应用概率统计　第二十四卷
第二期　2008年4月

Chinese Journal of Applied Probability
and Statistics Vol.24 No.2 Apr. 2008

几类正态过程的CUSUM图

涂玉娟[1,2]　　冯士雍[1]

([1]中国科学院数学与系统科学研究院, 北京, 100080;　[2]中国科学院研究生院, 北京, 100080)

摘　要

本文针对呈正态分布的过程数据, 讨论当过程参数(μ,σ^2)已知或未知时单参数变化、双参数变化的过程控制问题. 针对均值未知情况, 利用Quesenberry提出的Q统计量构造了一系列标准化的CUSUM图; 对于方差未知时的复杂情形, 给出了一种构造更为简单的CUSUM图的方法; 对于未知(或不关心) (μ,σ^2)中何者变化的特殊情形, 文章提出了纯变化的概念, 给出了相应统计量及由该统计量构造的CUSUM-D图, 针对每个考虑的情形, 通过模拟计算本文给出了相应于统计量的条件期望延时(CED), 同时在文章最后给出了本文提出的CUSUM图与已有的CUSUM图(Q图)的模拟比较, 结果表明新的CUSUM图是可行的.

关键词:　累积和, Q统计量, 单(双)参数变化, 纯变化问题, 条件期望延时.

学科分类号:　O213.1.

§1.　引　　言

自1954年Page[1]提出CUSUM(累积和)图, 并指出该图比常规(休哈特)控制图能更灵敏的检测过程中的小偏移问题以来, 各种CUSUM图相继出现, 从不同的方面对过程偏移检测问题进行改进, [2]、[3]回顾并总结了这些CUSUM图. 然而以往的CUSUM图往往都假定过程初始均处于受控状态, 并持续足够长时间, 在欲考察的过程性能参数未知时, 可以通过过程正常时的数据估计该参数[4]. 但实际操作中, 上述假定并不一定都能够满足. 此外, 通常情况下, 只对过程均值与过程方差只有一个发生变化的情况下进行讨论. 但在实际问题中, 两者同时发生变化的情形也经常发生; 有时人们可能并不关心均值和方差中究竟何者发生了变化, 只关心过程是否发生异常, 即纯变化问题. 本文即是在过程为正态分布的情况下考虑过程参数(μ,σ^2)已知或未知时两参数中只有一个参数变化、两个参数同时变化问题以及纯变化问题.

特别指出, 本文中"参数已知"是指: 过程参数的真实值在问题提出时已知; 或过程的真实参数虽未知, 但过程初始处于受控状态并持续足够长时间, 因而可以利用这部份数据得到参数的有效估计; 反之, 则"参数未知".

设随机变量X为正态过程的某种质量特性, $X\sim N(\mu,\sigma^2)$, 其中μ、σ^2可已知、可未知. $x_1,x_2,\cdots$为从过程中抽取的一系列独立观测值, 并假定过程在某个未知时刻$t\ (t\geq 1)$后突然

本文2004年7月19日收到, 2005年12月12日收到修改稿.

发生变化, 即观测值$x_1, x_2, \cdots, x_t$独立同分布于某一正态分布: $x_1, x_2, \cdots, x_t \sim N(\mu_0, \sigma_0^2)$, i.i.d.; 而观测值$x_{t+1}, x_{t+2}, \cdots$独立同分布于另一正态分布: $x_{t+1}, x_{t+2}, \cdots \sim N(\mu_1, \sigma_1^2)$, i.i.d., 且$\mu_1 \neq \mu_0$或$\sigma_1^2 \neq \sigma_0^2$中至少有一个成立. 显然, 若$t = \infty$, 则说明过程一直处于受控状态, CUSUM图应该显示出此状态. 如果$t < \infty$, 则说明过程在t时刻以后失控, CUSUM图应该尽快的给出警报.

假设已抽取m个观测值, 用假设检验的形式描述上述问题:

$$\begin{aligned} H_0 &: X_i \sim N(\mu_0, \sigma_0^2), \qquad i = 1, 2, \cdots, m; \\ H_t &: X_i \sim N(\mu_0, \sigma_0^2), \qquad i = 1, 2, \cdots, t; \\ &\quad\ X_i \sim N(\mu_1, \sigma_1^2), \qquad i = t+1, t+2, \cdots, m, \qquad 1 \leq t \leq m. \end{aligned}$$

利用序贯概率比检验理论, 序贯概率比为:

$$P = \frac{L_t(\mu_1, \sigma_1^2)}{L_0(\mu_0, \sigma_0^2)} = \frac{\prod\limits_{i=1}^{t} f_0(x_i) \prod\limits_{i=t+1}^{m} f_1(x_i)}{\prod\limits_{i=1}^{m} f_0(x_i)},$$

其中$f_0(\cdot)$为$N(\mu_0, \sigma_0^2)$的密度函数, $f_1(\cdot)$为$N(\mu_1, \sigma_1^2)$的密度函数.

确定常数A和B, 使得:

若$P \geq A$, 则相应的备择假设成立, 过程失控;

若$P \leq B$, 则原假设成立, 过程受控;

若$B < P < A$, 则不能给出定论, 待抽取下一观测值, 重复进行以上检验.

根据$P \geq A$可以得到用于构造CUSUM图的统计量及其递推公式. 通常A通过以下方法确定: 令ϑ表示第一类错误的概率, β表示第二类错误的概率, 取$A = (1-\beta)/\vartheta$; 或首先确定所要求的过程平稳时的平均链长(ARL_0), 根据ARL_0确定A的大小.

