

文章编号: 1001-0920(2011)04-0619-04

二阶自相关过程 Shewhart 型控制图的性能评价

孙秋霞^{1,2}, 高齐圣¹

(1. 青岛大学 经济学院, 山东 青岛 266071; 2. 山东科技大学 理学院, 山东 青岛 266051)

摘要: 针对二阶自相关过程, 分别采用有限马氏链内嵌法和积分法给出了 Shewhart 型修正控制图和残差控制图平均运行链长的计算方法, 并通过其数值结果的分析, 得出结论: 当自相关过程系数均为正值时, 修正图的性能较好; 当过程系数均取负值时, 残差图较为适用; 当过程系数符号相异时, 两图性能可采用所给方法具体比较. 该结论为控制图的选择和应用提供了理论依据.

关键词: 统计过程控制; 修正 Shewhart 图; Shewhart 残差图; 平均运行链长; 自相关过程

中图分类号: O213.1

文献标识码: A

Performance evaluation of Shewhart type charts for AR(2) process

SUN Qiu-xia^{1,2}, GAO Qi-sheng¹

(1. College of Economics, Qingdao University, Qingdao 266071, China; 2. College of Science, Shandong University of Science and Technology, Qingdao 266051, China. Correspondent: SUN Qiu-xia, E-mail: qiuxiasun@163.com)

Abstract: Limited Markov chain embedded method and integration method are given to calculate the respective average run length of modified Shewhart chart and Shewhart residual chart for 2-order autoregressive process. Performance of both charts with different parameters are discussed and compared by numerical results. Some conclusions show that, modified chart is more applicable when both parameters are positive, residual chart is more applicable when both parameters are negative, otherwise both charts need to be compared through the proposed method. The conclusion provides academic basis to choose and apply control charts.

Key words: statistical process control; modified Shewhart chart; residual Shewhart chart; average run length; autoregressive process

1 引言

统计过程控制 (SPC) 已在制造业得到了广泛的应用, 相关研究已比较成熟. 其目的是用来监控过程, 检测异常波动, 寻找并消除产生异常波动的异因. 控制图是实施过程监控的有效工具, 它的使用基于一个重要的假设, 即过程受控时产生的数据是独立且服从正态分布的. 但实际中许多过程数据通常表现出某种程度的自相关性^[1], 如在化学工业等相关连续型工业的质量特性的连续观测值、宏观经济或金融数据等均表现出不同程度的相关性. 在违背独立性假设的情况下, 如果在原始数据上直接应用传统控制图, 则会给控制图的性能造成不利影响, 如误警率增加等. 为更有效地监控过程, 众多学者对自相关数据控制图进行了大量研究^[2-4], 总体可分为两种方法: 一是修正控制图, 二是残差控制图. 前者的主要思想是: 将过程自

身的相关性考虑到控制限的设计中, 通过修正传统控制图的控制限来达到有效控制的目的; 后者先拟合自相关过程的时间序列模型, 再利用拟合模型的残差相互独立且服从正态分布的特性, 以传统控制图直接监控残差序列值.

大量自相关过程控制图的研究表明^[5-6], 没有一种适用于所有自相关过程的最优控制图, 当过程的自相关结构或模型的某个(些)参数发生变化时, 或过程中异常波动的类型和大小不同时, 原有最优控制图的性能很可能会发生变化, 而其他类型的控制图却能够达到更好的监控性能. 平均运行链长 (ARL) 是控制图使用最广泛的性能评价标准, 对于受控过程而言, 控制图的 ARL 越大, 说明其性能越好; 而对于失控过程, 其值越小越好, 即表明异常波动可以被及时发现. 以往文献大多集中于一阶自相关过程即 AR(1) 的监

收稿日期: 2010-04-23; 修回日期: 2010-07-25.

基金项目: 国家自然科学基金项目(70571041).

作者简介: 孙秋霞(1976—), 女, 博士生, 从事系统工程、过程控制等研究; 高齐圣(1966—), 男, 教授, 博士生导师, 从事系统工程、质量管理等研究.

控,而对二阶(AR(2))等较高阶过程的监控控制图的研究^[7]则较少.

