038	5	1	5	1	0.2895	0.2883
15	0.91	3.00	0.20	15	3	0.1442	0.1970	8	2	6	1	0.1931	0.1969
16	0.80	3.00	0.05	17	7	0.0377	0.0348	13	6	2	0	0.0549	0.0426
17	0.95	3.00	0.30	16	2	0.1817	0.2995	6	1	6	1	0.2890	0.2929
18	0.92	3.00	0.20	17	3	0.1481	0.2021	9	2	7	1	0.1962	0.1976
19	0.85	3.00	0.10	16	5	0.0790	0.0853	12	4	3	1	0.1023	0.0963
20	0.90	2.00	0.30	18	3	0.2584	0.3035	10	2	6	1	0.3037	0.2926
21	0.80	2.00	0.20	16	5	0.1956	0.1951	11	4	4	1	0.2012	0.2032

The Two-Step Sequential Mesh Test for A Proportion

Pu Xiaolong[1]　Yan Zhanggeng[2]　Mao Shisong[1]　Zhang Yingshan[1]　Li Yan[1]
([1] *Department of Statistics, East China Normal University, Shanghai, 200062*)
([2] *Huayin Weapon Test Center, Huayin, 714200*)

The sequential probability ratio test (SPRT) is widely used, while the sequential mesh test is much powerful. In this article we propose the two-step sequential mesh test. The results show that it is better than the sequential mesh test.