在计算CUSUM图的质量指标平均链长(ARL)时, 无论是Hoteling统计量方法, 还是Markov链方法, 都需要知道控制图统计量的分布. 但本文所用到的大部分统计量在过程异常时的分布并不是常见或已知的分布. 因此, 在给出一定的已知的统计量的分布同时, 所有CUSUM的性能指标和应用范围皆通过条件期望延时(conditional expected delay, 简作CED,[2])表示.

若RL为过程警报首次发生时的样本点数, τ $(\tau > 0)$为过程开始发生异常的时刻, 则条件期望延时(CED)定义为:

$$\mathsf{E}\{\mathrm{RL} - \tau | \mathrm{RL} > \tau\}.$$

CED与ARL有如下关系:

当过程参数变化γ时, $\mathrm{CED}_{(\tau=0)} = \mathrm{ARL}_\gamma$.

当过程一直处于正常状态, 且样本量为n, 此时$\tau > n$, CED无意义,

$$\mathrm{ARL}_0 = \mathsf{E}\{\mathrm{RL}\}.$$

CED的模拟计算结果和某些给定的常数在相应的部分给出, 当参数未产生变化时给出的计算结果的意义类似于ARL_0.

§2. (μ_0, σ_0^2)已知情形

将X_i标准化, $Y_i = (X_i - \mu_0)/\sigma_0$, 令$\mu_1 - \mu_0 = \delta\sigma_0$, $\sigma_1^2 = \alpha^2\sigma_0^2$, $\alpha > 0$, 其中$\delta = 0$, $\alpha^2 = 1$不同时成立. 又因为当方差变小时, 意味着过程性能的提高, 此时CUSUM图无需报警, 因此本文假定$\alpha^2 \geq 1$. 上述假设检验问题可表述为:

$$\begin{aligned}
H_0 &: Y_i \sim N(0,1), \qquad i = 1, 2, \cdots, m;\\
H_t &: Y_i \sim N(0,1), \qquad i = 1, 2, \cdots, t;\\
&\quad\; Y_i \sim N(\delta, \alpha^2), \qquad i = t+1, t+2, \cdots, m, \qquad 1 \leq t \leq m.
\end{aligned}$$

相应的序贯概率比P为:

$$\begin{aligned}
P &= \frac{\prod\limits_{i=1}^{t} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{y_i^2}{2}\Big) \prod\limits_{i=t+1}^{m} \frac{1}{\sqrt{2\pi}\alpha} \exp\Big(-\frac{(y_i-\delta)^2}{2\alpha^2}\Big)}{\prod\limits_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{y_i^2}{2}\Big)}\\
&= \Big(\frac{1}{\alpha}\Big)^{m-t} \exp\Big[-\frac{1}{2\alpha^2}\sum_{i=t+1}^{m}(y_i-\delta)^2 + \frac{1}{2}\sum_{i=t+1}^{m} y_i^2\Big].
\end{aligned}$$

(i) 当$\delta \neq 0$, $\alpha^2 = 1$时, 即过程均值发生偏移, 方差未变, 此时P的形式和已有结果[5]一致:

$$P = \exp\Big[\delta \sum_{i=t+1}^{m}\Big(y_i - \frac{\delta}{2}\Big)\Big];$$

失控判据为:

$$\begin{aligned}
&\sum_{i=t+1}^{m}(y_i - \delta/2) \geq \ln A/\delta, \qquad 若\ \delta > 0;\\
&\sum_{i=t+1}^{m}(y_i - \delta/2) \leq \ln A/\delta, \qquad 若\ \delta < 0.
\end{aligned}$$

用于构造CUSUM图的统计量递推公式为:

$$S_{1,(i+1)} = S_{1,i} + \Big(y_{i+1} - \frac{\delta}{2}\Big), \qquad (S_{1,0} = 0). \tag{2.1}$$

其中, 过程受控时$y_i - \delta/2 \sim N(-\delta/2, 1)$, 过程失控时$y_i - \delta/2 \sim N(\delta/2, \alpha^2)$.

(ii) 当$\delta=0$, $\alpha^2>1$时, 说明过程均值未变, 仅方差变大, 所得P和已有结果一致[5]:

$$P=\Big(\frac{1}{\alpha}\Big)^{m-t}\exp\Big(-\frac{1-\alpha^2}{2\alpha^2}\sum_{i=t+1}^{m}y_i^2\Big);$$

失控判据为:

$$\sum_{i=t+1}^{m}y_i^2\geq\frac{2\alpha^2}{\alpha^2-1}\ln A+(m-t)\frac{2\alpha^2}{\alpha^2-1}\ln\alpha;$$

CUSUM统计量的递推公式为:

$$S_{2,(i+1)}=S_{2,i}+y_{i+1}^2,\qquad(S_{2,0}=0),\tag{2.2}$$

其中, 过程受控时$y_i^2\sim\chi^2(1)$, 即中心卡方分布; 过程失控时$y_i^2/\alpha^2\sim\chi^2(1,\delta^2/\alpha^2)$, 即非中心卡方分布.

