基于贝叶斯动态模型的自相关控制图

崔敬巍 谢里阳

(东北大学 机械工程与自动化学院，沈阳 110004)

摘 要：传统控制图应用的假设前提是从过程得到的观测值彼此独立.但许多过程存在自相关现象,并且这种自相关现象对控制图性能有严重影响.从工程实际出发,对 AR(1)过程模型提出了基于贝叶斯动态模型的自相关控制图,阐明了此自相关控制图的原理和方法;并将该控制图与基于过程模型的自相关控制图的性能进行比较.此控制图借助贝叶斯先验估计,建模迅速,不以时间序列模型为基础,对强自相关模型具有很好的性能.对强自相关过程,并要求迅速建立过程控制时,可以使用此自相关控制图,进行过程监视.

关 键 词：统计过程控制；控制图；时间序列；贝叶斯方法

中图分类号：TB 114.2

文献标识码：A **文 章 编 号**：1001-5965(2007)03-0375-04

Auto-correlated charts based on dynamic Bayesian model

Cui Jingwei Xie Liyang

(School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110004, China)

Abstract: In statistical process control it is usually assumed that the observations taken from the process are independent. However, for many processes the observations are autocorrelated, and this autocorrelation can have a significant effect on the performance of traditional control chart. The problem of monitoring the mean of process in which the observations can be modeled as an AR(1) process was considered. A one-step forecasting autocorrelated control charts based on dynamic Bayesian model was presented for engineering practices. The principle and methodology to build autocorrelated control charts was described in detail. This control chart's performance was compared to the performance of control chart based on process model, and the effect of process parameter estimation on the control chart was also investigated. When the process model is strong autocorrelation AR(1), the autocorrelated control chart has good performances with few samples. In this case, autocorrelated control chart was suggested instead of traditional control charts.

Key words: statistical process control; control charts; time series; Bayes methods

质量控制图是现代质量管理中应用广泛的方法和工具,具有很重要的作用.经典控制图的理论前提是假定过程观测值服从独立、同分布.然而在实际生产过程中自相关现象普遍存在,违背了观测值之间相互独立这一基本假设[1].如果使用常规控制图进行过程控制,会错误理解数据中隐含的信息,造成虚报并不存在的特殊原因或实际存在的特殊原因不能检出,这使控制图的应用失去意义[2-3].目前,对自相关过程的质量控制已经成为国内外质量控制领域研究的热门课题[4-5].

1 自相关控制图

20 世纪七八十年代以来人们越来越关注如何对自相关过程进行有效统计过程控制.国外很多学者在这一领域作了大量研究工作,取得了丰硕的成果.主要处理方法有 3 类:第 1 类是 Vasi-

收稿日期：2006-04-20

基金项目：国家自然科学基金资助项目(50275025)

作者简介：崔敬巍(1976—),女,辽宁新民人,博士生,cuijingwei1976@126.com.

lopost 和 Stambonlis[6] 提出修正控制界限的控制图;第 2 类是 Wardell 提出 Special Cause 控制图[7];第 3 类是由 Jiang 和 Tsui 提出的 ARMA 控制图[8].

上述控制图方法建立在过程的时间序列模型基础上,即首先对过程模型进行识别和参数估计.在已知模型的基础上,对过程实施统计过程控制,保证过程处于统计控制状态,并对过程的失控状态进行预警.因此,这些方法都需要大量的样本支持,在需要迅速对过程实施监控或小样本条件下,建立在模型基础上的控制图方法显得无能为力.本文提出一种基于贝叶斯动态模型的自相关控制图,在要求对过程实施迅速监视时,有很强的优势.

2 贝叶斯动态模型的自相关控制图

2.1 过程模型

对一个生产过程,可用贝叶斯动态模型理论中的线性模型[9]来刻画过程的质量波动:

观测方程:

$$y_t = u_t + v_t \qquad v_t \sim N(0, V_t) \tag{1}$$

状态方程:

$$u_t = u_{t-1} + w_t \qquad w_t \sim N(0, W_t) \tag{2}$$

初始信息:

$$u_0 \mid D_0 \sim N(f_0, C_0) \tag{3}$$

其中,y_t 为 t 时刻过程质量特性的观测值;u_t 为 t 时刻过程质量特性的水平;v_t 为 t 时刻观测误差项,假定服从正态分布;w_t 为 t 时刻状态误差项,假定服从正态分布;D_0 为 t 时刻先验信息集合;f_0 为过程质量特性 u_0 的一个估计值;C_0 为关于水平 m_0 的一种不确定性估计.还假定对任意时刻 t 和 s,当 $t \neq s$ 时,v_t 和 v_s,w_t 和 w_s 相互独立,且 v_t 和 w_s 相互独立.

