排队论与随机过程

2019.12.13

引 言

排队论是什么？解决什么问题？

排队：研究一个排队系统

论：一种理论的分析模型

系统的性能/评价指标

从概率论与随机过程的角度给出严格的理论推导

引 言

排队论有什么用？

对现实物理场景的理论建模

分析性质能够用于系统设计与优化

排队论适用场景

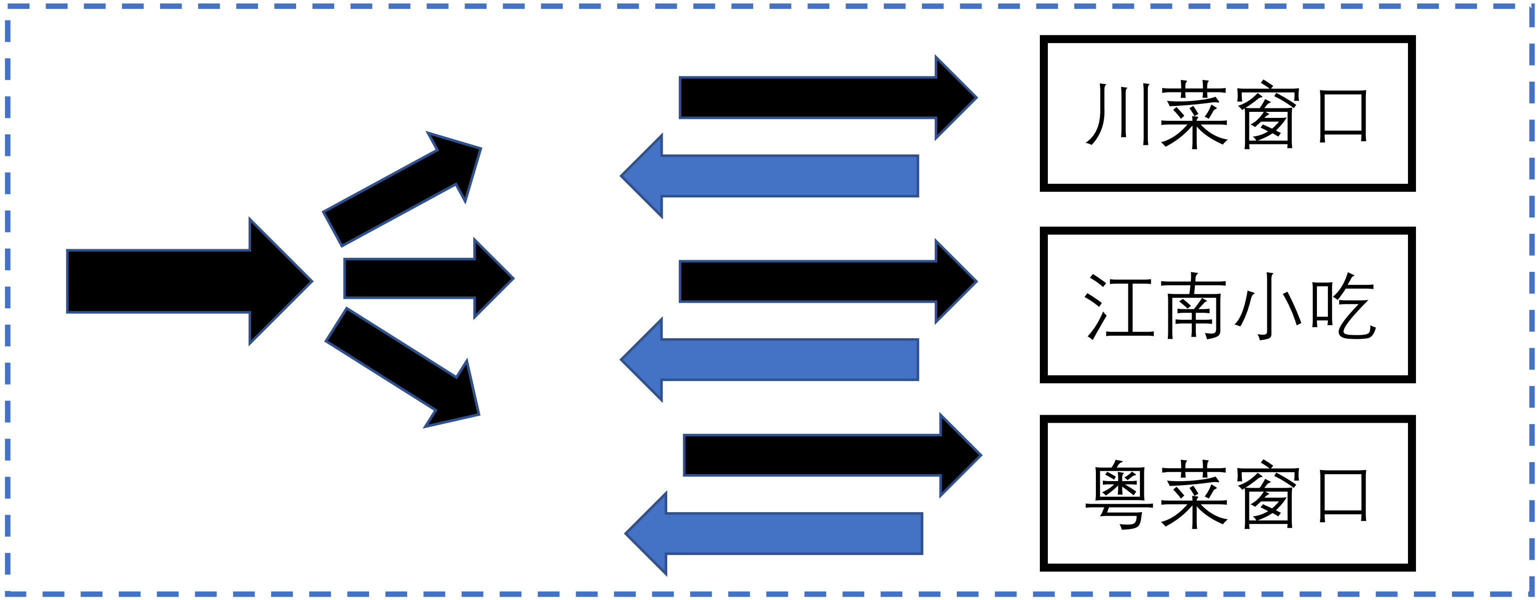
实际物理场景中的排队系统：游乐场，机场，收费站，医院

虚拟场景中的排队系统：打印请求序列，网络的收发信号

往其他场景上类比：一个有滞留/阻碍现象的系统

引 言

简单的例子：食堂排队就餐

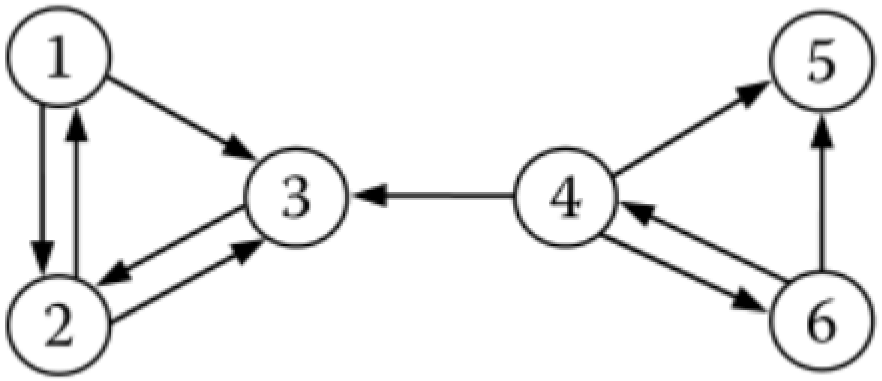


变量：进入的学生，选择概率，窗口服务效率……

性能指标：系统吞吐率，等待时间，平均服务人数……

基本知识扩充

从马尔科夫链/PageRank说起



$$A=\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

$$a_{ij}=P(s_{t+1}=j\mid s_t=i)$$

$$A\pi^t=\pi^{t+1}$$

$$\sum a_{ij}\pi_i^t=\pi_j^{t+1}$$

马尔科夫过程：这一时刻的状态只依赖于上一时刻的状态

基本知识扩充

指数分布

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases} \qquad F(x;\lambda)=\begin{cases}1-e^{-\lambda x} & , x\ge 0\\ 0 & , x<0\end{cases}$$

