

具有两类优先权顾客的 M/M/1 排队的优化分析*

张怡通

(燕山大学经济管理学院, 秦皇岛, 066004)

徐秀丽*

(燕山大学理学院, 秦皇岛, 066004)

摘要: 本文研究了一个带有强占优先权和非强占优先权的 M/M/1 排队模型, 顾客分为三个优先等级, 第一类顾客享有强占优先权, 第二类顾客享有非强占优先权, 第三类顾客无优先权. 三类顾客具有不同的到达率, 当第一类顾客到达后它将打断正在接受服务的第二类或第三类顾客立即接受服务; 当第二类顾客到达时, 若系统中只有第三类顾客, 则此顾客必须等待当前服务完成, 才能接受服务, 否则排队等待. 同一类顾客遵循 FCFS 的排队规则. 利用补充变量法构造多维向量马尔可夫过程并对此排队系统的状态转移方程进行分析, 得到三类顾客队长联合分布的概率母函数, 进而得出了每类顾客各自的平均队长以及服务台被三类顾客占有和闲置的概率. 利用 Matlab 进行数值计算, 主要考察服务率的变化对系统中各类顾客平均队长的影响. 最后构建不同成本费用函数进行优化分析.

关键词: 优先权; 强占; 非强占; 母函数

中图分类号: O226

英文引用格式: ZHANG Y T, XU X L. Optimization analysis of M/M/1 queue with two types of priority customers [J]. Chinese J Appl Probab Statist, 2021, 37(5): 449-460. (in Chinese)

§1. 引言

优先权排队在现实生活中一直具有广泛的应用. 例如在医院中急诊病人的救治, 超市中的易腐烂商品的处理等都体现出了优先权的排队思想. Miller^[1] 利用矩阵几何解, 分别推导了带有强占型的和非强占型的 M/M/1 排队系统稳态分布的递推计算公式. Lee^[2] 分析了一个带有阈值的优先权轮询排队系统, 得到了队列长度和逗留时间的分布变换. Feng 等^[3] 进一步推广了文献 [2] 的轮询模型, 在两个队列分别加入阈值 M 和 N , 提出了一种计算服务完成时刻平稳队长分布的母函数的算法. Sleptchenko 等^[4] 研究了一个具有强占优先级的单服务台排队系统, 该排队系统可以用多维马尔可夫过程来描述, 通过矩阵分解, 提出了一种确定联合队长分布的递推方法. 张宏波和史定华^[5] 讨论了一个带有强占优先权的 M/M/1 排队模型, 该模型可以用一个拟生灭过程来描述, 证明了低优先权顾客的平稳队长和平稳逗留时间分别具有几何衰减和指数衰减的特性. 张宏波等^[6] 推广了文献 [5] 的模型, 考虑了一个低优先级顾客数量是有限并且是可缓冲的排队模型, 该排队模型可以用有限多个拟生灭过程来描述. Fajardo 和 Drekić^[7] 研究了一个强占累积优先级队列, 其中

*河北省自然科学基金项目 (批准号: A2019203313) 和河北省高等学校科学研究重点项目 (批准号: ZD2019079) 资助.

*通讯作者, E-mail: xxl-ysu@163.com.

本文 2019 年 11 月 26 日收到, 2020 年 7 月 13 日收到修改稿.

一个服务器处理带有 N 个优先级的顾客. 顾客到达系统后会以不同速率累积优先级, 在每一个服务瞬间, 服务台都会选择累积优先级最高的顾客进行服务, 最终得到每一类顾客的等待时间分布. Gao^[8] 考虑了一类具有强占优先权客户的重试队列系统. 如果到达的普通客户发现服务器正忙, 则会进入一个重试组, 只有位于重试队列前面的普通客户才能访问服务器. 利用马尔可夫链和辅助变量法得到了系统稳定的充要条件和平稳概率分布等性能指标, 最后对任意普通客户系统中的逗留时间进行了分析. 对于非强占型优先权的排队系统, 赵国喜等^[9] 利用补充变量法, 得到两类顾客队长联合分布的概率母函数, 并得出了系统被每类顾客占有和闲置的概率. 黄业文和邝神芬^[10] 提出了一种处理非强占有限优先权 $M/M/n/m$ 排队模型的方法, 通过引入控制完全优先权的参数 r 使其成为有限优先权, 这样考虑了不同优先级队伍的公平性, 使得网络系统在有限优先权下有较好的稳定性. 周杰等^[11] 考虑 CT 室具有非强占优先权病人和常规病人容量有限的排队系统. 运用矩阵几何解的方法获得平稳分布, 运用全概率分解技术获得处于等待队列中常规病人的等待时间和他被拒绝进入的概率. 最后通过数值算例, 分析可变参数对常规病人队列等待时间, 损失率和医疗资源利用率的影响.