当观测误差方差和状态误差方差都是常数时的模型称为常量模型.在一般情况下,对一个稳态生产过程,可设定方程(1)和方程(2)的方差为常数,即 $V_t = V, W_t = W$,并设 $r = W/V$.

当观测误差方差 V 未知时,令 $\varphi = V^{-1}$,称 φ 为精度.根据贝叶斯共轭推断理论,可设初始信息 D_0 时,精度 φ 的先验分布为

$$(\varphi \mid D_0) \sim \Gamma(n_0/2, d_0/2) \tag{4}$$

其中,先验参数 n_0, d_0 都是已知的正常数.精度的均值 $E(\varphi \mid D_0) = n_0/d_0 = 1/S_0$,$S_0$ 是 V 的一个先验点估计.

2.2 模型修正方程

由贝叶斯动态预测模型理论,一步预测残差 e_t 的分布为 $e_t \sim T_{n_t}(0, Q_t)$.当 $n_t \to \infty$ 时,$e_t \sim N(0, Q_t)$.一般的情况下,T 分布的自由度大于 30 时,就可将 T 分布视为正态分布.预测残差的递推方程为

$$e_t = y_t - f_{t-1} \tag{5}$$

$$f_t = f_{t-1} + A_t e_t = A_t y_t + (1 - A_t) f_{t-1} \tag{6}$$

其中,$A_t = R_t/Q_t$ 为 u_t 与 y_t 的回归系数,$R_t = C_{t-1} + W_t$(C_t 为 u_t 的后验正态分布的方差),$Q_t = C_{t-1} + W_t + V_t$;$n_t = n_{t-1} + 1$,$d_t = d_{t-1} + S_{t-1} e_t^2/Q_t$,$S_{t-1} = d_{t-1}/n_{t-1}$;$r = W_t/V_t$.

2.3 模型极限性与收敛性

定理 1 在封闭常量模型中,当 $t \to \infty$ 时,$A_t \to A$:

$$A = r(\sqrt{1 + 4/r} - 1)/2 \tag{7}$$

其中,$r = W/V$;$V > 0$.一步预测值为 $f_t = f_{t-1} + Ae_t = Ay_t + (1 - A) f_{t-1}$.

推论 1 在封闭常量模型中,当 $t \to \infty$ 时,$Q_t \to Q$:

$$Q = V/(1 - A) \tag{8}$$

由式(5)、式(6)可得 $y_t - y_{t-1} = e_t - (1 - A_{t-1}) \cdot e_{t-1}$ 的极限形式为

$$y_t - y_{t-1} = a_t - \delta a_{t-1} \qquad a_t \sim N(0, Q) \tag{9}$$

其中 $\delta = 1 - A$.这种形式为时间序列模型 ARIMA(0,1,1).

2.4 自相关控制图的建立

当 $t \to \infty$ 时,观测值的极限形式为 ARIMA(0,1,1),随机序列 $e_t \sim N(0, Q)$ 是零均值且相互独立的随机序列.因此,可根据休哈特单值控制图,对一步预测的残差序列进行监视.理论上,e_t 的均值为 0,从而可得 e_t 控制图的上下界限.

上界限(UCL):$K\sqrt{Q}$

下界线(LCL):$-K\sqrt{Q}$

其中 K 为控制系数,$K = U_{2/\alpha}$,一般取为 3.

当过程模型不是 IMA(1,1),而是其它 ARIMA 模型时,根据大量研究表明[10],EWMA 预测仍是最优一步预测值的较好近似,从而基于贝叶斯模型的一步预测自相关控制图不失为较好的一步预测控制方法.由于封闭常量模型的收敛性,只要获得大约 30 的观测值时,就可将获得的 Q_t 作为稳态值 Q,并延长作为以后的控制界限.

3 参数的讨论

3.1 先验 f_0, C_0 影响分析

初始信息(f_0, C_0)反映了 $t = 0$ 时刻,主观对过程状态的一个估计,可包括工程判断、工程知

识、主观意见、历史过程或相似过程,它是创建质量控制图的基础. f_0 代表对过程水平的估计,估计越准确,则对建模越有利. 由预测方程(6)可知,随着新观测值出现,f_0 的影响会变小.

C_0 是过程水平 m_0 估计的方差. 由于常量模型的收敛性质,C_t 很快趋近于 0. 因此,只要 $C_0>0$,C_0 的取值对控制图的性能没有影响,这一点可以从图 1 得出. C_0 越大,表示对过程开始时的确定性越小.