(iii) 当$\delta\neq0$, $\alpha^2>1$时, 即过程均值与方差均发生变化(也即双参数变化), 此时P为:

$$P=\Big(\frac{1}{\alpha}\Big)^{m-t}\exp\Big[-\frac{1-\alpha^2}{2\alpha^2}\sum_{i=t+1}^{m}\Big(y_i-\frac{\delta}{1-\alpha^2}\Big)^2+\frac{(m-t)\delta^2}{2(1-\alpha^2)}\Big];$$

失控判据为:

$$\sum_{i=t+1}^{m}\Big(y_i+\frac{\delta}{\alpha^2-1}\Big)^2\geq\frac{2\alpha^2}{\alpha^2-1}\ln A+(m-t)\frac{2\alpha^2}{\alpha^2-1}\Big(\ln\alpha+\frac{\delta^2}{2(\alpha^2-1)}\Big);$$

CUSUM统计量的递推公式为:

$$S_{3,(i+1)}=S_{3,i}+\Big(y_{i+1}+\frac{\delta}{\alpha^2-1}\Big)^2,\qquad(S_{3,0}=0),\tag{2.3}$$

其中, 过程受控时$(y_i+\delta/(\alpha^2-1))^2\sim\chi^2(1,\delta^2/(\alpha^2-1)^2)$; 过程失控时$(y_i+\delta/(\alpha^2-1))^2/\alpha^2\sim\chi^2(1,\alpha^2\delta^2/(\alpha^2-1)^2)$.

相应的CED模拟计算结果见表1. 从表1可以看出, 若过程参数已知:

1) 当δ和α都非常小, 即均值和方差变化皆非常小时, CED偏大;

2) 当δ和α之一较大, 即均值偏移或方差变化较大时, CED较小, 即报警及时;

3) 同时考察均值和方差变化比考察单参数变化报警更及时.

表1 (μ_0,σ_0^2)已知时CED的模拟结果

CED	$\alpha^2=1$	$\alpha^2=1.25^2$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	427.36[a]	24.63	11.30	7.83	6.19
$\delta=0.25$	27.12	18.51	10.53	7.42	6.16
$\delta=0.5$	10.10	12.16	8.91	6.95	5.94
$\delta=0.75$	6.68	8.94	7.56	6.28	5.45
$\delta=1$	5.30	7.32	6.21	5.66	5.10

每次试验样本量n为500, 试验次数为10000次, $A=1.82$.

[a]表示过程一直处于受控状态下, 用S_3监测过程是否发生$(\delta,\alpha^2)=(0.5,1.5^2)$变化时首次警报时刻.

§3. μ_0未知、σ_0^2已知情形

利用Q统计量$Z_n = [(n-1)/n]^{1/2} \cdot [(x_n - \overline{x}_{n-1})/\sigma_0]$ ([6]), 其中, $\overline{x}_{n-1} = [1/(n-1)] \cdot \sum\limits_{i=1}^{n-1} x_i$.

设过程当中存在时间点$t > 1$, 且过程自t后失控, 则可以证明Z_n有以下性质:

1) 过程受控即: $X_i \sim N(\mu_0, \sigma_0^2)$, $i = 1, \cdots, t$时

$$Z_i \sim N(0,1), \qquad i = 2, \cdots, t;$$

2) 过程失控后, 即$X_i \sim N(\mu_0 + \delta, \alpha^2\sigma_0^2)$, $i > t$,

$$Z_i \sim N\Big(\frac{t\delta}{\sqrt{i(i-1)}}, \frac{i(i-1)\alpha^2 - (\alpha^2-1)t}{i(i-1)}\Big),$$

其中$i = t+1, \cdots, m$; $2 \le t \le m$.

此时的检验问题与引言中所述问题有所不同: 当过程失控后, 所得观测值非独立同分布.

记$\xi_{it} = t\delta/\sqrt{i(i-1)}$, $\eta_{it}^2 = [i(i-1)\alpha^2 - (\alpha^2-1)t]/[i(i-1)]$, $\eta_{it} > 0$ (以下结果的推导过程与第2节类似):

(iv) 当$\delta \ne 0$, $\alpha^2 = 1$时, $\eta_{it}^2 = 1$, 失控判据为:

$$\sum_{i=t+1}^{m} \xi_{it}(z_i - \frac{\xi_{it}}{2}) \ge \ln A.$$

CUSUM统计量的递推公式为:

$$S_{4,(i+1)} = S_{4,i} + \xi_{(i+1)t}(z_{(i+1)} - \frac{\xi_{(i+1)t}}{2}), \qquad (S_{4,1} = 0). \tag{3.1}$$

(v) 当$\delta = 0$, $\alpha^2 > 1$时, $\xi_{it} = 0$, 失控判据为:

$$\sum_{i=t+1}^{m} \frac{\eta_{it}^2 - 1}{2\eta_{it}^2} z_i^2 \ge \ln A + \sum_{i=t+1}^{m} \ln(\eta_{it}).$$

CUSUM统计量的递推公式为:

$$S_{5,(i+1)} = S_{5,i} + \frac{\eta_{(i+1)t}^2 - 1}{2\eta_{(i+1)t}^2} z_{i+1}^2, \qquad (S_{5,1} = 0). \tag{3.2}$$

(vi) 当$\delta \ne 0$, $\alpha^2 > 1$时, 失控判据为:

$$\sum_{i=t+1}^{m} \Big[\frac{\eta_{it}^2 - 1}{2\eta_{it}^2}\Big(z_i + \frac{\xi_{it}}{\eta_{it}^2 - 1}\Big)^2 - \frac{\xi_{it}^2}{2(\eta_{it}^2 - 1)}\Big] \ge \ln A + \sum_{i=t+1}^{m} \ln(\eta_{it}).$$