在网络资源共享和信息传输系统中, 两类优先权的排队模型已经不足以模拟综合信息传输环境中的多种信息类型. Landry 和 Stavrakakis^[12] 引入了控制信息流的第三类信息流. 该类信息流包括关键性的网络控制流, 预留信息和一些一旦延迟就会造成整个传输系统崩溃的信息流. 虽然该类信息很少, 但是一旦到达就要立即处理, 一般赋予该类信息流以队列首优先权 (head-of-line priority, 简称 HoL). Liu 等^[13] 在此基础上研究了一个基于阈值的 $c-G/L/HoL$ 的轮询排队模型. 排队系统由 3 个 $M/M/1$ 队列和一个服务台组成, 采用阈值服务策略, 队列 1 和队列 2 的阈值分别为 1 和 N , 队列 3 无阈值. 利用 Kernel 方法导出了联合平稳队长分布函数的概率, 得到了平均队列长度和平均逗留时间. Alawneh^[14] 提出了一种基于递推公式方法的扩展算法, 该方法可用于求解具有一台服务器的三个强占优先级排队系统稳态概率的一般解. 在 5G 中三个不同的网络片以几乎独立的切片方式共享网络资源, 这些切片处在不同的优先级类中. 实际上, 需要优先调度算法来保证其他所有优先级较低的切片的服务级别. AlQahtani 和 Altamrah^[15] 针对来自多个 5G 网络片的数据流量, 提出了一种基于优先级的多调度轮询方案分析模型, 并从平均等待时间和利用率两个方面给出了性能指标.

通常, 为了促进消费, 挖掘顾客的消费潜能, 各企业会根据客户的消费习惯或消费水平自行定制顾客等级. 例如, 中国农村信用社规定在同一银行账户号下, 满足一年内活期存款月均余额达 50 万元或定期存款实时余额达 100 万元条件之一者, 即可申领 VIP 白金卡; 四川海底捞餐饮股份有限公司规定在两个月内, 海底捞消费积分达到 12000 分, 可以通过相应的渠道去申请会员黑卡. 而这些拥有会员卡的顾客, 在到达系统时可以无需排队等待, 立刻获得服务, 这类顾客即具有强占优先权. 此外, 某些顾客由于自身或外界原因, 在号码到达时不能及时到达获得服务, 导致过号现象发生. 针对这种情况, 商家们为了平衡普通排队顾客与过号顾客的需求, 会令错过叫号的顾客排队等待几个普通排队顾客后再为其提

供服务. 那么过号顾客具有非强占式优先权, 其余普通排队顾客则无优先权. 若到达为会员顾客, 无论服务台正在服务哪类顾客, 都需要立刻为其提供服务; 若到达为过号顾客, 则可排在其他普通排队顾客之前等待服务. 各类顾客的服务服从先到先服务规则. 基于此, 本文进行带有两类优先权的排队模型的分析 and 优化. 通过补充变量法构造一个四维马尔可夫过程, 建立了系统稳态概率满足的方程组. 利用母函数法得到服务台分别被三类顾客占有和闲置的概率以及各类顾客的平均队长. 通过具体数值分析研究服务率的变化对系统的影响, 并选取合理的成本费用函数模型, 得到系统的最大效益, 为系统的资源配置提供参考.

§2. 模型假设

- (i) 到达过程: 三类顾客的输入为泊松过程, 参数为 λ_i , $i = 1, 2, 3$.
- (ii) 服务过程: 服务台对三类顾客服务时间相同且服从指数分布, 参数为 μ .
- (iii) 系统有且只有一个服务台, 且系统容量无限.
- (iv) 排队规则: 第一类顾客较第二类和第三类顾客具有强占型优先权, 第二类顾客较第三类顾客具有非强占型优先权. 服务过程与输入过程相互独立.
- (v) 假设服务台在各队列之间的转换时间可以忽略.

设在时刻 t 系统中第 i 类顾客的队长为 $L_i(t)$, $i = 1, 2, 3$. 由于 $\{L_1(t), L_2(t), L_3(t), t \geq 0\}$ 不是一个马尔可夫过程. 引入变量 $\xi(t) = 0, 1, 2, 3$ 分别表示时刻 t 服务台处于闲置, 被第一类顾客占据, 被第二类顾客占据, 被第三类顾客占据的状态, 则 $\{L_1(t), L_2(t), L_3(t), \xi(t), t \geq 0\}$ 是一个四维马尔可夫过程. 其状态空间为

$$\Omega = \{(0, 0, 0, 0)\} \cup \{(l, i, j, 1) \mid l \geq 1, i, j \geq 0\} \cup \{(0, i, j, 2) \mid i \geq 1, j \geq 0\} \\ \cup \{(0, i, j, 3) \mid i \geq 0, j \geq 1\}.$$

§3. 系统的稳态分布

三类顾客构成了一个输入流参数为 $\lambda_1 + \lambda_2 + \lambda_3$ 的泊松流的 M/M/1 排队系统, 当系统负载 $\rho = (\lambda_1 + \lambda_2 + \lambda_3)/\mu < 1$ 时, 稳态分布存在. 记系统的稳态分布为

$$P_{lijk} = \lim_{t \rightarrow \infty} P = \{L_1(t) = l, L_2(t) = i, L_3(t) = j, \xi(t) = k\}, \quad (l, i, j, k) \in \Omega,$$

则系统状态转移方程组如下:

服务台闲置:

$$P_{0000}(\lambda_1 + \lambda_2 + \lambda_3) = \mu P_{1001} + \mu P_{0102} + \mu P_{0013}. \quad (1)$$

服务台被第三类顾客占据:

$$P_{0013}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{1011} + \mu P_{0112} + \mu P_{0023} + \lambda_3 P_{0000}; \quad (2)$$

$$P_{00j3}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{10j1} + \mu P_{01j2} + \mu P_{00(j+1)3} + \lambda_3 P_{00(j-1)3}, \quad j = 2, 3, \dots; \quad (3)$$

$$P_{0113}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \lambda_2 P_{0013}; \quad (4)$$

$$P_{01j3}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \lambda_2 P_{00j3} + \lambda_3 P_{01(j-1)3}, \quad j = 2, 3, \dots; \quad (5)$$

$$P_{0i13}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \lambda_2 P_{0(i-1)13}, \quad i = 2, 3, \dots; \quad (6)$$

$$P_{0ij3}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \lambda_2 P_{0(i-1)j3} + \lambda_3 P_{0i(j-1)3}, \quad i = 2, 3, \dots, j = 2, 3, \dots. \quad (7)$$