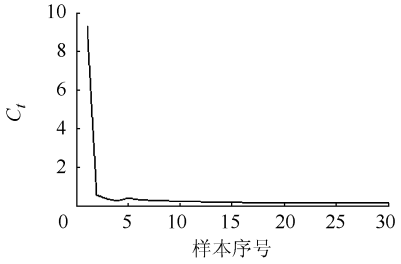

图 1　C_t 的衰减图

3.2　先验 S_0 影响分析

一般情况下,取 $n_0=1$,$d_0=S_0$. S_0 是观测方差 V 的先验估计,由封闭常量模型的极限特性,先验方差很快趋近于平稳值. 图 2 为仿真分析中,分别取 $S_0=0.5,1,3$ 时,Q 的比较图. 由图 2 可见,S_0 的选择对 Q_t 趋近稳定值的快慢是有影响的. 在工程背景下,一般保证 S_0 是理论值的 10 倍以内是完全可以的. 在这种条件下,经过贝叶斯估计 Q_t 都会在 30 个观测值内达到稳定值.

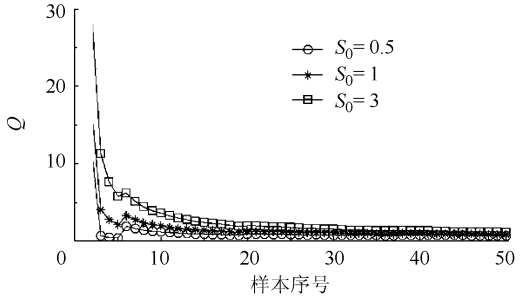

图 2　Q 在不同 S_0 下的比较图

3.3　参数 r 的选择

当 $t\to\infty$ 时,观测值的极限形式为 ARIMA(0,1,1)时,基于贝叶斯动态模型的残差是相互独立的. 因此,为了有效使用控制图,过程符合模型 AR(1):$y_t=\varphi y_{t-1}+\varepsilon_t$,$\varphi$ 为强自相关参数时,EWMA 能给出较好的预测,同时考虑到残差控制的性能,通过统计模拟给出参数 r 的参考原则为:当 $\varphi\geqslant 0.9$ 时,$5\leqslant r<8$;当 $0.8\leqslant\varphi<0.9$ 时,$3\leqslant r<5$;φ 值越大,参数 r 也越大.

4　性能分析及仿真

平均运行长度(ARL,Average Run Length)作为评价控制图性能的一个指标在学术界和实际工作中广泛采用. 对监视过程均值偏移的 ARL 计算,采用模拟仿真法,仿真次数为 10 000.

设过程均值变化以过程标准偏差的倍数来表示. 当 AR(1)模型的参数 $\varphi=0.8,0.9,0.95$ 时,贝叶斯自相关控制图的 ARL 值模拟结果见表 1.

表 1　不同 φ 值时,贝叶斯自相关控制图 ARL 值

偏移量 δ	$\varphi=0.8$	$\varphi=0.9$	$\varphi=0.95$
0	381.03	379.85	374.09
0.5	370.45	362.26	359.02
1	354.16	315.92	175.69
2	136.80	7.44	1.00
3	4.54	1.00	1.00

表 1 可见,过程稳态时,此自相关控制图与传统控制图的稳态 ARL 相同. 但这种控制图对小偏移变化不灵敏,对大偏移较灵敏. 表 2 为 $\varphi=0.8,0.9,0.95$ 时,基于过程模型的自相关控制图的 ARL 值. 对比两表可以看出,两者都存在对过程均值小偏移不灵敏的弱点. 随着系数 φ 的增大,基于贝叶斯自相关控制图对大偏移的检出能力,逐渐好于基于过程模型的自相关控制图.

表 2　不同 φ 值时,基于模型自相关控制图 ARL 值

偏移量 δ	$\varphi=0.8$	$\varphi=0.9$	$\varphi=0.95$
0	370.37	371.89	363.18
0.5	307.42	322.60	318.11
1	203.77	206.56	149.04
2	42.23	22.68	4.88
3	3.90	1.00	1.00

自相关过程的统计稳态是指过程期望不随时间的推移而变化,且 2 个时刻的线性相关性只与时间的间隔有关[11],因此,可由时间序列中 ARMA 模型来描述. 假设某质量过程处于稳态,且可由 AR(1)模型描述为 $y_t=\mu+0.9(y_{t-1}-\mu)+\varepsilon_t$. 不失一般性,取 $\mu=0$,$\varepsilon_t\sim N(0,1)$. 通过 Matlab

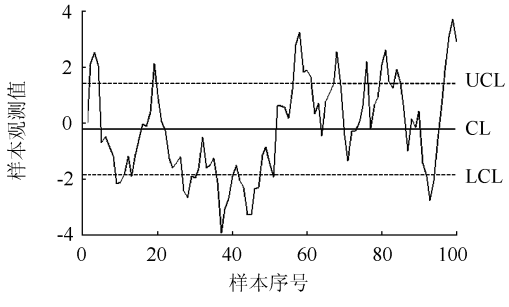

图 3　传统单值 X 控制图

产生100个仿真数据.建立传统休哈特型单值控制图,如图3所示.图中显示有许多点都超出控制界限,与过程稳态不相符.因此,上述判断都属误判,虚发报警.

