混合正态分布下的似然比累积和控制图及改进

吴纯杰*, 魏一青, 郁淼淼

上海财经大学统计与管理学院, 上海 200433

E-mail: wumaths@mail.shufe.edu.cn, wishwyq@126.com, yumiaomiao_01@163.com

收稿日期: 2016-03-02; 接受日期: 2016-07-02; 网络出版日期: 2017-05-18; * 通信作者

国家自然科学基金 (批准号: 11301323) 资助项目

摘要 累积和控制图主要用于监测正态分布过程中均值的中小漂移. 它对于厚尾分布过程和大漂移的监测效果十分不理想. 本文研究了在混合正态分布下的稳健似然比累积和 (RLCUSUM) 控制图的应用. 该控制图通过构造似然比检验函数作为监测指标, 改善了其对于大漂移和厚尾分布过程的监测效果. 比较控制图的平均运行长度, 发现稳健似然比累积和控制图比累积和控制图更灵敏. 最后提出了水平线段截断的方法来改进稳健似然比累积和控制图的稳健性.

关键词 稳健似然比累积和控制图 平均运行长度 水平线段截断 混合正态分布

MSC (2010) 主题分类 62P30

1 引言

控制图是一种基于假设检验构造的监测过程是否处于受控状态的重要统计工具. Page^[1] 提出了累积和检验 (cumulative sum, CUSUM) 控制图, 该控制图用于监测正态分布下的生产过程发生中小漂移的情形. 但是随着生产过程的日益复杂化, 生产过程数据的分布形态不仅仅局限于正态分布, 或者说绝大部分的数据都不再服从于正态分布. CUSUM 控制图对于生产过程中常见的厚尾分布过程的监测, 存在报警延迟和反应不灵敏问题, 并且 CUSUM 控制图对于大漂移的检验不是十分灵敏. 因此, 为了满足生产过程中监测的需求, 有必要在非正态分布下对控制图进行重新设计. 本文主要研究在厚尾分布下 (以混合正态分布为例) CUSUM 控制图的改进, 克服了 CUSUM 控制图在监测过程中使用的局限性.

对于 CUSUM 控制图在大漂移监测上的改进有了许多理论研究. Lucas^[2] 提出了 Shewhart-CUSUM 联合控制图, 用 CUSUM 控制图来检测小漂移, 在 CUSUM 控制图上增加 Shewhart 控制限来加快对大漂移的检测速度, 从而增大了 CUSUM 控制图对于大漂移监测的灵敏度. 对于 CUSUM 控制图对厚尾分布的监测, 亦有所突破. Vardeman 和 Ray^[3] 介绍了指数分布下的 CUSUM 控制图的设计方案, 并通过模拟验证了其在指数分布下优良的监测效果. 似然比统计量作为检测指标被广泛用于 CUSUM 控制图. 代毅等^[4] 根据似然比统计量的性质, 提出了两种基于似然比统计量的用于预分析的累积和控制图, 分别用于监测均值的漂移与监测均值和方差的同时漂移. 关于监测的稳健性也

英文引用格式: Wu C J, Wei Y Q, Yu M M. Improved robust-likelihood cumulative sum chart for the contaminated normal distributions (in Chinese). Sci Sin Math, 2017, 47: 853–868, doi: 10.1360/N012016-00048

被更多地采纳到 CUSUM 控制图设计中. Lucas 和 Crosier^[5] 提出了设计的稳健 CUSUM 控制图基于 3σ 原理, 即一个观测值落在控制限外的概率为 0.23%, 而连续两个观测值落在控制限外的概率为 $5.29 \times 10^{-4}\%$, 为极小概率事件, 当此事件发生时说明观测值异常. 故其指出忽略第一个异常值并对连续出现的两个异常值给出警报的控制图表现稳健. Hawkins^[6] 提出了对于监测数据进行截断的方法构建稳健 CUSUM 控制图. 事先设定一个截断常数 W (文中对不同情形下的数据均采用截断值 $W = 2$), 具体步骤如下:

- (1) 使用样本 $\{X_n\}$ 的均值和方差作为初始的估计值 $\hat{\mu}$ 和 $\hat{\sigma}^2$.
- (2) 若 $(X_n - \hat{\mu})/\hat{\sigma} > W$, 则 $Y_n = \hat{\mu} + W\hat{\sigma}$; 若 $(X_n - \hat{\mu})/\hat{\sigma} < -W$, 则 $Y_n = \hat{\mu} - W\hat{\sigma}$; 反之, $Y_n = X_n$.
- (3) 重新计算样本 $\{Y_n\}$ 的 $\hat{\mu}$ 和 $\hat{\sigma}$.
- (4) 重复第 (2) 步直至 $\hat{\mu}$ 和 $\hat{\sigma}$ 的值稳定.

Stoumbos 和 Reynolds^[7] 分析了正态分布下和双指数分布下 CUSUM 控制图的稳健性, 研究显示当样本容量 $n = 1$ 时, CUSUM 控制图能够被设计得非常稳健, 且对不同大小的漂移均有较好的检测效果. Reynolds 和 Stoumbos^[8] 研究了两种稳健的联合 CUSUM 控制图, 其中一种是检测均值的标准 CUSUM 控制图与检测标准差的绝对偏差 CUSUM 控制图的联合, 另一种是检测均值的标准 CUSUM 控制图与检测标准差的方差 CUSUM 控制图, 前者在检测小漂移时稳健, 后者适用于不需对单个异常值报警的情形. 许多工业和服务业应用中, 我们所关心的产品质量已经不是简单的某个或某几个指标的均值或方差 (或分布), 而是需要用一些自变量与响应之间的某种关系来更好地刻画或衡量. 因此, 控制图的监测指标也日趋多样化. 邹长亮^[9] 归纳了复杂数据过程和多元过程等过程控制中的监测指标, 引入并开发了各种新的统计技术和紧密结合计算算法, 解决了这些当前质量控制领域研究的难点问题. Hafiz 等^[10] 提出了在不同分布情形下, 稳健 CUSUM 控制图将均值、中位数和 Hodges-Lehmann 等位置统计量作为监测指标的适用情形. 将似然比统计量作为控制图监测指标成为近几年研究热点. 似然比函数与对数似然比函数在渐进性和相合性等方面有着很好的性质. 张久军^[11] 归纳总结了基于似然比检验的若干控制图及其相关性质. MacEachern 等^[12] 提出了稳健似然比累积和 (robust-likelihood cumulative sum, RLCUSUM) 控制图. 他们对 Hawkins^[6] 提出的截断方法进行了拓展, 对对数似然比函数进行水平射线截断, 由此提出了针对 t 分布非单调的对数似然比函数改造成单调的似然比函数的方法. 对于似然比函数的截断, 其截断点需根据似然函数而改变, 相较于 Hawkins 的截断更加灵活, 提高了控制图的效率. 截断后的似然比函数称为稳健对数似然比 (robust log-likelihood ratio, RLLR), 并将 RLLR 作为控制图的检验统计量. 该控制图对异常值不敏感, 又能提高监测大漂移的效率. Liu 等^[13] 提出了对于稳健似然比改进的方法, 使得不同控制限下可控平均运行长度不存在间断点.