λ称为率参数（rate parameter），通常表示单位时间事件发生的次数

统计量 $E(X)=\frac{1}{\lambda}$ $\quad D(X)=Var(X)=\frac{1}{\lambda^2}$

无记忆性 $P(T>s+t|T>t)=P(T>s)$

唯一满足此性质的分布

泊松分布

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,\cdots$$

适用于描述单位时间内随机事件发生的次数

随机过程

计数过程

$\{N(t), t \geq 0\}$ 表示到事件t为止发生的事件的总数

eg：进入商店的人数，小孩的诞生，足球比赛进球个数……

独立增量

发生在不相交的时间区间中的事件的个数是彼此独立的

平稳增量

在时间区间(t, s+t)中的事件个数的分布对于一切t都相同

$$P(T > s + t | T > t) = P(T > s)$$

泊松过程

定义：计数过程$\{N(t), t \geq 0\}$满足：

1）$N(0)=0$

2）过程有独立增量

3）长度为t的任意时间区间的事件个数服从均值为λt的**泊松分布**

$$P\{N(s+t)-N(s)=n\}=e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

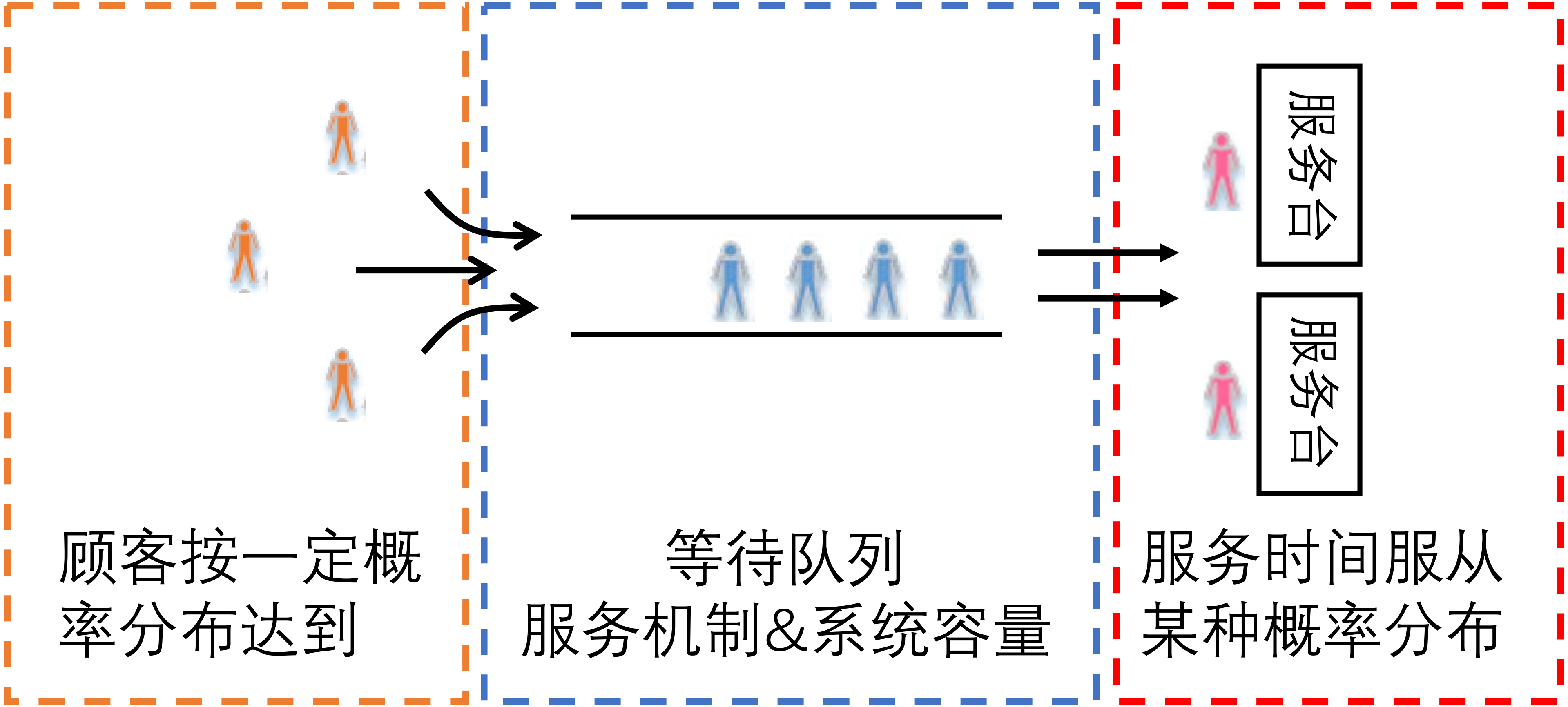
重要性质和概念

$E\{N(t)\}=\lambda t$　　λ 称为泊松过程的**速率**（单位时间发生的次数）

$P\{两次事件之间的间隔\}=e^{-\lambda t}$　服从均值为$\frac{1}{\lambda}$的**指数分布**

排队理论

基本模型



排队理论

排队论肯德尔记号

一般的排队模型可以表示为模板“X/Y/Z/A/B/C”：

X—顾客相继到达的间隔时间的分布；

Y—服务时间的分布；

（X和Y的取值可以为：M-指数分布，G-一般分布）

Z—服务台个数；

A—系统容量限制（默认为∞）；

B—顾客源数目（默认∞）；

C—服务规则（默认为FCFS"First Come, First Service"）。

经典排队论类型：M/M/1, M/M/k, M/G/1……

排队理论

基本量与价格方程

L，系统中顾客的平均数；

L_Q，队列中平均等待顾客数；

W，一个顾客在系统中所耗的平均时间；

W_Q，一个顾客在队列中等待的平均时间。

时间意义上的平均

强制进入系统的顾客向系统付钱（按某种规则）

系统赚钱的平均速率$=\lambda_a$*进入系统的顾客所支付的平均金额

其中，顾客进入系统的速率 $\lambda_a = \lim_{t\to\infty} \frac{N(t)}{t}$

排队理论

基本量与价格方程

L，系统中顾客的平均数；

L_Q，队列中平均等待顾客数；

W，一个顾客在系统中所耗的平均时间；

W_Q，一个顾客在队列中等待的平均时间。

系统赚钱的平均速率=λ_a*进入系统的顾客所支付的平均金额

规则一：每个客户只要处于系统中一个单位时间，就需要支付一个单位金钱。

$$L = \lambda_a * W$$

排队理论

基本量与价格方程

L，系统中顾客的平均数；

L_Q，队列中平均等待顾客数；

W，一个顾客在系统中所耗的平均时间；

W_Q，一个顾客在队列中等待的平均时间。

系统赚钱的平均速率=λ_a*进入系统的顾客所支付的平均金额

规则二：每个客户只要其处于等待队列中一个单位时间，就需要支付一个单位金钱。

$$L_Q = \lambda_a * W_Q$$

排队理论

基本量与价格方程

L，系统中顾客的平均数；

L_Q，队列中平均等待顾客数；

W，一个顾客在系统中所耗的平均时间；

W_Q，一个顾客在队列中等待的平均时间。

系统赚钱的平均速率=λ_a*进入系统的顾客所支付的平均金额

规则三：每个客户只要其处于被服务状态中一个单位时间，就需要支付一个单位金钱。E(S)为一个顾客被服务平均时间

$$1 - P_0 = \lambda_a * E(S)$$

价格方程对所有模型都成立

稳态概率

系统中有n个顾客的长程概率 $P_n = \lim_{n\to\infty} P\{X(t)=n\}$

等于系统恰好包含n个顾客的时间（长程）比例

P_0 表示系统处于空闲的概率

发现n人的到达者速率=留下n人的离开者速率

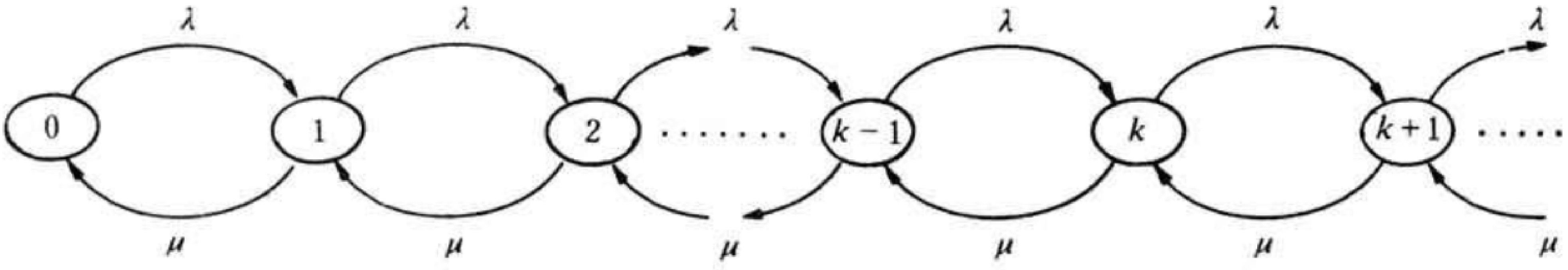
两边同除以总达到速率

发现n人的到达者比例=留下n人的离开者比例

对于泊松到达者，P_n=发现n人到达者比例

平衡方程

进入状态n速率=离开状态n的速率

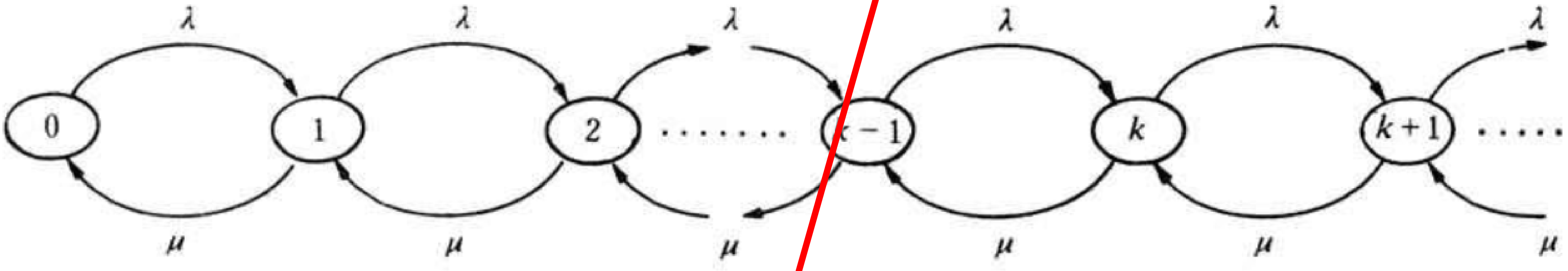


λ 顾客进入平均速率，μ 顾客离开平均速率/服务速率

排队理论

M/M/1型

$$P_n=\left(\frac{\lambda}{\mu}\right)^n P_0 \Rightarrow \sum P_n=\sum\left(\frac{\lambda}{\mu}\right)^n P_0=1 \Rightarrow P_0=1-\frac{\lambda}{\mu}$$