服务台被第二类顾客占据:

$$P_{0102}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{1101} + \mu P_{0202} + \mu P_{0113} + \lambda_2 P_{0000}; \quad (8)$$

$$P_{01j2}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{11j1} + \mu P_{02j2} + \mu P_{01(j+1)3} + \lambda_3 P_{01(j-1)2}, \quad j = 1, 2, \dots; \quad (9)$$

$$P_{0i02}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{1i01} + \mu P_{0(i+1)02} + \mu P_{0i13} + \lambda_2 P_{0(i-1)02}, \quad i = 2, 3, \dots; \quad (10)$$

$$P_{0ij2}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{1ij1} + \mu P_{0(i+1)j2} + \mu P_{0i(j+1)3} + \lambda_2 P_{0(i-1)j2} + \lambda_3 P_{0i(j-1)2}, \quad i = 2, 3, \dots, j = 1, 2, \dots. \quad (11)$$

服务台被第一类顾客占据:

$$P_{1001}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{2001} + \lambda_1 P_{0000}; \quad (12)$$

$$P_{10j1}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{20j1} + \lambda_1 P_{00j3} + \lambda_3 P_{10(j-1)1}, \quad j = 1, 2, \dots; \quad (13)$$

$$P_{1i01}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{2i01} + \lambda_1 P_{0i02} + \lambda_2 P_{1(i-1)01}, \quad i = 1, 2, \dots; \quad (14)$$

$$P_{1ij1}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{2ij1} + \lambda_1 P_{0ij2} + \lambda_1 P_{0ij3} + \lambda_2 P_{1(i-1)j1} + \lambda_3 P_{1i(j-1)1}, \quad i = 1, 2, \dots, j = 1, 2, \dots; \quad (15)$$

$$P_{l001}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{(l+1)001} + \lambda_1 P_{(l-1)001}, \quad l = 2, 3, \dots; \quad (16)$$

$$P_{l0j1}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{(l+1)0j1} + \lambda_1 P_{(l-1)0j1} + \lambda_3 P_{l0(j-1)1}, \quad l = 2, 3, \dots, j = 1, 2, \dots; \quad (17)$$

$$P_{li01}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{(l+1)i01} + \lambda_1 P_{(l-1)i01} + \lambda_2 P_{l(i-1)01}, \quad l = 2, 3, \dots, i = 1, 2, \dots; \quad (18)$$

$$P_{lij1}(\lambda_1 + \lambda_2 + \lambda_3 + \mu) = \mu P_{(l+1)ij1} + \lambda_1 P_{(l-1)ij1} + \lambda_2 P_{l(i-1)j1} + \lambda_3 P_{li(j-1)1}, \quad l = 2, 3, \dots, i = 1, 2, \dots, j = 1, 2, \dots. \quad (19)$$

为求解上述稳态方程组, 引入如下记号:

$$\begin{aligned}
 Q_{li}^1(z_3) &= \sum_{j=0}^{\infty} P_{lij1} z_3^j, \quad l = 1, 2, \dots, i = 0, 1, \dots; \\
 Q_{0i}^2(z_3) &= \sum_{j=0}^{\infty} P_{0ij2} z_3^j, \quad i = 1, 2, \dots; \\
 Q_{0i}^3(z_3) &= \sum_{j=1}^{\infty} P_{0ij3} z_3^j, \quad i = 0, 1, \dots; \\
 Q_l^1(z_2, z_3) &= \sum_{i=0}^{\infty} Q_{li}^1(z_3) z_2^i, \quad l = 1, 2, \dots; \\
 Q_0^2(z_2, z_3) &= \sum_{i=1}^{\infty} Q_{0i}^2(z_3) z_2^i, \quad Q_0^3(z_2, z_3) = \sum_{i=0}^{\infty} Q_{0i}^3(z_3) z_2^i, \\
 Q^1(z_1, z_2, z_3) &= \sum_{l=1}^{\infty} Q_l^1(z_2, z_3) z_1^l.
 \end{aligned}$$

定理 1 $\rho < 1$ 时, 第一类顾客的概率母函数为

$$\begin{aligned}
 Q^1(z_1, z_2, z_3) &= \left\{ [\lambda_1(1 - z_1) + \lambda_2(1 - z_2) + \lambda_3(1 - z_3)] P_{0000} \right. \\
 &\quad - \left[\mu \left(\frac{1}{z_2} - 1 \right) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q_0^2(z_2, z_3) \\
 &\quad - \left[\mu \left(\frac{1}{z_3} - 1 \right) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q_0^3(z_2, z_3) \left. \right\} \\
 &\quad \div \left[\mu \left(\frac{1}{z_1} - 1 \right) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right].
 \end{aligned}$$

证明: 对 (1)、(2) 和 (3), (4) 和 (5), (6) 和 (7) 两边分别同时乘 z_3^j 并关于 j 求和, 再对求得的三个式子两边同乘 z_2^i 并关于 i 求和, 整理得:

$$[\lambda_1 + \lambda_2(1 - z_2) + \lambda_3(1 - z_3) + \mu] Q_0^3(z_2, z_3) = [\lambda_1 + \lambda_2 + \lambda_3(1 - z_3) + \mu] Q_{00}^3(z_3). \quad (20)$$