本文主要研究混合正态分布下的稳健似然比累积和控制图及其改进. 本文深入研究稳健似然比累积和控制图的性质, 证明 ARL_0 (average run length) 曲线的间断点的存在性. 模拟计算证明混合正态分布下稳健似然比累积和控制图表现的稳定性. 本文发现稳健似然比累积和控制图检测效果不佳及失灵的情形, 提出了改进的稳健似然比累积和控制图的水平线段截断法, 证明了它的优良性质和合理性.

2 RLCUSUM 控制图

2.1 CUSUM 控制图和序贯概率比检验

与传统的假设检验不同, 序贯概率比检验 (sequential probability ratio test, SPRT) 并没有固定的样本容量. 其假设为 $H_0: x_1, x_2, \dots \stackrel{iid}{\sim} f_0(x)$ vs. $H_1: x_1, x_2, \dots \stackrel{iid}{\sim} f_1(x)$. 抽得样本 $X_1, X_2, \dots, X_n$, 似然

比统计量为 $\lambda_n = \frac{\prod_{i=1}^n f_1(x_i)}{\prod_{i=1}^n f_0(x_i)}$. 对数似然比函数为

$$\ln \lambda_n = \sum_{i=1}^n \ln \frac{f_1(x_i)}{f_0(x_i)} = \sum_{i=1}^n z_i, \quad z_i = \ln \frac{f_1(x_i)}{f_0(x_i)}.$$

对于事先给定的两个常数 $0 < A < 1 < B < +\infty$, 当 $\lambda_n \geq B$ 时, 停止抽样, 拒绝原假设; 当 $\lambda_n \leq A$ 时, 停止抽样, 接受原假设; 当 $A < \lambda_n < B$ 时, 继续抽取第 $n+1$ 个样本并重复以上步骤. 设停止抽样时间 (简称停时) 为 $\tau^* = \min\{n : \lambda_n \leq A \text{ 或 } \lambda_n \geq B, n \geq 1\}$. 当 $P(f_0(x) = f_1(x)) < 1$ 时, 有 $P_{H_1}(\tau^* < \infty) = 1$, $E_{H_1} \exp\{\lambda \tau^*\} < \infty$, $E_{H_1}(\tau^*)^k < \infty$, $\forall \lambda > 0, k > 0, i = 0, 1$.

CUSUM 控制图的提出正是基于 SPRT. 它们的主要区别是 CUSUM 控制图中受控的假设 H_0 一直不被接受. 设 $x_1, x_2, \dots, x_n$ 是来自某生产过程的质量特性 X 的一列随机抽样样本, 可控状态下 X 服从密度函数为 $f_0(x)$ 的分布, 失控状态下 X 服从密度函数为 $f_1(x)$ 的分布, 且 $f_0(x) \neq f_1(x)$. 建立如下变点模型:

$$X_i \sim \begin{cases} f_0(x), & i = 1, 2, \dots, n, \\ f_1(x), & i = \nu + 1, \nu + 2, \dots, n, \quad \nu \text{ 为变点时刻.} \end{cases}$$

该模型等价表述形式为

$$H_0 : \nu = n \quad \text{vs.} \quad H_1 : \nu < n,$$

其中 ν 为非负整数, 则参数空间 $\Theta_0 = \{n\}$, $\Theta_1 = \{0, 1, \dots, n-1\}$, $\Theta = \{0, 1, \dots, n\}$. 似然比统计量为

$$\lambda_n = \frac{\sup_{\theta \in \Theta} f(x, \theta)}{\sup_{\theta \in \Theta_0} f(x, \theta)} = \begin{cases} \max \left\{ \frac{\prod_{i=1}^{\nu} f_0(x_i) \prod_{i=\nu+1}^n f_1(x_i)}{\prod_{i=1}^n f_0(x_i)} \right\} = \max \left\{ \frac{\prod_{i=\nu+1}^n f_1(x_i)}{\prod_{i=\nu+1}^n f_0(x_i)} \right\}, & \nu < n, \\ 1, & \nu = n. \end{cases}$$

对数似然比统计量为

$$S_n = \ln \lambda_n = \begin{cases} \max \left\{ \sum_{i=\nu+1}^n \ln \frac{f_1(x_i)}{f_0(x_i)} \right\} = \max \left\{ \sum_{i=\nu+1}^n z_i \right\}, & \nu < n, \\ 0, & \nu = n, \end{cases}$$

其中 $z_i = \ln \frac{f_1(x_i)}{f_0(x_i)}$, S_n 称为 CUSUM 统计量, 也可等价表示为 $S_n = \max\{0, S_{n-1} + z_n\}$.

似然比检验的拒绝域形式为 $\{x : \lambda(x) > c\}$, 即 CUSUM 检验统计量的拒绝域为 $\{x : S_n > h\}$, 其中控制限 $h > 0$. $S_n = \max\{0, S_{n-1} + z_n\}$ 是对数似然比函数的累加迭代, 它描述了如下计算过程: 假设在时刻 n , 过去时刻的累积和统计量为 S_{n-1} , 且满足 $0 \leq S_{n-1} \leq h$. 令 $S_n = S_{n-1} + z_n$, 若 $S_n < 0$, 即 $z_n < -S_{n-1}$, 则将检验统计量清零, 即令 $S_n = 0$; 若 $S_n > h$, 即 $z_n > h - S_{n-1}$, 则认为过程失控, 停止抽样并检查生产过程; 若 $0 \leq S_n \leq h$, 即 $-S_{n-1} \leq z_n \leq h - S_{n-1}$, 则继续抽取第 $n+1$ 个观测, 并重复以上过程.