对数据构建贝叶斯自相关控制图(图4),选定$r=5$,这是根据工程经验和主观判断得出所监视过程符合AR(1)($\varphi\approx0.9$)模型.控制图的初始值为$f_0=0$,$C_0=10$,$d_0=1$,$n_0=1$.由于封闭常量模型的收敛性,在大约30个观测值后,就可以获得稳态的Q值,可延长作为控制界限,建立贝叶斯自相关控制图,对其余70个观测值进行监视.对于受控过程,自相关控制图显示过程一直处于统计受控状态,没有异常报警点.

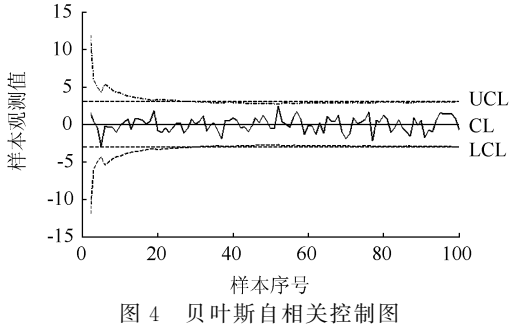

图4 贝叶斯自相关控制图

若采用基于模型的自相关控制图方法,通过成熟的软件和方法需要用100个观测值进行模型识别和参数估计.在100个观测值后,才能建立控制图.这样建立的自相关控制图与贝叶斯自相关控制图从理论上性能很接近.若减少样本数量(例如30个观测值)进行模型识别和参数估计,由于样本不足,得出模型及参数的误差很大,或者是错误的,不能建立有效控制图对过程实施监视.因此,综合考虑先验信息与客观数据的贝叶斯自相关控制图,可以快速建模,对过程的稳态、失控状态进行控制.

5 结束语

在生产过程中,只依据客观数据来监视质量波动常常是被动的,又需要大量样本支持.而操作者或质量工程师对过程的主现经验和判断很有价值,从某种意义上说是概率的一部分,是很重要的信息,应充分利用.本文引入的贝叶斯动态模型预测理论,则是充分建立在客观数据和主观信息有机结合基础上,快速建立起对过程质量监视的方法.避免了常规自相关控制图,在收集到存在大量自相关现象的数据,进行模型识别后,才能建立控制图的弊端.研究表明,贝叶斯自相关控制图对AR(1)($\varphi>0.8$)模型有很好的稳态性能和较好的预警能力.

基于贝叶斯动态模型的自相关控制图既克服了使用传统控制图出现虚警告过多的问题,又避免了一般ARIMA模型识别和参数估计的复杂计算,在现代质量控制中具有良好的应用前景.

参考文献(References)

[1] Montgomery D C. Introduction to statistical quality control [M]. 3rd ed. New York:John Wiley & Sons,1996

[2] Awin R W, Lee S J. The effects of autocorrelation and outliers on two-sided tolerance limits[J]. Technometrics,1998,40:24-38

[3] Atienza O O, Tang L C, Ang B W. A SPC procedure for detecting level shifts of autorrelated process[J]. Journal of Quality Technology,1998,30:340-351

[4] Dyer J N, Adams B M, Conerly M D. The reverse moving average control chart for monitoring auto-correlated process [J]. Journal of Quality Technology,2003,35:139-152

[5] Wang F K. A simple data transformation of auto-correlated data for SPC[J]. International Journal of Production Research,2005, 43(3):981-989

[6] Vasilopoulos A V, Stamboulis A P. Modification of control chart limits in the resence of data correlation[J]. Journal of Quality Technology,1978,10(1):20-30

[7] Wardell D G, Moskowitz H, Plantz R D. Run length distribution of special-cause control charts for correlated observations[J]. Technometrics,1994,36:3-17

[8] Jiang Wei, Tsui K L. A new SPC monitoring method: the ARMA chart [J]. Technometrics,2000,42(4):399-410

[9] 张孝令,刘福升,张承进,等. 贝叶斯动态模型及其预测[M]. 济南:山东科学技术出版社,1992
Zhang Xiaoling, Liu Fusheng, Zhang Chengjin, et al. Bayesian forecasting and dynamic models[M]. Jinan: Shandong Science and Technology Press, 1992(in Chinese)

[10] Cox D R. Prediction by exponentially weighted moving average and related methods[J]. Journal of the Royal Statistical Society,1961,23:414-412

[11] 孙静. 自相关过程的自相关控制图[J]. 清华大学学报,2002,42(6):735-738
Sunjing. Residual charts for autocorrelated processes[J]. Journal of Tsinghua University, 2002,42(6):735-738(in Chinese)

(责任编辑:文丽芳)