



中国地质大学暑期建模专题培训

# 数学建模 —— 排队论

付丽华

2019.08

# 排队论

## 排队论 (queuing theory) 研究内容包括三个部分：

- (1) 排队系统的性态问题
- (2) 排队系统的最优化问题
- (3) 排队系统的统计推断问题

⑩ 性态问题，即研究各种排队系统的概率规律性，主要研究队长分布、等待时间分布和忙期分布等。

⑩ 最优化，又分静态最优和动态最优，前者指最优设计（现有设备如何达到最优），后者指现有排队系统的最优运营。

系统符合哪种模型，以便根据排队理论进行研究。

解排队问题的目的，是研究排队系统运行的效率，估计服务质量，确定系统参数的最优值，以决定系统结构是否合理，研究设计改进措施等。



# 提纲

---

- ❖ 基本概念
- ❖ M/M/1模型
- ❖ M/M/C模型
- ❖ 排队系统的最优化问题
- ❖ 分析排队系统的随机模拟法



# 提纲

---

- ❖ 基本概念
- ❖ M/M/1模型
- ❖ M/M/C模型
- ❖ 排队系统的最优化问题
- ❖ 分析排队系统的随机模拟法



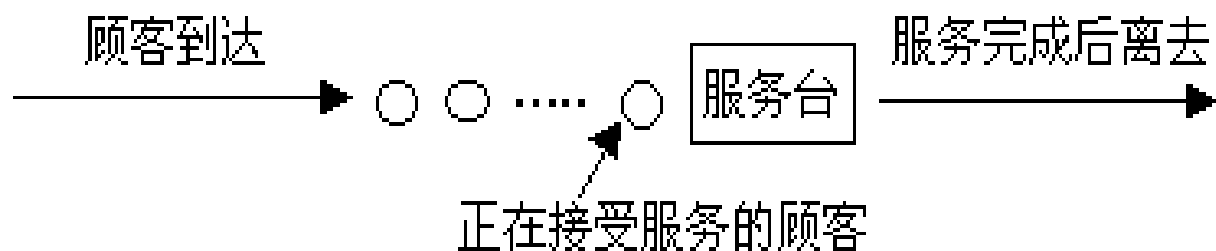
# 第1节 基本概念

---

- ❖ 1.1 排队过程的一般表示
- ❖ 1.2 排队系统的组织和特征
- ❖ 1.3 排队模型的分类
- ❖ 1.4 排队问题的求解

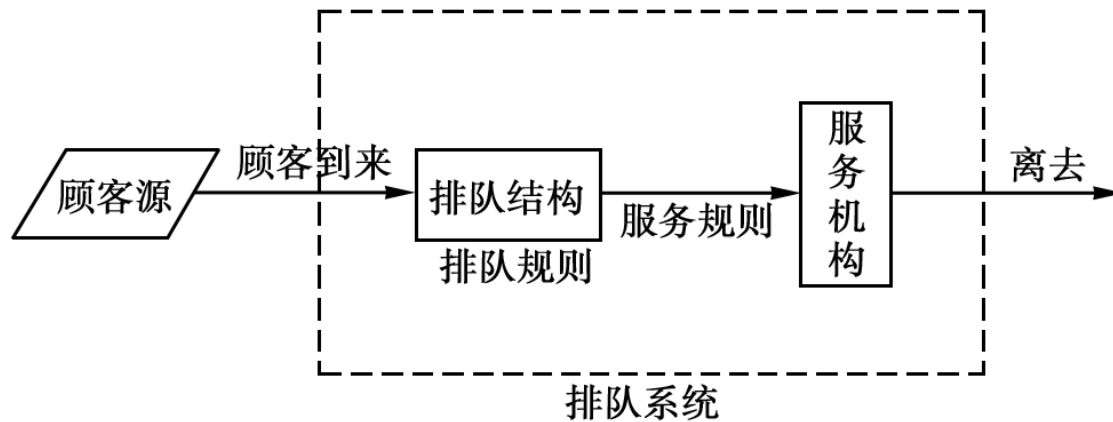
## 1.1 排队过程的一般表示

不同的顾客与服务组成了各式各样的服务系统。顾客为了得到某种服务而到达系统、若不能立即获得服务而又允许排队等待，则加入队列排队等待接受服务，然后服务台按一定规则从队列中选择顾客进行服务，获得服务的顾客立即离开系统。



# 1.1 排队过程的一般表示

排队过程的一般模型



- ❖ 各个顾客由顾客源(总体)出发，到达服务机构(服务台、服务员)前排队等候接受服务，服务完成后离开。
- ❖ 排队结构指队列的数目和排列方式，排队规则和服务规则是说明顾客在排队系统中按怎样的规则、次序接受服务的。

## 1.1 排队过程的一般表示

### 形形色色的排队系统

到达的顾客	要求服务内容	服务机构
1.不能运转的机器 2.修理技工	修理 领取修配零件	修理技工 发放修配零件的管 理员
3.病人 4.电话呼唤 5.文件稿 6.提货单 7.到达机场上空的飞机 8.驶入港口的货船 9.上游河水进入水库 10.进入我方阵地的敌机	诊断或动手术 通话 打字 提取存货 降落 装(卸)货装(卸) 放水, 调整水位 我方高射炮进行射击	医生(或包括手术台) 交换台 打字员 仓库管理员 跑道 货码头(泊位) 水闸管理员 我方高射炮



## 1.2 排队系统的组成和特征

实际的排队系统虽然千差万别，但是它们有以下的共同特征：

- (1) 有请求服务的人或物——顾客；
- (2) 有为顾客服务的人或物，即服务员或服务台；
- (3) 顾客到达系统的时刻是随机的，为每一位顾客提供服务的时间是随机的，因而整个排队系统的状态也是随机的。排队系统的这种随机性造成某个阶段顾客排队较长，而另外一些时候服务员(台)又空闲无事。



## 1.2 排队系统的组成和特征

---

- ❖ 排队系统由三个基本部分组成：
  - ① 输入过程
  - ② 排队规则
  - ③ 服务机构

## 1.2 排队系统的组成和特征

### 输入过程

❖ 输入即指顾客到达排队系统。输入过程是指要求服务的顾客是按怎样的规律到达排队系统的过程，有时也把它称为顾客流。

❖ 一般可以从以下几个方面来描述一个输入过程

(1) 顾客的总体数，又称顾客源、输入源。这是指顾客的来源。

顾客源可以是有限的，也可以是无限的。

例如，到售票处购票的顾客总数可以认为是无限的；上游河水流入水库可以认为顾客总体是无限的

例如，某个工厂因故障待修的机床则是有限的。

## 1.2 排队系统的组成和特征

### 输入过程

- (2) 顾客到来的方式。这是描述顾客是怎样来到系统的，他们是单个到达，还是成批到达。

病人到医院看病是顾客单个到达的例子。

在库存问题中如将生产器材进货或产品入库看作是顾客，那么这种顾客则是成批到达的。

## 1.2 排队系统的组成和特征

### 输入过程

(3) 顾客流的概率分布，或称相继顾客到达的时间间隔的分布。这是求解排队系统有关运行指标问题时，首先需要确定的指标。这也可以理解为在一定的时间间隔内到达 $K$ 个顾客( $K=1, 2, \dots$ )的概率是多大。