CUSUM统计量的递推公式为:

$$S_{6,(i+1)} = S_{6,i} + \frac{\eta_{(i+1)t}^2 - 1}{2\eta_{(i+1)t}^2}\Big(z_{i+1} + \frac{\xi_{(i+1)t}}{\eta_{(i+1)t}^2 - 1}\Big)^2 - \frac{\xi_{(i+1)t}^2}{2(\eta_{(i+1)t}^2 - 1)}, \qquad (S_{6,1} = 0). \tag{3.3}$$

其中,

$$\frac{\xi_{it}^2}{\eta_{it}^2-1}=\frac{t^2\delta^2}{(\alpha^2-1)[i(i-1)-t]},$$
$$\frac{\xi_{it}}{\eta_{it}^2-1}=\frac{t\sqrt{i(i-1)}\delta}{(\alpha^2-1)[i(i-1)-t]},$$
$$\frac{\eta_{it}^2-1}{\eta_{it}^2}=\frac{(\alpha^2-1)[i(i-1)-t]}{\alpha^2[i(i-1)-t]+t}.$$

显然, 使用S_4,S_5,S_6需明确t的大小, 即知道变化发生的时刻, 这对于本文要解决的问题是不可能做到的. 因为如果t事先已知, 就无谓再用CUSUM图进行过程监控了. 因此, 用δ/i代替ξ_{it}, α^2代替η_{it}^2, 并对上述统计量作适当处理:

1. $S_{4,(i+1)}=S_{4,i}+(z_{(i+1)}-\delta/i)$, 相应判据为$S_{4,i}\geq\ln A/\delta$;

2. $S_{5,(i+1)}=S_{5,i}+z_{(i+1)}^2$, 相应判据为$S_{5,i}\geq[2(\alpha^2)/(\alpha^2-1)]\cdot\ln A+(i-1)\cdot[2\alpha^2/(\alpha^2-1)]\cdot\ln\alpha$;

3. $S_{6,(i+1)}=S_{6,i}+[(\alpha^2-1)/(2\alpha^2)]\cdot[z_{(i+1)}+\delta/[i(\alpha^2-1)]]^2-\delta^2/[2i^2(\alpha^2-1)]$, 相应判据为$S_{6,i}\geq\ln A+(i-1)\ln\alpha$.

替代后, 相应的CED模拟计算结果见表2, 其中t是预设的变化时刻, 统计量中并未使用. 从表2中可以看出:

1) 当α较小时, CED偏大;

2) 对α, t的大小对CED的影响不明显; 对δ, t的大小对CED的影响较为显著, 且当只有均值偏移时, CED在t较小时较大, 说明此统计量不能及时报警;

3) 同时考察均值和方差变异比考察单参数变异报警更及时.

表2 μ_0未知、σ_0^2已知时CED的模拟结果

CED (t^c= 2)	$\alpha^2=1$	$\alpha^2=1.25^2$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	407.12[b]	25.46	11.87	8.25	6.67
$\delta=0.25$	85.30	24.51	11.59	8.19	6.53
$\delta=0.5$	43.61	23.22	11.02	7.83	6.38
$\delta=0.75$	29.01	20.69	10.21	7.75	6.27
$\delta=1$	21.45	17.54	9.40	7.17	5.89
CED (t^c= 5)	$\alpha^2=1$	$\alpha^2=1.25^2$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	406.37[b]	26.90	13.06	9.18	7.47
$\delta=0.25$	67.34	26.25	12.64	9.16	7.25
$\delta=0.5$	26.86	24.05	11.70	8.58	7.15
$\delta=0.75$	14.92	20.01	10.85	8.17	6.92
$\delta=1$	8.95	16.70	9.67	7.50	6.35

CED (t^c= 10)	$\alpha^2=1$	$\alpha^2=1.25^2$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	402.66[b]	29.43	14.97	10.89	8.95
$\delta=0.25$	51.42	28.32	14.17	10.79	8.63
$\delta=0.5$	17.26	24.94	13.35	10.05	8.39
$\delta=0.75$	6.96	20.56	11.54	9.39	7.89
$\delta=1$	3.59	15.24	9.90	8.08	7.15
对每个确定的t, 样本量500, 试验10000次, $A=1.82$.					

§4. μ_0已知、σ_0^2未知情形

当方差未知时, Quesenberry提出的Q统计量在理论上是非常有意义的, 但在实际操作中, 直接利用该统计量很难及时进行过程监控. 因为使用Q统计量需要先计算一个变量的t分布函数, 再将之转化为服从正态分布的变量值, 这几乎不可能在短时间内完成(濮晓龙, [7]). 此时, 利用Arthur B.Yeh等人[4]的想法, 本文构造了一系列基于Q统计量的简单CUSUM统计量. 因Q统计量的限制, 将过程控制分为均值监控与方差监控两个情形分别进行考虑.

4.1 均值监控

利用Q统计量$Q_n=\Phi^{-1}\{T_{n-1}((X_n-\mu_0)/S_{0,n-1})\}^{\rm d}$, $n=2,3,\cdots$, 其中, $S_{0,n}^2=(1/n)\cdot\sum\limits_{i=1}^{n}(x_i-\mu_0)^2$. 从Quesenberry的文章([6])知: 当过程受控时, Q_n服从标准正态分布, 此时令$W_n=\Phi(Q_n)$:

$$\forall\, 0<y<1,\qquad \mathsf{P}(w_n\le y)=\mathsf{P}(\Phi(Q_n)\le y)=\mathsf{P}(Q_n\le\Phi^{-1}(y))=\Phi(\Phi^{-1}(y))=y.$$

由此可见, 在过程受控时, $W_n\sim U(0,1)^{\rm e}$.