事实上, CUSUM 检验统计量具有无后效性. 当 S_{n-1} 已知时, CUSUM 统计量 S_n 的分布仅与 S_{n-1} 有关, 与 $S_1, S_2, \dots, S_{n-2}$ 无关. 给定控制限 $h > 0$, 当 $S_{n-1} > h$ 时, 我们认为 $S_n > h$. 因此, 当 $0 \leq x \leq h$ 时,

$$P(S_n \leq x | S_1, S_2, \dots, S_{n-1}) = \begin{cases} P(z_n \leq x - S_{n-1} | S_{n-1}), & S_{n-1} \leq h, \\ 0, & S_{n-1} > h. \end{cases}$$

所以, CUSUM 序列 $\{S_n, n = 0, 1, 2, \dots\}$ 可以看作一个 Markov 链.

2.2 混合正态分布和对数似然比函数

设一系列观测值为 $x_1, x_2, \dots, x_n$, 多数观测值服从 $N(\mu, \sigma_0^2)$, 少数观测值由于受到某些因素的影响而发生变化, 成为异常观测值, 服从 $N(\mu, \sigma^2)$. 设 ϵ ($0 < \epsilon < 1$) 为异常观测值在全部观测值中所占的比例, 当 n 较大时, 我们可以认为总体的分布函数为

$$F(x) = P(X \leq x) = (1 - \epsilon)\Phi\left(\frac{x - \mu}{\sigma_0}\right) + \epsilon\Phi\left(\frac{x - \mu}{\sigma}\right),$$

其中, $\Phi(\cdot)$ 表示标准正态分布函数. 称上式表示的分布为混合正态分布, 或受污染的正态分布 (contaminated normal distribution), 记为 $X \sim (1 - \epsilon)N(\mu, \sigma_0^2) + \epsilon N(\mu, \sigma^2)$, 称 ϵ 为污染程度. 混合正态分布的密度函数为

$$f(x) = (1 - \epsilon)\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(x - \mu)^2}{2\sigma_0^2}\right) + \epsilon\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

检验均值发生漂移的情形, 构造假设检验: $H_0: \mu = \mu_0$ vs. $H_1: \mu = \mu_1$, 则对数似然比为

$$z_i = \ln \frac{f(x | \mu_1)}{f(x | \mu_0)} = \ln \frac{(1 - \epsilon)\frac{1}{\sigma_0} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_0^2}\right) + \epsilon\frac{1}{\sigma} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma^2}\right)}{(1 - \epsilon)\frac{1}{\sigma_0} \exp\left(-\frac{(x - \mu_0)^2}{2\sigma_0^2}\right) + \epsilon\frac{1}{\sigma} \exp\left(-\frac{(x - \mu_0)^2}{2\sigma^2}\right)}.$$

将 z_i 视为关于质量特性 X 的函数, 为了研究不同参数 μ_1 、 σ^2 和 ϵ 下对数似然比函数的变化趋势, 固定 $\mu_0 = 0$, 见图 1-3.

图 1 不同 μ_1 下的对数似然比函数, 其中 $\mu_0 = 0$, $\sigma_0^2 = 1$, $\sigma^2 = 9$, $\epsilon = 0.05$

图 2 不同 σ^2 下的对数似然比函数, 其中 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0^2 = 1$, $\epsilon = 0.05$

图 3 不同污染程度下的对数似然比函数, 其中 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0^2 = 1$, $\sigma^2 = 9$

由图 1 可知, 当 $\mu_1 > 0$, 也就是检测均值向上漂移时, z_i 的总趋势是随着 X 的增大而增大; 当 $\mu_1 < 0$, 也就是检测均值向下漂移时, z_i 的总趋势是随着 X 的增大而减小. 随着 μ_1 ($\mu_1 > 0$) 的增大, 对数似然比的中心点右移, 总趋势的斜率增大, 非单调区间的跨度逐渐变大. 我们有 $\ln \frac{f(\mu_0 + \mu_1 - x | \mu_1)}{f(\mu_0 + \mu_1 - x | \mu_0)} = -\ln \frac{f(x | \mu_1)}{f(x | \mu_0)}$, 对数似然比函数的中心点为 $(\frac{\mu_0 + \mu_1}{2}, 0)$.

从图 2 可以看到, 当 $\sigma^2 > \sigma_0^2$ 时, 随着 σ^2 的增大, 对数似然比函数在非单调区间上的波动幅度增大, 两侧变得相对平坦; 但当 $\sigma^2 < \sigma_0^2$ 时, 对数似然比函数几乎是一条直线. 因此, 本文主要关注 $\sigma^2 > \sigma_0^2$ 的情形.

由图 3 可知, 当 $\sigma^2 > \sigma_0^2$ 时, 随着污染程度的增加, 对数似然比函数的非单调区间变窄, 函数在非单调区间上的波动幅度减小, 即当异常观测值的方差相对较大时, 少量的异常观测值对对数似然比函数有较大的扰动.

2.3 RLCUSUM 控制图

由 Neyman-Pearson 引理可知, 对于单边检验问题, 若似然比是单调似然比 (monotone likelihood ratio, MLR), 则可以构造出一个水平为 α 的一致最优势检验 (uniformly most powerful test, UMPT). 但是当似然比函数不单调时, 由此构造的似然比检验并不是 UMPT, 检验效果表现不理想. 由图 1-3 可知, 混合正态分布下的似然比函数是非单调的.

用 Markov 链法计算 CUSUM 控制图的平均运行长度 (具体见附录 A), 可得 $ARL_0 = 1,000$ 时, $h = 5.08$. 令 $h = 5.08$, 用不同均值 μ 的混合正态分布模拟真实数据, 求得控制图的 ARL_1 , 其中 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0 = 1$, $\sigma = 3$, $\epsilon = 0.05$, 结果如图 4 所示. 由图 4 可知, 混合正态分布的 CUSUM 控制图的 ARL_1 并没有随着 μ 的增大而下降. 当 $\mu < 3$ 时, ARL_1 迅速下降; 当 $\mu > 3$ 时, ARL_1 开始上升, 并在 $\mu = 6$ 处达到最大值, 之后再缓慢下降. 与漂移为 3 时相比, 漂移为 6 时 CUSUM 控制图的反应更加迟钝. 这种情形并不理想, 我们往往希望漂移越大 CUSUM 控制图报警越迅速. 在 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0 = 1$, $\sigma = 3$ 以及 $\epsilon = 0.05$ 时, $\Pr(x > 3) = 1 - \Pr(x \leq 3) = 1 - (0.95\Phi(3) + 0.05\Phi(\frac{2}{3})) = 0.0139$, 则对于 $\mu > 3$ 的大漂移情形是存在的, 即以上 ARL_1 不合理的漂移区间内的均值漂移是无法被忽略的.