平衡方程

$$\lambda P_0=\mu P_1$$

$$(\lambda+\mu)P_n=\lambda P_{n-1}+\mu P_{n+1}, n\geq 1$$

约束条件

$$\sum_{n=0}^{\infty} P_n=1$$

$$P_0=1-\frac{\lambda}{\mu}$$

$$P_n=\left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right), n\geq 1$$

排队理论

M/M/1型

性能指标计算

$$L=\frac{\mu-\lambda}{\mu}\frac{\lambda}{\mu}\sum n\left(\frac{\lambda}{\mu}\right)^{n-1}=\frac{\mu-\lambda}{\mu}\frac{\lambda}{\mu}\left(\sum\left(\frac{\lambda}{\mu}\right)^{n}\right)'=\frac{\mu-\lambda}{\mu}\frac{\lambda}{\mu}\frac{\mu^2}{(\mu-\lambda)^2}$$

$$L=\sum_{n=0}^{\infty} nP_n=\frac{\mu-\lambda}{\mu}\sum_{n=0}^{\infty} n(\frac{\lambda}{\mu})^n$$

$$L=\frac{\mu-\lambda}{\mu}\frac{\lambda/\mu}{(1-\lambda/\mu)^2}=\frac{\lambda}{\mu-\lambda}$$

$$W=\frac{L}{\lambda}=\frac{1}{\mu-\lambda}$$

$$W_Q=W-E(S)=\frac{\lambda}{\mu(\mu-\lambda)}$$

$$L_Q=\lambda W_Q=\frac{\lambda^2}{\mu(\mu-\lambda)}$$

这里 $E(S)=\dfrac{1}{\mu}$

E(S)为一个顾客被服务平均时间

Q&A

变形1：有限容量的M/M/1型

最后一个平衡方程 $\mu P_N = \lambda P_{N-1}$

修正约束条件 $\sum_{n=0}^{N} P_n = 1$

性能指标

$$W = \frac{L}{\lambda_a}$$

$\lambda_a = \lambda(1 - P_N)$ 表示实际进入系统的速率

变形2：（生灭排队模型）到达率和离开率不是定值

平衡方程逐差

$$\lambda_0 P_0 = \mu_1 P_1$$

$$\lambda_1 P_1 = \mu_2 P_2$$

......

$$\lambda_n P_n = \mu_{n+1} P_{n+1}$$

$$\Rightarrow \quad P_0 = \left(\left(\sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}\right)+1\right)^{-1}$$

$$P_n = \frac{\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}}{\sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}+1}$$

要求 $\sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i} < \infty$

变形3：M/M/k型

调整系统变量

$$\lambda_n = \lambda$$

$$\mu_n = \begin{cases} n\mu & \mathrm{n} < \mathrm{k} \\ k\mu & \mathrm{n} \geq \mathrm{k} \end{cases}$$

代入上一页式子

对于$n < k$:

$$P_n = P_0 \frac{\lambda^n}{\mu^n n!}$$

对于$n \geq k$:

$$P_n = P_0 \frac{\lambda^n}{\mu^n k^{n-k} k!}$$

排队理论

排队论理论分析思路

Step1：分析系统状态空间及转换关系

Step2：根据转换关系列出平衡方程

Step3：求解平衡方程（逐差法、观察法）

Step4：计算性能指标（级数求和、价格方程）

排队理论

其他变形

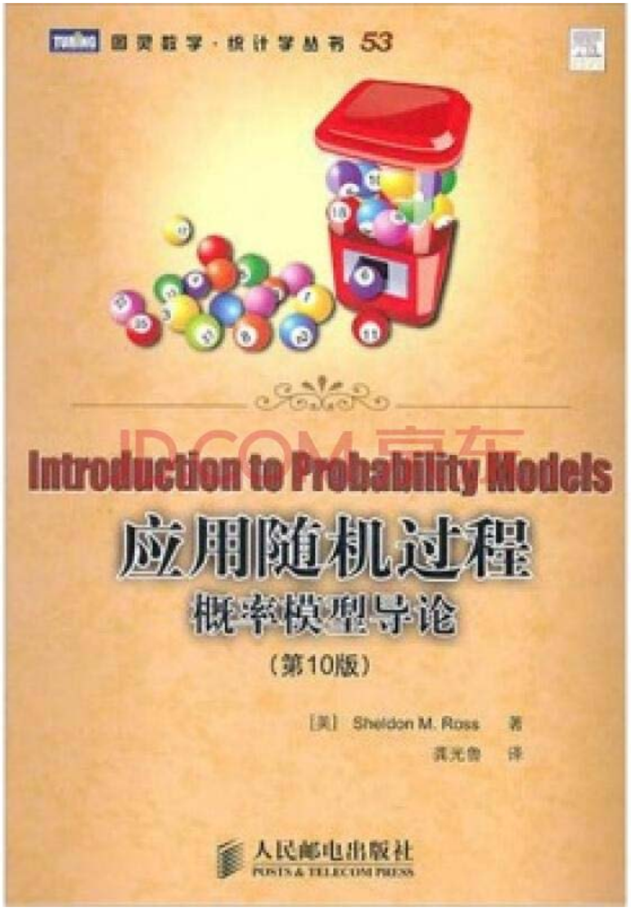
M/G/1

G/M/1

批量到达

优先级队列

中断服务



举例1：擦鞋店

考虑一个擦鞋店，每个顾客要进行两项服务A与B，两项服务必须连续完成，且每项服务同一时间最多只能服务一人。做出以下假设：

1）顾客的到来服从均值 λ 指数分布，每项服务的服务时间也服从均值分别为 μ_1, μ_2 指数分布。

2）只有当完成A服务的顾客才能享受B服务。

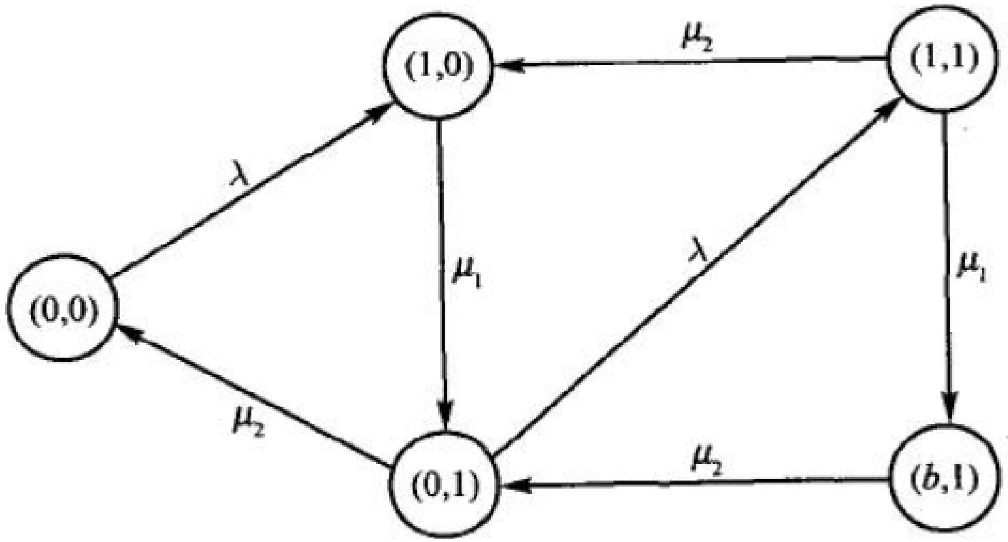
3）当有顾客在享受B服务时，已经完成A服务的顾客需要在其座位上等待，直到享受B服务的顾客完成服务后才离开座位将A服务让权给后来的顾客。