顾客相继到达的间隔时间可以是确定型的，也可以是随机型的。

例如：在流水线上装配的各部件必须按确定的时间间隔到达装配点，定点运行的列车、班机的到达也都是确定的；

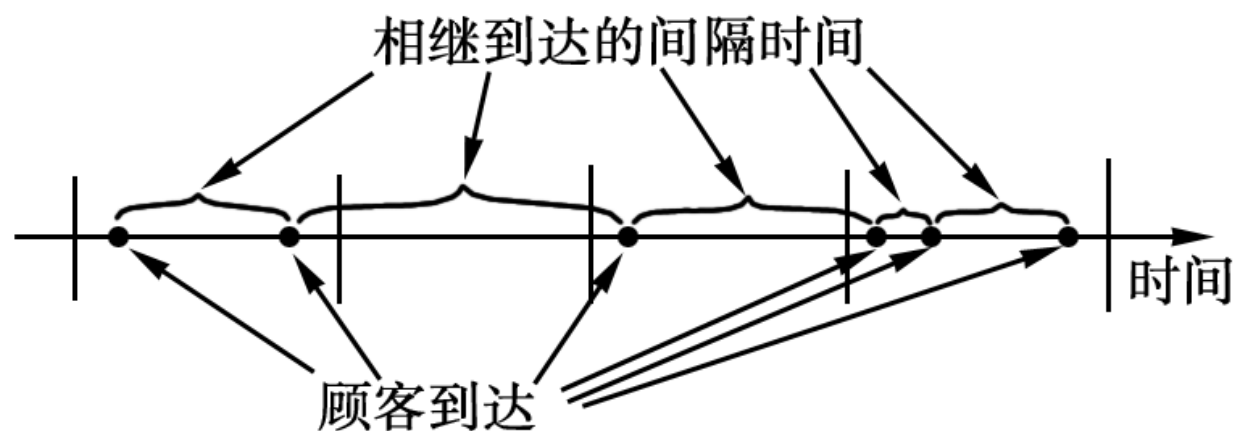
例如：物流配送等待的顾客、办理出关手续的顾客、通过路口的车辆的到达都是随机的。

## 1.2 排队系统的组成和特征

### 输入过程

对于随机的情形，必须了解单位时间的顾客到达数或相继到达的时间间隔的概率分布。

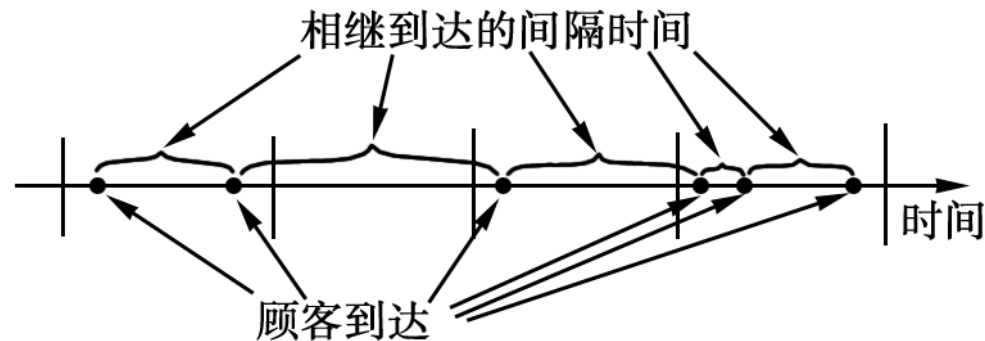
顾客流的概率分布一般有定长分布、二项分布、泊松流（最简单流）、爱尔朗分布等若干种。



## 1.2 排队系统的组成和特征

### 输入过程

- (4) 顾客的到达可以是相互独立的。
- (5) 输入过程可以是平稳的，或称对时间是齐次的，即描述相继到达的间隔时间分布和所含参数(如期望值、方差等)都是与时间无关的。



## 1.2 排队系统的组成和特征

### 排队规则

- ❖ 这是指服务台从队列中选取顾客进行服务的顺序。一般可以分为损失制、等待制和混合制等3大类。
- ❖ (1) 损失制。这是指如果顾客到达排队系统时，所有服务台都已被先来的顾客占用，那么他们就自动离开系统永不再来。

例如：电话拨号后出现忙音，顾客不愿等待而自动挂断电话，如要再打，就需重新拨号，这种服务规则即为损失制。



## 1.2 排队系统的组成和特征

### 排队规则

(2) 等待制。这是指当顾客来到系统时，所有服务台都不空，顾客加入排队行列等待服务。

例如：排队等待售票，故障设备等待维修等。

对于等待制，为顾客进行服务的次序可以采用下列各种规则：

- 先到先服务(FCFS)
- 后到先服务(LCFS)
- 随机服务(RS)
- 有优先权的服务

## 1.2 排队系统的组成和特征

### 排队规则

#### (2) 等待制（续）。

① 先到先服务。按顾客到达的先后顺序对顾客进行服务，这是最普遍的情形。

② 后到先服务。

例如：仓库中迭放的钢材，后迭放上去的都先被领走。

③ 随机服务。即当服务台空闲时，不按照排队序列而随意指定某个顾客去接受服务。

例如：电话交换台接通呼叫电话。

④ 优先权服务。

例如：老人、儿童先进车站；

危重病员先就诊；

遇到重要数据需要处理计算机立即中断其他数据的处理等。

## 1.2 排队系统的组成和特征

### 排队规则

(3) 混合制. 这是等待制与损失制相结合的一种服务规则，一般是指允许排队，但又不允许队列无限长下去。具体说来，大致有三种：

- ① 队长有限。
- ② 等待时间有限。
- ③ 逗留时间有限。

## 1.2 排队系统的组成和特征

### 排队规则

#### (3) 混合制

① 队长有限。当排队等待服务的顾客人数超过规定数量时，后来的顾客就自动离去，另求服务，即系统的等待空间是有限的。

具体地，最多只能容纳 $K$ 个顾客在系统中，当新顾客到达时，若系统中的顾客数（又称为队长）小于 $K$ ，则可进入系统排队或接受服务；否则，便离开系统，并不再回来。

例如：水库的库容是有限的，旅馆的床位是有限的。

## 1.2 排队系统的组成和特征

### 排队规则

#### (3) 混合制

- ① 队长有限。
- ② 等待时间有限。即顾客在系统中的等待时间不超过某一给定的长度  $T$ ，当等待时间超过  $T$  时，顾客将自动离去，并不再回来。

例如：易损坏的电子元器件的库存问题，超过一定存储时间的元器件被自动认为失效。

例如：顾客到饭馆就餐，等了一定时间后不愿再等而自动离去另找饭店用餐。

## 1.2 排队系统的组成和特征

### 排队规则

#### (3) 混合制

- ① 队长有限。
- ② 等待时间有限。
- ③ 逗留时间(等待时间与服务时间之和)有限。

例如：用高射炮射击敌机，当敌机飞越高射炮射击有效区域的时间为  $t$  时，若在这个时间内未被击落，也就不可能再被击落了。

不难注意到，损失制和等待制可看成是混合制的特殊情形，如记  $s$  为系统中服务台的个数，则当  $K=s$  时，混合制即成为损失制；当  $K=\infty$  时，混合制即成为等待制。

## 1.2 排队系统的组成和特征

### 排队规则 (续)

- ❖ 从允许排队的空间看
  - 队列可以排在具体的处所，也可以是抽象的。
  - 排队空间可以有限，也可以无限。
- ❖ 从排队的队列数目看，可以是单列，也可以是多列。
  - 在多列的情形，各列间的顾客有的可以互相转移，有的不能。
  - 有的排队顾客因等候时间过长而中途退出，有的不能退出，必须坚持到被服务为止。