图 4 不同真实均值 μ 下 CUSUM 控制图的 ARL_1

为了解决上述两个问题, 构造稳健对数似然比 (RLLR) 函数 (参见文献 [12])

$$l^*(x) = \begin{cases} \ln \frac{F(a | \mu_1)}{F(a | \mu_0)}, & x \leq a, \\ \ln \frac{f(x | \mu_1)}{f(x | \mu_0)}, & a < x < b, \\ \ln \frac{1 - F(b | \mu_1)}{1 - F(b | \mu_0)}, & x \geq b. \end{cases}$$

为保持单调性和连续性, 则应有

$$\ln \frac{F(a | \mu_1)}{F(a | \mu_0)} = \ln \frac{f(a | \mu_1)}{f(a | \mu_0)}, \quad \ln \frac{1 - F(b | \mu_1)}{1 - F(b | \mu_0)} = \ln \frac{f(b | \mu_1)}{f(b | \mu_0)}, \quad (2.1)$$

且 $a, b \in [x_2, x_3]$, 即 a 和 b 位于第二和三极值点之间. 令 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0^2 = 1$, $\sigma^2 = 9$, $\epsilon = 0.05$, 解方程得 $a = -1.251$, $b = 2.251$. 由图 5(a) 可见, 如此构造的稳健对数似然比函数保持了单调性和连续性. 左截断点在第二个极值点的右上方, 右截断点在第三个极值点的左下方, 但都相距不远, 使控制图对大漂移有较灵敏反应的同时对异常点稳健.

用稳健对数似然比 $l^*(x)$ 代替 CUSUM 统计量中的 z_i 得到 RLCUSUM 统计量

$$S_0^* = 0, \quad S_n^* = \max\{0, S_{n-1}^* + l^*(x_n)\}.$$

当 $\mu_1 > 0$ 时, 该控制图对均值向上漂移进行检测; 当 $\mu_1 < 0$ 时, 该控制图对均值向下漂移进行检测. 用 Markov 链法计算混合正态分布 CUSUM 控制图和 RLCUSUM 控制图不同控制限 h 下的 ARL_0 , 如图 6

图 5 当 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0^2 = 1$, $\sigma^2 = 9$, $\epsilon = 0.05$ 时, RLLR 和改进的 RLLR

 图 6 混合正态分布 RLCUSUM 控制图不同控制限 h 下的 ARL_0

所示, 其中 $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0^2 = 1$, $\sigma^2 = 9$, $\epsilon = 0.05$. 由图 6 可知, RLCUSUM 控制图的 ARL_0 关于 h 不连续, 在 $h = 1.63$ 和 $h = 3.26$ 处有明显的间断点, 而

$$l^*(b) = \ln \frac{f(b | \mu_1)}{f(b | \mu_0)} = 1.63,$$

其中 $b = 2.251$ 为右截断点. 这是因为当 h 从稍小于 $k \times l^*(b)$ (其中 $k = 1, 2, \dots$) 的某个位置移动到 $k \times l^*(b)$ 时, 所有本应该报警的使得 $S_n^* = k \times l^*(b)$ 的点列不再报警, 而这些点列取值的概率大于 0, ARL_0 在 $h = k \times l^*(b)$ 处向上跳跃. MacEachern 等^[12] 已发现了 ARL_0 的不连续性, 本文对此性质给出详细证明.

定理 1 RLCUSUM 控制图的 ARL_0 在 $h = k \times l^*(b)$ 处间断, 其中 $k = 1, 2, \dots$, $l^*(b)$ 为对数似然比函数在右截断点的取值.

证明 对 $\forall \epsilon > 0$, 有

$$\begin{aligned} & E(\tau^* | h = k \times l^*(b)) - E(\tau^* | h = k \times l^*(b) - \epsilon) \\ &= \sum_{m=0}^{\infty} P(T_i > m | h = k \times l^*(b)) - \sum_{m=0}^{\infty} P(T_i > m | h = k \times l^*(b) - \epsilon) \\ &= \sum_{m=1}^{k-1} P(S_1 \leq k \times l^*(b), S_2 \leq k \times l^*(b), \dots, S_m \leq k \times l^*(b)) \\ &\quad - \sum_{m=1}^{k-1} P(S_1 \leq k \times l^*(b) - \epsilon, S_2 \leq k \times l^*(b) - \epsilon, \dots, S_m \leq k \times l^*(b) - \epsilon) \\ &\quad + \sum_{m=k}^{\infty} P(S_k \leq k \times l^*(b), S_{k+1} \leq k \times l^*(b), \dots, S_m \leq k \times l^*(b)) \\ &\quad - \sum_{m=k}^{\infty} P(S_k \leq k \times l^*(b) - \epsilon, S_{k+1} \leq k \times l^*(b) - \epsilon, \dots, S_m \leq k \times l^*(b) - \epsilon). \end{aligned}$$

因为

$$\bigcap_{m=1}^i \{S_m \leq k \times l^*(b)\} \supseteq \bigcap_{m=1}^i \{S_m \leq k \times l^*(b) - \epsilon\}, \quad i < k,$$

所以,

$$\begin{aligned} & \sum_{m=1}^{k-1} P(S_1 \leq k \times l^*(b), S_2 \leq k \times l^*(b), \dots, S_m \leq k \times l^*(b)) \\ & - \sum_{m=1}^{k-1} P(S_1 \leq k \times l^*(b) - \epsilon, S_2 \leq k \times l^*(b) - \epsilon, \dots, S_m \leq k \times l^*(b) - \epsilon) \geq 0. \end{aligned}$$

因为

$$\begin{aligned} & \bigcap_{m=1}^i \{S_m \leq k \times l^*(b) - \epsilon\} \cap \{l^*(x_1) = 0, l^*(x_2) = 0, \dots, l^*(x_{m-k}) = 0, l^*(x_{m-k+1}) = l^*(b), \\ & l^*(x_{m-k+2}) = l^*(b), \dots, l^*(x_m) = l^*(b)\} \subseteq \bigcap_{m=1}^i \{S_m \leq k \times l^*(b)\}, \quad i \geq k. \end{aligned}$$