问题关键：需要确定系统可能存在的所有状态

举例1：擦鞋店

Step1：分析系统状态空间与状态转移关系

状态	解释
(0,0)	在系统中没有顾客
(1,0)	在系统中有一个顾客，且他在椅子 1 上
(0,1)	在系统中有一个顾客，且他在椅子 2 上
(1,1)	在系统中有两个顾客，都在接受服务
(b,1)	在系统中有两个顾客，椅子 1 上的顾客已经完成了其接受的服务且在等待椅子 2 空出来



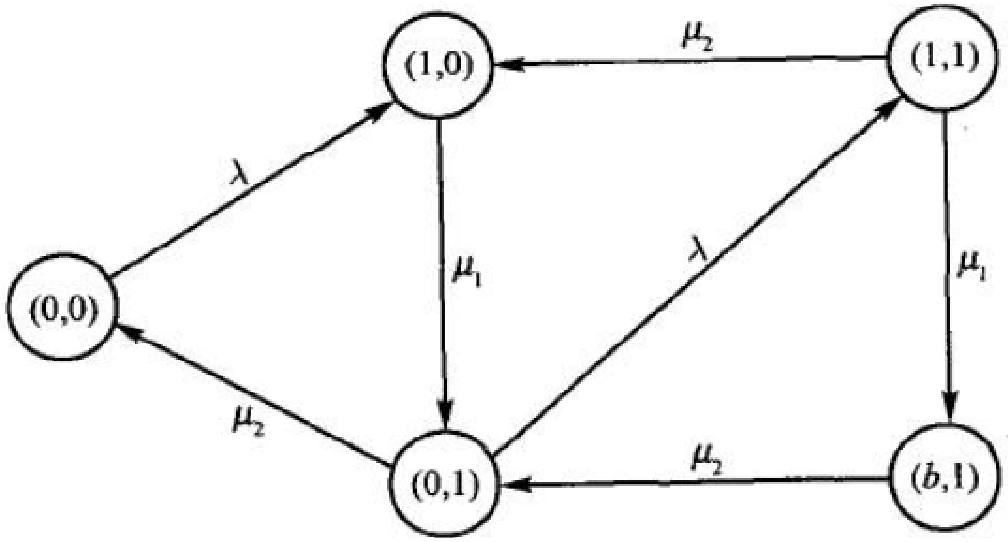
排队理论

举例1：擦鞋店

Step2：根据每种状态列出平衡方程

状态	过程离开的速率 = 进入的速率
(0,0)	$\lambda P_{00} = \mu_2 P_{01}$
(1,0)	$\mu_1 P_{10} = \lambda P_{00} + \mu_2 P_{11}$
(0,1)	$(\lambda + \mu_2) P_{01} = \mu_1 P_{10} + \mu_2 P_{b1}$
(1,1)	$(\mu_1 + \mu_2) P_{11} = \lambda P_{01}$
(b,1)	$\mu_2 P_{b1} = \mu_1 P_{11}$

$$P_{00} + P_{10} + P_{01} + P_{11} + P_{b1} = 1$$



举例1：擦鞋店

Step3&4：求解方程，计算各性能指标

系统平均顾客数 $L = P_{01} + P_{10} + 2(P_{11} + P_{b1})$

由价格方程 $W = L/\lambda_a$

及 $\lambda_a = \lambda(P_{00} + P_{01})$

一个顾客在系统中所耗平均时间 $$W = \frac{P_{01} + P_{10} + 2(P_{11} + P_{b1})}{\lambda(P_{00} + P_{01})}$$

排队理论

举例1：擦鞋店

代入具体数值进行结果分析比较

情况一：$\lambda = 1, \quad \mu_1 = 1, \quad \mu_2 = 2$

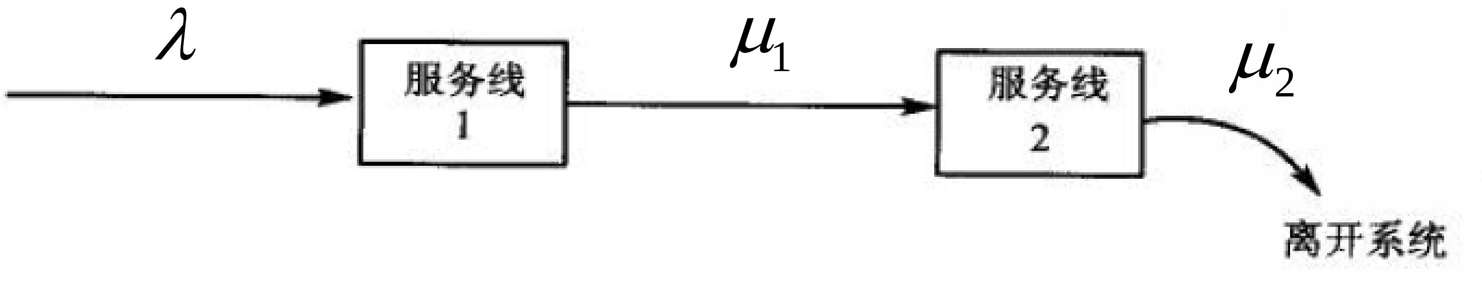
$$L = \frac{28}{37} \qquad W = \frac{28}{18}$$

情况二：$\lambda = 1, \quad \mu_1 = 2, \quad \mu_2 = 1$

$$L = 1 \qquad W = \frac{11}{6}$$

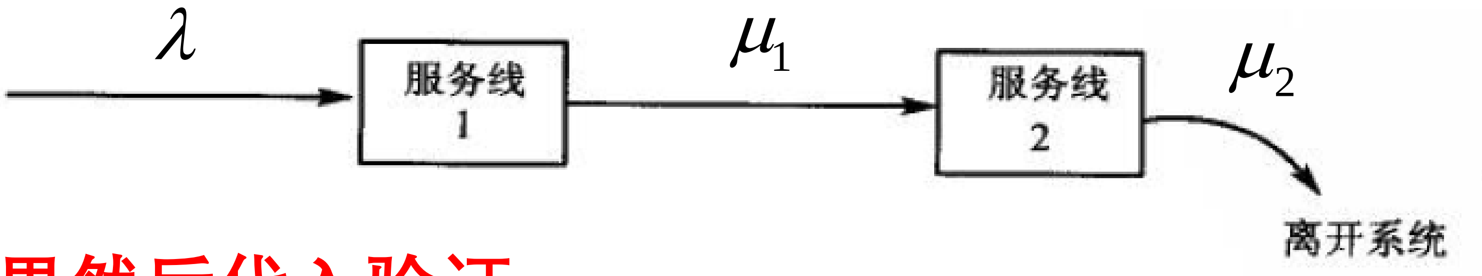
在投入总成本相同的情况下，比较如何分配投入成本使得期望效益最大化

举例2：串联排队系统



状态	过程离开的速率 = 进入的速率
$0,0$	$\lambda P_{0,0}=\mu_2 P_{0,1}$
$n,0;n>0$	$(\lambda+\mu_1)P_{n,0}=\mu_2 P_{n,1}+\lambda P_{n-1,0}$
$0,m;m>0$	$(\lambda+\mu_2)P_{0,m}=\mu_2 P_{0,m+1}+\mu_1 P_{1,m-1}$
$n,m;nm>0$	$(\lambda+\mu_1+\mu_2)P_{n,m}=\mu_2 P_{n,m+1}+\mu_1 P_{n+1,m-1}+\lambda P_{n-1,m}$