令

$$S_{7,(i+1)}=\sum_{j=1}^{i+1}\Big(w_j-\frac{1}{2}\Big)=S_{7,i}+\Big(w_{i+1}-\frac{1}{2}\Big),\qquad (S_{7,1}=0). \tag{4.1}$$

若过程均值增大, 即$x_i-\mu_0>0$, 则$w_i-1/2>0$, 最终导致$S_{7,i}$值的增加, 故可预先给出某一定值H_u, 如果$S_{7,i}>H_u$, 则认为过程已经失控; 若过程均值减小, 即$x_i-\mu_0<0$, 则$w_i-1/2<0$, 最终导致$S_{7,i}$值的减少, 类似地, 预先确定某值H_l, 如果$S_{7,i}<H_l$, 则认为过程已经失控. 又因为$\forall\, i$, $S_{7,i}$关于0对称, 因此可取$|H_l|=H_u$.

[b]表示过程一直处于受控状态下, 用S_6监测过程是否发生$(\delta,\alpha^2)=(0.5,1.5^2)$变化时首次警报时刻.

[c]表示过程在t时刻后发生相应变化.

[d]$\Phi(\cdot)$, 标准正态分布的分布函数; $\Phi^{-1}(\cdot)$, 标准正态分布的分布函数的逆函数; $T_n(\cdot)$, 自由度为n的t分布的分布函数; $F_{m,n}(\cdot)$, 自由度为(m,n)的F分布的分布函数. 下同.

[e]同理可证下文出现的V_n、W_n^*、V_n^*服从$(0,1)$上的均匀分布.

由于均值增大与均值减小情形类似, 我们进行模拟计算只考虑均值增大情形. 又因计算量的问题, 因此适当调整试验次数. 通过模拟计算, 过程发生变化的时刻对CED没有太大影响. 从表3可以看出:

1) 当过程正常时, CED并不十分令人满意;

2) 当均值偏移较大时, CED与已有结果具可比性.

表3 μ_0已知、σ_0^2未知且均值变化时CED的模拟结果

CED	$\delta=0$	$\delta=0.25$	$\delta=0.5$	$\delta=0.75$	$\delta=1$
$\alpha^2=1$	213.18[f]	32.72	16.75	12.29	10.27
每次试验样本量500, 试验次数为1000次, $H_u=2$.					

4.2 方差监控

利用Q统计量$Q_n=\Phi^{-1}\{F_{1,n-1}\cdot[(X_n-\mu_0)^2/S_{0,n-1}^2]\}$, $n=2,3,\cdots$, 当过程受控时服从标准正态分布. 令$V_n=\Phi(Q_n)$, 同样可以证明: 当过程受控时, $V_n\sim U(0,1)$.

令

$$S_{8,(i+1)}=\sum_{j=1}^{i+1}\Big(v_j-\frac{1}{2}\Big)=S_{8,i}+\Big(v_{i+1}-\frac{1}{2}\Big),\qquad (S_{8,1}=0).\tag{4.2}$$

若过程方差变大, 即过程趋于不稳定, 当i较大时, 有$v_i-1/2>0$, 最终导致S_{8i}值的增加. 故可预先给出某一定值H_u, 如果$S_{8,i}>H_u$, 则认为过程已经失控; 若过程方差变小, 即过程越来越稳定, 因此可不予以考虑, 但理论上有类似结果: $v_i-1/2<0$, 最终导致$S_{8,i}$值的减少, 类似地, 预先确定某值H_l, 如果$S_{8,i}<H_l$, 则认为过程稳定性增强.

从表4可以看出, 当过程方差增大, 尤其增加程度较大时, 该统计量能够较好的监控过程; 但当过程正常, 该统计量的CED偏小, 即易发生错误报警.

表4 μ_0已知、σ_0^2未知且方差变化时CED的模拟结果

CED	$\alpha^2=1^2$	$\alpha^2=1.25^2$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	129.20[g]	35.03	21.46	16.03	10.33
每次试验样本数为500, 试验次数为1000次, $H_u=2$.					

§5. (μ_0,σ_0^2)未知情形

依照第4节, 将过程控制分为均值控制与方差控制两个问题分别进行考虑.

[f]表示过程一直处于受控状态下, 用S_7监测过程是否发生变化$(\delta,\alpha^2)=(0.5,1)$时首次警报时刻.

[g]表示过程一直处于受控状态下, 用S_8监测过程是否发生变化$(\delta,\alpha^2)=(0,1.5^2)$时首次警报时刻.

5.1 均值监控

当过程受控时, Q统计量$Q_n = \Phi^{-1}\{T_{n-2}[((n-1)/n)^{1/2}\cdot((X_n-\overline{X}_{n-1})/S_{n-1})]\}$ $(n=3,4,\cdots)$服从标准正态分布, 其中$S_n^2=[1/(n-1)]\cdot\sum\limits_{i=1}^{n}(X_i-\overline{X}_n)^2$, $\overline{X}_n=(1/n)\cdot\sum\limits_{i=1}^{n}X_i$.

令$W_n^*=\Phi(Q_n)$, 可以证明: 当过程受控时, $W_n^*\sim U(0,1)$.