所以,

$$\begin{aligned} & \sum_{m=k}^{\infty} P(S_1 \leq k \times l^*(b), S_2 \leq k \times l^*(b), \dots, S_m \leq k \times l^*(b)) \\ & - \sum_{m=k}^{\infty} P(S_1 \leq k \times l^*(b) - \epsilon, S_2 \leq k \times l^*(b) - \epsilon, \dots, S_m \leq k \times l^*(b) - \epsilon) \\ & \geq \sum_{m=k}^{\infty} P(l^*(x) \leq 0)^{m-k} P(l^*(x) = b)^k \\ & = \frac{(1 - F(b | \mu_0))^k}{1 - F(\frac{\mu_0 + \mu_1}{2} | \mu_0)}, \quad i \geq k. \end{aligned}$$

所以,

$$E(\tau^* | h = k \times l^*(b)) - E(\tau^* | h = k \times l^*(b) - \epsilon) \geq \frac{(1 - F(b | \mu_0))^k}{1 - F(\frac{\mu_0 + \mu_1}{2} | \mu_0)} > 0,$$

即存在间断点. □

混合正态分布的 RLCUSUM 控制图与 CUSUM 控制图的 ARL_0 十分接近. 令 $h = 5.08$, 用不同均值 μ 的混合正态分布模拟真实数据. 分别计算 RLCUSUM 控制图和 CUSUM 控制图的 ARL_1 , 结果如图 7(a) 所示. 由图 7(a) 可知, 虽然 RLCUSUM 控制图并没有在所有情形下都优于 CUSUM 控制图, 但它却十分稳定, 克服了 CUSUM 控制图在较大漂移下不灵敏的缺点. 同时, 在对数似然比函数尾部进行截断, 使得 RLCUSUM 控制图不容易受到异常值的影响. 因此可以认为混合正态分布下的 RLCUSUM 控制图是对 CUSUM 控制图的一种优化, 使控制图在实际应用中对较大漂移灵敏的同时不容易受到异常值的影响, 表现更加稳健.

 图 7 当 $ARL_0 = 1,000$ 时, RLCUSUM 控制图和 CUSUM 控制图的 ARL_1

3 改进的稳健似然比累积和控制图

3.1 RLCUSUM 控制图存在的问题

在构造 RLLR 时可能存在无解的情形, 例如, 尝试对 $\mu_0 = 0, \mu_1 = 1, \sigma_0^2 = 1, \sigma^2 = 4, \epsilon = 0.5$ 的混合正态分布构造 RLLR, 却发现方程 (2.1) 在 $[x_2, x_3]$ 上无解, 其中 (x_2, y_2) 和 (x_3, y_3) 分别是对数似然比函数的第二和三个极值点.

当 σ^2 较大且 ϵ 较小时, RLCUSUM 控制图的表现远优于 CUSUM 控制图; 但当 σ^2 ($\sigma^2 > \sigma_0^2$) 较小且 ϵ 较大时, RLCUSUM 控制图的表现并不好, 如图 7(b) 所示.

3.2 改进的稳健对数似然比

由图 5(a) 可知, 通过水平射线截断的方法构造的稳健似然比函数, 其尾部单调递增的性质被弱化, 影响了对于大漂移的监测效率. 故保留似然比函数尾部的严格单调递增的性质, 对其进行水平线段截断, 从而改进 RLCUSUM 控制图的监测表现是一种可行的方法.

改进 RLLR, 令

$$l^\#(x) = \begin{cases} \ln \frac{F(b | \mu_1) - F(a | \mu_1)}{F(b | \mu_0) - F(a | \mu_0)}, & a \leq x \leq b, \\ \ln \frac{F(d | \mu_1) - F(c | \mu_1)}{F(d | \mu_0) - F(c | \mu_0)}, & c \leq x \leq d, \\ \ln \frac{f(x | \mu_1)}{f(x | \mu_0)}, & \text{其他.} \end{cases}$$

为保持单调性和连续性, 则应有

$$\ln \frac{F(b|\mu_1) - F(a|\mu_1)}{F(b|\mu_0) - F(a|\mu_0)} = \ln \frac{f(a|\mu_1)}{f(a|\mu_0)} = \ln \frac{f(b|\mu_1)}{f(b|\mu_0)}, \quad (3.1)$$

$$\ln \frac{F(d|\mu_1) - F(c|\mu_1)}{F(d|\mu_0) - F(c|\mu_0)} = \ln \frac{f(c|\mu_1)}{f(c|\mu_0)} = \ln \frac{f(d|\mu_1)}{f(d|\mu_0)}. \quad (3.2)$$

3.2.1 改进的 RLLR 的性质

相比于水平射线截断后的 RLLR, 改进的 RLLR 具有优良的性质. 它在任何情形下均可被构建, 并且在不同的非单调区间内的截断不会存在相交的情形. 以下给出证明.

令 $z(x) = \ln \frac{f(x|\mu_1)}{f(x|\mu_0)}$, $W_1 = \{(a, b) : z(a) = z(b)\}$, $W_2 = \{(c, d) : z(c) = z(d)\}$, 则值域为

$$z(W_1) = [\min(y_1, y_2), \max(y_1, y_2)], \quad z(W_2) = [\min(y_3, y_4), \max(y_3, y_4)].$$

命题 1 $z(W_1)$ 与 $z(W_2)$ 不相交.

证明 平移后的混合正态分布密度函数与原来的密度函数只有一个交点 $x = (\mu_0 + \mu_1)/2$, 此时 $z(x) = 0$. 当 $\mu_1 > \mu_0$ 时, $\forall x > (\mu_0 + \mu_1)/2$, 有 $f(x|\mu_1) > f(x|\mu_0)$, 从而 $z(x) > 0$; $\forall x < (\mu_0 + \mu_1)/2$, 有 $f(x|\mu_1) < f(x|\mu_0)$, 从而 $z(x) < 0$. 所以必然有 $\min(y_1, y_2) > 0 > \max(y_3, y_4)$, 所以, $z(W_1)$ 和 $z(W_2)$ 不相交.