举例2：串联排队系统



猜测结果然后代入验证

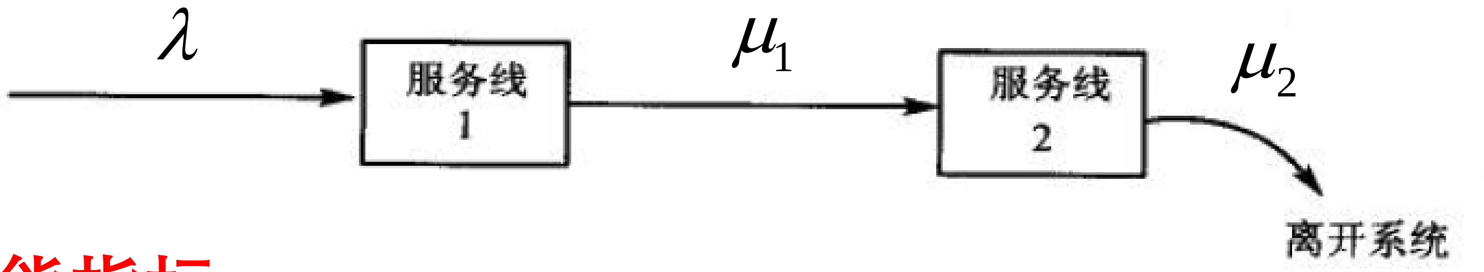
$$P\{n\ 个顾客在服务线\ 1\} = \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right)$$

$$P\{m\ 个顾客在服务线\ 2\} = \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right)$$

$$P_{n,m} = \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right) \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right)$$

满足上一页的解是唯一的

举例2：串联排队系统



计算性能指标

$$
\begin{aligned}
L &= \sum_{n,m}(n+m)P_{n,m} \\
&= \sum_{n} n\left(\frac{\lambda}{\mu_1}\right)^n\left(1-\frac{\lambda}{\mu_1}\right)+\sum_{m} m\left(\frac{\lambda}{\mu_2}\right)^m\left(1-\frac{\lambda}{\mu_2}\right) \\
&= \frac{\lambda}{\mu_1-\lambda}+\frac{\lambda}{\mu_2-\lambda}
\end{aligned}
$$

$$
W=\frac{L}{\lambda}=\frac{1}{\mu_1-\lambda}+\frac{1}{\mu_2-\lambda}
$$

推广：含有k条服务线的排队网络

假设顾客以均值为λ_i的指数分布加入第i条服务线，第i条服务线的服务时间服从均值为μ_i的指数分布，在第i条服务线完成服务的顾客会以P_{ij}的概率加入第j条服务线，以 $1-\sum_{j=1}^{k} P_{ij}$ 的概率离开系统

$$\lambda_j = r_j + \sum_{i=1}^{k} \lambda_i P_{ij}, \qquad i = 1, \cdots, k$$

$$P\{n \text{ 个顾客在服务线 } j\} = \left(\frac{\lambda_j}{\mu_j}\right)^n \left(1 - \frac{\lambda_j}{\mu_j}\right), \qquad n \geqslant 1$$

排队理论

推广：含有k条服务线的排队网络

$$P(n_1, n_2, \cdots, n_k) = \prod_{j=1}^{k} \left(\frac{\lambda_j}{\mu_j}\right)^{n_j} \left(1 - \frac{\lambda_j}{\mu_j}\right)$$

$$L = \sum_{j=1}^{n} \text{在服务线 } j \text{ 的平均数} = \sum_{j=1}^{k} \frac{\lambda_j}{\mu_j - \lambda_j}$$

$$\lambda = \sum\nolimits_{j=1}^{k} r_j \qquad L = \lambda W$$

$$W = \frac{\sum_{j=1}^{k} \lambda_j / (\mu_j - \lambda_j)}{\sum_{j=1}^{k} r_j}$$

案例分析

MCM2003 ICM C题：机场安检扫描机安置问题

问题（其中一个） 机场需要购置多少新型的安检扫描机？

主要思路 将行李包视为排队系统的顾客，基于平均意义分析

变量设定

到达率：先生成一个小于1随机数，乘以一个航班的座位数得到本航班总达到人数，然后除以两小时，得到顾客到达速率，再考虑每个顾客平均携带的行李数，作为行李到达速率

服务率：扫描机扫描行李的服务速率

服务数：扫描机数目

MCM2003 ICM C题：机场安检扫描机安置问题

创新点 将总顾客数按航班起飞时间分配到不同时间段，对每个航班对应的顾客用一个泊松过程来刻画，总的系统则是各个子系统的叠加

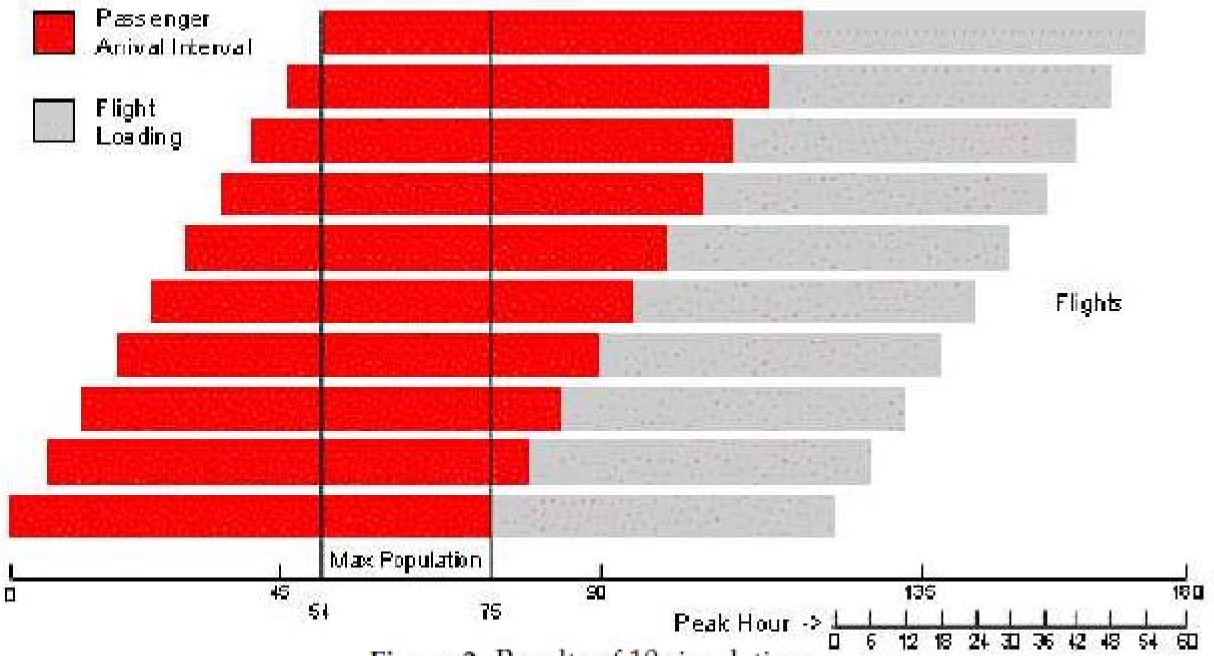


Figure 2. Results of 10 simulations.