同样的, 对 (8) 和 (9), (10) 和 (11) 做上述处理得:

$$\begin{aligned}
 &\left[\mu \left(\frac{1}{z_2} - 1 \right) - \lambda_1 - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q_0^2(z_2, z_3) \\
 &= \mu Q_{10}^1(z_3) + \mu Q_{01}^2(z_3) - \mu Q_1^1(z_2, z_3) - \frac{\mu}{z_3} Q_0^3(z_2, z_3) + \frac{\mu}{z_3} Q_{00}^3(z_3) - \lambda_2 z_2 P_{0000}. \quad (21)
 \end{aligned}$$

对 (12) 和 (13), (14) 和 (15) 做上述处理得:

$$\begin{aligned}
 &[\lambda_1 + \lambda_2(1 - z_2) + \lambda_3(1 - z_3) + \mu] Q_1^1(z_2, z_3) \\
 &= \mu Q_2^1(z_2, z_3) + \lambda_1 Q_0^2(z_2, z_3) + \lambda_1 Q_0^3(z_2, z_3) + \lambda_1 P_{0000}. \quad (22)
 \end{aligned}$$

对 (16) 和 (17), (18) 和 (19) 做上述处理得:

$$[\lambda_1 + \lambda_2(1 - z_2) + \lambda_3(1 - z_3) + \mu] Q_l^1(z_2, z_3)$$

$$= \mu Q_{l+1}^1(z_2, z_3) + \lambda_1 Q_{l-1}^1(z_2, z_3), \quad l = 2, 3, \dots \quad (23)$$

将 (20) 代入 (21) 中得:

$$\begin{aligned} & \left[\mu \left(\frac{1}{z_2} - 1 \right) - \lambda_1 - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q_0^2(z_2, z_3) \\ &= [\lambda_1 + \lambda_2(1 - z_2) + \lambda_3(1 - z_3)] P_{0000} \\ & \quad - \left[\mu \left(\frac{1}{z_3} - 1 \right) - \lambda_1 - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q_0^3(z_2, z_3) - \mu Q_1^1(z_2, z_3). \end{aligned} \quad (24)$$

对 (22) 和 (23) 方程两边同乘 z_1^l 并关于 l 求和得:

$$\begin{aligned} & \left[\mu \left(\frac{1}{z_1} - 1 \right) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q^1(z_1, z_2, z_3) \\ &= \mu Q_1^1(z_2, z_3) - \lambda_1 z_1 Q_0^2(z_2, z_3) - \lambda_1 z_1 Q_0^3(z_2, z_3) - \lambda_1 z_1 P_{0000}. \end{aligned} \quad (25)$$

将 (24) 代入 (25) 中整理得到定理 1. 定理 1 证毕. \square

定理 2 $\rho < 1, |z_1| < 1$ 时, 第二类顾客的概率母函数为

$$\begin{aligned} Q_0^2(z_2, z_3) &= \left\{ [\lambda_1(1 - z_1) + \lambda_2(1 - z_2) + \lambda_3(1 - z_3)] P_{0000} \right. \\ & \quad \left. - \left[\mu \left(\frac{1}{z_3} - 1 \right) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right] Q_0^3(z_2, z_3) \right\} \\ & \quad \div \left[\mu \left(\frac{1}{z_2} - 1 \right) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) \right], \quad |z_2| < 1, |z_3| < 1. \end{aligned}$$

证明: 考察方程 $\mu(1/z_1 - 1) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) = 0$, 整理可得

$$\lambda_1 z_1^2 - z_1(\lambda_1 + \lambda_2 + \lambda_3 - \lambda_2 z_2 - \lambda_3 z_3 + \mu) + \mu = 0, \quad (26)$$

若 $|z_1| = 1, |z_2| < 1, |z_3| < 1$, 则

$$\begin{aligned} |\lambda_1 z_1^2 + \mu| &\leq \lambda_1 + \mu < |\lambda_1 + \lambda_2 + \lambda_3 - \lambda_2 z_2 - \lambda_3 z_3 + \mu| \\ &= |\lambda_1 + \lambda_2 + \lambda_3 - \lambda_2 z_2 - \lambda_3 z_3 + \mu| z_1. \end{aligned}$$

当 $|z_1| < 1$ 时, 由 Rouché 定理, 方程 (26) 有唯一的根且为 z_2, z_3 的函数, 记为 $z_1 = f(z_2, z_3)$. 由于 $Q^1(z_1, z_2, z_3)$ 在 $|z_1| < 1, |z_2| < 1, |z_3| < 1$ 范围内收敛, 所以 $f(z_2, z_3)$ 也是定理 1 右侧分子的根, 代入得到定理 2. \square

定理 3 $\rho < 1, |z_1| < 1$ 时, 第三类顾客的概率母函数为

$$Q_0^3(z_2, z_3) = \frac{[\lambda_1(1 - z_1) + \lambda_2(1 - z_2) + \lambda_3(1 - z_3)] P_{0000}}{\mu(1/z_3 - 1) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3)}, \quad |z_2| < 1, |z_3| < 1.$$

类似的, 考察方程 $\mu(1/z_2 - 1) - \lambda_1(1 - z_1) - \lambda_2(1 - z_2) - \lambda_3(1 - z_3) = 0$, 由 Rouché 定理可以证明定理 2 右侧分母与分子在 $|z_1| < 1, |z_2| < 1, |z_3| < 1$ 时有唯一且相同的零点 z_2 , 且满足表达式 $z_2 = g(z_3)$, 并由此可求得第三类顾客的概率母函数为定理 3.