令

$$S_{9,(i+1)}=\sum_{j=1}^{i+1}\Big(w_j^*-\frac{1}{2}\Big)=S_{9,i}+\Big(w_{i+1}^*-\frac{1}{2}\Big),\qquad i\geq 2, \tag{5.1}$$

其中$S_{9,2}=0$. 若过程均值变大, 则$w_i^*-1/2>0$, 最终导致$S_{9,i}$值的增加, 给定H_u, 如果$S_{9,i}>H_u$, 则认为过程已经失控; 若过程均值变小, 则$w_i^*-1/2<0$, 最终导致$S_{9,i}$值的减少, 给定H_l, 如果$S_{9,i}<H_l$, 则认为过程已经失控, 亦可取$|H_l|=H_u$ (原因同上).

表5是模拟结果. 从表中可以看出, 该统计量不适用δ较小情形.

表5 (μ_0,σ_0^2)未知且均值变化时CED的模拟结果

CED	$\delta=0$	$\delta=0.25$	$\delta=0.5$	$\delta=0.75$	$\delta=1$
$\alpha^2=1^2$	71.36[h]	49.74	34.05	23.35	16.18
每次试验样本数为100, 试验次数为1000次, $H_u=2$.					

5.2 方差监控

记$R_n=X_n-X_{n-1}$. Q统计量为: $Q_n=\Phi^{-1}\{F_{1,\nu}[\nu R_n^2/(R_2^2+R_4^2+\cdots+R_{n-2}^2)]\}$, $n=4,6,\cdots,\nu$; $\nu=n/2-1$.

当过程受控时, 该统计量也服从标准正态分布(见[6]). 令$V_n^*=\Phi(Q_n)$, 有结论: 当过程受控时, $V_n^*\sim U(0,1)$.

令

$$S_{10,2(i+1)}=\sum_{j=1}^{i+1}\Big(v_{2j}^*-\frac{1}{2}\Big)=S_{10,2i}+\Big(v_{2(i+1)}^*-\frac{1}{2}\Big),\qquad i\geq 1, \tag{5.2}$$

其中$S_{10,2}=0$. 其它讨论与4.2类似.

事实上, 可令$S_{10,2(i+1)}=\sum\limits_{j=1}^{i+1}(v_j-k)=S_{10,2i}+(v_{2(i+1)}-k)$, k根据对ARL_0的要求或模拟计算确定. 本文中, k取$1/2$, 是基于模拟计算而得.

从表6数据可以看出, 该统计量总体来说并不好, k和相应H_u的确定有待进一步研究.

表6 (μ_0,σ_0^2)未知且方差变化时CED的模拟结果

CED	$\alpha^2=1^2$	$\alpha^2=1.25^2$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	82.13[i]	30.57	27.78	24.54	22.98
每次试验样本数为100, 试验次数为1000次, $H_u=2$.					

[h] 表示过程一直处于受控状态下, 用S_9监测过程是否发生变化$(\delta,\alpha^2)=(0.5,1)$时首次警报时刻.

[i] 表示过程一直处于受控状态下, 用S_{10}监测过程是否发生变化$(\delta,\alpha^2)=(0,1.5^2)$时首次警报时刻.

§6. 纯变化问题

当实际问题只是关心过程是否发生异常时, 若按以上方法, 则需用两个判据来分别考虑过程可能发生的变化: 均值变化和方差变化. 但纯变化问题并不关心均值和方差中究竟何者发生了变化, 因此可以改进以上方法, 以节省时间和费用, 尤其在考察多个指标、或多个过程时, 可以减少计算量和绘制CUSUM图的费用. 本节综合利用了统计假设检验中的极大似然比([8])和序贯概率比检验的原理, 给出了一种新的考察方法以及用于解决此问题的判据、递推公式和用于构造控制图的统计量.

设过程自时间$t+1$发生变化, 即检验问题是:

$$\begin{aligned}
H_0\ &:\ \mu_i=\mu_0,\quad \sigma_i^2=\sigma_0^2,\qquad 1\le i\le m;\\
H_{tm}&:\ \mu_1=\cdots=\mu_t=\mu_0,\quad \mu_{t+1}=\cdots=\mu_m\ne\mu_0,\\
&\ \text{or}\quad \sigma_1^2=\cdots=\sigma_t^2=\sigma_0^2,\quad \sigma_{t+1}^2=\cdots=\sigma_m^2\ne\sigma_0^2.
\end{aligned}$$

极大似然比为:

$$\begin{aligned}
\frac{\max\limits_{\theta\in\Theta_t}L_{tm}}{\max\limits_{\theta\in\Theta_0}L_0}
&=\frac{\max\limits_{\theta\in\Theta_t}\prod\limits_{i=1}^{m}\dfrac{1}{\sqrt{2\pi}\sigma_i}\exp\Big\{-\dfrac{(x_i-\mu_i)^2}{2\sigma_i^2}\Big\}}{\prod\limits_{i=1}^{m}\dfrac{1}{\sqrt{2\pi}\sigma_0}\exp\Big\{-\dfrac{(x_i-\mu_0)^2}{2\sigma_0^2}\Big\}}\\
&=\frac{(2\pi)^{-(m-t)/2}\Big(\sum\limits_{i=t+1}^{m}(x_i-\overline{x}_{t,m})^2\Big)^{-(m-t)/2}(m-t)^{(m-t)/2}e^{-(m-t)/2}}{\Big(\dfrac{1}{\sqrt{2\pi}\sigma_0}\Big)^{m-t}\exp\Big\{-\sum\limits_{i=t+1}^{m}(x_i-\mu_0)^2\Big/(2\sigma_0^2)\Big\}}\\
&=\frac{\sigma_0^{m-t}\Big(\sum\limits_{i=t+1}^{m}(x_i-\overline{x}_{t,m})^2\Big)^{-(m-t)/2}(m-t)^{(m-t)/2}e^{-(m-t)/2}}{\exp\Big\{-\sum\limits_{i=t+1}^{m}(x_i-\mu_0)^2\Big/(2\sigma_0^2)\Big\}}\\
&=\Big[\sum_{i=t+1}^{m}\Big(\frac{x_i-\overline{x}_{t,m}}{\sigma_0}\Big)^2\Big/(m-t)\Big]^{-(m-t)/2}\exp\Big\{\frac12\Big[\sum_{i=t+1}^{m}\Big(\frac{x_i-\mu_0}{\sigma_0}\Big)^2-(m-t)\Big]\Big\}.
\end{aligned}$$

