MEWMA 控制图 ARL 计算及参数优化

杜福洲 唐晓青 孙 静

(北京航空航天大学 机械工程及自动化学院, 北京 100083) (清华大学 管理科学与工程系, 北京 100084)

摘 要: 平均运行链长(ARL)是控制图性能评价的一个重要指标,深入研究 ARL 是控制图参数优化、性能评价和实际应用的重要工作. 在采样数据标准化处理的基础上,应用马尔可夫链进一步研究了多元指数移动平均(MEWMA)控制图中 ARL 计算的数学模型,采用 Matlab 实现了算法,就算法的收敛性及计算结果进行分析. 以 ARL 为优化指标,给出了 MEWMA 控制图参数优化的具体步骤、计算实例和参数优化结果.

关 键 词: 质量控制; 统计过程控制; 多变量控制系统; 马尔可夫过程; 优化控制系统

中图分类号: TH 165$^+$.4; C 931.1

文献标识码: A **文 章 编 号**: 1001-5965(2006)08-0974-05

ARL computation and parameters optimization for MEWMA control chart based on the Markov chain

Du Fuzhou Tang Xiaoqing

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Sun Jing

(Department of Management Science and Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Average run length(ARL) is one of the important indices in performance evaluation on quality control charts, the research on which will set up the foundation of quality control chart parameters optimization, performance evaluation and application. An ARL computation model based on the Markov chain was studied for the multivariate exponentially weighted moving average(MEWMA) control chart, the Matlab program was designed and implemented. The convergence of the algorithm and the performance of the MEWMA control chart in different conditions through the analysis of the computation results were analyzed. A procedure for parameters optimization of MEWMA control chart was proposed. An computation example and some optimization results was presented.

Key words: quality control; statistical process control; multivariable control systems; Markov processes; optimal control systems

多元质量控制图是通过监控过程中多个变量来控制产品质量常用的方法,对于控制图的性能,学术界和实际应用中常用 ARL(Average Run Length)作为其评价指标. ARL 是指对给定的质量水平,控制图从开始应用到发出警报平均抽取的样本数. 理想的控制图为:过程受控时,ARL 值尽可能大;过程失控时,ARL 值尽可能小.

应用马尔可夫链来计算控制图 ARL 的思想最早是由 Brook 和 Evans 在 1972 年提出的[1],用于计算 CUSUM 控制图的 ARL,随后马尔可夫链法在单变量控制图 ARL 计算中得到了广泛的应用. Runger 和 Prabhu 将其扩展到多元质量控制图的性能评价中,并提出了基于马尔可夫链的多元指数移动平均(MEWMA)控制图 ARL 的计算方法[2]. 本文基于马尔可夫链的 MEWMA 控制图 ARL 计算的数学模型,采用 Matlab 平台,对该模

收稿日期: 2005-09-14

基金项目: 国家自然科学基金资助项目(50475005;70321001)

作者简介: 杜福洲(1976-),男,山东潍坊人,博士生, du_fuzhou@163.com.

型算法的实现及其收敛性进行了研究. 最后针对计算结果分析了 MEWMA 控制图的性能,以 ARL 为优化指标,提出了一种 MEWMA 控制图参数优化方法,给出了具体实现步骤、算例及优化结果.

1　MEWMA 控制图

设 t 时刻采样数据为 $\boldsymbol{X}_t=[x_{1t}\quad x_{2t}\quad\cdots\quad x_{pt}]$,$\boldsymbol{X}_t$ 服从多元正态分布,均值向量为 $\boldsymbol{\mu}_0$,协方差矩阵为 $\boldsymbol{\Sigma}_0$,p 为过程变量维数. 设 MEWMA 控制图的统计量为 $\boldsymbol{W}$,其表达式为[3]

$$\left.\begin{aligned}\boldsymbol{W}_t&=r\boldsymbol{Y}_t+(1-r)\boldsymbol{W}_{t-1}\\ \boldsymbol{W}_0&=\boldsymbol{0}\end{aligned}\right\}\tag{1}$$

式中,$\boldsymbol{Y}_t$ 为 t 时刻标准化处理后的过程采样值,即 $\boldsymbol{Y}_t=\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{X}_t-\boldsymbol{\mu}_0)$;$r$ 为 MEWMA 控制图的平滑系数. 当过程发生偏移时,其偏移系数为

$$\delta=(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\tag{2}$$

式中,$\boldsymbol{\mu}$ 为偏移后的均值向量. 已证明,MEWMA 控制图的 ARL 只与过程偏移系数有关,而与过程偏移后的均值向量无关[4];$\boldsymbol{Y}_t$ 与原始变量 $\boldsymbol{X}_t$ 具有相同的偏移系数,因此两者具有相同的 ARL.