本文着力于AR(2)过程控制图的研究,针对其修正Shewhart图和Shewhart残差图的性能指标平均运行链长,分别给出其计算方法,并通过对相应数值结果的详细分析和比较,为控制图的选择和应用给出了相关建议.鉴于过程参数对控制图的性能有显著影响,为使所给控制图性能具有可比性,假设所给过程中的参数均已知,待估情况及其影响将另文讨论.

2 预备知识

2.1 过程模型和符号假设

假设 $\{X_t\}$ 为观测值过程,它是一随机过程;目标过程设为 $\{Y_t\}$,此处考虑它是具有常均值 $E(Y_t) = \mu_0$ 的(弱)平稳过程,不失一般性,令 $\mu_0 = 0$.两过程的关系为

$$X_t = Y_t + al_{q,q+1,\dots}(t), t \geq 1. \quad (1)$$

其中: l_A 表示集合A在时刻t时的示性函数,a表示观测值过程在偏移点q处的偏移大小.本文设定如果观测值过程有偏移发生,则均在时刻 $q = 1$ 处发生.

假设目标过程 $\{Y_t\}$ 为AR(2)过程,且满足

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t. \quad (2)$$

其中: ε_t 是均值为零、方差为 σ_ε^2 的独立且服从正态分布的随机误差,即 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$; α_1, α_2 表示自相关参数,且满足以下约束:

$$\alpha_1 + \alpha_2 < 1, \alpha_2 - \alpha_1 < 1, |\alpha_2| < 1. \quad (3)$$

另外假设随机变量 Y_1 和 Y_2 服从二维正态分布,边际均值和方差分别为0和 $(1 - \alpha_2)/[(1 + \alpha_2)(1 - \alpha_2 + \alpha_1)(1 - \alpha_2 - \alpha_1)]\sigma_\varepsilon^2$,相关系数 $\rho = \alpha_1/(1 - \alpha_2)$.当 $t \geq 1$ 时, $Y_t \sim N(0, \sigma_Y^2)$,其中

$$\sigma_Y^2 = \gamma_0 \sigma_\varepsilon^2, \\ \gamma_0 = \frac{1 - \alpha_2}{(1 + \alpha_2)(1 - \alpha_2 + \alpha_1)(1 - \alpha_2 - \alpha_1)}.$$

2.2 有限马氏链内嵌法

设 S_t 为某一控制图统计量,控制图平均运行链长(ARL)可定义为

$$N = \inf\{t \in \mathbf{N} : S_t \geq \text{UCL} \text{ or } S_t \leq \text{LCL} | S_0 = 0\},$$

其中UCL和LCL分别为控制图的上、下控制限.

在正态性假设下,Fu等人^[8-9]采用有限马氏链内嵌法计算出所给控制图的平均运行链长.其主要思想^[10]是:

1) 将控制图的双侧或单侧分别划分为 $2m + 3$ 或 $m + 2$ 个状态,即

$$\Omega_1 = a_0, \pm a_1, \dots, \pm a_m, \pm a_{m+1},$$

$$\Omega_2 = a_0, a_1, \dots, a_m, a_{m+1}.$$

其中: $a_0 = 0, a_{m+1} = \text{UCL}, -a_{m+1} = \text{LCL}, a_i = (i - 0.5)d, i = 1, 2, \dots, m, d = h/(m + 1), h = \text{UCL} - |\text{LCL}|$,从而产生离散的统计量序列 $S_t(m)$.

2) 在 $S_t(m)$ 的基础上,构建相应状态空间上的马氏链 $Z_t(m)$,该链的转移概率矩阵定义为

$$M(m) = \begin{bmatrix} A(m) & B(m) \\ \mathbf{0} & 1 \end{bmatrix}.$$

其中

$$A(m) = (p_{ij})_{(m+1) \times (m+1)},$$

$$p_{ij} =$$

$$P\{Z_t(m) = j | Z_t(m) = i, i, j = 0, 1, \dots, m\},$$

$$B(m) = (p_{i(m+1)})_{(m+1) \times 1},$$

$$p_{i(m+1)} =$$

$$P\{Z_t(m) = m + 1 | Z_t(m) = i, i = 0, 1, \dots, m\},$$

且 $\mathbf{0}$ 为 $m + 1$ 维的零行向量.