由此可得判据为:

$$\sum_{i=t+1}^{m}\Big(\frac{x_i-\mu_0}{\sigma_0}\Big)^2-(m-t)\Big[\ln\Big(\sum_{i=t+1}^{m}\Big(\frac{x_i-\overline{x}_{t,m}}{\sigma_0}\Big)^2\Big/(m-t)\Big)+1\Big]\ge 2\ln A.$$

其中, $\overline{x}_{t,m}=[1/(m-t)]\cdot\sum\limits_{i=t+1}^{m}x_i$.

记$\lambda_i=\sum\limits_{j=1}^{i}((x_j-\mu_0)/\sigma_0)^2$, $\tau_i=i\Big[\ln\Big((1/i)\cdot\sum\limits_{j=1}^{i}((x_j-\overline{x}_{0,i})/\sigma_0)^2\Big)+1\Big]$, 其中, $\overline{x}_{0,i}=(1/i)\cdot\sum\limits_{j=1}^{i}x_j$.

因此有以下递推公式:

$$\lambda_i=\lambda_{i-1}+\Big(\frac{x_i-\mu_0}{\sigma_0}\Big)^2,$$
$$\tau_i=i\Big\{\ln\Big(\frac{i-1}{i^2}\Big[i\cdot\exp\Big\{\frac{\tau_{i-1}}{i-1}-1\Big\}+\Big(\frac{x_i-\overline{x}_{0,i-1}}{\sigma_0}\Big)^2\Big]\Big)+1\Big\},\qquad i\geq 2.$$

用于构造控制图的统计量为:

$$S_{11,i}=\lambda_i-\tau_i,$$
$$S_{11,1}=\Big(\frac{x_1-\mu_0}{\sigma_0}\Big)^2-1. \tag{6.1}$$

$S_{11,i}$的形式与通常的CUSUM统计量稍不同, 为区别通常的CUSUM图, 将此控制图命名为CUSUM-D图.

CED的模拟结果见表7. 从表中可以看出, 该方法能够比以上方法更快地给出警报.

表7　纯变化问题的条件期望延时(CED)的模拟结果

CED	$\alpha^2=1$	$\alpha^2=1.25$	$\alpha^2=1.5^2$	$\alpha^2=1.75^2$	$\alpha^2=2^2$
$\delta=0$	60.53[j]	12.06	6.81	5.13	4.33
$\delta=0.25$	16.01	9.73	6.49	4.99	4.31
$\delta=0.5$	8.56	7.45	5.88	4.90	4.22
$\delta=0.75$	6.27	5.91	5.15	4.56	4.02
$\delta=1$	5.11	4.95	4.55	4.09	3.83

每次试验样本数为100, 试验次数为10000次, $\ln A=2$.

§7.　各CUSUM图之间的比较

表8给出了S_3、S_{11}与S_1、S_2 (即最常见的CUSUM图, S_1用于控制均值, S_2用于控制方差)的比较, 其中, 样本数为500 (如无特别声明, 下列计算的样本数皆为500), "—"表示该统计量不用于检验此种过程偏移(下同). 从结果可以看出, 本文提出的S_3、S_{11}比S_1、S_2的CED要小; 在过程受控下, S_{11}的CED偏小. 但总体而言, S_3、S_{11}的结果是令人满意的.

表8　S_1、S_2与S_3、S_{11}的CED比较

(δ,α)	$(0,1)$	$(0.5,1)$	$(0,1.5)$	$(0.25,1.25)$	$(0.5,1.5)$
$S_{1(\delta,1)}$	467.03	10.10	—	27.12	11.47
$S_{2(0,\alpha)}$	423.79	—	11.30	21.17	9.08
$S_{3(\delta,\alpha)}$	427.36	—	15.45	18.51	8.91
$S_{11(\delta,\alpha)}$	270.27	8.56	6.81	9.73	5.88

[j]表示过程一直处于受控状态下, 用S_{11}监测过程是否发生变化$(\delta,\alpha^2)=(0.5,1.5^2)$时首次警报时刻.

表9给出了S_4、S_5、S_6与Q统计量(图)的比较，其中，

$$Q_1=\Big(\frac{n-1}{n}\Big)^{1/2}\frac{(x_n-\overline{x}_{n-1})}{\sigma_0},\qquad \text{用于控制过程均值},$$

$$Q_2=\Phi^{-1}\Big\{H_1\Big(\frac{R_r^2}{2\sigma_0^2}\Big)\Big\},\qquad \text{用于控制过程方差}.$$

从模拟结果可以看出，S_4的CED稍大于Q_1，但当均值偏移较大时，二者相差不大；S_5与Q_2的CED差别不大、S_6的CED与S_5几乎一致.