§4. 系统的性能指标

1) 各类顾客的占有率和系统闲置概率

(i) 第一类顾客的占有率

由于第一类顾客具有强占优先权, 所以第一类顾客的到达可视为一个经典的 M/M/1 排队^[1]. 其中输入为泊松过程, 参数为 λ_1 , 服务时间服从指数分布, 参数为 μ . 设平稳状态下第一类顾客的稳态分布为 $\pi = \{\pi_l, l = 0, 1, 2, \dots\}$. 可知系统被第一类顾客占有的概率是

$$Q^1(1, 1, 1) = 1 - \pi_0 = \frac{\lambda_1}{\mu}. \quad (27)$$

(ii) 第二三类顾客的占有率和系统闲置概率

由于 $z_1 = f(z_2, z_3) = f(g(z_3), z_3)$, 所以记 $z_1 = h(z_3)$. 将定理 2 右侧分母化简得:

$$\lambda_2 z_2^2 - z_2[\lambda_1 + \lambda_2 + \lambda_3 - \lambda_1 f(z_2, z_3) - \lambda_3 z_3 + \mu] + \mu = 0, \quad (28)$$

把 $z_1 = h(z_3)$, $z_2 = g(z_3)$ 分别代入 (26) 和 (28) 得:

$$\lambda_1 h(z_3)^2 - h(z_3)[\lambda_1 + \lambda_2 + \lambda_3 - \lambda_2 g(z_3) - \lambda_3 z_3 + \mu] + \mu = 0, \quad (29)$$

$$\lambda_2 g(z_3)^2 - g(z_3)[\lambda_1 + \lambda_2 + \lambda_3 - \lambda_1 h(z_3) - \lambda_3 z_3 + \mu] + \mu = 0, \quad (30)$$

令 $z_3 \rightarrow 1$, 有

$$\lambda_1 h(1)^2 - h(1)[\lambda_1 + \lambda_2 - \lambda_2 g(1) + \mu] + \mu = 0,$$

$$\lambda_2 g(1)^2 - g(1)[\lambda_1 + \lambda_2 - \lambda_1 h(1) + \mu] + \mu = 0,$$

得 $h(1) = g(1) = 1$.

下面将 $Q_0^2(z_2, z_3)$ 和 $Q_0^3(z_2, z_3)$ 都写成含 P_{0000} 的表达式. 由于 $h(z_3)$, $g(z_3)$ 分别是关于 z_1 和 z_2 的常系数一元二次方程 (26) 和 (28) 的根, 必为连续函数, 分别对 (29) 和 (30) 两端关于 z_3 求导数, 并令 $z_3 \rightarrow 1$, 得

$$h'(1) = g'(1) = \frac{\lambda_3}{\mu - \lambda_1 - \lambda_2}, \quad h''(1) = g''(1) = \frac{2\lambda_3^2\mu}{(\mu - \lambda_1 - \lambda_2)^3}. \quad (31)$$

将 (31) 代入定理 3, 令 $z_3 \rightarrow 1$, 得

$$Q_0^3(z_2, z_3) \Big|_{z_3=1} = Q_0^3(1, 1) = \frac{\lambda_3}{\mu - \lambda_1 - \lambda_2 - \lambda_3} P_{0000}. \quad (32)$$

将 (31) 代入定理 2 中, 令 $z_3 \rightarrow 1$, 得

$$Q_0^2(z_2, z_3) \Big|_{z_3=1} = Q_0^2(1, 1) = \frac{2\lambda_3(\mu - \lambda_1 - \lambda_2)}{(\mu - \lambda_1 - \lambda_2 - \lambda_3)^2} P_{0000}. \quad (33)$$

利用归一化条件:

$$Q^1(1, 1, 1) + Q_0^2(1, 1) + Q_0^3(1, 1) + P_{0000} = 1, \quad (34)$$

将 (27)、(32)、(33) 代入 (34) 得系统闲置的概率:

$$P_{0000} = \frac{(\mu - \lambda_1)(\mu - \lambda_1 - \lambda_2 - \lambda_3)^2}{\mu(\mu - \lambda_1 - \lambda_2)(\mu - \lambda_1 - \lambda_2 + \lambda_3)}, \quad (35)$$

将 (35) 代入到 (32) 和 (33) 分别得到系统被第三类和第二类顾客占有的概率:

$$Q_0^3(1, 1) = \frac{\lambda_3(\mu - \lambda_1)(\mu - \lambda_1 - \lambda_2 - \lambda_3)}{\mu(\mu - \lambda_1 - \lambda_2)(\mu - \lambda_1 - \lambda_2 + \lambda_3)}, \quad Q_0^2(1, 1) = \frac{2\lambda_3(\mu - \lambda_1)}{\mu(\mu - \lambda_1 - \lambda_2 + \lambda_3)}.$$

2) 顾客的平均队长

(i) 第一类顾客的平均队长为

$$E(L_1) = \frac{dQ^1(z_1, 1, 1)}{dz_1} \Big|_{z_1=1} = \frac{\lambda_1}{\mu - \lambda_1}.$$