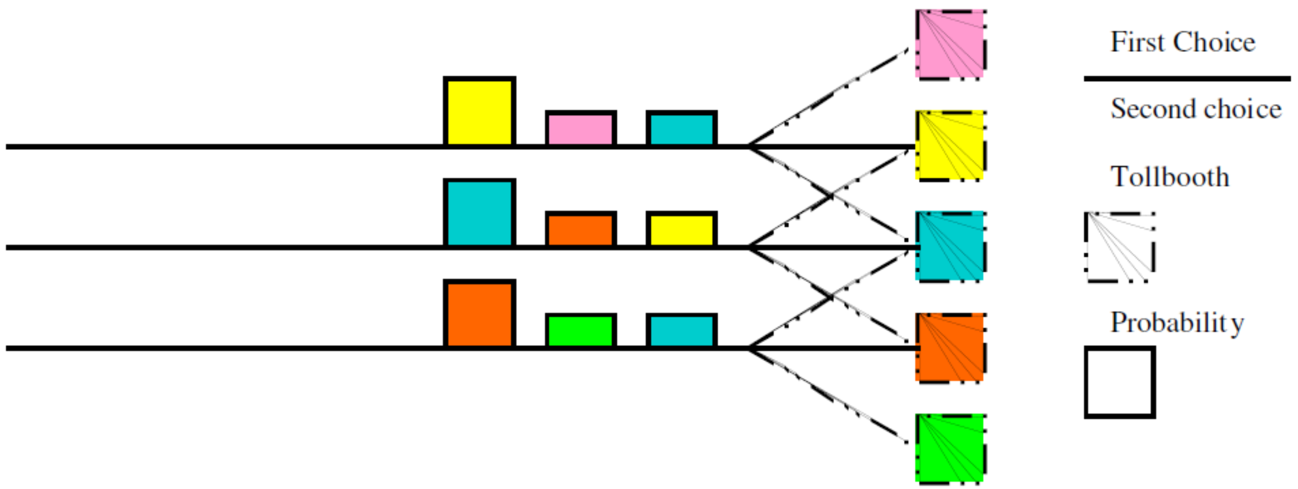
案例分析

MCM2005 MCM B题：高速公路收费站设计

问题 优化收费站窗口的数目

主要思路 将进入收费站和离开收费站分别视为两个排队系统

创新点 将收费窗口视为一个multiple single server



实验仿真

银行ATM机设置问题

问题 A型机的服务效率是B型机的两倍，问购置一台A型机好还是购置两台B型机好？（M/M/1和M/M/2）

性能指标 等待队列中一个顾客的平均等待时间

MATLAB实现思路

1）生成n个顾客到达时间间隔与他们的服务时间（指数分布）

2）得到每个顾客的到达时间、离开时间、等待时间

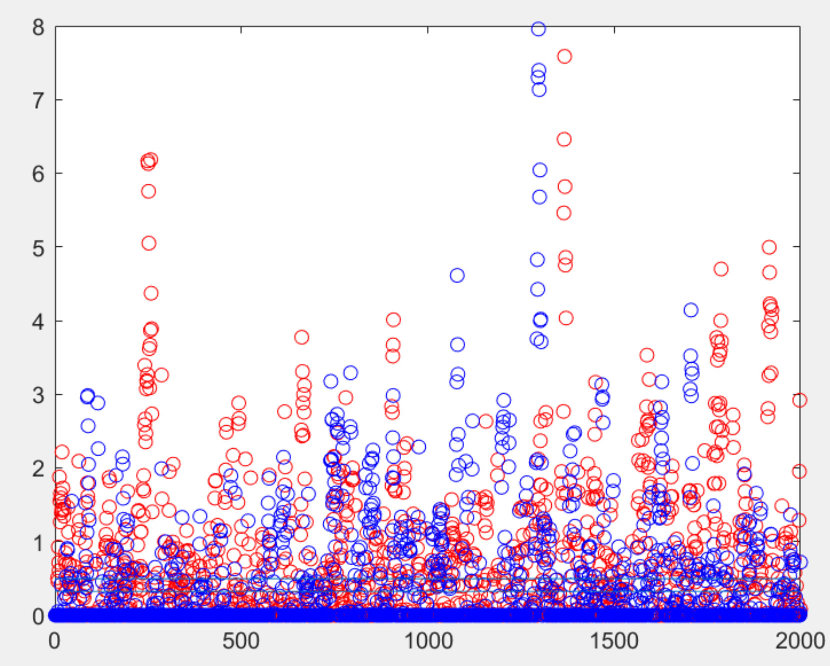
3）计算平均等待时间，再重复m次实验取平均值

实验仿真

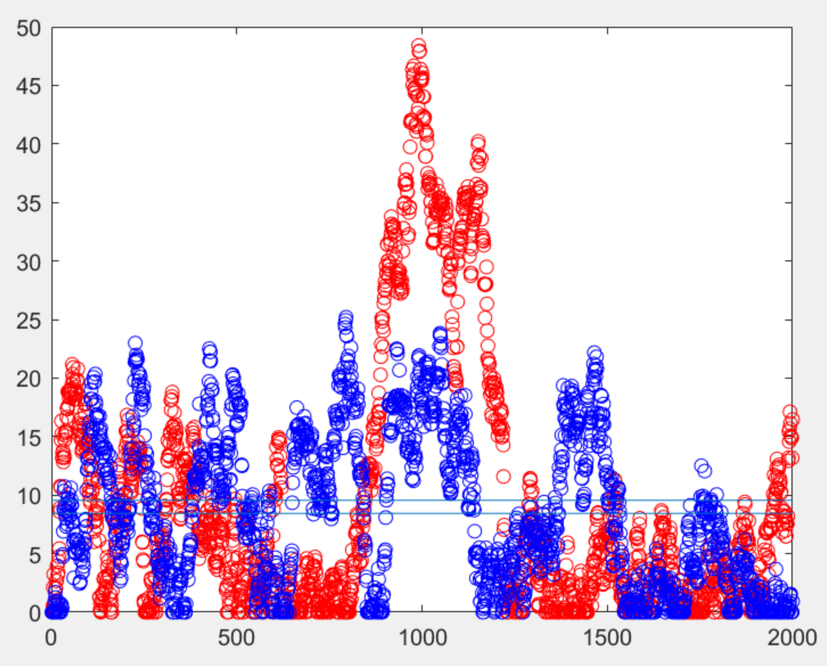
银行ATM机设置问题

实验结果展示：2000个顾客的等待时间

