

卡方分布下FSI CUSUM和VSI CUSUM控制图的比较

李 妍

燕山大学理学院, 河北 秦皇岛

收稿日期: 2022年3月29日; 录用日期: 2022年4月18日; 发布日期: 2022年4月28日

摘 要

本文选用平均报警时间来衡量VSI CUSUM控制图的监测效率, 在标准正态分布下分别对FSI CUSUM控制图和VSI CUSUM控制图的参数进行调整优化, 对卡方分布下FSI CUSUM控制图和VSI CUSUM控制图的监控效率进行比较。

关键词

可变抽样区间, 平均报警时间, CUSUM, 卡方分布

Comparison of FSI CUSUM and VSI CUSUM Control Charts under Chi-Square Distribution

Yan Li

School of Science, Yanshan University, Qinhuangdao Hebei

Received: Mar. 29th, 2022; accepted: Apr. 18th, 2022; published: Apr. 28th, 2022

Abstract

In this paper, the average time to signal is chosen to measure the monitoring efficiency of VSI CUSUM control chart, and the parameters of FSI CUSUM control chart and VSI CUSUM control chart are adjusted and optimized under the standard normal distribution, and the monitoring efficiency of FSI CUSUM control chart and VSI CUSUM control chart under the chi-square distribution are compared.

Keywords

Variable Sampling Interval, Average Time to Signal, CUSUM, Chi-Square Distribution

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

统计质量控制方法可以通过科学合理的手段, 实时监测生产过程中的产品质量问题, 分析异常原因, 以便及时采取措施, 保证生产过程的正常运行。质量控制图技术作为统计质量控制方法中一个主要技术, 一直以来对生产过程进行有效直观地监控, 得到各行各业的广泛运用。

随着质量控制理论中常规控制图的普遍使用, 其缺点也逐渐显现出来, 其中之一就是对过程的小偏移不灵敏, 由此, 累积和(Cumulative Sum, 简记为 CUSUM)控制图随之诞生, 并得到了极大的发展。近几年来, 国内外不少学者对这种控制图的控制限、监控效率做了大量的研究与改进, 例如, 2021 年, 许志良和朱永忠[1]对多元变异系数的自适应 CUSUM 控制图进行研究, 当多元变异系数发生不同程度的偏移时, 对偏移量进行无偏估计, 提出了两种新的具有多元变异系数的自适应 CUSUM 控制图; 最后通过蒙特卡洛模拟计算的方法, 根据运行的平均链长的值对控制图的灵敏度进行判断, 最终验证了自适应控制图在实际生产过程中具有极大的优越性。同年, 王芝珺[2]等人针对热点研究课题——对多元计数型数据的产品质量进行监控, 在对多元 CUSUM 控制图研究的基础上, 引入了加权对数似然比, 提升了多元累积和控制图的监控效果, 并且通过验证发现其在偏移程度较小的情况下, 稳健性更好。

另外, 对 CUSUM 控制图的研究大多数都是集中在正态分布条件下进行的。然而, 在实际生产过程中, 大部分的监控过程都不是服从正态总体分布的, 传统的质量控制图则容易出现漏报或者误报情况, 进而导致无效监控。因此, 近几年, 学者对非正态分布情况也进行了丰富扩展。2015 年, 毛樛[3]就针对实际生产中很难满足正态分布这一 CUSUM 控制图的前提条件提出了建议和向导。他利用随机模拟的方法, 对几种典型的非正态分布做了正态性检验, 并且对其进行正态变换; 然后在传统的 CUSUM 控制图的基础上对其进行变换。2021 年, 朱永忠和王室壹[4]同样就实际生产中控制图不服从正态分布这一问题进行了补充, 利用反正弦方法将二项分布近似成了正态分布, 提出了检验过程方差无参数的累计和控制图, 根据马尔可夫链方法计算出来的平均链长, 证明了控制图的可行性, 通过了在多元正态分布和多元 T 分布下的模拟验证。

众所周知, 传统的质量控制图一般都是静态的, 即抽样区间长度及抽样样本容量的大小都是固定的, 并且控制限也是固定的, 这些静态的质量控制图往往不能及时地监控到过程的偏移, 也不适用于当今社会及经济的快速发展。为了能够适应社会和经济的快速发展, 1988 年, Reynolds 首次提出了具有可变抽样区间(Variable Sampling Interval, 简称为 VSI)的均值控制图, 并由此开启了控制图方向的研究新领域——动态质量控制图的研究, 使得动态控制图的研究成果越来越丰富。