(ii) 第二类顾客的平均队长为

$$\begin{aligned} E(L_2) &= \frac{d(Q^1(1, z_2, 1) + Q_0^2(z_2, 1) + Q_0^3(z_2, 1))}{dz_2} \Big|_{z_2=1} \\ &= \frac{2\lambda_3(\mu - \lambda_1)}{\lambda_2(\mu - \lambda_1 - \lambda_2 + \lambda_3)} + \frac{\mu(\mu - \lambda_1)(\mu - \lambda_1 - \lambda_2 - \lambda_3)}{(\mu - \lambda_2)^2(\mu - \lambda_1 - \lambda_2 + \lambda_3)}. \end{aligned}$$

(iii) 第三类顾客的平均队长为

$$\begin{aligned} E(L_3) &= \frac{d(Q^1(1, 1, z_3) + Q_0^2(1, z_3) + Q_0^3(1, z_3))}{dz_3} \Big|_{z_3=1} \\ &= \frac{(\mu - \lambda_1)(3\mu - \lambda_2 - \lambda_3)(\mu - \lambda_1 - \lambda_2 - \lambda_3)^2}{(\mu - \lambda_3)^2(\mu - \lambda_1 - \lambda_2)(\mu - \lambda_1 - \lambda_2 + \lambda_3)} + \frac{3(\mu - \lambda_1)(\mu - \lambda_1 - \lambda_2 - \lambda_3)}{(\mu - \lambda_1 - \lambda_2)(\mu - \lambda_1 - \lambda_2 + \lambda_3)}. \end{aligned}$$

注记 4 若服务时间推广到服从不同参数的指数分布时, 采用同样的方法, 可以得到三类顾客队长联合分布的概率母函数及每类顾客的平均队长等系统性能指标, 只是计算过程变得更加复杂, 计算量变大.

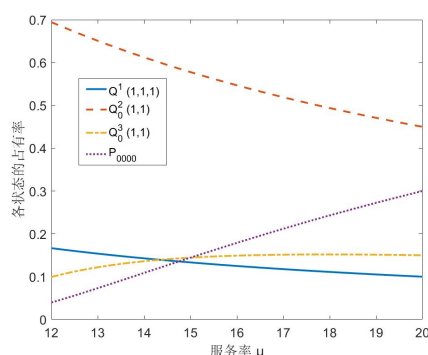
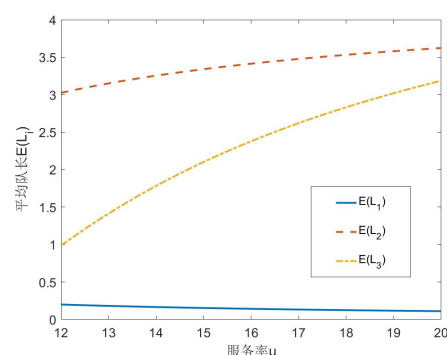
§5. 数值分析

在某大型餐饮公司中, 假设各类顾客到达均为泊松过程, 服务时间均服从指数分布, 且相互独立. 会员顾客的到达率为 $\lambda_1 = 2$, 过号顾客的到达率 $\lambda_2 = 3$, 普通排队顾客的到达率为 $\lambda_3 = 5$, 各类顾客的服务率均为 $\mu = 12$. 会员顾客较过号顾客和普通排队顾客有强占优先权, 即只有当会员顾客全部被服务完毕后才开始服务其余顾客. 过号顾客较普通排队顾客有非强占型优先权. 根据以上条件, 得到服务台被各类顾客占有以及闲置的概率, 并且得到各类顾客的平均等待队长见表 1.

表1 系统的性能指标

性能指标 \ 顾客类别	1	2	3	系统闲置
第 i 类顾客的占有率	0.1667	0.6944	0.0992	0.0397
平均队长 $E(L_i)$	0.2000	3.0247	0.9864	0.0000

为了合理利用资源, 提高服务台的效率, 研究某些参数对一些主要性能指标的影响也尤为重要. 接下来通过比较在服务率不同的情况下, 服务台被各类顾客占有以及闲置的概率以及各类顾客平均队长的变化情况, 从而为系统提供更便捷的策略. 图1为服务率不同的情况下服务台被各类顾客占有以及闲置概率的变化情况. 图2为服务率不同的情况下对各类顾客平均队长的影响.

图1 各状态的占有率随服务率 μ 的变化图2 平均队长 $E(L_i)$ 随服务率 μ 的变化情况

由图1所示, $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 5$, $\mu \in (12, 20)$ 时, 随服务率 μ 的增加, 服务台被会员顾客和过号顾客占有的概率减少, 但服务台被普通排队顾客占有的概率以及闲置的概率在逐渐增加. 这也符合实际情况, 当增设多个服务台窗口时, 先服务有优先权的顾客, 当有优先权的顾客全部服务完毕后, 服务台才开始服务无优先权的顾客或者进入闲置状态.

由图2所示, $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 5$, $\mu \in (12, 20)$ 时, 会员顾客的平均队长 $E(L_1)$ 随服务率 μ 的变化在逐渐下降, 但并不明显. 过号顾客的平均队长 $E(L_2)$ 与普通排队顾客的平均队长 $E(L_3)$ 则变化显著, 呈现上升的趋势. 由此, 可以通过增设服务台的个数选择合适的服务率, 有效的设计系统进程, 尽可能使系统达到最优.