MEWMA 控制图打点量为

$$Q_t=\boldsymbol{W}_t'\boldsymbol{\Sigma}_W^{-1}\boldsymbol{W}_t\tag{3}$$

式中,$\boldsymbol{\Sigma}_W=[r/(2-r)]\boldsymbol{\Sigma}_Y$;$\boldsymbol{\Sigma}_Y$ 为 $\boldsymbol{Y}_t$ 的协方差矩阵,其为单位阵. 所以式(3)变为

$$Q_t=b\parallel\boldsymbol{W}_t\parallel^2\tag{4}$$

式中,$b=r/(2-r)$;当 $Q_t>H$ 时,表示过程出现异常;H 为给定初始 ARL 时的控制线,通常取此时的平均运行链长 $L_0=200$,即控制图的置信度为 0.005.

2　基于马尔可夫链的 ARL 计算

2.1　数学模型

利用马尔可夫链的无后效性,将 MEWMA 控制图的打点过程近似为一个马尔可夫过程,过程受控时有以下定理[2]:

定理 1　对于给定 $\parallel\boldsymbol{W}_1\parallel$,$\parallel\boldsymbol{W}_2\parallel$,…,$\parallel\boldsymbol{W}_t\parallel$ 的 $\boldsymbol{W}_t$,其概率分布函数为球面 $\boldsymbol{S}(\parallel\boldsymbol{W}_t\parallel)$,其中 $\parallel\boldsymbol{W}_t\parallel$ 为半径.

定理 2　$\boldsymbol{Q}_t$ 打点过程可近似为马尔可夫链.

选用变量 $q_t=\parallel\boldsymbol{W}_t\parallel$ 作为打点量,由定理 2 可知 q_t 可近似为马尔可夫链,其上控线 U_{cl} 为

$$U_{\mathrm{cl}}=b^{-1/2}H^{1/2}\tag{5}$$

对于转换后的 MEWMA 控制图的合格区域可以看作半径为 U_{cl} 的超球体,将该超球体划分为 $m+1$ 个部分,每部分代表一个马尔可夫链的状态,表示为 S_i,$i=0,1,\cdots,m$,S_0 为半径 $r=0.5g$ 的超球体,而 $S_i(i=1,2,\cdots,m)$ 均为厚度为 g 的超球壳,$g=2U_{\mathrm{cl}}/(2m+1)$. 图 1 为受控状态下,$p=2$ 时 q_t 马尔可夫链状态空间划分示意图. 当 $\parallel\boldsymbol{W}_t\parallel>U_{\mathrm{cl}}$ 时,过程失控,设此时的状态 S_{m+1} 为马尔可夫链的吸收态,这样 MEWMA 控制图的 ARL 就转化为该马尔可夫链从非吸收态转移到吸收态所经历状态的平均次数.

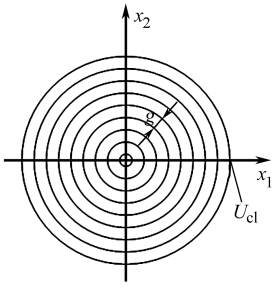


图 1　受控状态马尔可夫链状态空间划分示意图

当过程均值发生偏移时,因为 MEWMA 控制图的运行链长只与过程偏移系数 δ 有关,不失一般性,设其均值偏移向量为

$$\left.\begin{aligned}\boldsymbol{\mu}-\boldsymbol{\mu}_0&=\delta\boldsymbol{e}\\ \boldsymbol{e}&=[1\quad 0\quad\cdots\quad 0]\end{aligned}\right\}\tag{6}$$

将 $\boldsymbol{W}_i$ 分为 2 部分,均值为 δ 的一维随机向量 $\boldsymbol{W}_{t1}$ 和均值为 0 的 $p-1$ 维的随机向量 $\boldsymbol{W}_{t2}$,即

$$q_t=\parallel\boldsymbol{W}_t\parallel=\left(\boldsymbol{W}_{t1}^2+\boldsymbol{W}_{t2}'\boldsymbol{W}_{t2}\right)^{1/2}\tag{7}$$

q_t 近似为一个二维的马尔可夫链,此时状态空间的划分可以用一个二维的坐标系来表示,如图 2 所示,其中 $\boldsymbol{W}_{t1}$ 为马尔可夫链 $\boldsymbol{W}_{t1}$ 的状态划分,$\parallel\boldsymbol{W}_{t2}\parallel$ 为马尔可夫链 $\boldsymbol{W}_{t2}$ 的状态划分.