$$3) \quad P\{Z_0 = a_0\} = 1.$$

因此基于马氏链 $Z_t(m)$ 的平均运行链长为

$$N(m) = \inf\{t \geq 1 : Z_t(m) = a_{m+1} | Z_0(m) = a_0\}.$$

其重要性质有^[8]:

定理1 若有限马氏链 $Z_t(m)$ 满足以上3个条件,则对于给定正整数 m ,有如下关系式成立:

$$P[N(m) = n] =$$

$$\pi_0(m) A^{n-1}(m) [I - A(m)] U^T(m), \quad (4)$$

$$E[N(m)] = \pi_0(m) [I - A(m)]^{-1} U^T(m), \quad (5)$$

$$E[N^2(m)] =$$

$$\pi_0(m) [I + A(m)] [I - A(m)]^{-2} U^T(m). \quad (6)$$

其中: $\pi_0(m) = (1, 0, \dots, 0)$,且1对应于给定的初始状态;矩阵I为 $m + 1$ 维的单位矩阵; $U^T(m)$ 为 $m + 1$ 维行向量 $U(m) = (1, 1, \dots, 1)$ 的转置矩阵.

3 AR(2)过程的Shewhart型控制图

针对AR(2)过程构建Shewhart修正控制图和Shewhart残差图,并分析所给控制图的平均运行链长的计算方法,为进一步的性能比较作准备.

3.1 Shewhart修正控制图

Shewhart修正图中统计量 S_t 直接采用观测值 X_t ,控制限采用 $|X_t - \mu_0| > c\sqrt{\text{Var}(X_t)}$.给定正整数 m ,取步长 $\Delta = h/(m + 0.5) > 0$.控制图的每一侧等分成 $m + 1$ 个状态,其中分别有一个状态处于控制限(CL)之外.由于初始变量 Y_1 和 Y_2 服从二维正态分布,给出虚拟变量 Y_{-1} 和 Y_0 以构成马氏链.记 DY_t 为随机变量 Y_t 的离散变量,并规定初始态满足

$P[(DY_{-1}, DY_0) = (\emptyset, \emptyset)] = 1$, 故整个状态空间为

$$\Omega = \{(\emptyset, \emptyset), (\emptyset, e), (e_1, e_2), \bar{e}\}.$$

其中: $e, e_1, e_2 = -m, \dots, m$; \bar{e} 表示吸收态. 定义马氏链 $\{Z_t(m)\} : Z_t(m) = [DY_{t-1}, DY_t], t \geq 0$. 从状态 $u = (i_{t-2}, i_{t-1})$ 到状态 $v = (i_{t-1}, i_t)$ 的转移概率记为 p_{uv} , 于是内嵌式马氏链 $\{Z_t(m)\}$ 的转移概率为:

1) 若 $u = (\emptyset, \emptyset), v = (\emptyset, i_0), i_0 = -m, \dots, m$, 则

$$p_{uv} = P\{(i_0 - 0.5)\Delta \leq Y_1 \leq (i_0 + 0.5)\Delta\}, \quad (7)$$

其中随机变量 $Y_1 \sim N(0, \sigma_Y^2)$;

2) 若 $u = (\emptyset, i_0), v = (i_0, i_1), i_0, i_1 = -m, \dots, m$, 则

$$p_{uv} = P\{(i_1 - 0.5)\Delta \leq Y_2 \leq (i_1 + 0.5)\Delta | Y_0 = i_0\Delta\}, \quad (8)$$

其中条件随机变量 $Y_2 | Y_1 = i_0\Delta$ 服从正态分布

$$N\left(\frac{\alpha_1}{1 - \alpha_2}(i_0\Delta), \left[1 - \left(\frac{\alpha_1}{1 - \alpha_2}\right)^2\right]\sigma_Y^2\right);$$

3) 当 $t \geq 3$ 时, 若 $i_{t-2}, i_{t-1}, i_t = -m, \dots, m$, 则

$$p_{uv} = P\{(i_t - 0.5 - \alpha_1 i_{t-1} - \alpha_2 i_{t-2})\Delta \leq \varepsilon_t \leq (i_t + 0.5 - \alpha_1 i_{t-1} - \alpha_2 i_{t-2})\Delta\}, \quad (9)$$

其中随机误差变量 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

由式 (7) ~ (9), 可得该马氏链的转移概率矩阵 $A(m)$, 修正图的受控 ARL 可由式 (5) 计算. 当过程有偏移发生时, 利用方程 (1) 变换式 (7) ~ (9), 进而可求得该图的失控 ARL.