## 1.2 排队系统的组成和特征

### 服务机构 (服务台情况)

服务台可以从以下3方面来描述:

#### (1) 服务台数量及构成形式。

服务机构可以没有服务员,也可以有一个或多个服务员(服务台、通道、窗口等)。

从数量上说,服务台有单服务台和多服务台之分。在有多个服务台的情形中,可以是平行排列的,也可以是前后排列的,或混合排列的。



## 1.2 排队系统的组成和特征

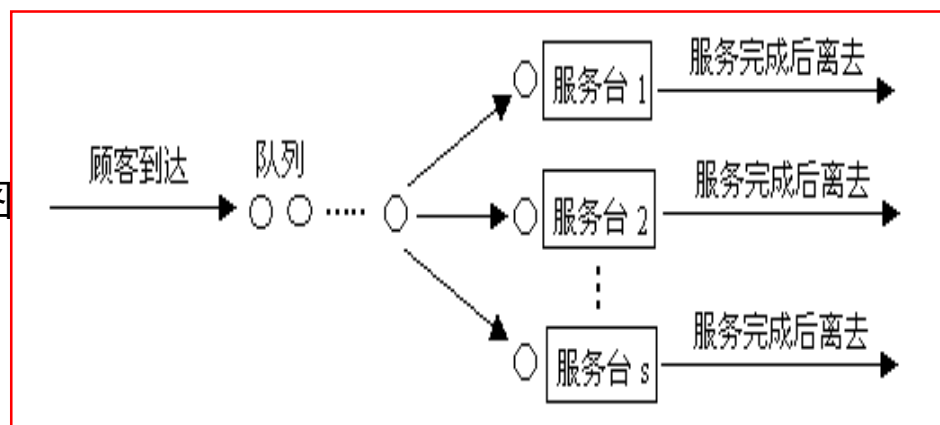
### 服务机构(服务台情况)

服务台可以从以下3方面来描述:

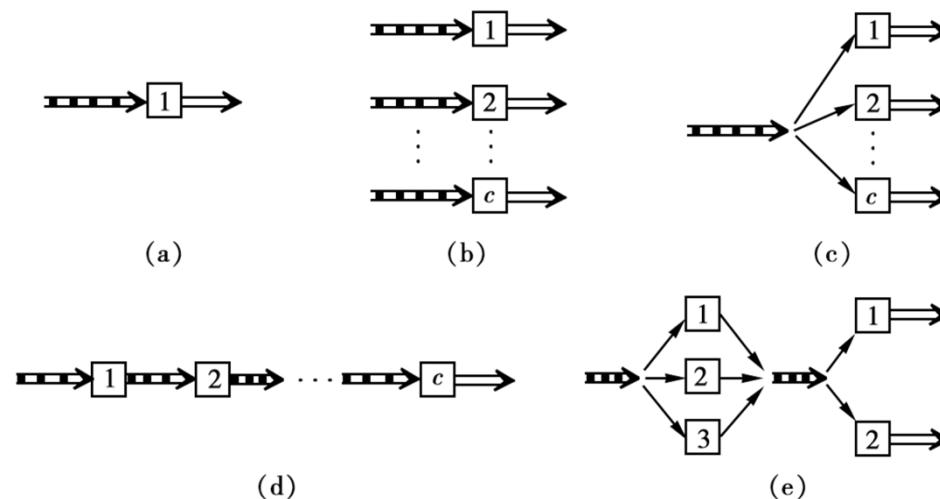
(1) 服务台数量及构成形式。

从构成形式上看, 服务台有:

- ①单队——单服务台式; 如(a)图
- ②单队——多服务台并联式; 如(c)图



服务台的各种排列方式



## 1.2 排队系统的组成和特征

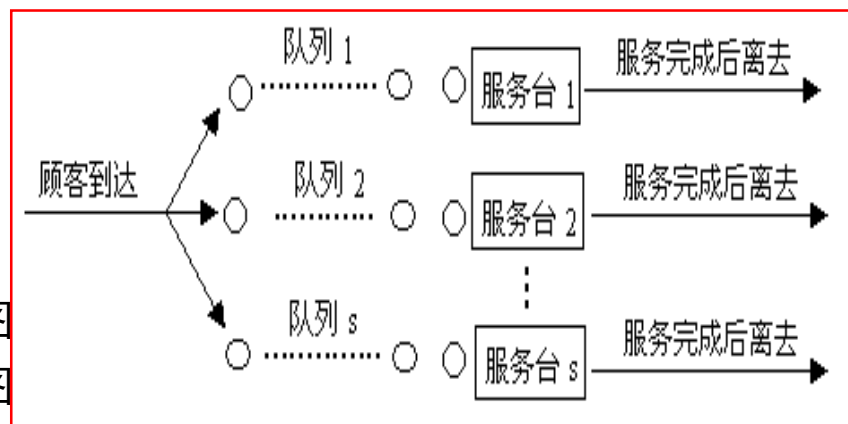
### 服务机构(服务台情况)

服务台可以从以下3方面来描述:

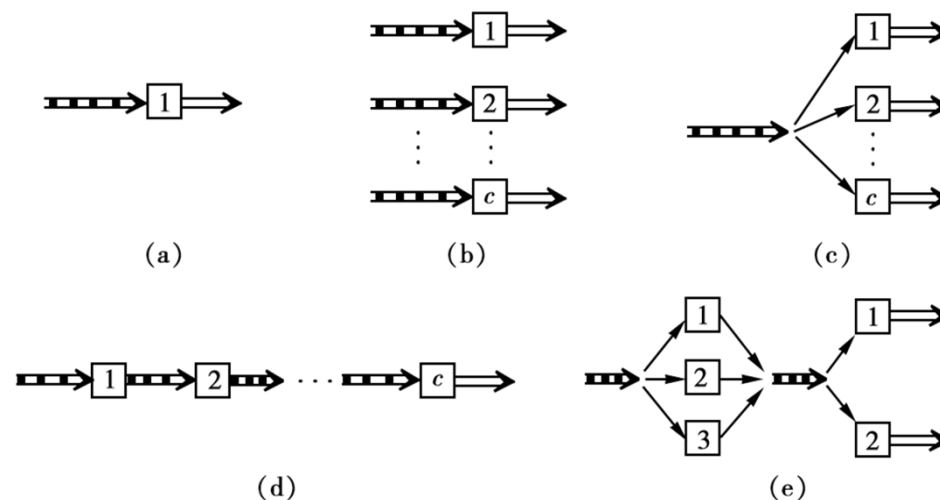
(1) 服务台数量及构成形式。

从构成形式上看, 服务台有:

- ①单队——单服务台式; 如(a)图
- ②单队——多服务台并联式; 如(c)图
- ③多队——多服务台并联式; 如(b)图



服务台的各  
种排列方式



## 1.2 排队系统的组成和特征

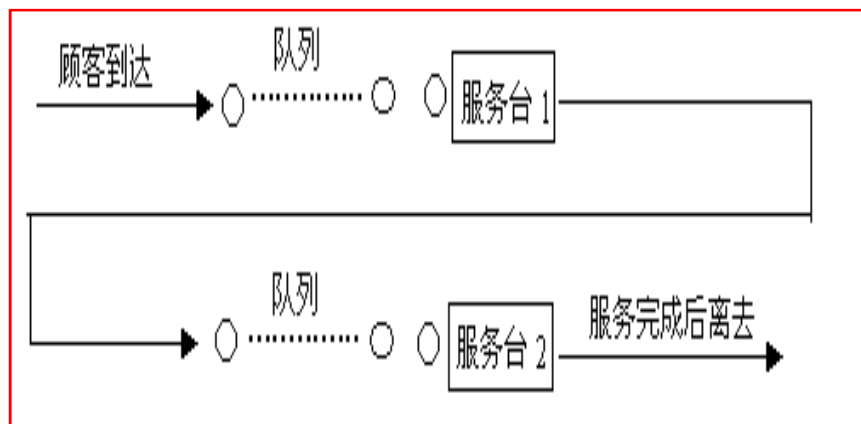
### 服务机构(服务台情况)