根据系统的性能指标, 构造成本-费用函数, 求出不同效益函数下系统所得的收益, 进而为系统选取更合适的成本费用方案. 常见的效益函数为指数函数或幂函数的形式. 下面考虑上述大型餐饮公司在 (a)、(b) 两种利润策略下单位时间的收益. 其中,

$$(a) R_i = e^{\lambda_i}, \quad C_i = e^{-\lambda_i}, \quad i = 1, 2, 3. \quad (b) R_i = \lambda_i^2, \quad C_i = \lambda_i^{-2}, \quad i = 1, 2, 3.$$

假设该餐饮公司的效益函数为

$$F = (R_1 - C_1)Q^1(1, 1, 1) + (R_2 - C_2)Q_0^2(1, 1) + (R_3 - C_3)Q_0^3(1, 1) - C_4P_{0000},$$

其中 R_i ($i = 1, 2, 3$) 分别表示该公司服务会员顾客、过号顾客、普通排队顾客单位时间所获利润; C_i ($i = 1, 2, 3$) 分别表示该公司服务会员顾客、过号顾客、普通排队顾客单位时间所耗费用; C_4 表示该公司空闲时单位时间所消耗的费用, 不妨设 $C_4 = 1$.

接下来考虑不同利润策略下该餐饮公司的收益随各类顾客的到达率和服务率的变化情况.

(i) 假设参数 $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$, 选取 $\mu \in (6, 20)$, 图 3 展示了在到达率均相同的情况下, 该餐饮公司在不同成本费用函数下的收益随服务率 μ 的变化情况. 从图 3 中观察到随着服务率 μ 的增加, 该餐饮公司的收益均呈现先增加后减少的趋势. 并且当利润策略为 (a) 时, 公司收益在 $\mu = 7.6$ 附近取得最大值, 当利润策略为 (b) 时, 公司收益在 $\mu = 7.2$ 附近取得最大值.

(ii) 假设参数 $\lambda_2 = 3, \lambda_3 = 5, \mu = 10$, 选取 $\lambda_1 \in (1, 2)$, 图 4 展示了在部分到达率 λ_2, λ_3 和服务率 μ 相同的情况下, 该餐饮公司在不同成本费用函数下的收益随会员顾客的到达率 λ_1 的变化情况. 从图 4 可以看出, 随着会员顾客的到达率 λ_1 的增加, 该餐饮公司收益均逐渐减少, 并且在利润策略为 (a) 时的变化更显著些.

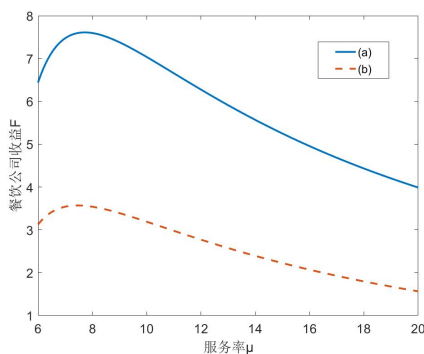


图 3 餐饮公司收益随服务率 μ 的变化情况

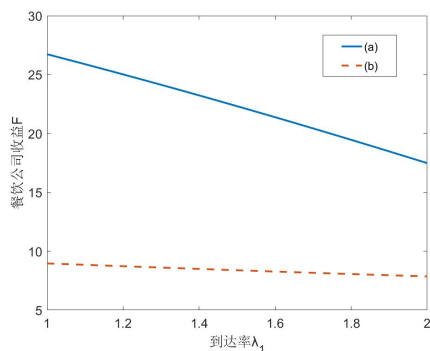
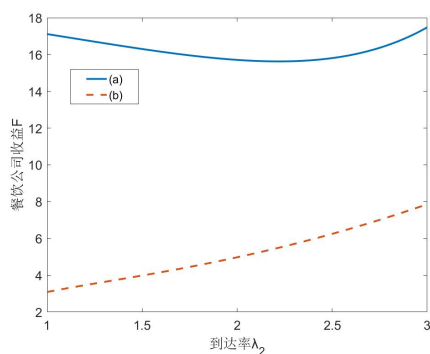
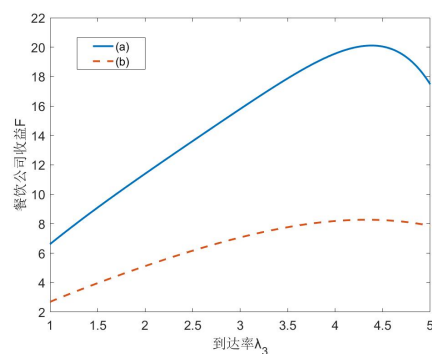


图 4 餐饮公司收益随到达率 λ_1 的变化情况

(iii) 假设参数 $\lambda_1 = 2, \lambda_3 = 5, \mu = 10$, 选取 $\lambda_2 \in (1, 3)$, 图 5 展示了在部分到达率 λ_1, λ_3 和服务率 μ 相同的情况下, 该餐饮公司在不同成本费用函数下的收益随过号顾客的到达率 λ_2 的变化情况. 从图 5 可以看出, 随着过号顾客的到达率 λ_2 的增加, 公司收益在利润策略为 (a) 时呈现先减少后增加的趋势, 在 $\lambda_2 = 2.3$ 附近取得最小值, 公司收益在利润策略为 (b) 时却呈现缓慢上升趋势.

(iv) 假设参数 $\lambda_1 = 2, \lambda_2 = 3, \mu = 10$, 选取 $\lambda_3 \in (1, 5)$, 图 6 展示了在部分到达率 λ_1, λ_2 和服务率 μ 相同的情况下, 该餐饮公司在不同成本费用函数下的收益随着普通排队顾客的到达率 λ_3 的变化情况. 从图 6 中观察到随着普通排队顾客的到达率 λ_3 的增加, 公司收益均呈现先增加后减少的趋势. 当利润策略为 (a) 时, 公司收益在 $\lambda_3 = 4.4$ 附近取得最大值, 当利润策略为 (b) 时, 公司收益在 $\lambda_3 = 4.1$ 附近取得最大值.