3.2 Shewhart 残差控制图

Shewhart 残差控制图通常是利用观测值与其预测值的差值来实现监控目的, 其控制限通常采用 $|X_t - \hat{X}_t| > c\sqrt{\text{Var}(X_t - \hat{X}_t)}$. 简记 $\Delta_t = X_t - \hat{X}_t, \sigma = \text{Var}(X_t - \hat{X}_t)$, 则 Shewhart 残差图的平均运行链长为

$$N_r(c) = \inf\{t \in N : |\Delta_t| > c\sigma\}.$$

对于模型 (1), (2), 取:

当 $t = 1$ 时

$$\Delta_1 = \frac{1}{\sqrt{\gamma_0}}X_1 = \frac{1}{\sqrt{\gamma_0}}Y_1 + \frac{1}{\sqrt{\gamma_0}}a;$$

当 $t = 2$ 时

$$\Delta_2 = \frac{1}{\sqrt{\gamma_0}}X_2 = \frac{1}{\sqrt{\gamma_0}}Y_2 + \frac{1}{\sqrt{\gamma_0}}a;$$

当 $t \geq 3$ 时, 一步预测值取 $\hat{X}_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2}$, 故

$$\Delta_t = (Y_t - \alpha_1 Y_{t-1} - \alpha_2 Y_{t-2}) + (1 - \alpha_1 - \alpha_2)a = \varepsilon_t + (1 - \alpha_1 - \alpha_2)a,$$

显然 $\text{Var}(\Delta_t) = \sigma_\varepsilon^2, \forall t \geq 1$.

因为 $E_a[N_r(c)] = \sum_{k=1}^{\infty} k \cdot P[N_r(c) = k]$, 且

$$P[N_r(c) > k] =$$

$$P[|\Delta_1| \leq c\sigma, |\Delta_2| \leq c\sigma, \dots, |\Delta_k| \leq c\sigma] =$$

$$P[|\Delta_1| \leq c\sigma, |\Delta_2| \leq c\sigma] \cdot \prod_{i=3}^k P[|\Delta_i| \leq c\sigma],$$

故有

$$E_a[N_r(c)] = 1 + (\Phi_3 - \Phi_4) + \frac{1}{1 - (\Phi_1 - \Phi_2)} \times \int_{-c - \frac{1}{\sqrt{\gamma_0}} \frac{a}{\sigma_\varepsilon}}^{c - \frac{1}{\sqrt{\gamma_0}} \frac{a}{\sigma_\varepsilon}} \int_{-c - \frac{1}{\sqrt{\gamma_0}} \frac{a}{\sigma_\varepsilon}}^{c - \frac{1}{\sqrt{\gamma_0}} \frac{a}{\sigma_\varepsilon}} f(x, y) dx dy. \quad (10)$$

其中

$$\Phi_1 = \Phi\left(c - \frac{a}{\sigma_\varepsilon}(1 - \alpha_1 - \alpha_2)\right),$$

$$\Phi_2 = \Phi\left(-c - \frac{a}{\sigma_\varepsilon}(1 - \alpha_1 - \alpha_2)\right),$$

$$\Phi_3 = \Phi\left(c - \frac{1}{\sqrt{\gamma_0}} \frac{a}{\sigma_\varepsilon}\right),$$

$$\Phi_4 = \Phi\left(-c - \frac{1}{\sqrt{\gamma_0}} \frac{a}{\sigma_\varepsilon}\right),$$

符号 Φ 和 $f(x, y)$ 分别表示标准正态分布函数和二维正态分布 $N(0, 1; 0, 1; \rho)$ 的联合概率密度函数. 当 $a = 0$ 时, 即 $E_0[N_r(c)]$ 表示无异常发生时, 该图的受控平均运行链长的值可由式 (5) 的简化式计算. 经分析, 受控 ARL 仅与相邻随机变量的相关系数 ρ 有关, 与参数 $\alpha_1, \alpha_2, \sigma_\varepsilon$ 无关; 而失控 ARL 与 $\alpha_1, \alpha_2, \rho, \frac{a}{\sqrt{\gamma_0}\sigma_\varepsilon}$ 均有关.