服务台可以从以下3方面来描述:

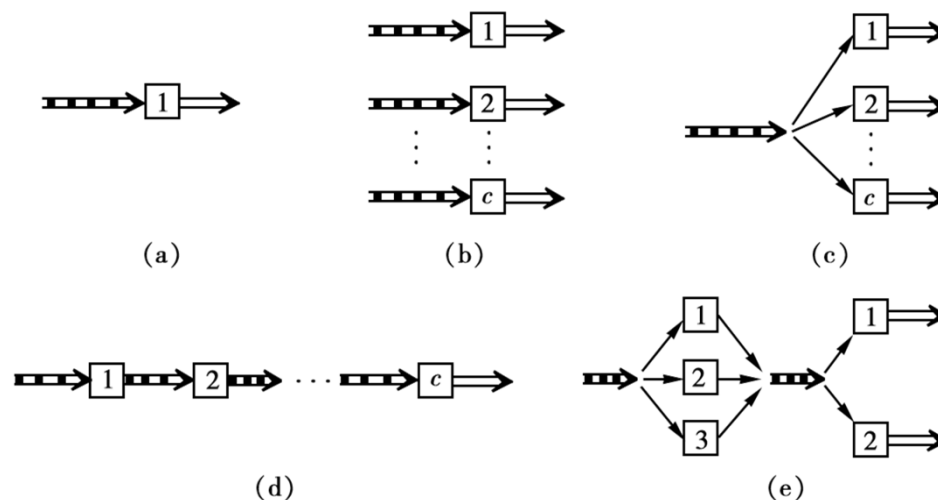
(1) 服务台数量及构成形式。

从构成形式上看, 服务台有:

- ①单队——单服务台式; 如(a)图
- ②单队——多服务台并联式; 如(c)图
- ③多队——多服务台并联式; 如(b)图
- ④单队——多服务台串联式; 如(d)图



服务台的各  
种排列方式



## 1.2 排队系统的组成和特征

### 服务机构(服务台情况)

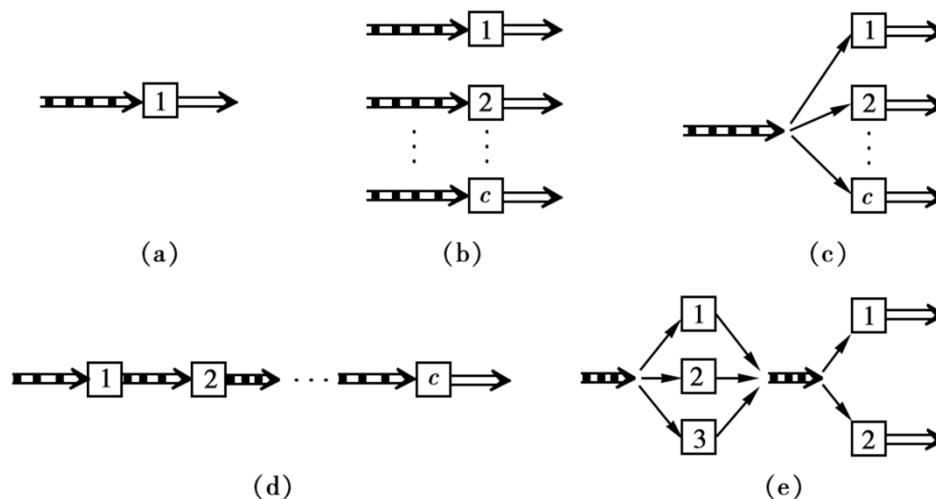
服务台可以从以下3方面来描述:

(1) 服务台数量及构成形式。

从构成形式上看, 服务台有:

- ①单队——单服务台式; 如(a)图
- ②单队——多服务台并联式; 如(c)图
- ③多队——多服务台并联式; 如(b)图
- ④单队——多服务台串联式; 如(d)图
- ⑤单队——多服务台并串联混合式;

服务台的各  
种排列方式



## 1.2 排队系统的组成和特征

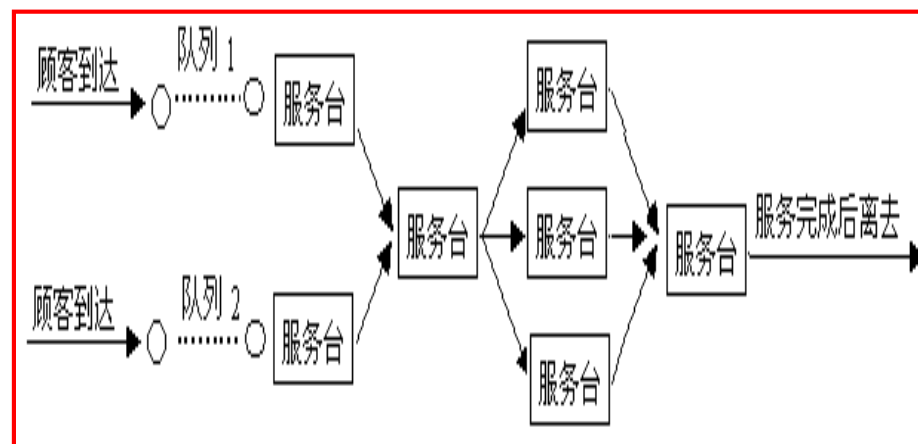
### 服务机构(服务台情况)

服务台可以从以下3方面来描述:

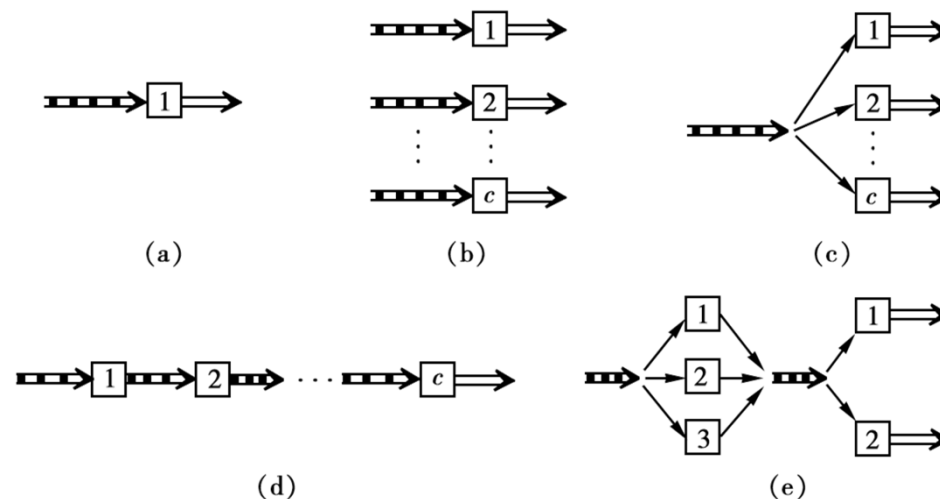
#### (1) 服务台数量及构成形式。

从构成形式上看, 服务台有:

