

# 监控M/M/1排队系统的AEWMA控制图

赵 薇, 齐德全

长春理工大学数学与统计学院, 吉林 长春

收稿日期: 2025年7月21日; 录用日期: 2025年8月14日; 发布日期: 2025年8月22日

## 摘 要

监控排队系统的到达率、服务率或系统利用率等指标, 可以有针对性地增加或减少服务台, 提高服务效率, 预防系统拥塞, 避免资源浪费。从统计过程控制的角度, 给出了在线监控M/M/1排队系统的一般框架。将常用的相关的队列长度数据转化为独立的到达人数数据。基于似然比检验构建自适应的EWMA控制图(AEWMA\*), 根据漂移的估计量计算权重函数, 提高控制图的性能。引入相对平均指标(RMI)和平均样本量(ANOS)作为评价准则, 实验结果显示: 与对比方法相比, AEWMA\*具有较低的RMI值, 说明AEWMA\*控制图在不同漂移场景下均具备较快的异常响应速度和综合检测优势。

## 关键词

自适应控制图, M/M/1系统, 似然比检验, 平均运行长度

# AEWMA Control Chart for Monitoring the M/M/1 Queuing System

Wei Zhao, Dequan Qi

School of Mathematics and Statistics, Changchun University of Science and Technology, Changchun Jilin

Received: Jul. 21<sup>st</sup>, 2025; accepted: Aug. 14<sup>th</sup>, 2025; published: Aug. 22<sup>nd</sup>, 2025

## Abstract

Monitoring the arrival rate, service rate, or system utilization of the queuing system can increase or decrease service counters in a targeted manner, improve service efficiency, prevent system congestion, and avoid resource waste. From the perspective of statistical process control, a general framework for online monitoring of M/M/1 queuing systems is presented. Convert commonly used queue length data into independent arrival count data. Constructing an adaptive EWMA control chart (AEWMA\*) based on likelihood ratio test, calculating weight functions according to the estimated drift, and improving the performance of the control chart. By introducing the Relative Mean Index

(RMI) and Average Sample Size (ANOS) as evaluation criteria, the experimental results show that compared with the comparative method, AEWMA\* has a lower RMI value, indicating that the AEWMA\* control chart has faster abnormal response speed and comprehensive detection advantages in different drift scenarios.

## Keywords

Adaptive Control Chart, M/M/1 System, Likelihood Ratio Test, Average Run Length

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

排队问题广泛存在于通信[1]和服务行业[2]。排队系统的监测与优化是其中的重要研究内容[3]。在排队系统的监测问题中,常监控到达率、服务率和系统利用率的漂移。系统性能指标(如平均队列长度、平均等待时间)对到达率与服务强度的漂移极为敏感,持续微小的漂移即可导致稳态队长指数级攀升,进而触发系统崩溃[4]。这种固有的强非线性特性,使得对系统运行状态的实时监控成为保障其鲁棒性的核心需求。尤其在5G低时延通信、实时物流调度等高并发、低容忍场景中,异常漂移检测能力决定服务等级协议的合规性,并对运维成本的影响巨大[5]。

监控排队系统的现有文献讨论了两种抽样方案。一种是部分抽样方案[6]-[8],即只在顾客离开时观察队列长度。另一种是完全抽样方案[9],在顾客到达和离开的时刻观察队列长度和系统时间。Chen等人[8]给出了计算WZ控制图和nL控制图平均样本量的马尔可夫链方法。WZ控制图和nL控制图适合对较大漂移的监控。Chen和Zhou[9]提出了适合监控中小漂移的累积和(CUSUM)控制图。但CUSUM控制图的监控效果受到设计参数的影响,如果真实的到达率、服务率或系统利用率偏离设计参数较远,CUSUM控制图的表现很差。Qi等人[10]考虑到最新的数据可能携带更多的最新信息,提出了一种基于加权似然比检验(WLRT)的控制图监控系统利用率的漂移。但与指数加权滑动平均(EWMA)控制图类似,WLRT控制图的监控效果受到光滑参数的影响,较大的光滑参数对监控中等漂移更有效,较小的光滑参数对监控小漂移更有效。