在此背景下, 本文在非正态分布中的卡方分布的前提下, 对固定样本抽样区间(Fixed Sampling Interval, 简称为 FSI)的 CUSUM 控制图和可变样本抽样区间(Variable Sampling Interval, 简称为 VSI)的 CUSUM 控制图的监控效率进行比较。

2. CUSUM 控制图的参数确定

2.1. CUSUM 控制图的设计

一般将观测值设定在理想条件下, 即观测值 $X_i \sim (\mu_0, \sigma_0^2)$, 其中 μ_0 和 σ_0 分别表示均值和标准差的目标值。当生产过程处于受控状态下时, 样本的均值和样本的标准差分别为 μ_0 和 σ_0 , 当生产过程处于失控状态下时, 均值由 μ_0 变为 μ_1 , 且满足等式 $\mu_1 = \mu_0 + \alpha \cdot \mu_0$, 其中 α 表示偏移参数。在 t 时刻取一个样本量为 n 的样本, 所取样本可表示为: $X_{t1}, X_{t2}, \dots, X_{tm}$, 只考虑向上偏移的情况下, CUSUM 控制图的统计量为:

$$Z_j = \max\{0, Z_{j-1} + X_j + k\}$$

其中 k 为参考值, X_j 表示连续的单个观测值。

在控制限的选取上, 以往的研究中大多都是利用 3σ 原则来确定的, 而本文将选取在实际应用中更为方便的截顶 V 型模板。

那么, 对于 FSI CUSUM 控制图来说, 判定线的斜率 F 为:

$$F = f \cdot \sigma_0$$

其中 f 为可变参数。

判定区间 H 满足等式:

$$H = h \cdot \sigma_0$$

其中 h 为可变参数。

当过程处于受控状态时, 统计量 $Z_t = \max\{0, Z_{t-1} + X_t - F - \mu_0\} < H$, 当过程处于失控状态时, 统计量 $Z_t = \max\{0, Z_{t-1} + X_t - F - \mu_0\} > H$ 。

在 VSI CUSUM 控制图中, 我们将控制限斜率 F_1 定为:

$$F_1 = f_1 \cdot \sigma_0$$

其中 f_1 为可变参数。

控制区间 H_1 满足等式:

$$H_1 = h_1 \cdot \sigma_0$$

其中 h_1 为可变参数。

VSI CUSUM 控制图中还需要设定上下警戒限, 斜率 F_2 可以写为:

$$F_2 = f_2 \cdot \sigma_0$$

其中 f_2 为可变参数, $0 < f_2 < f_1$ 。

控制区间 H_2 满足等式:

$$H_2 = h_2 \cdot \sigma_0$$

其中 h_2 为可变参数, $0 < h_2 < h_1$ 。

如此, 将控制图分为中心域和警戒域, 有:

$I_1 = (0, H_2)$ 为中心域;

$I_2 = (H_2, H_1)$ 为警戒域。

对控制图进行可变抽样区间设计时, 选取两个样本抽样间隔 T_1 和 T_2 , 且满足 $T_1 > T_2 > 0$ 。

检验方案为:

若 $0 < |Z_i| \leq H_2$ ，即观测点落在中心域，说明产品质量良好，此时可以采取宽松的政策，也就是延长下一次抽样的时间，增大样本抽样间隔，即下一个抽样间隔为 T_1 ；

若 $H_2 < |Z_i| \leq H_1$ ，即观测点落在警戒域，说明产品质量情况不稳定，有落在控制限之外的可能性，此时需要引起警惕，为了尽快发现偏差出现的原因是否具有随机性，需要缩短下一次抽样的时间，减小样本抽样间隔，即下一个抽样间隔为 T_2 ；

若 $|Z_i| > H_1$ ，即观测点落在控制限之外，说明产品生产过程中出现了严重问题，此时报警。

2.2. 控制图比较标准

衡量静态控制图监测效率的优劣，一般通过产品的平均链长(Average RunLength, 简记为 ARL)的大小来比较，但是根据 ARL 的定义式，我们可以了解到在可变抽样区间的动态控制图中，产品平均链长 ARL 无法作为其比较的指标。因此，我们可以采用平均报警时间(Average Time to Signal, 简记为 ATS)，表示从生产过程开始被检测到报警所需要的平均时间，作为可变抽样区间控制图的指标，计算公式为：