- ①单队——单服务台式; 如(a)图
- ②单队——多服务台并联式; 如(c)图
- ③多队——多服务台并联式; 如(b)图
- ④单队——多服务台串联式; 如(d)图
- ⑤单队——多服务台并串联混合式;
- ⑥多队——多服务台并串联混合式等等。



### 服务台的各种排列方式



## 1.2 排队系统的组成和特征

### 服务机构(服务台情况)

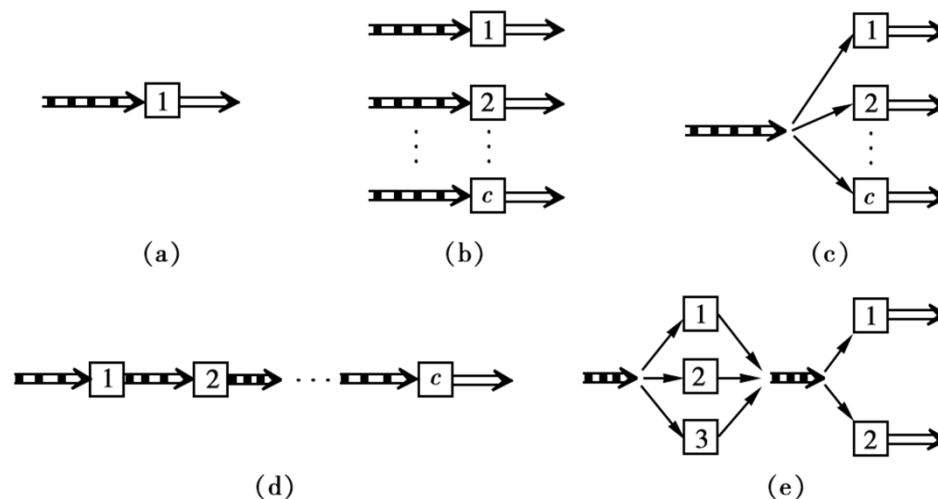
服务台可以从以下3方面来描述:

(1) 服务台数量及构成形式。

从构成形式上看, 服务台有:

- ①单队——单服务台式; 如(a)图
- ②单队——多服务台并联式; 如(c)图
- ③多队——多服务台并联式; 如(b)图
- ④单队——多服务台串联式; 如(d)图
- ⑤单队——多服务台并串联混合式;
- ⑥多队——多服务台并串联混合式等等。

服务台的各  
种排列方式



## 1.2 排队系统的组成和特征

### 服务机构 (服务台情况)

- (2) **服务方式**。这是指在某一时刻接受服务的顾客数，它有单个服务和成批服务两种。如公共汽车一次就可装载一批乘客就属于成批服务。
  - (3) **服务时间的分布**。服务时间可分为**确定型**和**随机型**。一般来说，在多数情况下，对每一个顾客的服务时间是一随机变量，其概率分布有定长分布、负指数分布、 $K$ 级爱尔良分布、一般分布(所有顾客的服务时间都是独立同分布的)等等。
- ❖ 服务时间的分布通常假定是平稳的。  
指时间间隔分布及其特征参数（数学期望、方差等）不随时间的变化而变化。

## 1.3 排队模型的分类

### 排队模型分类方法——D.G.Kendall, 1953

- 构成排队模型的三个主要特征指标
  - (1) 相继顾客到达间隔时间的分布;
  - (2) 服务时间的分布;
  - (3) 服务台的个数。
- 根据这三个特征对排队模型进行分类的Kendall记号:

**$X/Y/Z$**

- **X**: 表示相继到达间隔时间的分布;
- **Y**: 表示服务时间的分布;
- **Z**: 并列的服务台的数目。



## 1.3 排队模型的分类

表示相继到达间隔时间和服务时间的各种分布符号

- ❖ M——负指数分布(M是Markov的字头, 因为负指数分布具有无记忆性, 即Markov性)
- ❖ D——确定型(deterministic)
- ❖  $E_k$ ——k阶爱尔朗(erlang)分布
- ❖ GI——一般相互独立(general independent)的时间间隔的分布
- ❖ G——一般(general)服务时间的分布

## 1.3 排队模型的分类

### ❖ Kendall符号的扩充

**X/Y/Z/A/B/C**

其中前三项的意义不变，后三项的意义分别是：

- **A**: 系统容量限制 $N$ ，或称等待空间容量。如系统有 $N$ 个等待位子，则  $0 < N < \infty$ 。当  $N = 0$  时，说明系统不允许等待，即为损失制。 $N = \infty$  时为等待制系统，此时 $\infty$ 般省略不写。 $N$ 为有限整数时，表示为混合制系统。
- **B**: 顾客源数目 $m$ 。分有限与无限两种， $\infty$ 表示顾客源无限，此时一般 $\infty$ 也可省略不写。
- **C**: 服务规则，如先到先服务(FCFS)，后到后服务(LCFS)，优先权服务(PR)等。

## 1.3 排队模型的分类

### ❖ Kendall符号的扩充

**$X/Y/Z/A/B/C$**

某些情况下，排队问题仅用上述表达形式中的前3个、4个、5个符号。如不特别说明则均理解为系统等待空间容量无限；顾客源无限，先到先服务，单个服务的等待制系统。

约定：如略去后三项，即指 $X/Y/Z/\infty/\infty/\text{FCFS}$ 的情形。

例如：某排队问题为 $M/M/S/\infty/\infty/\text{FCFS}/$ ，则表示顾客到达间隔时间为负指数分布(泊松流)；服务时间为负指数分布；有 $s(s > 1)$ 个服务台；系统等待空间容量无限(等待制)；顾客源无限，采用先到先服务规则。

## 1.3 排队模型的分类

### ❖ Kendall符号的扩充

**X/Y/Z/A/B/C**

- ❖ M/M/c/ $\infty$ 表示输入过程是负指数分布，服务时间服从负指数分布，系统有c个服务台平行服务（ $0 < c \leq \infty$ ），系统容量为无穷，系统是等待制系统。
- ❖ M/G/1/ $\infty$ 表示输入过程是负指数分布，顾客所需的服务时间为独立、服从一般概率分布，系统中只有一个服务台，容量为无穷的等待制系统。
- ❖ GI/M/1/ $\infty$ 表示输入过程是负指数分布，顾客独立到达且相继到达的间隔时间服从一般概率分布，服务时间是相互独立、服从负指数分布，系统中只有一个服务台，容量为无穷的等待制系统。
- ❖ Ek/G/1/K表示相继到达的间隔时间独立、服从k阶爱尔朗分布，服务时间为独立、服从一般概率分布，系统中只有一个服务台，容量为K（ $1 \leq K < \infty$ ）的混合制系统；
- ❖ D/M/c/K表示相继到达的间隔时间独立、服从定长分布，服务时间相互独立、服从负指数分布，系统中有c个服务台平行服务，容量为K（ $c \leq K < \infty$ ）的混合制系统。

## 1.4 排队问题的求解

- ❖ 首先需要确定属于哪种排队模型，其中**顾客到达的间隔时间分布和服务时间的分布**需要实测的数据来确定
- ❖ 确定用以判断系统运行优劣的基本数量指标，求出这些数量指标的概率分布或特征数