综上,WZ控制图和nL控制图对监控大漂移更有效,CUSUM控制图和WLRT控制图对监控中小漂移更有效。而实际应用当中,过程发生的漂移是未知的。因此,本文基于似然比检验建立了一个新的自适应EWMA控制图(称为AEWMA\*控制图),在部分抽样下监控M/M/1排队系统的系统利用率的漂移。为提升监控方法对过程变化的适应能力,本文定义了新的权重函数,通过引入两个调节参数 $\theta$ 和 $k$ ,实现光滑参数值的动态调整,确保控制图对不同幅度的漂移均能有效识别。通过蒙特卡罗模拟实验,将AEWMA\*控制图与AEWMA控制图、EWMA控制图进行性能比较,同时计算RMI值来评估控制图的整体性能。结果显示,与对比方法相比,AEWMA\*具有较低的RMI值,说明AEWMA\*控制图的监控效果更好,整体性能更优异。

## 2. 变点模型

M/M/1排队系统假设顾客到达时间服从泊松分布,服务时间服从指数分布,只有1个服务台。令 $\lambda$ 、 $\mu$ 和 $\rho$ 分别表示系统的到达率、服务率和系统利用率,这里 $\rho = \frac{\lambda}{\mu}$ 。假设在某个未知时刻 $\tau$ ,系统利用率

$\rho$  从  $\rho_0$  变为  $\rho_1$ , 该变点模型等价于在每个监控时刻检验如下假设

$$\begin{cases} H_0: \rho = \rho_0, \\ H_1: \rho = \rho_1. \end{cases}$$

### 3. AEWMA 控制图

在排队论中, 若  $Q_i$  表示第  $i$  个顾客离开时的队列长度, 则  $Q_i$  和  $Q_{i-1}$  是相关的。如果忽略这种相关性, 直接利用队列长度数据建立控制图监控排队系统, 会使得监控效果不理想[9] [10]。因此引入以下变量, 即观察第  $i$  个服务时间段内的到达人数

$$A_i = Q_i - Q_{i-1} + 1 - I_{i-1}$$

其中  $I_{i-1} = \begin{cases} 1, & \text{若 } Q_{i-1} = 0 \\ 0, & \text{若 } Q_{i-1} \neq 0 \end{cases}$ 。由排队论得, 随机变量序列  $\{A_i\}$  是独立且同分布的, 其分布律为

$$P\{A_i = k\} = \frac{1}{1+\rho} \cdot \left(\frac{\rho}{1+\rho}\right)^k, k = 0, 1, 2, \dots$$

于是利用以上的到达人数数据基于似然比检验建立如下自适应控制图监控系统利用率  $\rho$  的漂移。假设在  $t$  时刻抽取  $n$  个样本  $A_t = (A_{t,1}, A_{t,2}, \dots, A_{t,n})$ , 由该样本得对数似然函数为

$$l_t(\rho) = \ln \rho \cdot \sum_{i=1}^n A_{t,i} - \ln(1+\rho) \cdot \left(n + \sum_{i=1}^n A_{t,i}\right)$$

从而得到如下似然比检验统计量

$$LRT_t = 2[l_t(\hat{\rho}_t) - l_t(\rho_0)] = 2n \left[ \hat{\rho}_t \cdot \ln \frac{\hat{\rho}_t(1+\rho_0)}{\rho_0(1+\hat{\rho}_t)} - \ln \frac{1+\hat{\rho}_t}{1+\rho_0} \right]$$

其中,  $\hat{\rho}_t = \arg \max_{\rho} l_t(\rho) = \frac{1}{n} \sum_{i=1}^n A_{t,i}$  是  $\rho$  的极大似然估计。进一步计算自适应的监控统计量

$$Z_t = (1 - g(\hat{\delta}_t))Z_{t-1} + g(\hat{\delta}_t)LRT_t \quad (1)$$

其中, 初始值  $Z_0 = 0$ ,  $g(\hat{\delta}_t)$  是权重函数, 这里

$$g(\hat{\delta}_t) = \begin{cases} \theta/4, & \hat{\delta}_t \leq k, \\ \theta/2, & \hat{\delta}_t \leq 2k, \\ \theta, & \hat{\delta}_t > 2k. \end{cases}$$