$$ATS = E(T) \cdot ARL$$

其中， $E(T)$ 为平均样本抽样间隔，计算公式为：

$$E(T) = T_1 \cdot p + T_2 \cdot (1 - p)$$

其中 p 为观测点落入警戒域的概率。

当过程发生偏移的时候，ATS 越小越好，这样就说明其能很快地发现偏移，及时发出警报。

要比较两种质量控制图的监测效率，一方面需使得它们在无偏移时 FSI CUSUM 控制图的 ARL 与 VSI CUSUM 控制图的 ATS 大致相等，另一方面需要保证两种控制图的成本大致相等，即保证 VSI CUSUM 控制图的平均样本抽样间隔 $E(T)$ 与 FSI CUSUM 控制图的样本抽样间隔 T 大致相等。

2.3. CUSUM 控制图参数的确定

对于 FSI CUSUM 控制图来说，可以给定平均链长 ARL 值为 370，抽样间隔 T 为 1 单位，在重视 0.5 倍偏移量时，利用国标中的常用参数 $f = 0.25$ ，经过随机模拟计算可得 h 值，见表 1。

Table 1. ARL values corresponding to different h values
表 1. 不同 h 值对应的 ARL 值

h	ARL
5	142.6887
6	248.2994
7	428.2408
8	738.8482
9	1257.5

从表 1 中可以清晰的看出， h 取值不同时得到的平均链长的值都不同，并且随着参数 h 选取的值的不断增大，平均链长也随之变大。通过对表 1 的观察，发现要使 ARL 值为 370， h 取值在 6 到 7 之间，且更接近 7，因此可以对 h 取值进一步精确，计算结果见表 2。

通过表 2 可以得到，在误差允许的范围内，受平均链长为 370 的条件约束， h 取值为 6.7。因此，FSI CUSUM 控制图的参数为 $f = 0.25$ ， $h = 6.7$ 。

Table 2. ARL values corresponding to different h values
表 2. 不同 h 值对应的 ARL 值

h	ARL
6.5	334.0950
6.6	350.6511
6.7	369.9284
6.8	388.8791
6.9	413.2099

对于 VSI CUSUM 控制图来说, 在保证两种控制图指标值相等的原则下, 要使平均报警时间 ATS 与平均链长 ARL 值同为 370, 同样在重视 0.5 倍偏移量时, 利用国标中的常用参数数 $f_1 = 0.25$, $f_2 = 0.25$ 。而参数 h_1 , h_2 , 需满足 $h_1 > h > h_2$ 的关系; 选取两个样本抽样间隔 T_1 , T_2 , 需满足 $T_1 > T_2 > 0$ 的关系。

当 $h_1 = 6.75$, $h_2 = 6.65$ 时, 利用随机模拟可以计算观测点落入警戒域的概率 p 为 0.0032, 在 ATS 值为 370 和 $E(T)$ 值接近 1 的条件下, 可得 $T_1 = 0.99$, $T_2 = 0.4685$, 此时 $E(T) = 0.9884$ 与固定样本抽样区间控制图的抽样间隔 1 大致相等, 满足生产成本要求。

因此, VSI CUSUM 控制图的参数为 $f_1 = 0.25$, $f_2 = 0.25$, $h_1 = 6.75$, $h_2 = 6.65$, $T_1 = 0.99$, $T_2 = 0.4685$ 。

3. 卡方分布下两种控制图的比较

3.1. CUSUM 控制图

卡方分布同正态分布一样, 在质量控制这一方面, 主要是描述控制实验中的测量误差的分布。从理论上讲, 正常情况下测量误差应该是服从正态分布的, 然而, 实际状况总会与理论存在一定的冲突。本章节是在结合了张华初[5]等人对卡方分布模拟的基础上, 假设实验中测量误差服从了卡方分布, 对两种控制图的监测效率进行比较。

假定观测值 $X_i \sim \chi^2(v)$, 其中 v 代表自由度。令 μ_0 和 σ_0 分别表示均值和标准差的目标值。根据卡方分布的性质, 我们可以得到:

$$E(X_i) = \mu_0 = v$$

$$\text{Var}(X_i) = \sigma_0^2 = 2v$$

CUSUM 控制图的统计量可以写为:

$$Z_t = \max\{0, Z_{t-1} + X_t + k\}$$

在控制限的选取上, 利用截顶 V 型模板来确定。当过程处于受控状态时, 统计量 $Z_t = \max\{0, Z_{t-1} + X_t - F - v\} < H$; 当过程处于失控状态时, 统计量 $Z_t = \max\{0, Z_{t-1} + X_t - F - v\} > H$ 。