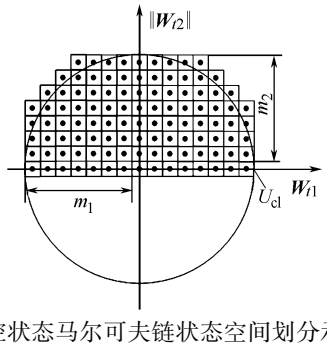


图 2　失控状态马尔可夫链状态空间划分示意图

$\boldsymbol{W}_{t1}$ 所对应的马尔可夫链,沿着 $\boldsymbol{W}_{t1}$ 轴的方向,将受控区域 $[-U_{\mathrm{cl}},+U_{\mathrm{cl}}]$ 划分为 $2m_1+1$ 个等距离的区间,每个区间代表马尔可夫链的一个状态,设其状态空间为 h_i,$i=1,2,\cdots,2m_1+1$. $\boldsymbol{W}_{t2}$ 所对应的马尔可夫链,参照受控状态下马尔可夫链状态空间的划分方法,设其状态空间为 v_j,$j=0,1,2,\cdots,m_2$. q_t 所对应的二维马尔可夫链的状态空间

为

$$S_{i,j}\left(\sqrt{h_i^2+v_i^2}\leqslant U_{\mathrm{cl}}\right) \tag{8}$$

式中,$i=1,2,\cdots,2m_1+1$,$j=0,1,2,\cdots,m_2$. 设状态 $S_{i,j}\left(\sqrt{h_i^2+v_i^2}>U_{\mathrm{cl}}\right)$ 为 q_t 所对应的马尔可夫链的吸收态,这样失控状态下 MEWMA 控制图的 ARL 就转化为该二维马尔可夫链从非吸收态转移到吸收态所经历状态的平均次数.

2.2 状态转移概率矩阵计算

过程受控时,设马尔可夫链由状态 i 到状态 j 的转移概率为 $p(i,j)$,则

$$p(i,j)=P\{(j-0.5)^2g^2/r^2<\chi^2(p,c)<(j+0.5)^2g^2/r^2\} \tag{9}$$

式中 $c=[(1-r)ige/r]^2$

当 $j=0$ 时

$$p(i,0)=P\{\chi^2(p,c)<(0.5)^2g^2/r^2\} \tag{10}$$

过程失控时,设 $\boldsymbol{W}_{t1}$ 所对应的马尔可夫链从状态 i 到状态 j 的转移概率为 $h(i,j)$,则

$$\begin{aligned}h(i,j)&=P(\boldsymbol{W}_{t1}\text{ 状态 }j\mid \boldsymbol{W}_{t-1,1}\text{ 状态 }i)=\\&P\{-U_{\mathrm{cl}}+(j-1)g<rY_{t1}+(1-r)W_{t-1,1}<-U_{\mathrm{cl}}+jg\mid W_{t-1,1}=c_i\}=\\&P\{(-U_{\mathrm{cl}}+(j-1)g-(1-r)c_i)/r-\delta<Y_{t1}-\delta<(-U_{\mathrm{cl}}+jg-(1-r)c_i)/r-\delta\}\end{aligned} \tag{11}$$

式中 $c_i=-U_{\mathrm{cl}}+(i-0.5)g$

参照受控状态下,马尔可夫链状态转移概率矩阵的计算方法,设 $\boldsymbol{W}_{t2}$ 所对应的马尔可夫链从状态 i 到状态 j 的转移概率为 $v(i,j)$,则

$$v(i,j)=P\{(j-0.5)^2g^2/r^2<\chi^2(p-1,c)<(j+0.5)^2g^2/r^2\} \tag{12}$$

式中 $c=[(1-r)ige/r]^2$

当 $j=0$ 时

$$v(i,0)=P\{\chi^2(p-1,c)<(0.5)^2g^2/r^2\} \tag{13}$$

因为 $\boldsymbol{W}_{t1}$ 与 $\boldsymbol{W}_{t2}$ 相互独立,则由 $\boldsymbol{W}_{t1}$ 与 $\boldsymbol{W}_{t2}$ 组成的二维马尔可夫链 $\{\boldsymbol{W}_{t1},\|\boldsymbol{W}_{t2}\|\}$ 从状态 (i_x,i_y) 到状态 (j_x,j_y) 的概率为