权重函数  $g(\hat{\delta}_t)$  中的  $\hat{\delta}_t = |\hat{\rho}_t - \rho_0|$  是漂移的估计量,  $\theta$  和  $k$  是调节参数。 $\theta$  与 EWMA 控制图的光滑参数类似。EWMA 控制图中光滑参数  $0 < \theta \leq 1$ , 通常可取值为 0.05、0.1 或 0.2。本文建议上面的自适应控制图中  $\theta = 0.2$ , 使得权重函数  $g(\hat{\delta}_t)$  根据漂移估计量的不同情况分别取值为 0.05、0.1 和 0.2。 $k$  作为漂移估计量大小的临界值, 可通过使 RMI 达到最小来确定, 使得该控制图有较好的综合性能。参考 Han 和 Tsung 等人[11]文献, RMI 指数定义为

$$\text{RMI} = \frac{1}{M} \sum_{l=1}^M \frac{\text{ARL}_{\Delta l} - \text{MARL}_{\Delta l}}{\text{MARL}_{\Delta l}}$$

其中,  $M$  表示漂移量分类情况的个数,  $\text{ARL}_{\Delta l}$  表示控制图在检测到漂移大小为  $\Delta l$  时的  $\text{ARL}_1$ ,  $\text{MARL}_{\Delta l}$

表示检测到漂移大小为  $\Delta I$  时所有控制图的  $ARL_1$  的最小值。当监控统计量  $Z_t$  大于预先设定的控制线  $CL$  时, AEWMA\*控制图发出过程失控的警报。

#### 4. 统计模拟

本小节通过统计模拟评估所提出的 AEWMA\*控制图的性能。构建了两种对比方案, 一种对比方案构建的是 EWMA 控制图, 另一种对比方案是采用贡平邨等人[12]提出的权重函数构建的 AEWMA 控制图。为了公平起见, 将平均运行长度(ARL)转换为平均样本量(ANOS), 同时调整不同控制图的控制线  $CL$ , 使  $ANOS_0$  值尽可能接近 370。通过蒙特卡罗模拟方法, 比较不同控制图的 ANOS 和 RMI。本节所有结果均基于 10,000 次重复实验得出。

在各种控制图方法的比较中, ARL 或 ANOS 是常用的标准。当监控过程受控(IC)时, 具有较大 IC ARL (简记为  $ARL_0$ )或 ANOS (简记为  $ANOS_0$ )的控制图比其他控制图具有更低的虚警率。当监控过程失控(OC)时, 具有较小 OC ARL (简记为  $ARL_1$ )或 ANOS (简记为  $ANOS_1$ )的控制图报警更早, 对过程漂移的检测能力优于其他控制图。为了评估不同控制图的整体性能, 比较 RMI 值。RMI 值被认为是所有相对效率值的平均值, RMI 值越小的控制图被认为整体性能越好[11]。

在生成 M/M/1 队列模拟数据时, 参考 Chen 等人[8]文献, 假设被监控的系统从一个空队列开始, 即  $Q_0 = 0$ 。AEWMA\*控制图的样本容量  $n = 5$ 。为了在实际应用中, 便捷地使用 AEWMA\*控制图, 表 1 提供了由二分法得出的  $\rho_0 = 0.3, 0.5, 0.7$  和  $\theta = 0.05, 0.1, 0.2$  时, 不同  $ANOS_0$  下 AEWMA\*控制图的控制线  $CL$ 。在对离散型数据的监控问题中, 有的控制图因为监控统计量取值比较分散而很难得到满足指定  $ANOS_0$  的控制线。从表 1 可以看出, 本文所提出的 AEWMA\*控制图针对不同的  $ANOS_0$  都能得出对应的控制线。

**Table 1.** The CL value of the simulated AEWMA\* control chart

**表 1.** 模拟 AEWMA\*控制图的控制线 CL 值

$\rho_0$	ANOS						
	200	300	370	500	800	1000	2000
$\theta = 0.05$							
0.3	0.9430	1.1810	1.3160	1.5130	1.8090	1.9310	2.2383
0.5	0.5875	0.6260	0.6485	0.6832	0.7387	0.7650	0.8337
0.7	0.7182	0.7428	0.7606	0.7950	0.8510	0.8742	0.9540
$\theta = 0.1$							
0.3	1.3360	1.6280	1.7850	2.0065	2.2988	2.4164	2.7327
0.5	0.6600	0.7170	0.7465	0.7900	0.8553	0.8825	0.9586
0.7	0.6860	0.6920	0.6970	0.7085	0.7308	0.7461	0.7972
$\theta = 0.2$							
0.3	1.8250	2.2200	2.4345	2.7070	3.0280	3.1600	3.5523
0.5	0.7792	0.8557	0.8957	0.9496	1.0264	1.0590	1.1524
0.7	0.6785	0.6965	0.7068	0.7200	0.7469	0.7606	0.8054