4 数值结果与分析

给出一个满足方程 (2) 的 AR(2) 平稳过程, 通过给定不同参数值 (满足条件 (3)), 计算所给两种控制图 ARL 的数值结果. 临界值取 $c = 3.09$ 以使各图受控 ARL 与变量独立时的结果保持一致. 不失一般性, 取 $\sigma_\varepsilon = 1$.

利用第 3 节所给方法得到两图 ARL 的数值结果, 见表 1 和表 2. 表 1 和表 2 的数据显示, AR(2) 过程的系数参数 $\alpha_1 = \alpha_2 = 0$ 时, 该过程即为随机变量相互独立的情形, 两种方法的数值结果完全一致, 此时两图的性能相当. 当 $\alpha_2 = 0$ 时, 该过程为 AR(1) 过程, 修正图的受控 ARL 随着相关性的增强而变小; 残差图受控 ARL 的变化方向与之相反, 但其变化幅度极其微弱.

对于 AR(2) 过程的一般情形, 当其中一个系数确定而另外一个系数增大时, 修正图的受控 ARL 随着增大, 且变化幅度较大. 残差图的受控 ARL 正如理论分析, 其值仅与相邻变量的相关系数 ρ 有关, 且随着

表1 修正 Shewhart 图的 ARL

α_1	α_2	a				
		0	0.5	1	1.5	2
0	0	370.398 1	155.224 2	43.894 7	14.967 7	6.303 0
	0.2	364.639 3	156.528 0	46.209 5	16.705 1	7.451 7
	0.4	349.487 4	158.828 9	50.631 6	19.690 1	9.243 2
	0.6	328.898 6	165.843 5	59.608 4	25.410 0	12.573 6
	0.8	304.307 9	186.483 0	82.859 4	40.507 0	21.673 0
0.2	0	364.433 6	156.134 7	45.702 8	16.197 7	7.019 6
	0.2	357.580 2	159.862 0	50.007 5	19.096 0	8.804 6
	0.4	344.006 2	169.035 8	59.390 7	24.963 4	12.230 0
	0.6	338.313 3	200.858 4	86.655 7	41.834 7	22.273 9
0.4	0	348.735 5	157.794 9	49.454 4	18.570 0	8.309 7
	0.2	342.325 3	168.990 0	59.142 8	24.491 4	11.749 9
	0.4	343.779 5	204.077 5	87.441 6	41.684 5	21.875 4
0.6	0	327.275 0	163.822 9	57.404 2	23.349 9	10.872 8
	0.1	326.065 8	175.204 6	66.638 5	28.905 9	14.082 2
	0.3	341.040 6	242.518 5	127.290 7	68.303 2	38.488 9
0.8	0	300.723 1	182.269 7	78.301 8	36.229 7	18.097 9
	0.1	344.485 8	244.086 8	126.631 4	66.452 2	35.987 6
	-0.2	299.165 5	153.575 1	54.255 4	21.508 6	9.532 2
	-0.6	258.611 2	125.173 8	41.545 4	13.993 8	4.796 1
-0.2	0	364.433 6	152.492 5	42.948 6	14.439 2	5.943 9
	0.2	357.580 2	153.197 2	44.964 6	15.909 3	6.890 0
	0.4	344.006 2	156.020 1	49.397 2	18.577 4	8.327 9
	0.6	338.313 3	170.266 7	61.280 8	24.759 7	11.168 0
	-0.2	359.577 6	150.440 9	41.983 2	13.618 4	5.248 0
	-0.4	343.899 5	145.379 2	41.511 7	13.260 0	4.682 1
	-0.8	282.736 9	123.111 4	44.235 3	13.982 2	3.161 1
-0.6	0	327.275 0	137.161 8	42.267 4	15.418 0	6.385 7
	0.2	326.858 8	144.527 4	51.214 8	20.470 1	8.728 4
	-0.2	326.048 6	136.163 2	39.742 4	13.598 1	5.307 7
	-0.6	281.739 5	129.704 8	40.648 9	13.063 4	4.078 4