$$p[(i_x,i_y),(j_x,j_y)]=h(i_x,j_x)\cdot v(i_y,j_y) \tag{14}$$

2.3 ARL 计算

设 $\boldsymbol{P}$ 为 q_t 所对应的马尔可夫链状态转移概率矩阵,其计算方法如2.2节所述,设 $\boldsymbol{P}$ 的维数为 p_m,$\boldsymbol{E}'=(1,1,\cdots,1)$ 为所有元素为1的 p_m 维列向量,$\boldsymbol{I}$ 为 $p_m\times p_m$ 维的单位矩阵,$\boldsymbol{S}_m$ 为 p_m 维的行向量,其中有一个对应马尔可夫链初始状态的元素为1,其余元素为0. 则 MEWMA 控制图 ARL 计算公式为

$$L=\boldsymbol{S}_m(\boldsymbol{I}-\boldsymbol{P})^{-1}\boldsymbol{E} \tag{15}$$

式中,L 为控制图的平均运行链长. 如果 N 表示控制图的运行链长,则 MEWMA 控制图运行链长概率分布函数为

$$P(N>n)=\boldsymbol{S}_m\boldsymbol{P}^n\boldsymbol{E} \tag{16}$$

3 算法实现、结果分析及收敛性

3.1 算法实现及结果分析

基于上述 MEWMA 控制图 ARL 计算的数学模型,采用 Matlab 平台实现了该算法,给出了计算程序,其输入参数包括过程变量维数 p、偏移系数 δ(过程受控时,$\delta=0$)、平滑系数 r、控制线 H、状态空间计算参数 m(对于存在过程偏移时,取 $m_1=m_2=m$). 与 Lowry 等人采用仿真法、Rigdon 采用积分法、Runger 和 Prabhu 采用基于马尔可夫链的 Fortran 法的计算结果[2]进行了比较,计算条件均为 $p=4$,$r=0.1$,$H=12.73$,计算结果比较见表1.

表1 不同方法 MEWMA 控制图 ARL 计算结果比较

δ	仿真法	积分法	马尔可夫链法	
			Fortran	Matlab
0.0	201.00		199.41	198.7442
0.5	34.70	34.50	33.95	33.9554
1.0	12.10	12.00	12.02	12.0222
1.5	7.20		7.17	7.1659
2.0	5.18	5.15	5.16	5.1620
3.0	3.14	3.39	3.41	3.4058

表1为相同条件下,采用不同方法时 MEWMA 控制图 ARL 计算结果的比较,通过比较可以看出,各种方法的计算结果基本相同,说明本文采用的算法是正确的、有效的. 利用该程序,绘制了过程受控时 $L_0=200$,即置信度 $\alpha=0.005$,MEWMA 控制图中 H 与 p 之间的关系图,如图3所示.

图3给出了不同变量维数控制图控制线 H 的取值,为 MEWMA 控制图参数优化设计及应用提供了参考. 从图3可以看出,H 随着 p 的增大而增大,在相同的维数时,r 的取值越大,其控制线

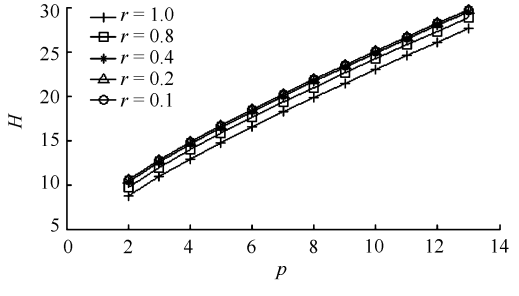


图3 MEWMA 控制图 H 与 p 关系图

也越大.

当过程均值发生偏移时,MEWMA 控制图 L 与平滑系统 r 之间的关系如图 4 所示.