根据前文论述可知, VSI CUSUM 控制图的控制限和警戒限的斜率以及控制区间和警戒区间等式可写为:

$$F_1 = f_1 \cdot \sigma_0 = f_1 \cdot \sqrt{2v}$$

$$H_1 = h_1 \cdot \sigma_0 = h_1 \cdot \sqrt{2v}$$

$$F_2 = f_2 \cdot \sigma_0 = f_2 \cdot \sqrt{2v}$$

$$H_2 = h_2 \cdot \sigma_0 = h_2 \cdot \sqrt{2v}$$

3.2. 两种控制图的比较

对于固定样本抽样区间的 CUSUM 控制图来说, 取参数值 $f = 0.25$, $h = 6.7$, 在发生偏移时得到不同的平均链长 ARL。对于可变样本抽样区间的 CUSUM 控制图来说, 取参数值 $f_1 = 0.25$, $f_2 = 0.25$, $h_1 = 6.75$, $h_2 = 6.65$, $T_1 = 0.99$, $T_2 = 0.4685$, 在发生偏移时得到不同的平均报警时间 ATS。

在服从自由度为 1 的卡方分布下, 计算 ARL 和 ATS 结果见表 3。

Table 3. ARL and ATS values at different offsets

表 3. 不同偏移下 ARL 和 ATS 值

α	ARL	ATS
0	177.8245	177.8588
0.25	91.9763	67.1829
0.5	45.6391	27.4106
0.75	25.1016	14.7561
1	16.8407	9.6519

通过观察表 3 可以发现, 当过程失控, 出现偏移时, VSI CUSUM 控制图能更快的发现异常, 发出警报。在偏移量大小相同时, FSI CUSUM 和 VSI CUSUM 控制图的表现各有不同, VSI CUSUM 控制图的平均报警时间更小, 尤其是当发生 0.5 倍偏移量时, FSI CUSUM 控制图和 VSI CUSUM 控制图的平均链长和平均报警时间分别是 45.6391 和 27.4106, VSI CUSUM 控制图具有明显优势。另外, 随着偏移量的增大, 平均链长和平均报警时间都明显降低, 而且平均报警时间明显小于平均链长。为了更加清晰直观地观察, 根据表 3 绘制折线图, 见图 1。

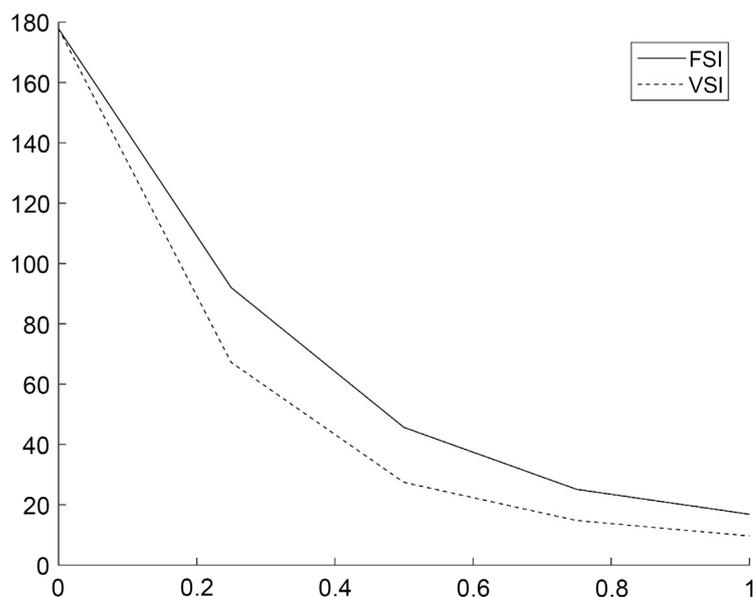


Figure 1. ARL and ATS values at different offsets

图 1. 不同偏移下 ARL 和 ATS 值

从图 1 中可以清晰地看出 FSI CUSUM 控制图和 VSI CUSUM 控制图的平均链长和平均报警时间之间

的差距, 显而易见, 平均报警时间明显小于平均链长, 说明在服从自由度为 1 的卡方分布下, VSI CUSUM 控制图比 FSI CUSUM 控制图更能及时发现偏移, 发出警报, 减小企业损失。