## 1.4 排队问题的求解

### ❖ 用于描述排队系统运行状况的基本数量指标

- (1) 队长：系统中的顾客数，是排队等待的顾客数与正在接受服务的顾客数之和，期望值记作 $L_s$ ；

排队长（队列长）：系统中排队等待服务的顾客数，期望值记作 $L_q$ ；

$$\begin{bmatrix} \text{系统中} \\ \text{顾客数} \end{bmatrix} = \begin{bmatrix} \text{在队列中等待} \\ \text{服务的顾客数} \end{bmatrix} + \begin{bmatrix} \text{正在服务} \\ \text{的顾客数} \end{bmatrix}$$

队长和排队长一般都是随机变量。我们希望能确定它们的分布，或至少能确定它们的平均值（即平均队长和平均排队长）及有关的矩（如方差等）。队长的分布是顾客和服务员都关心的，特别是对系统设计人员来说，如果能知道队长的分布，就能确定队长超过某个数的概率，从而确定合理的等待空间。

## 1.4 排队问题的求解

### ❖ 用于描述排队系统运行状况的基本数量指标

(2) **逗留时间**：顾客在系统中的停留时间，从顾客到达时刻起到他接受服务完成止这段时间，期望值记作 $W_s$ ；是随机变量，也是顾客最关心的指标之一。

**等待时间**：顾客在系统中排队等待的时间，从顾客到达时刻起到他开始接受服务止这段时间，期望值记作 $W_q$ ，是随机变量，也是顾客最关心的指标，因为顾客通常希望等待时间越短越好。

$$[\text{逗留时间}] = [\text{等待时间}] + [\text{服务时间}]$$

对这两个指标的研究当然是希望能确定它们的分布，或至少能知道顾客的平均等待时间和平均逗留时间。

## 1.4 排队问题的求解

### ❖ 用于描述排队系统运行状况的基本数量指标

(3) **忙期**：指从顾客到达空闲服务机构起，到服务机构再次为空闲止的时间长度，即服务机构连续忙的时间。

这是个随机变量，是服务员最为关心的指标，因为它关系到服务员的服务强度。

**闲期**：即服务机构连续保持空闲的时间。

在排队系统中，忙期和闲期总是交替出现的。

其他一些指标，如在损失制或系统容量有限的情况下，由于顾客被拒绝，而使服务系统受到损失的**顾客损失率及服务强度**等



## 1.4 排队问题的求解

### ❖ 系统状态的概率分布

- 系统的状态即指系统中的顾客数，它的可能取值是：
  - (1) 队长没有限制时， $n=0, 1, 2, \dots$
  - (2) 队长有限制、最大数为 $N$ 时， $n=0, 1, 2, \dots, N$
  - (3) 即时制且服务台个数为 $c$ 时， $n=0, 1, 2, \dots, c$
- 系统处于这些状态的概率一般是随时间 $t$ 变化的，所以在时刻 $t$ 、系统状态为 $n$ 的概率可以用 $P_n(t)$ 表示。

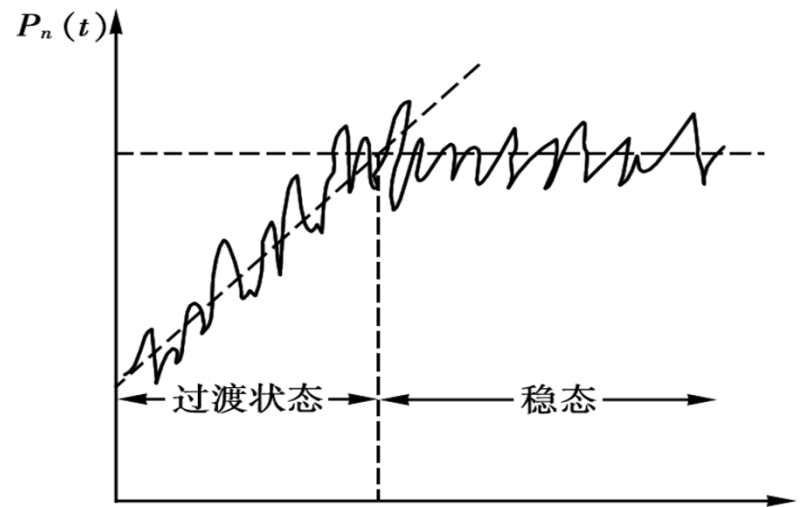
## 1.4 排队问题的求解

❖ 求状态概率 $P_n(t)$ 的方法：

- 建立含 $P_n(t)$ 的关系式，该关系式一般是包含 $P_n(t)$ 的微分差分方程(关于 $t$ 的微分方程，关于 $n$ 的差分方程)。
- 该方程的解称为瞬态(或称过渡状态)(transient state)解。
- 它的极限

$$\lim_{t \rightarrow \infty} P_n(t) = P_n$$

称为稳态(steady state)解，或称统计平衡状态(statistical equilibrium state)解。



## 1.4 排队问题的求解

### ❖ 稳态解的物理意义

- 当系统运行了无限长时间后，初始( $t=0$ )状态的概率分布( $P_n(0)$ ,  $n \geq 0$ )的影响将消失，系统状态的概率分布不再随时间而变化。
- 在实际应用中，大多数系统会很快趋于稳态，而无需等到 $t \rightarrow \infty$ 以后。

## 1.4 排队问题的求解

上面给出的这些数量指标一般都是和系统运行的时间有关的随机变量，求这些随机变量的瞬时分布一般是很困难的。

为了分析上的简便，并注意到相当一部分排队系统在运行了一定时间后，都会趋于一个平衡状态(或称平稳状态)。在平衡状态下，队长的分布、等待时间的分布和忙期的分布都和系统所处的时刻无关，而且系统的初始状态的影响也会消失。因此，将主要讨论与系统所处时刻无关的性质，即统计平衡性质。

# 排队论

- ①系统中顾客数(队长)的期望值 $L$ 或 $L_s$ ;
- ②排队等待的顾客数(排队长)的期望值 $L_q$ ;
- ③顾客在系统中全部时间(逗留时间)的期望值 $W$ 或 $W_s$ ;
- ④顾客排队等待时间的期望值 $W_q$ 。

排队系统中, 由于顾客到达分布和服务时间分布是多种多样的, 加之服务台数。顾客源有限无限, 排队容量有限无限等的不同组合, 就会有不胜枚举的不同排队模型, 若对所有排队模型都进行分析与计算, 不但十分繁杂而且也没有必要。

下面拟分析两种常见排队系统模型。

# 提纲

---

- ❖ 基本概念
- ❖ M/M/1模型
- ❖ M/M/C模型
- ❖ 排队系统的最优化问题
- ❖ 分析排队系统的随机模拟法

## 2 标准的M/M/1模型(M/M/1/ $\infty$ / $\infty$ )

### ❖ 标准的M/M/1模型

- (1) 输入过程——顾客源是无限的，顾客单个到来，相互独立，一定时间内的到达数服从泊松分布，到达过程是平稳的。
- (2) 排队规则——单队，且对队长没有限制，先到先服务。
- (3) 服务机构——单服务台，各顾客的服务时间相互独立，服从相同的负指数分布。
- 此外，还假定到达间隔时间和服务时间是相互独立的。