同理可证, 当 $\mu_1 < \mu_0$ 时, $z(W_1)$ 与 $z(W_2)$ 不相交. □

命题 2 必然可以找到 $(a, b) \in W_1$, $(c, d) \in W_2$, 使得 (3.1) 和 (3.2) 成立.

证明 令

$$g_1(y) = \ln \frac{F(b|\mu_1) - F(a|\mu_1)}{F(b|\mu_0) - F(a|\mu_0)}, \quad g_2(y) = \ln \frac{F(d|\mu_1) - F(c|\mu_1)}{F(d|\mu_0) - F(c|\mu_0)}.$$

由于 $z(a) = z(b) = y$, 且 $z(a)$ 和 $z(b)$ 都在 W_1 上严格单调连续, 则 $z(a)$ 和 $z(b)$ 在 W_1 上存在逆函数 $a = z_a^{-1}(y)$, $b = z_b^{-1}(y)$, 且 $z_a^{-1}(y)$ 和 $z_b^{-1}(y)$ 都在 $[\min(y_1, y_2), \max(y_1, y_2)]$ 上连续. 由于 $F(a|\mu)$ 和 $F(b|\mu)$ 在 W_1 上连续, $\ln(x)$ 在 $\mathbb{R}^+$ 上连续, 则 $g_1(y)$ 是在 $z(W_1)$ 上的连续函数.

$F(x|\mu_0)$ 和 $F(x|\mu_1)$ 满足 Cauchy 中值定理的四个条件: (i) 在 $[a, b]$ 上都连续; (ii) 在 (a, b) 上都可导; (iii) $f(x|\mu_1)$ 和 $f(x|\mu_0)$ 不同时为零, 事实上总有 $f(x|\mu_1) > 0$, $f(x|\mu_0) > 0$; (iv) $F(a|\mu_0) \neq F(b|\mu_0)$. 所以存在 $\xi \in (a, b)$, 使得

$$z(\xi) = \ln \frac{f(\xi|\mu_1)}{f(\xi|\mu_0)} = \ln \frac{F(b|\mu_1) - F(a|\mu_1)}{F(b|\mu_0) - F(a|\mu_0)}.$$

当 $y = \max(y_1, y_2)$ 时, $a = z_a^{-1}(y)$, $b = z_b^{-1}(y)$, $\forall x \in (a, b)$ 有 $z(x) < y$, 则存在 $\xi \in (a, b)$, 使得

$$g_1(y) = \ln \frac{F(b|\mu_1) - F(a|\mu_1)}{F(b|\mu_0) - F(a|\mu_0)} = z(\xi) < y,$$

即 $G_1(y) = g_1(y) - y < 0$. 同理, 当 $y = \min(y_1, y_2)$ 时, 存在 $\xi \in (a, b)$, 使得

$$g_1(y) = \ln \frac{F(b|\mu_1) - F(a|\mu_1)}{F(b|\mu_0) - F(a|\mu_0)} = z(\xi) > y,$$

即 $G_1(y) = g_1(y) - y > 0$.

由于 $g_1(y)$ 在 $z(W_1)$ 上连续, 所以, $G_1(y)$ 也在 $z(W_1)$ 上连续. 由根的存在定理得, 至少存在一点 $y_0 \in z(W_1)$, 使得 $G_1(y_0) = 0$, 即 $g_1(y_0) = y_0$. 所以必然可以找到 $(a, b) = (z_a^{-1}(y_0), z_b^{-1}(y_0)) \in W_1$, 使得 (3.1) 成立.

同理可证必然存在 $(c, d) \in W_2$, 使得 (3.2) 成立. $\square$

3.2.2 改进的 RLLR 的求解

改进的 RLLR 截断点的求解计算方法为遍历 y 分别搜寻方程 $G_1(y) = 0$ 和 $G_2(y) = 0$ 的根. 若方程有多个解, 可选取令 $|g_1(y)|$ 和 $|g_2(y)|$ 最大的解, 以提高控制图的灵敏度. 当 $\mu_0 = 0, \mu_1 = 1, \sigma_0^2 = 1, \sigma^2 = 9, \epsilon = 0.05$ 时, 存在唯一解 $(a, b) = (-14.17, -1.25)$, $(c, d) = (2.25, 15.17)$, 如图 5(b) 所示. 由图 5(b) 可知, 在状态受控时, 观测值的绝对值大于 15 的可能性非常小, 对数似然比函数尾部的变化对截断点选择的影响也非常小, 因而, 改进的 RLLR 与原始的 RLLR 几乎在同一处截断, 但是改进的 RLLR 保留了对数似然比函数的尾部严格单增性.

3.3 改进的 RLCUSUM 控制图

构造改进的 RLCUSUM 统计量

$$S_0^\# = 0, \quad S_n^\# = \max\{0, S_{n-1}^\# + l^\#(x_n)\},$$

选择合适的 h 值, 令 ARL_0 为某一固定长度, 当 $S_n^\# > h$ 时, 控制图报警. 当 $h = 5.08$ 时, 改进的 RLCUSUM 控制图的 ARL_0 大约为 1,000. 令 $h = 5.08$, 用不同均值 μ 的混合正态分布模拟真实观测值, 计算 ARL_1 并与 CUSUM 控制图和 RLCUSUM 控制图进行比较, 结果如图 8 所示. 由图 8(a) 可知, 当 μ 小于 14 时, 改进的 RLCUSUM 控制图的表现与 RLCUSUM 控制图相似; 当 μ 大于 14 时, 改进的 RLCUSUM 控制图的表现与 CUSUM 控制图相似. 改进的 RLCUSUM 控制图结合了两者的优点, 在只有真实均值变化的情形下, 其表现优于 CUSUM 控制图和传统的 RLCUSUM 控制图. 当 σ^2

图 8 三种控制图在不同真实均值下 μ 的 ARL_1 , 其中 $ARL_0 = 1,000$

($\sigma^2 > \sigma_0^2$) 较大且 ϵ 较小时, 改进的 RLCUSUM 控制图明显提高了 RLCUSUM 控制图对中等漂移的灵敏度. 改进的 RLCUSUM 控制图利用了对数似然比函数尾部的信息, 在保证 ARL_1 函数单调的同时解决了 σ^2 ($\sigma^2 > \sigma_0^2$) 较小且 ϵ 较大时 RLCUSUM 控制图对中等漂移不够灵敏的问题, 如图 8(b) 所示.