表9 S_4、S_5、S_6与Q_1、Q_2的比较

(δ,α)	$(0,1)$	$(0.5,1)$	$(0,1.5)$	$(0.25,1.25)$	$(0.5,1.5)$
$Q_{1(\delta,1)}$	431.03	15.94	—	37.90	13.05
$S_{4(\delta,1)}$	301.79	17.26	—	39.13	11.04
$Q_{2(0,\alpha)}$	392.16	—	12.43	25.23	11.29
$S_{5(0,\alpha)}$	412.35	—	14.97	28.57	13.79
$S_{6(\delta,\alpha)}$	402.66	—	15.09	28.32	13.35

表10将S_7、S_9与Q_3、Q_4比较，其中，

$$Q_3=\Phi^{-1}\Big\{T_{n-1}\Big(\frac{X_n-\mu_0}{S_{0,n-1}}\Big)\Big\},\qquad Q_4=\Phi^{-1}\Big\{T_{n-2}\Big[\Big(\frac{n-1}{n}\Big)^{1/2}\Big(\frac{X_n-\overline{X}_{n-1}}{S_{n-1}}\Big)\Big]\Big\}.$$

表10 S_7、S_9与Q_3、Q_4的比较

(δ,α)	$(0,1)$	$(0.25,1)$	$(0.5,1)$	$(0.75,1)$	$(1,1)$
Q_3	304.33	23.60	11.79	10.76	9.10
S_7	213.18	32.72	16.75	12.29	10.27
Q_4	346.65	28.36	19.01	17.83	14.57
S_9	218.02	49.74	34.05	22.35	16.18

表11给出S_8、S_{10}与Q_5、Q_6的比较结果(样本数为100)，其中，

$$Q_5=\Phi^{-1}\Big\{F_{1,n-1}\frac{(X_n-\mu_0)^2}{S_{0,n-1}^2}\Big\},\qquad Q_6=\Phi^{-1}\Big\{F_{1,\nu}\Big(\frac{\nu R_n^2}{R_2^2+R_4^2+\cdots+R_{n-2}^2}\Big)\Big\}.$$

表11 S_8、S_{10}与Q_5、Q_6的比较

(δ,α)	$(0,1)$	$(0,1.25)$	$(0,1.5)$	$(0,1.75)$	$(0,2)$
Q_5	89.22	27.91	19.56	17.84	13.00
S_8	90.37	35.03	21.46	16.03	10.33
Q_6	91.05	28.80	23.78	20.35	18.91
S_{10}	82.13	30.57	27.19	24.54	22.98

从表10、表11可以看出, σ^2未知时, 特别是偏移较小时, 新统计量的CED稍大于相应的Q统计量; 在偏移较大时, $S_7,\cdots,S_{10}$的CED与Q统计量的CED相差较小. 但正如前问所述, $S_7,\cdots,S_{10}$更易计算, 因此, 在偏移较大时, 相应的CUSUM图具有一定优势.

特别指出, 如无特别说明, 本节计算所用参数及计算结果的意义与前文计算相应参数时一致.

参 考 文 献

[1] Page, E.S., Continuous inspection schemes, *Biometrika*, **41**(1954), 100–115.

[2] Kenett, R. & Zacks, S., *Modern Industrial Statistics-Design and Control of Quality and Reliability*, 1st edition, Boston: Duxbury Press, 1998.

[3] ISO/TR 7871: 1997(E), *Cumulative Sum Charts-Guidance on Quality Control and Data Analysis Using CUSUM Techniques.*

[4] Yeh, A.B., Lin, D.K.J. & Venkataramani, C., Unified CUSUM charts for monitoring process mean and variability, *Quality Technology and Quantitative Management*, **1**(2004), 65–86.

[5] 王淑君, 常规控制图与累积和控制图, 北京: 国防工业出版社, 1990.

[6] Quesenberry, C.P., SPC Q charts for start-up processes and short or long runs, *Journal of Quality Technology*, **23**(1991), 213–224.

[7] 濮晓龙, 过程参数未知时的连续检验问题, 应用概率统计, **17**(2001), 437–447.

[8] 方开泰, 实用多元统计分析, 上海: 华东师范大学出版社, 1989.

Several Types of CUSUM Charts for Normal Distribution Process

Tu Yujuan[1,2] Feng Shiyong[1]

([1]*Academy of Mathematics and Systems Science, CAS, Beijing, 100080*)

([2]*Graduate School of the Chinese Academy of Sciences, Beijing, 100080*)

According to different cases of normal-distribution process parameters (μ,σ^2), several problems are studied in this paper, for monitoring the process with parameters known or unknown, with single-parameter-deviation or with double-parameter-deviation. A series of CUSUM Charts are proposed in standardized or simpler forms mainly based on Q-statistics proposed by Quesenberry. In the last part of the paper, the concept of deviation-interested-only (DIO) is put forward and the relevant statistics are given where there is no need to know which has deviated, process mean or process variance. The simulation results of Conditional Expected Delay (CED) are shown in each case, and the comparisons in the end of the paper show which new CUSUM charts act well.

Keywords: CUSUM, Q-Statistics, Single(Double)-parameter-deviation, DIO, CED.

AMS Subject Classification: 62N10.