进一步, 在  $\tau = 0$  时, 分  $\rho_0 = 0.5$  和  $\rho_0 = 0.7$  两种情况, 假设到达率  $\lambda$  以不同的幅度变化导致系统利用率  $\rho$  发生相应的漂移, 模拟 AEWMA\*控制图、EWMA 控制图和 AEWMA 控制图的 ANOS 和 RMI。

在  $\rho_0 = 0.7$  时, 不同控制图的 ANOS 和 RMI 值见表 2。从表 2 可以看出, 不管  $\rho$  发生的是小漂移、中等漂移还是大漂移, AEWMA\* 控制图的 ANOS<sub>1</sub> 值都最小。例如, 当  $\rho = 0.7935$  时, AEWMA\* 与 AEWMA、EWMA-1、EWMA-2 和 EWMA-3 控制图对应的 ANOS<sub>1</sub> 分别为 150.139、231.60、194.65、208.69 和 226.28。从而 AEWMA\* 控制图的 RMI 值最小。模拟结果表明, AEWMA\* 控制图的监控效果在小、中、大漂移的场景下均要优于其他控制方法, 综合表现最好。

**Table 2.** Comparison of ANOS<sub>1</sub> when  $\rho_0 = 0.7$

**表 2.** 当  $\rho_0 = 0.7$  时 ANOS<sub>1</sub> 的比较

$\rho$	EWMA-1 控制图	EWMA-2 控制图	EWMA-3 控制图	AEWMA 控制图	AEWMA* 控制图
	$\theta = 0.05$	0.1	0.2		
	CL = 1.1687	1.377	1.6527	1.2811	0.70675
0.7	370.09	370.05	370.01	370.13	369.82
0.7177	321.63	328.40	334.61	331.47	301.204
0.7585	241.41	251.65	270.76	271.73	200.164
0.7935	194.65	208.69	226.28	231.60	150.139
0.85	145.32	158.61	180.02	180.49	100.5125
0.945	99.22	108.41	125.38	129.26	60.687
1.145	57.09	61.77	72.96	78.31	30.4265
2.2	22.37	20.77	21.77	24.64	10.2795
15	11.20	9.15	8.14	10.74	5.38
RMI	0.5983	0.6067	0.7321	0.8616	0.0000

当  $\rho_0 = 0.5$  时, 不同控制图的 ANOS 和 RMI 值如表 3 所示。当  $\rho$  发生小漂移时, AEWMA\* 控制图的 ANOS<sub>1</sub> 值较小。例如, 当  $\rho = 0.5263$  时, AEWMA\* 与 AEWMA、EWMA-1、EWMA-2 和 EWMA-3 控制图对应的 ANOS<sub>1</sub> 分别为 288.06、300.37、286.53、299.72 和 310.49。此外, 与 AEWMA 控制图相比, AEWMA\* 控制图在检测中、小漂移方面表现较好。例如, 当  $\rho = 0.7430$  时, AEWMA\* 与 AEWMA 控制图的 ANOS<sub>1</sub> 分别为 93.47、99.96。从 RMI 值来看, AEWMA\* 比 AEWMA 控制图的监控效果更好。从表 3 还可以看出, EWMA 控制图的表现受光滑参数  $\theta$  的影响。光滑参数  $\theta$  越小, EWMA 控制图对小漂移表现越好, 光滑参数  $\theta$  越大, EWMA 控制图对大漂移表现越好。

**Table 3.** Comparison of ANOS<sub>1</sub> when  $\rho_0 = 0.5$

**表 3.** 当  $\rho_0 = 0.5$  时 ANOS<sub>1</sub> 的比较

$\rho$	EWMA-1 控制图	EWMA-2 控制图	EWMA-3 控制图	AEWMA 控制图	AEWMA* 控制图
	$\theta = 0.05$	0.1	0.2		
	CL = 1.003	1.2195	1.5195	1.9015	0.89565
0.5	369.88	369.89	370.10	369.65	370.05
0.5263	286.53	299.72	310.49	300.37	288.06