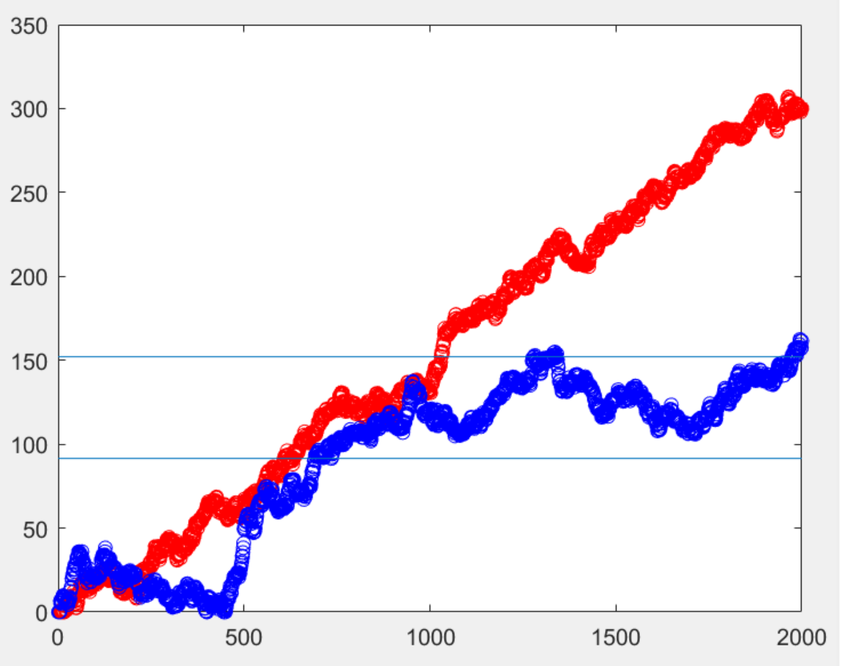
n=2000， m=1000， lamda=1



mu1=0.5， mu2=1



mu1=0.9， mu2=1.8



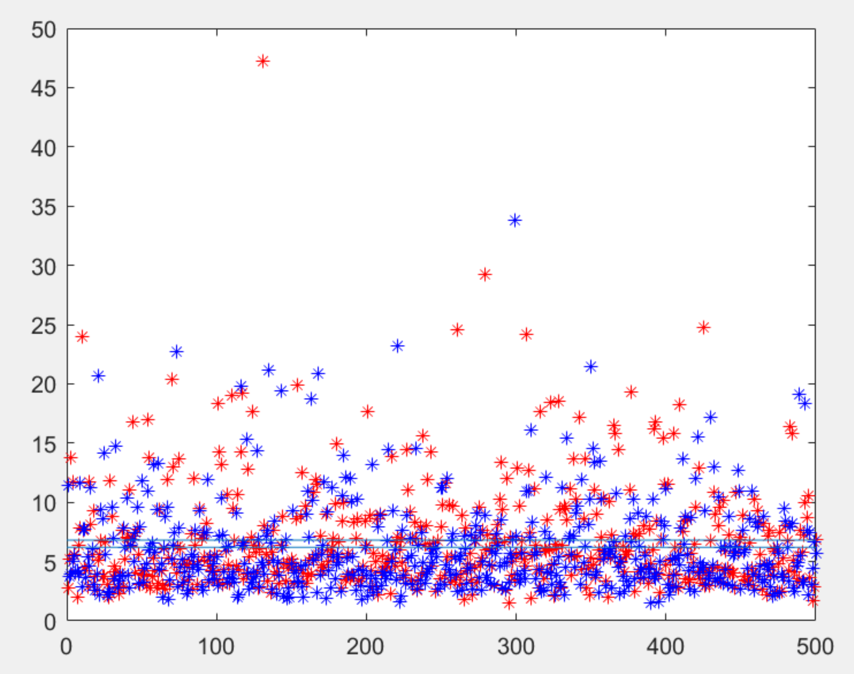
mu1=1.1， mu2=2.2

实验仿真

银行ATM机设置问题

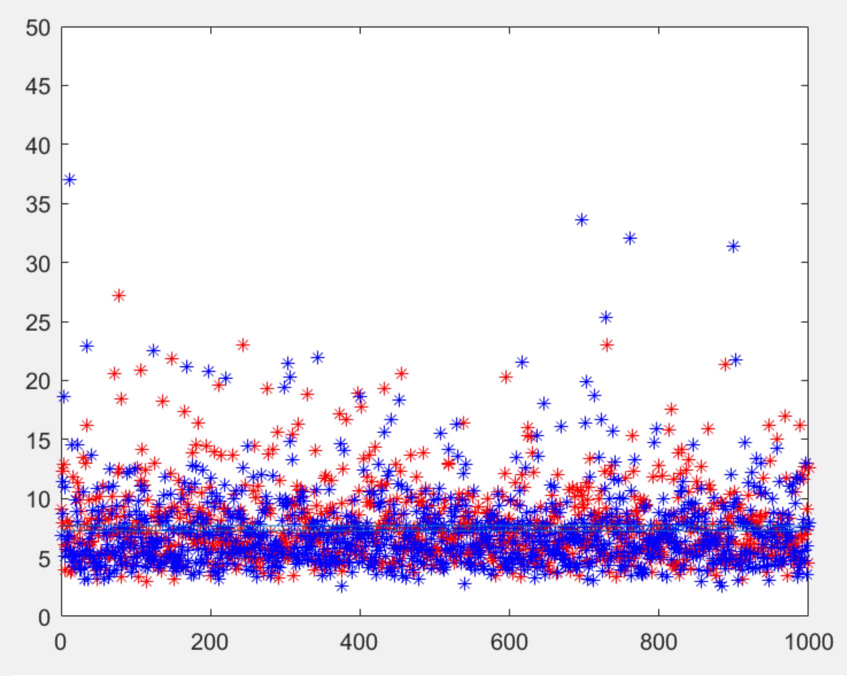
关于稳态的问题：单次实验的顾客数

n=500



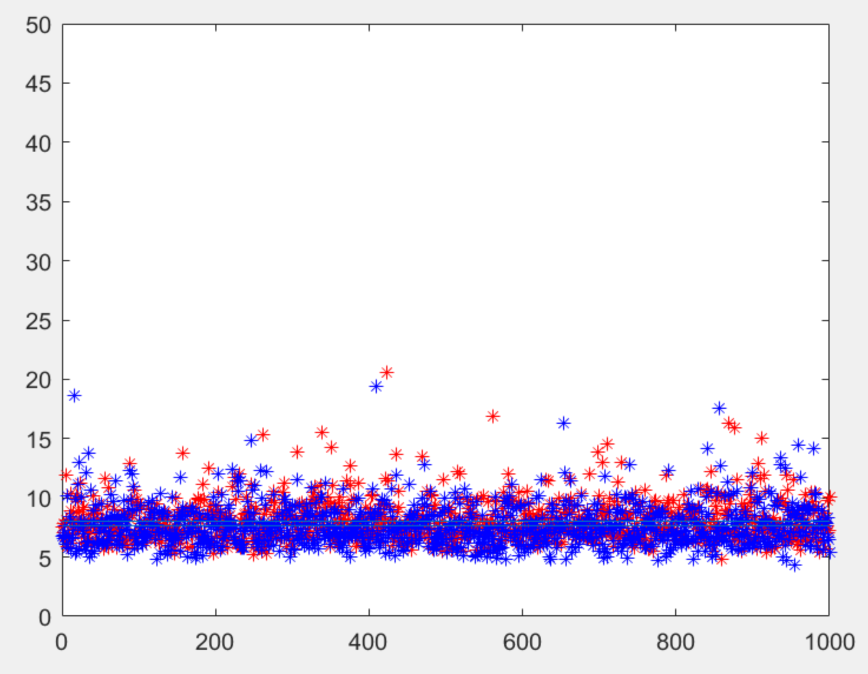
d1=16.6， d2=15.7

n=2000



d1=10.7， d2=12.5

n=10000



d1=2.8， d2=3.1

实验仿真

银行ATM机设置问题

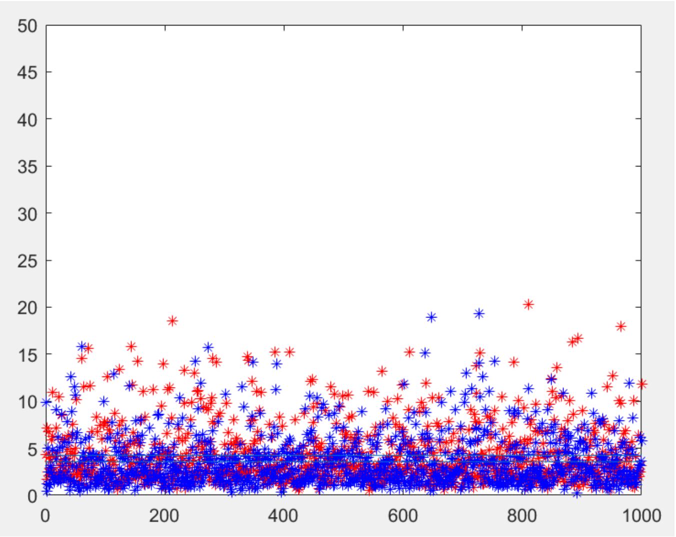
观测时间窗口

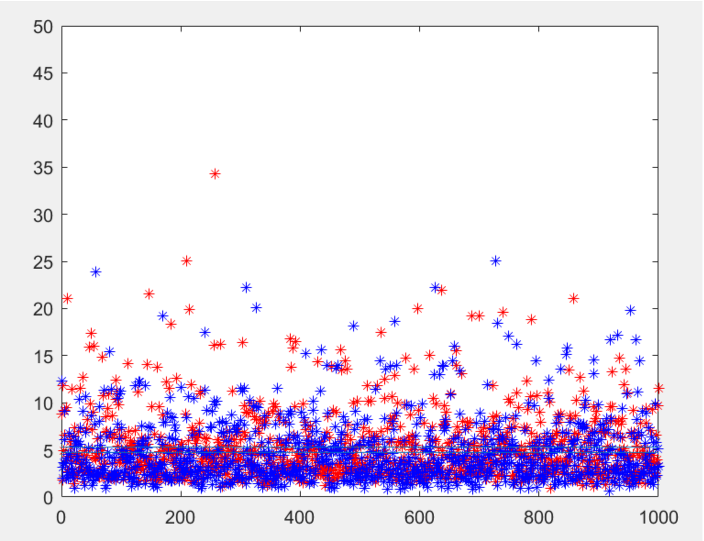
前100个顾客

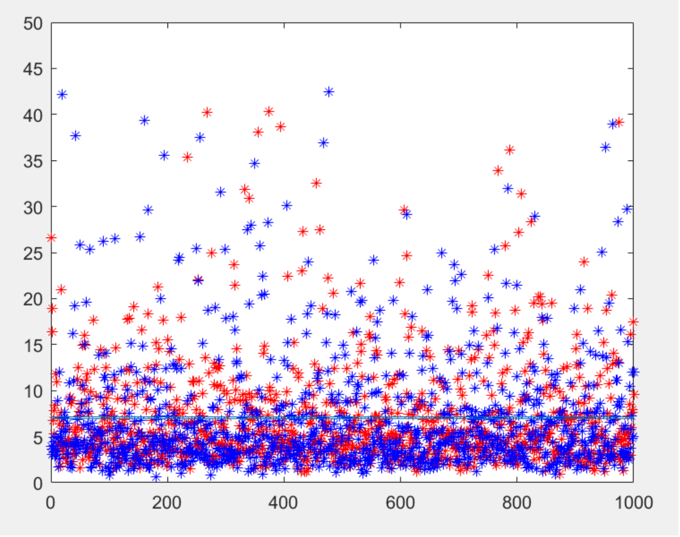