#### 即已知条件：

- ❖  $\lambda$ ：顾客进入系统的平均速度；单位时间到达的顾客数。
- ❖  $\mu$ ：顾客离开系统的平均速度；单位时间能被服务完成的顾客数。

对于M / M / 1模型有如下公式：

$$P_0 = 1 - \rho \quad P_n = \rho^n (1 - \rho)$$

$$L = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1 - \rho} \quad L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{1 - \rho} = L\rho$$

$$W = \frac{1}{\mu - \lambda} \quad W_q = \frac{\lambda}{\mu(\mu - \lambda)} = W\rho$$

$$P(N > k) = \rho^{k+1} \quad \rho = \frac{\lambda}{\mu}$$



四个指标的关系为(Little 公式):

$$L_s = \frac{\lambda}{\mu - \lambda}$$

$$L_q = L_s - \rho = \frac{\rho^2}{1 - \rho} = \frac{\rho \lambda}{\mu - \lambda} = \rho L_s$$

$$L_s = \lambda W_s$$

$$L_q = \lambda W_q$$

服务强度

### 3. 系统的忙期与闲期

系统处于空闲状态的概率:  $P_0 = 1 - \rho$

系统处于繁忙状态的概率:  $P(N > 0) = 1 - P_0 = \rho$

$$P(N > k) = \rho^{k+1}$$

## 2 标准的M/M/1模型( $M/M/1/\infty/\infty$ )

❖ 例1 某医院手术室根据病人来诊和完成手术时间的记录，任意抽查了100个工作小时，每小时来就诊的病人数 $n$ 的出现次数如下表所示；又任意抽查了100个完成手术的病历，所用时间 $v$ (单位：小时)出现的次数如下表所示。

到达的病人数 $n$	出现人数 $f_n$
0	10
1	28
2	29
3	16
4	10
5	6
6以上	1
合计	100

为病人完成手术时间 $v$ (小时)	出现人数 $f_v$
0.0~0.2	38
0.2~0.4	25
0.4~0.6	17
0.6~0.8	9
0.8~1.0	6
1.0~1.2	5
1.2以上	0
合计	100

## 2 标准的M/M/1模型(M/M/1/ $\infty$ / $\infty$ )

- ❖ (1) 每小时病人平均到达率  $= \frac{\sum n f_n}{100} = 2.1$  (人/小时)  
每次手术平均时间  $= \frac{\sum v f_v}{100} = 0.4$  (小时/人)  
每小时完成手术人数(平均服务率)  $= \frac{1}{0.4} = 2.5$  (人/小时)
- ❖ (2) 取  $\lambda = 2.1$ ,  $\mu = 2.5$  可以通过统计检验的方法, 认为病人到达数服从参数为2.1的泊松分布, 手术时间服从参数为2.5的负指数分布。
- ❖ (3)  $\rho = \lambda \mu = \frac{2.1}{2.5} = 0.84$  说明服务机构(手术室)有84%的时间是繁忙的, 有16%的时间是空闲的。
- ❖ (4) 依次代入(12-21)式, 算出各指标:
  - 在病房中病人数(期望值)  $L_s = \frac{2.1}{2.5 - 2.1} = 5.25$  (人)
  - 排队等待病人数(期望值)  $L_q = 0.84 \times 5.25 = 4.41$  (人)
  - 病人在病房中逗留时间(期望值)  $W_s = \frac{1}{2.5 - 2.1} = 2.5$  (小时)
  - 病人排队等待时间(期望值)  $W_q = \frac{0.84}{2.5 - 2.1} = 2.1$  (小时)

## 2 标准的M/M/1模型( $M/M/1/\infty/\infty$ )

**例2:** 某修理店只有一位修理工，来修理的顾客到达过程为Poisson流，平均每小时4人；修理时间服从负指数分布，平均需要6分钟。

试求：修理店空闲的概率；店内恰有3位顾客的概率；店内至少有一位顾客的概率；在店内平均顾客数；每位在店内平均逗留时间；等待服务的平均顾客数；每位顾客平均等待服务时间。

## 2 标准的M/M/1模型(M/M/1/ $\infty$ / $\infty$ )

解：本例可看成是一个M/M/1/ $\infty$ 排队问题，  
其中 $\lambda = 4$ ， $\mu = 1/0.1 = 10$ (人/小时)， $\rho = \lambda/\mu = 2/5 < 1$

- 修理店内空闲的概率

$$P_0 = 1 - \rho = (1 - 2/5) = 0.6$$

- 店内恰有3个顾客的概率

$$P_3 = \rho^3(1 - \rho) = (2/5)^3(1 - 2/5) = 0.038$$

## 2 标准的M/M/1模型(M/M/1/ $\infty$ / $\infty$ )

- 店内至少有1位顾客的概率

$$P\{N \geq 1\} = 1 - P_0 = 1 - (1 - \rho) = \rho = 2/5 = 0.4$$

- 在店内平均顾客数

$$L = \rho / (1 - \rho) = (2/5) / (1 - 2/5) = 0.67(\text{人})$$

- 每位顾客在店内平均逗留时间

$$W = L / \lambda = 0.67 / 4 = 10 \text{ 分钟}$$

## 2 标准的M/M/1模型(M/M/1/ $\infty$ / $\infty$ )

- 等待服务的平均顾客数

$$L_q = L - \rho = 0.67 - 2/5 = 0.27(\text{人})$$

- 每个顾客平均等待服务时间

$$W_q = L_q / \lambda = 0.27 / 4 = 0.0675 \text{ 小时} \\ = 4 \text{ 分钟}$$

## 2.2 系统的容量有限制的情况( $M/M/1/N/\infty$ )

- ❖ 如果系统的最大容量为 $N$ ，对于单服务台的情形，排队等待的顾客最多为 $N-1$ ，在某时刻一顾客到达时，如系统中已有 $N$ 个顾客，那么这个顾客就被拒绝进入系统。



- ❖ 当 $N=1$ 时为即时制的情形；当 $N \rightarrow \infty$ 时为容量无限制的情形。



## 2.2 系统的容量有限制的情况(M/M/1/N/∞)

❖ 令  $\rho = \lambda / \mu$  , 得到

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}} \quad \rho \neq 1$$

$$P_n = \frac{1-\rho}{1-\rho^{N+1}} \rho^n \quad n \leq N \quad (12-24)$$

❖ M/M/1/N/∞排队系统的各项指标:

❖ (1) 队长(期望值)

$$L_s = \sum_{n=0}^N n P_n = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} \quad \rho \neq 1$$

❖ (2) 队列长(期望值)

$$L_q = \sum_{n=1}^N (n-1) P_n = L_s - (1-P_0)$$

❖ (3) 顾客逗留时间(期望值)

$$W_s = \frac{L_s}{\mu(1-P_0)} = \frac{L_q}{\lambda(1-P_N)} + \frac{1}{\mu}$$

❖ (4) 顾客等待时间(期望值)

$$W_q = W_s - 1/\mu$$

❖ (5) 有效到达率

$$\lambda_e = \lambda(1-P_N)$$

## 2.2 系统的容量有限制的情况( $M/M/1/N/\infty$ )

例4. 某单人理发馆共有六把椅子接待顾客排队，无座时将离去，顾客平均到达率为3人/h，理发时间平均为15分钟，求：

- (1) 求某一顾客到达就能理发的概率；
- (2) 求需要等待的顾客数的期望值；
- (3) 求有效到达率；
- (4) 求一顾客在系统中的逗留时间和排队时间平均值？

## 2.2 系统的容量有限制的情况(M/M/1/N/∞)

(1) 求某一顾客到达就能理发的概率:

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{1-0.75}{1-0.75^8} = 0.2778$$