4 结论

本文构造了混合正态分布下的 RLCUSUM 控制图, 证明了它对不同大小的漂移的稳健性, 并对其性质进行了更深入的研究, 包括对 ARL_0 曲线间断点的证明以及对 ARL_1 的极限的讨论. 但是当 σ^2 ($\sigma^2 > \sigma_0^2$) 较小且 ϵ 较大时, RLCUSUM 控制图将对数似然比函数整个尾部截断造成了过多的信息丢失, 还可能无法构造出 RLCUSUM 控制图. 因此, 本文将对数似然比函数的非单调区间水平线段截断, 但保留尾部, 构造出改进的 RLCUSUM 控制图. 在混合正态分布下, 改进的 RLCUSUM 控制图在两个方面优于 RLCUSUM 控制图. 第一, 必能构造出改进的 RLCUSUM 控制图, 而 RLCUSUM 控制图在某些情形下是不存在的. 第二, 为了保护控制图免受异常值影响, RLCUSUM 控制图过多地牺牲了对中等漂移及大漂移的检测灵敏度; 改进的 RLCUSUM 控制图保留了对数似然比函数尾部的信息, 当漂移逐渐增大时, 控制图会越来越灵敏, 因此, 其监测效果优于 CUSUM 控制图和 RLCUSUM 控制图.

在其他厚尾分布过程中, 改进的 RLCUSUM 控制图比传统的 RLCUSUM 控制图更具有普适性. 对于有多个非单调区间的对数似然比函数, RLCUSUM 控制图可能不得不截断包含多个非单调区间的大片区域, 从而造成信息的过多损失; 而改进的 RLCUSUM 控制图只需在小范围内将非单调区间截断, 能有效保留非单调区间之间的信息, 最大限度地减少信息损失. 若两个非单调区间过于靠近而导致无法分别对它们进行截断, 或者两个截断区间之间有重合, 可以将两个非单调区间一同截断.

参考文献

- 1 Page E S. Continuous inspection schemes. *Biometrika*, 1954, 41: 100–115
- 2 Lucas J M. Combined Shewhart-CUSUM quality control schemes. *J Qual Technol*, 1982, 14: 51–59
- 3 Vardeman S, Ray D. Average run lengths for CUSUM schemes when observations are exponentially distributed. *Technometrics*, 1985, 27: 129–144
- 4 代毅, 王兆军, 邹长亮. 基于似然比统计量的预分析累积和控制图. *中国科学 A 辑*, 2006, 36: 984–1000
- 5 Lucas J M, Crosier R B. Robust CUSUM: A robustness study for CUSUM quality control schemes. *Comm Statist Theory Methods*, 1982, 11: 2669–2687
- 6 Hawkins D M. Robustification of cumulative sum charts by winsorization. *J Qual Technol*, 1993, 25: 248–261
- 7 Stoumbos Z G, Reynolds M R. The robustness and performance of CUSUM control charts based on the double-exponential and normal distributions. *Front Stat Qual Control*, 2004, 7: 79–100
- 8 Reynolds M R, Stoumbos Z G. Robustness to non-normality of CUSUM control charts for monitoring the process mean and variance. *Qual Reliab Eng Int*, 2010, 26: 453–473
- 9 邹长亮. 复杂数据统计过程的若干研究. *中国科学: 数学*, 2013, 43: 741–750
- 10 Hafiz Z N, Muhammad R, Ronald J M M, et al. Robust CUSUM control charting. *Qual Eng*, 2013, 25: 211–224
- 11 张久军. 基于似然比检验的若干控制图. 博士学位论文. 天津: 南开大学, 2009
- 12 MacEachern S, Rao Y L, Wu C J. A robust-likelihood cumulative sum chart. *J Amer Statist Assoc*, 2007, 102: 1440–1447
- 13 Liu G F, Pu X L, Wang L, et al. CUSUM chart for detecting range shifts when monotonicity of likelihood ratio is invalid. *J Appl Stat*, 2015, 42: 1635–1644
- 14 Li Z H, Zou C L, Gong Z, et al. The computation of average run length and average time to signal: An overview. *J Stat Comput Simul*, 2014, 84: 1779–1802
- 15 Brook D, Evans D A. An approach to the probability distribution of CUSUM length. *Bilmetrika*, 1972, 59: 539–549

附录 A 平均运行长度的计算

称控制图从监测开始到第一次报警抽取的样本个数 τ^* 为报警所需的运行时间. 对于 CUSUM 控制图, 有

$$\tau^* = \min\{n : S_n \geq h, h \geq 1\}.$$

当 $P(f_0(x) = f_1(x)) < 1$ 时, $\forall \lambda > 0, \forall k > 0, k$ 为整数, $i = 0, 1$, 有

$$P_{H_i}(\tau^* < \infty) = 1, \quad E_{H_i}(e^{\lambda \tau^*}) < \infty, \quad E_{H_i}(\tau^*)^k < \infty.$$

CUSUM 控制图永远不接受原假设, 直到拒绝原假设时才停止抽样. 由 $P_{H_i}(\tau^* < \infty) = 1$ 可知, CUSUM 控制图必将在某一时刻拒绝原假设; 由 $E_{H_i}(\tau^*)^k < \infty$ 可知, 运行时间的均值存在. 因此, CUSUM 控制图一般不通过第一、二类错误来评价其优劣, 而是将平均运行长度作为比较控制图有效性的重要指标.

平均运行长度 (average run length, ARL) 是控制图从监测开始到它发出警报为止所抽取的平均样本个数, 即为 $E_{H_i}(\tau^*)$. 则受控平均运行长度和失控平均运行长度分别为

$$ARL_0 = E_{H_0}(\tau^*) = E(\tau^* | \text{过程受控}),$$

$$ARL_1 = E_{H_1}(\tau^*) = E(\tau^* | \text{过程从开始检测时就失控}).$$

当过程受控时, 我们希望控制图有较小的误报率, ARL_0 越大越好; 当过程失控时, 我们希望控制图能尽快报警, ARL_1 越小越好. 通常, 比较控制图效果好坏的方法是, 选择合适的 h 值, 令 ARL_0 满足设定的要求, 例如, 固定为某一长度, ARL_1 越小, 说明该控制图越有效. Li 等^[14] 对于控制图平均运行长度的计算做了详细的总结归纳.