前200个顾客

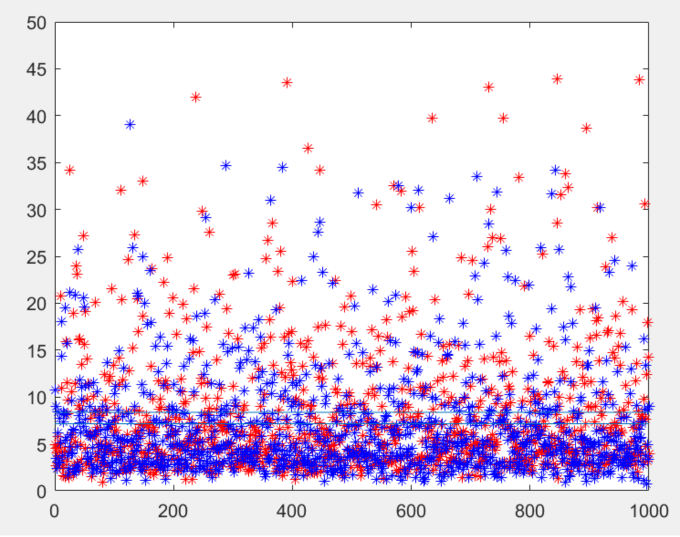
200-400

400-600









时间窗口越靠后，
系统不稳定性增加

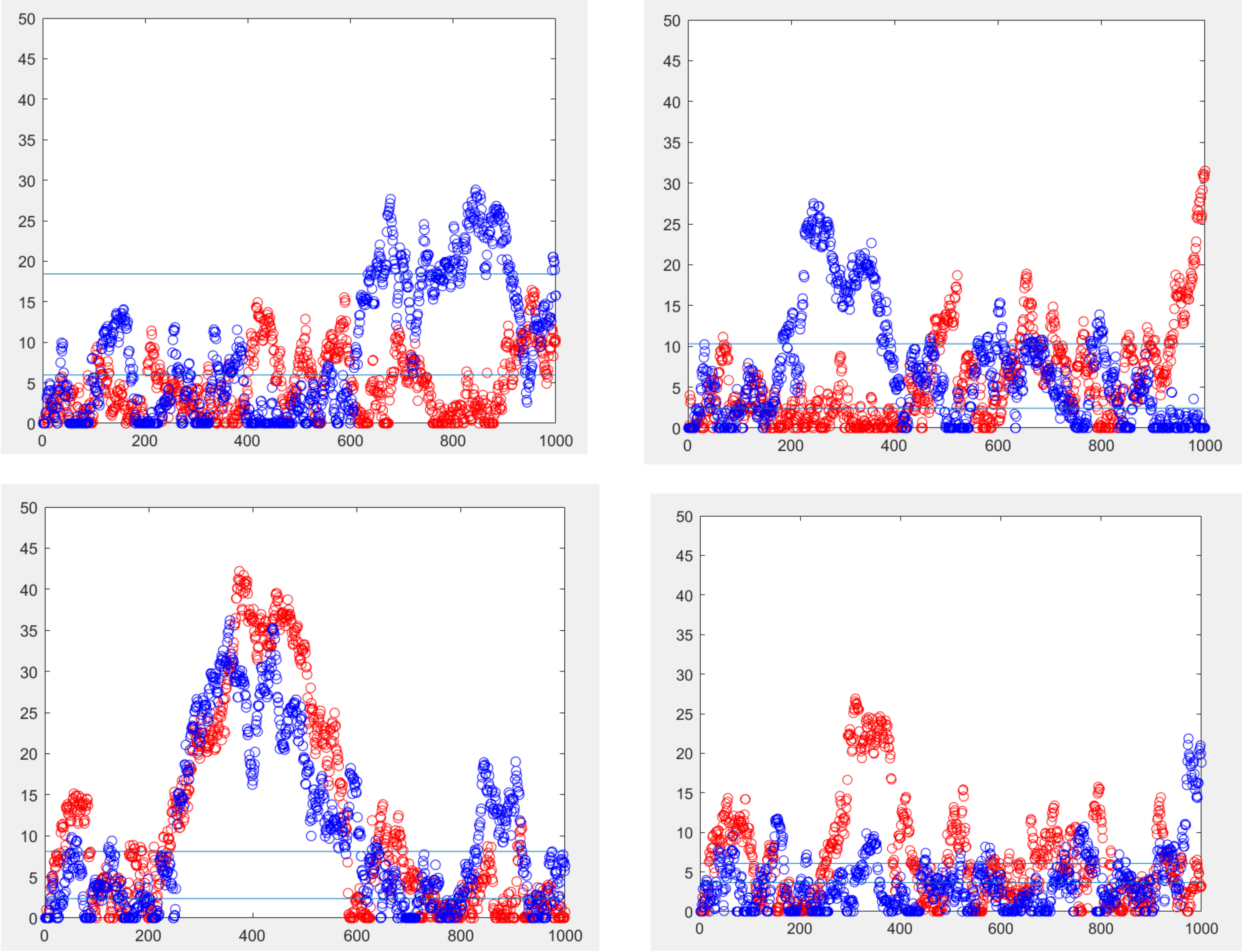
实验仿真

银行ATM机设置问题

观测时间窗口

平行对比四次独立实验的1000个到来顾客的等待时间分布

越靠后的时间窗口，事件随机性增加



模型延伸

分析角度 最优化/规划问题

如果最终求解的性能指标中含有待确定的未知参数或者优化变量，这时可以转化为规划问题

但要注意性能指标是平均意义下求得的

模拟角度 元胞自动机（可以视为一种有空间延伸的排队系统）

如果系统的服务机制较为复杂（比如需要考虑顾客的具体移动与空间位置），这时可使用元胞的观点

多次实验结果的误差分析

总 结

计数过程与泊松过程

基本排队模型M/M/1

三种基本变形：有限容量，生灭排队模型，M/M/k,

串联网络，排队网络

建模案例与实验仿真

Q&A