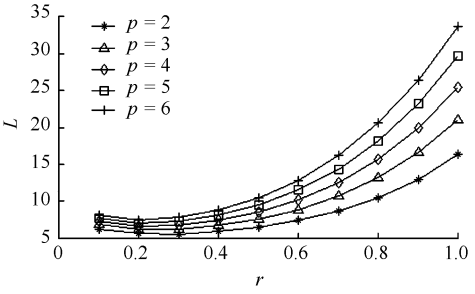


图 4　$\delta=1.5$ 时 L 与 r 的关系图

从图 4 可以看出,当过程均值发生偏移($\delta=1.5$)时,MEWMA 控制图的 r 存在一个最优值,当 r 取得最优值时,控制图对于该过程均值偏移最为敏感. 当 p 一定时,对于不同过程均值偏移,其 r 最优值不同. 通过对计算结果分析可以看出 MEWMA 对于持续小偏移的检出能力比 HotellingT^2 控制图要强,随着 δ 的增大其检出能力变得不那么明显,当 $\delta=3$ 时,2 种控制图的检出能力相当,而对于更大的过程偏移,HotellingT^2 控制图的优势反而渐渐明显. $p=4$ 时,MEWMA 控制图的 ARL 与 δ 的关系如图 5 所示.

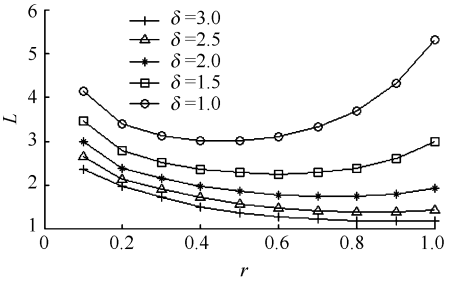


图 5　$p=4$ 时 L 与 r 的关系图

3.2　算法收敛性

算法收敛性主要是指在给定过程 δ,p,r 的情况下,MEWMA 控制图 ARL 随马尔可夫链状态空间计算参数 m 的收敛情况. 下面分 2 种情况对该算法的收敛性进行研究,即过程受控与过程失控.

通过分析可知,该算法的收敛性与过程的维数基本无关,因此选取 $p=3$,$r=0.4$ 作为示例进行研究,在此条件下 MEWMA 控制图 ARL 随马尔可夫链状态空间计算参数 m 的收敛过程如图 6 所示.

当过程均值发生偏移时,选取 $p=3$,$\delta=1.5$,$r=0.4$,在此条件下 MEWMA 控制图 ARL 随马尔可夫链状态空间计算参数 m 的收敛过程如图 7 所示.

由图 6、图 7 可以看出,本文采用的算法具有较好的收敛性,对于一般的 p,$m=10$ 基本能够得到较为理想的结果.

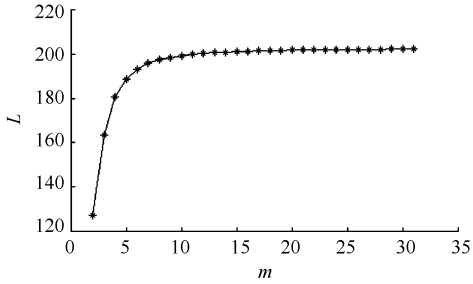


图 6　过程受控时 ARL 收敛曲线

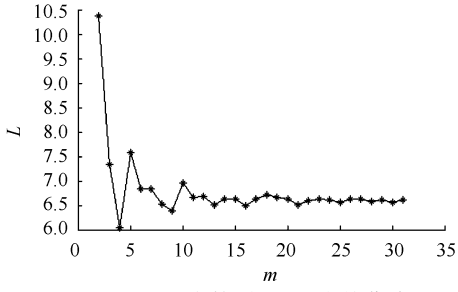


图 7　过程失控时 ARL 收敛曲线

4　MEWMA 控制图参数优化

控制图的实际应用中,一般过程 p,α 以及要监测的 δ 是已知的,根据上述的 MEWMA 控制图 ARL 的计算程序,以 ARL 为优化依据,提出了一种 MEWMA 控制图参数优化方法,其优化参数主要包括 H 和 r,其主要步骤为:

1) 根据控制图的置信度计算出过程受控时控制图的 $L_0=1/\alpha$. 通常选取 $\alpha=0.005$,即 $L_0=200$.

2) 令过程偏移系数 $\delta=0$,采用 ARL 计算程序计算出不同平滑系数下的控制线 H,通常选取 $r=0.1,0.2,\cdots,0.9,1.0$,然后采用线性回归方法求出控制图 H 与 r 之间关系式.

3) 选定要监测的过程偏移系数 δ,根据控制线与平滑系数的关系式,计算出 r 与控制图在此偏移下 ARL 的关系式.

4) 由 r 与控制图在此偏移下的 ARL 的关系式,寻找 L 最小的那一点,求出此点所对应的 r 及 H,即为该条件下 MEWMA 控制图的最优参数.