续表

0.5895	170.08	183.87	203.56	200.32	183.77
0.6470	117.90	128.36	147.05	150.16	134.91
0.7430	76.83	83.80	97.22	99.96	93.47
0.9350	45.81	45.83	53.47	60.14	58.07
1.45	24.32	22.21	23.84	29.84	32.29
30	9.48	7.43	6.43	10.29	10.94
RMI	0.0814	0.0661	0.1476	0.2939	0.2672

## 5. 结论

本文提出了一个新的监控 M/M/1 排队系统的自适应 EWMA 控制图(AEWMA\*)。将似然比检验统计量和 EWMA 相结合,使得离散数据监控统计量连续化,从而易得控制线。兼顾 EWMA 的常用光滑参数和相对平均指标(RMI)构造权重函数,适应不同的漂移量,综合表现最好。统计模拟表明,当系统利用率发生漂移时,AEWMA\*控制图具有较好的表现。此外,在似然已知的情况下,本文所提出的方法较易推广至完全抽样方案或 M/M/s、M/G/1 等排队系统的监控问题。M/M/s 排队系统可看做变服务率的 M/M/1 系统,M/G/1 排队系统的队长序列仍是马尔科夫链。针对这两个系统,可以采用完全抽样方案根据排队论求似然函数(参考 Chen 和 Zhou 2015 [9]),然后再使用本文的方法建立自适应 EWMA 控制图。

## 基金项目

国家自然科学基金面上项目(12271271)。

## 参考文献

- [1] 薛建彬, 郁柏文, 徐小凤, 等. 边缘计算网络中基于排队论的通信和计算资源联合优化[J]. 计算机科学, 2024, 51(S2): 704-712.
- [2] 曹石强. 智能排队取号系统设计与实现[J]. 工程技术研究, 2025, 10(8): 207-209.
- [3] 高琦. 基于 PEWMA 控制图的排队系统的监测与优化[D]: [硕士学位论文]. 济南: 山东财经大学, 2024.
- [4] Reed, J. and Zwart, B. (2014) Limit Theorems for Markovian Bandwidth-Sharing Networks with Rate Constraints. *Operations Research*, **62**, 1453-1466. <https://doi.org/10.1287/opre.2014.1321>
- [5] Narayanan, S., Arora, R.K., Gangwar, S., Kandhasamy, J.P., Ratheesh, T. and Murali, K. (2024) Ultra-Low Latency Communication Technique for Augmented Reality Application in Mobile Edge Computing. *International Journal of Engineering Systems Modelling and Simulation*, **15**, 97-105. <https://doi.org/10.1504/ijesms.2024.138288>
- [6] Bhat, U.N. and Rao, S.S. (1972) A Statistical Technique for the Control of Traffic Intensity in the Queuing Systems M/G/1 and GI/M/1. *Operations Research*, **20**, 955-966. <https://doi.org/10.1287/opre.20.5.955>
- [7] Shore, H. (2006) Control Charts for the Queue Length in a G/G/S System. *IIE Transactions*, **38**, 1117-1130. <https://doi.org/10.1080/07408170600737336>
- [8] Chen, N., Yuan, Y. and Zhou, S. (2011) Performance Analysis of Queue Length Monitoring of M/G/1 Systems. *Naval Research Logistics (NRL)*, **58**, 782-794. <https://doi.org/10.1002/nav.20483>
- [9] Chen, N. and Zhou, S. (2015) CUSUM Statistical Monitoring of M/M/1 Queues and Extensions. *Technometrics*, **57**, 245-256. <https://doi.org/10.1080/00401706.2014.923787>
- [10] Qi, D.Q., Li, Z.H., Zi, X.M., et al. (2017) Weighted Likelihood Ratio Chart for Statistical Monitoring of Queueing Systems. *Quality Technology & Quantitative Management*, **14**, 19-30. <https://doi.org/10.1080/16843703.2016.1189184>
- [11] Han, D. and Tsung, F. (2006) A Reference-Free Cuscore Chart for Dynamic Mean Change Detection and a Unified Framework for Charting Performance Comparison. *Journal of the American Statistical Association*, **101**, 368-386.

<https://doi.org/10.1198/016214505000000556>

- [12] 贡平邨, 郑冰静, 郭宝才. 基于加权似然比检验的自适应 EWMA 控制图设计[J]. 系统科学与数学, 2023, 43(6): 1587-1611.