由图 3 至图 6 可以观察到成本费用函数均为指数形式时该餐饮公司收益更高, 此外,

图5 餐饮公司收益随到达率 λ_2 的变化情况图6 餐饮公司收益随到达率 λ_3 的变化情况

并不是各类顾客的到达率或者服务率越大公司收益越高. 由此, 各企业可通过限流或调整服务台效率, 选取合适的成本费用函数, 使得效益最大.

§6. 结 论

本文阐述了一个带有强占优先权和非强占优先权的 M/M/1 排队系统, 通过三类顾客的平均队长, 结合实际情况, 进行数值算例, 分析服务率的变化对系统主要性能指标的影响, 最后通过优化分析选择合理的成本费用函数, 并找到最优的到达率和服务率, 丰富了排队论的应用. 随着人们对服务行业质量需求的提高, 优先权排队大大加快了各大企业的运行效率, 会员卡的出现不仅可以刺激顾客消费, 增加营业额, 同时也可以满足顾客的尊享需求, 为顾客提供更优质的消费体验, 赢得了越来越多企业的青睐.

参 考 文 献

- [1] MILLER D R. Computation of steady-state probabilities for M/M/1 priority queues [J]. *Oper Res*, 1981, **29**(5): 945–958.
- [2] LEE D S. A two-queue model with exhaustive and limited service disciplines [J]. *Comm Statist Stochastic Models*, 1996, **12**(2): 285–305.
- [3] FENG W, KOWADA M, ADACHI K. Performance analysis of a two-queue model with an (M, N) -threshold service schedule [J]. *J Oper Res Soc Japan*, 2001, **44**(2): 101–124.
- [4] SLEPTCHENKO A, SELEN J, ADAN I, et al. Joint queue length distribution of multi-class, single-server queues with preemptive priorities [J]. *Queueing Syst*, 2015, **81**(4): 379–395.
- [5] 张宏波, 史定华. M/M/1 抢占优先权排队平稳指标的尾部分析 [J]. *系统科学与数学*, 2014, **34**(1): 1–9.
- [6] 张宏波, 周高军, 封平华. 对一类等待空间有限的抢占优先权排队的分析 [J]. *运筹学学报*, 2016, **20**(3): 11–20.
- [7] FAJARDO V A, DREKIC S. Waiting time distributions in the preemptive accumulating priority queue [J]. *Methodol Comput Appl Probab*, 2017, **19**(1): 255–284.

- [8] GAO S. A preemptive priority retrial queue with two classes of customers and general retrial times [J]. *Oper Res-Ger*, 2015, **15**(2): 233–251.
- [9] 赵国喜, 陈燕, 薛晓东. 非强占优先权下的 M/M/1 排队 [J]. *大学数学*, 2006, **22**(1): 44–48.
- [10] 黄业文, 邝神芬. 非强占有限优先权 M/M/n/m 排队系统 [J]. *应用概率统计*, 2018, **34**(4): 364–380.
- [11] 周杰, 李军, 郭鹏, 等. 急诊非强占优先权下 Geom/NB/1 排队系统常规病人等待时间及损失率 [J]. *系统工程理论与实践*, 2016, **36**(8): 2135–2143.
- [12] LANDRY R, STAVRAKAKIS I. Queueing study of a 3-priority policy with distinct service strategies [J]. *IEEE ACM T Network*, 1993, **1**(5): 576–589.
- [13] LIU Z M, CHU Y Q, WU J B. On the three-queue priority polling system with threshold service policy [J]. *J Appl Math Comput*, 2017, **53**(1-2): 445–470.
- [14] ALAWNEH A J. Steady state probabilities of a three preemptive single server queue [J]. *JMASM*, 2012, **11**(1): 203–210.
- [15] ALQAHTANI S A, ALTAMRAH A S. Supporting QoS requirements provisions on 5G network slices using an efficient priority-based polling technique [J]. *Wirel Netw*, 2019, **25**(7): 3825–3838.

Optimization Analysis of M/M/1 Queue with Two Types of Priority Customers

ZHANG Yitong

(School of Economics and Management, Yanshan University, Qinhuangdao, 066004, China)

XU Xiuli

(School of Science, Yanshan University, Qinhuangdao, 066004, China)

Abstract: This paper considers an M/M/1 queue model with preemptive and non-preemptive priorities. Customers are divided into three priority levels. The first category of customers enjoy the priority of preemptive, the second category of customers enjoy the priority of non-preemptive, and the third category of customers have no priority. When the first kind of customers arrive, they will interrupt the second or third kind of customers who are receiving the service immediately; When the second type of customers arrive, if there are only the third type of customers in the system, the customer must wait for the current service to be completed before accepting the service. Customers of the same type follow the queuing rule of FCFS. A multi-dimensional vector Markov process is constructed here using the supplementary variable method, and the probability generating functions of three types of joint distribution of queue length are obtained. Then, each category of customer's average queue length and the probabilities in service are obtained. The influence of the change of service rate on the average queue length of all kinds of customers in the system is studied using MATLAB software. Finally, different cost functions are constructed for optimization analysis.

Keywords: priority; preemptive; non-preemptive; probability generation function

2020 Mathematics Subject Classification: 60K25; 90B22