表2 Shewhart 残差图的 ARL

α_1	α_2	ρ	a				
			0	0.5	1	1.5	2
0	0	0	370.398	155.224	43.894	14.967	6.303
	0.2	0	370.398	199.565	70.456	26.364	10.816
	0.4	0	370.398	252.211	117.225	52.113	23.344
	0.6	0	370.398	307.345	196.636	113.181	62.381
	0.8	0	370.398	352.143	305.284	246.133	187.915
0.2	0	0.20	370.402	199.586	70.513	26.470	10.965
	0.2	0.25	370.407	252.125	116.985	51.790	23.056
	0.4	0.33	370.417	307.152	195.773	111.814	60.600
	0.6	0.50	370.455	351.957	304.285	243.634	183.384
0.4	0	0.40	370.428	252.296	117.442	52.520	23.929
	0.2	0.50	370.455	307.301	196.230	112.697	61.872
	0.4	0.67	370.541	352.132	304.778	244.713	185.288
0.6	0	0.60	370.498	307.559	197.046	114.227	64.013
	0.1	0.67	370.541	332.251	249.779	169.655	109.137
	0.3	0.85	370.759	365.803	351.431	328.991	300.350
0.8	0	0.80	370.671	352.568	306.196	247.928	191.013
	0.1	0.89	370.817	365.944	351.846	329.941	302.180
	-0.2	0.67	370.541	307.843	197.953	115.889	66.341
	-0.6	0.50	370.455	200.387	72.119	28.534	13.175
-0.2	0	-0.2	370.402	120.159	28.594	9.546	4.456
	0.2	-0.25	370.407	155.350	44.108	15.205	6.517
	0.4	-0.33	370.417	199.912	71.191	27.324	11.794
	0.6	-0.5	370.455	252.943	119.118	55.014	26.808
	-0.2	-0.16	370.400	93.177	19.461	6.724	3.585
	-0.4	-0.14	370.400	72.729	13.914	5.231	3.216
	-0.8	-0.11	370.399	45.545	8.190	3.912	3.070
-0.6	0	-0.60	370.498	72.846	14.023	5.348	3.366
	0.2	-0.75	370.612	93.587	19.895	7.180	4.122
	-0.2	-0.50	370.455	57.235	10.400	4.355	3.017
	-0.6	-0.38	370.424	36.471	6.594	3.493	2.883

ρ 值的增大而增大, 但其增加幅度较小. 另外数据显示, 当过程参数相同时, 过程受控修正图的 ARL 比残差图的 ARL 值小, 且前者对过程系数变化的反应相对敏感; 而过程一旦出现均值偏移, 两图的失控 ARL 均会随着偏移程度的增大而减小, 即过程偏移越大, 两图的报警时间越短, 故两图均具有一定的有效性. 当相关系数值较大时, 如 $\rho \geq 0.5$, 两图对过程偏移大小的敏感性均较差, 修正图的性能相对残差图较好些. 这也说明, 两图关于过程偏移的敏感程度与该过程的相关系数密切相关.

通过两表数据的比较还可发现, 当过程系数均为正数时, 修正图的失控 ARL 普遍比残差图的失控 ARL 小; 过程系数均取负值(相关系数也为负数)时, 后者相对于前者较小, 即在此情形下残差控制图能够更及时地发现异常存在; 当两过程系数取值符号相异时, 两图的性能差异不尽相同.

5 结 论

为实现对已知参数的 AR(2) 平稳过程的监控, 本文给出两种不同的 Shewhart 型控制图, 且对所给控制图的性能——平均运行链长进行了相应的数值计算. 通过数值结果的分析发现: 当过程的系数参数 α_1, α_2 均为正数时, Shewhart 修正图的性能较好; 当两系数均为负值时, 残差图的性能比前者具有优势; 当两系数符号不同时, 两图性能规律不明显, 此时可通过所给方法, 针对系数的具体取值计算并比较两图 ARL 的结果, 进而进行控制图的选择和应用.

在实际情况中, 过程参数往往需要估计, 而估计值对控制图性能的影响必然应加以考虑, 这一问题有待于进一步研究.

参考文献(References)

- [1] Montgomery D C. Introduction to statistical quality control[M]. 6th ed. New York: Wiley, 2008.

(下转第628页)