根据以上所提出的 MEWMA 控制图参数优化方法,采用 Matlab 编制了相应的 MEWMA 控制图参数优化程序,其输入参数包括:p,L_0,δ,m,输出参数为 H,r 以及 L. 下面以 $p=3$,$L_0=200$,$\delta=1.5$为例,采用编制的优化程序,对 MEWMA 控制图进行了参数优化,其步骤如下:

1) 令 $\delta=0$,采用 ARL 计算程序计算出不同平滑系数下的 H,选取 $r=0.1,0.2,\cdots,0.9,1.0$,然后采用线性回归方法求出 H 与 r 之间的关系式为

$$H = 9.0156 + 26.2341r - 79.2579r^2 + 123.8555r^3 - 96.4895r^4 + 29.4846r^5 \quad (17)$$

其拟合程度如图 8 所示.

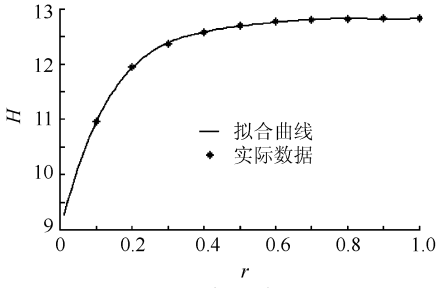


图 8 H 与 r 关系图

2) 选取马尔可夫链状态空间计算参数 $m=m_1=m_2=10$,根据步骤 1) 中计算出的 H 与 r 的关系式(17)计算出 r 与控制图在此偏移下的 ARL ($L_{1.5}$)的关系式,如图 9 所示.

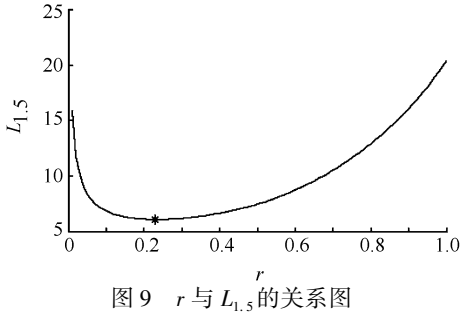


图 9 r 与 $L_{1.5}$ 的关系图

3) 选取图 9 中 ARL 的最小值点 $L=6.1266$,该点对应 $r=0.23$,然后通过式(17)求出所对应的 $H=12.1126$. 由此得出 $p=3,\delta=1.5,L_0=200$ 时,MEWMA 控制图的参数优化结果为 $H=12.1126,r=0.23$.

此外本文还对不同维数下 MEWMA 控制图的参数进行了优化,其优化结果如表 2 所示.

表 2 $\delta=1.5,L_0=200$ 时 MEWMA 控制图参数优化结果

p	r	H	$L_{1.5}$
2	0.25	9.9602	5.5287
3	0.23	12.1126	6.1266
4	0.22	14.0785	6.5825
5	0.22	15.9623	6.9608
6	0.21	17.6916	7.2943

5 结束语

ARL 作为评价控制图性能的一个重要指标,对其进行深入的研究是控制图参数优化设计、性能评价及应用的重要工作. 本文针对 MEWMA 控制图,基于马尔可夫链的 ARL 计算的数学模型,采用 Matlab 对其算法实现、算法收敛性及计算结果进行了研究与分析,进而以 ARL 作为优化指标,提出了一种 MEWMA 控制图参数优化方法,该方法为 MEWMA 控制图的实际应用奠定了基础,同时也为其他多元质量控制图的性能评价、参数优化及实际应用提供了参考.

参考文献 (References)

[1] Brook D, Evans D A. An approach to the probability distribution of CUSUM length [J]. Biometrika, 1972, 59:539-549

[2] Runger G C, Prabhu S S. A Markov chain model for the multivariate exponentially weighted moving averages control chart [J]. Journal of the American Statistical Association, 1996, 91(436):1701-1706

[3] Lowry C A, Woodall W H. A multivariate exponentially weighted moving average control chart[J]. Technometrics, 1992, 34(1):46-53

[4] Rigdon S E. A double-integral equation for the average run length of a multivariate exponentially weighted moving average control chart [J]. Statistics & Probability Letters, 1995,24:365-373

[5] Prabhu S S, Runger G C. Designing a multivariate EWMA control chart[J]. Journal of Quality Technology, 1997, 29(1):8-15

[6] Sonesson C. Evaluations of some exponentially weighted moving average methods [J]. Journal of Applied Statistics, 2003, 30(10):1115-1133

(责任编辑:张 嵘)