根据卡方分布特点, 增大自由度的值, 在服从自由度为 5 的卡方分布下, 计算 ARL 和 ATS, 结果见表 4。

Table 4. ARL and ATS values at different offsets

表 4. 不同偏移下 ARL 和 ATS 值

α	ARL	ATS
0	234.1975	234.5298
0.25	148.9077	63.1819
0.5	100.1242	24.6758
0.75	67.6225	13.7135
1	48.0725	9.4145

通过观察表 4 可以发现, 当过程失控, 出现偏移时, VSI CUSUM 控制图能更快的发现异常, 发出警报。在偏移量大小相同时, VSI CUSUM 控制图的平均报警时间更小, 尤其是当发生 0.5 倍偏移量时, FSI CUSUM 控制图和 VSI CUSUM 控制图的平均链长和平均报警时间分别是 100.1242 和 24.6758, VSI CUSUM 控制图具有明显优势。另外, 随着偏移量的增大, 平均链长和平均报警时间都明显降低, 而且平均报警时间明显小于平均链长。为了更加清晰直观地观察, 根据表 4 绘制折线图, 见图 2。

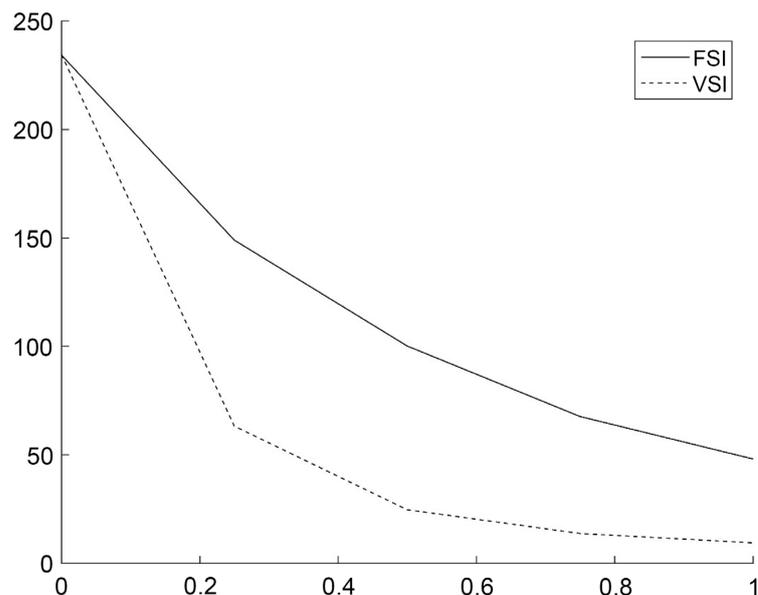


Figure 2. ARL and ATS values at different offsets

图 2. 不同偏移下 ARL 和 ATS 值

从图 2 中可以清晰地看出, FSI CUSUM 控制图和 VSI CUSUM 控制图的平均链长和平均报警时间之间的差距更加明显, 平均报警时间明显小于平均链长, 说明在服从自由度为 5 的卡方分布下, VSI CUSUM 控制图比 FSI CUSUM 控制图更能及时发现偏移, 发出警报, 减小企业损失。

4. 结论

综合上述,可以发现,随着自由度的不断增大,FSI CUSUM 控制图和 VSI CUSUM 控制图的平均链长和平均报警时间之间的差距也在不断增大。当然,无论自由度如何变化,都是 VSI CUSUM 控制图表现得更好,平均报警时间更小,监控效率更高,因此建议当观测值服从卡方分布时选择监控效率更高的 VSI CUSUM 控制图。

参考文献

- [1] 许志良,朱永忠. 两种多元变异系数的自适应 CUSUM 控制图[J]. 信息技术, 2021(7): 1-8, 14.
- [2] 王芝珺,雷骏峰,吴纯杰. 基于对数似然比的多元加权 Poisson CUSUM 控制图研究及应用[J]. 系统科学与数学, 2021, 41(3): 837-853.
- [3] 毛樛. 基于正态变换的 CUSUM 控制图应用研究[D]: [硕士学位论文]. 上海: 华东师范大学, 2015.
- [4] 朱永忠,王室壹. 监控过程方差无参数多变量 CUSUM 控制图设计[J]. 统计与决策, 2021, 37(5): 49-53.
- [5] 张华初,楚鹏飞,谢观霞. 统计分布和中心极限定理的随机模拟[J]. 统计与决策, 2021, 37(4): 69-72.