CUSUM 控制图计算 ARL 的方法主要有三种: 积分方程法、Markov 链法和随机模拟法. 以下介绍 Markov 链法. Brook 和 Evans^[15] 于 1972 年提出了用 Markov 链方法计算 CUSUM 的 ARL, 之后该方法被用于其他类型控制图 ARL 的计算. 监控的 CUSUM 序列 $\{S_n, n = 0, 1, 2, \dots\}$ 是一个 Markov 链. 首先构造一步转移概率矩阵 $\mathbf{P}$. 考虑连续情形. 将 $[0, h]$ 划分为如下 $t+1$ 个子区间:

$$\left[0, \frac{\omega}{2}\right] \cup \left(\omega - \frac{\omega}{2}, \omega + \frac{\omega}{2}\right) \cup \dots \cup \left((t-1)\omega - \frac{\omega}{2}, (t-1)\omega + \frac{\omega}{2}\right] \cup (h, \infty),$$

其中 $\omega = 2h/(2t-1)$. 当 $S_n \in [0, \omega/2]$ 时, 称过程属于状态 E_0 ; 当 $S_n \in [i\omega - \omega/2, i\omega + \omega/2]$ ($i = 1, 2, \dots, t-1$) 时, 称过程属于状态 E_i ; 当 $S_n \in (h, \infty)$ 时, 称过程属于吸收状态 E_t . 由于 $S_n \in E_i$ 处于一个区间内, 但是为了便于计算, 采用区间的中值作为替代. 该替代方法是由于区间被无限划分, 每个小区间足够小到可以用中点值近似状态表示. 对一步转移概率进行近似计算 (其中 $i = 0, 1, \dots, t-1$):

$$p_{ij} = P(S_{n+1} \in E_j | S_n = i\omega) = P\left((j-i)\omega - \frac{\omega}{2} < Z_{n+1} \leq (j-i)\omega + \frac{\omega}{2}\right), \quad j = 1, 2, \dots, t-1,$$

$$p_{i0} = P\left(S_n + Z_{n+1} \leq \frac{\omega}{2} \mid S_n = i\omega\right) = P\left(Z_{n+1} \leq -i\omega + \frac{\omega}{2}\right),$$

$$p_{it} = P(S_n + Z_{n+1} > h | S_n = i\omega) = P\left(Z_{n+1} > (t-i-1)\omega + \frac{\omega}{2}\right),$$

$$p_{tj} = 0, \quad j = 0, 1, \dots, t-1,$$

$$p_{tt} = 1.$$

令

$$p_r = P\left(r\omega - \frac{\omega}{2} < Z_{n+1} \leq r\omega + \frac{\omega}{2}\right), \quad F_r = P\left(Z_{n+1} \leq r\omega + \frac{\omega}{2}\right),$$

则得到一步转移概率矩阵

$$\mathbf{P} = \begin{bmatrix} F_0 & p_1 & p_2 & \cdots & p_j & \cdots & p_{t-1} & 1 - F_{t-1} \\ F_{-1} & p_0 & p_1 & \cdots & p_{j-1} & \cdots & p_{t-2} & 1 - F_{t-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ F_{-i} & p_{1-i} & p_{2-i} & \cdots & p_{j-i} & \cdots & p_{t-1-i} & 1 - F_{t-1-i} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ F_{1-t} & p_{2-t} & p_{3-t} & \cdots & p_{j-(t-1)} & \cdots & p_0 & 1 - F_0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

将一步转移概率矩阵 $\mathbf{P}$ 写成分块的形式:

$$\mathbf{P} = \begin{pmatrix} \mathbf{R} & (\mathbf{I} - \mathbf{R})\mathbf{1} \\ \mathbf{0}^T & 1 \end{pmatrix},$$

其中, $\mathbf{R}$ 为 $\mathbf{P}$ 去掉最后一行和最后一列后得到的 $t \times t$ 维矩阵, $\mathbf{1}$ 为由元素值 1 构成的长度为 t 的列向量, $\mathbf{I}$ 为 $t \times t$ 维单位阵. m 步转移概率矩阵为

$$\mathbf{P}^m = \begin{pmatrix} \mathbf{R}^m & (\mathbf{I} - \mathbf{R}^m)\mathbf{1} \\ \mathbf{0}^T & 1 \end{pmatrix}.$$

用 T_i 表示当 CUSUM 统计量初始值 $Z_0 \in E_i$ 时的运行时间, 即第一次达到状态 E_t 的步数. $\mathbf{T} = (T_0, T_1, \dots, T_{t-1})^T$, 对 $\mathbf{T}$ 求期望, 可得 $E(\mathbf{T})$, 其第 i 个分量表示初始状态 E_{i-1} 出发的 ARL. 对于任意的正整数 r , 有

$$\mathbf{F}_r = (P(T_0 \leq r), P(T_1 \leq r), \dots, P(T_{t-1} \leq r))^T = (\mathbf{I} - \mathbf{R}^r)\mathbf{1},$$

其中 $\mathbf{I}$ 为 t 阶单位阵.

由于

$$\begin{aligned} E(T_i) &= \sum_{m=1}^{\infty} mP(T_i = m) = \sum_{m=1}^{\infty} P(T_i \geq m), \\ (\mathbf{I} - \mathbf{R}) \left(\sum_{m=1}^{\infty} \mathbf{R}^m \right) &= \mathbf{I} \Rightarrow \sum_{m=1}^{\infty} \mathbf{R}^m = (\mathbf{I} - \mathbf{R})^{-1}, \end{aligned}$$

所以,

$$\overrightarrow{\text{ARL}} = E(\mathbf{T}) = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{1}.$$

Improved robust-likelihood cumulative sum chart for the contaminated normal distributions

WU ChunJie, WEI YiQing & YU MiaoMiao

Abstract The cumulative sum (CUSUM) chart is usually used for detecting small drifts in the mean of a normally distributed process, but it performs poorly for thick-tailed distributed processes and large drifts. This paper mainly studies performance of robust likelihood cumulative sum (RLCUSUM) chart when observations are contaminated normal distributions, which overcomes the shortcomings of CUSUM by using the likelihood ratio. Compared with the CUSUM chart in average run length, RLCUSUM chart is much more sensitive. To improve its robustness, the method of winsorization by horizontal segment is used.

Keywords robust likelihood cumulative sum chart, average run length, winsorization by horizontal segment, contaminated normal distribution

MSC(2010) 62P30

doi: 10.1360/N012016-00048