(2) 求需要等待的顾客数的期望值:

$$L_s = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} = \frac{0.75}{0.25} - \frac{8 \times 0.75^8}{1-0.75^8} = 2.11$$

$$L_q = L_s - \rho = 2.11 - (1 - P_0) = 2.11 - (1 - 0.2778) = 1.39$$

(3) 求有效到达率:

$$\lambda_e = \mu(1 - P_0) = 4 \times (1 - 0.2778) = 2.89$$

(4) 求一顾客在系统中的逗留时间和排队时间平均值:

## 2.2 系统的容量有限制的情况(M/M/1/N/∞)

$$W_s = \frac{L_s}{\lambda_e} = \frac{2.11}{2.89} = 0.73h = 43.8 \text{ min}$$

$$W_q = \frac{L_q}{\lambda_e} = \frac{1.39}{2.89} = 0.48h = 28.86 \text{ min}$$

## 2.2 系统的容量有限制的情况(M/M/1/N/∞)

❖ 例4 某单人理发馆为等待的顾客准备了6把椅子，当6把椅子都坐满时，再来的顾客将不进店而离开。顾客的平均到达率为3人/小时，理发平均需要15分钟。则系统的容量为 $N=6+1=7$ ， $\lambda = 3$ 人/小时， $\mu = 4$ 人/小时。

- (1) 某顾客一到达就能理发的概率

$$P_0 = \frac{1 - 3/4}{1 - (3/4)^8} = 0.2778$$

- (2) 需要等待的顾客数的期望值

$$L_s = \frac{3/4}{1 - 3/4} - \frac{8(3/4)^8}{1 - (3/4)^8} = 2.11$$

$$L_q = L_s - (1 - P_0) = 2.11 - (1 - 0.2778) = 1.39$$

- (3) 有效到达率

$$\lambda_e = \mu(1 - P_0) = 4(1 - 0.2778) = 2.89 \text{ (人/小时)}$$

- (4) 一顾客在理发馆内逗留的期望时间

$$W_s = L_s / \lambda_e = 2.11 / 2.89 = 0.73 \text{ (小时) (分钟)}$$

## 2.2 系统的容量有限制的情况(M/M/1/N/∞)

以本例为例，比较队长为有限和无限的情形：

$\lambda = 3$ 人/小时 $\mu = 4$ 人/小时	$L_s$	$L_q$	$W_s$	$W_q$	$P_0$	$P_7$
有限队长 $N = 7$	2.11	1.39	0.73	0.48	0.278	3.7
无限队长	3	2.25	1.0	0.75	0.25	0

# 提纲

---

- ❖ 基本概念
- ❖ M/M/1模型
- ❖ M/M/C模型
- ❖ 排队系统的最优化问题
- ❖ 分析排队系统的随机模拟法

## 第3节 多服务台负指数分布排队系统的分析

本节讨论单队、并列的多服务台（服务台数为 $c$ ）的情形。

- ❖ 3.1 标准的M/M/c模型
- ❖ 3.2 M/M/c型系统和 $c$ 个M/M/1型系统的比较
- ❖ 3.3 系统容量有限的情形（M/M/c/N/ $\infty$ ）
- ❖ 3.4 顾客源有限的情形（M/M/c/ $\infty$  /m）



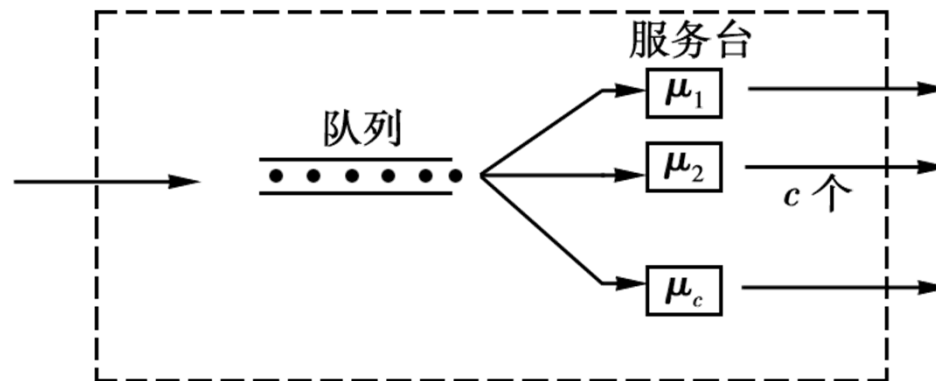
## 3.1 标准的M/M/c模型

- ❖ 标准的M/M/c模型各种特征的规定与标准的M/M/1模型的规定相同。
- ❖ 各服务台工作是相互独立的(不搞协作), 且平均服务率相同。

$$\mu_1 = \mu_2 = \cdots = \mu_c = \mu$$

- ❖ 整个服务机构的平均服务率为  $c\mu$  (当  $n \geq c$ ); 为  $n\mu$  (当  $n < c$ )。
- ❖ 令  $\rho = \frac{\lambda}{c\mu}$ , 只有当  $\frac{\lambda}{c\mu} < 1$  时才不会排成无限的队列

称  $\rho = \frac{\lambda}{c\mu}$  为系统的服务强度或服务机构的平均利用率。



对于M / M / C 模型有如下公式：

$$\left\{ \begin{array}{l} P_0 = \left[ \sum_{k=0}^{c-1} \frac{1}{k!} \left( \frac{\lambda}{\mu} \right)^k + \frac{1}{c!} \cdot \frac{1}{1-\rho} \cdot \left( \frac{\lambda}{\mu} \right)^c \right]^{-1} \\ P_n = \begin{cases} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n P_0 & (n \leq c) \\ \frac{1}{c! c^{n-c}} \left( \frac{\lambda}{\mu} \right)^n P_0 & (n > c) \end{cases} \end{array} \right.$$

## 对于M / M / C 模型有如下公式：

❖ 系统的运行指标为：

❖ 平均队长

$$\begin{cases} L_s = L_q + \frac{\lambda}{\mu} \\ L_q = \sum_{n=c+1}^{\infty} (n-c)P_n = \frac{(c\rho)^c \rho}{c!(1-\rho)^2} P_0 \end{cases} \quad (12-30)$$

$$\left( \text{因为 } \sum_{n=c+1}^{\infty} (n-c)P_n = \sum_{n'=1}^{\infty} n'P_{n'+c} = \sum_{n'=1}^{\infty} \frac{n'}{c!cn'} (c\rho)^{n'+c} P_0 = \text{右边} \right)$$

❖ 平均等待时间和逗留时间可由Little公式求得，

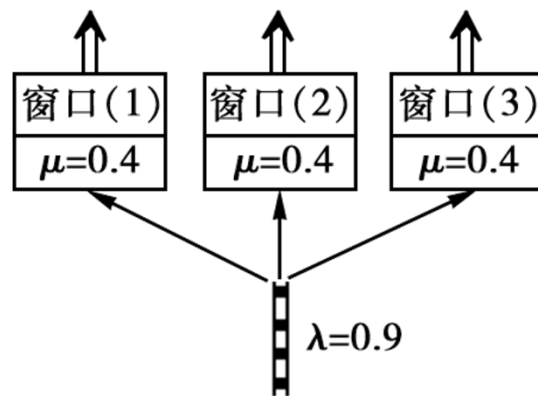
$$W_q = \frac{L_q}{\lambda}, \quad W_s = \frac{L_s}{\lambda}$$

❖ 顾客必须等待的概率（即所有服务台均被占用）为

$$p(n \geq c) = \sum_{k=c}^{\infty} p_k = \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!(1-\rho)} p_0$$

## 3.1 标准的M/M/c模型

- ❖ 例6 某售票处有三个窗口，顾客的到达服从泊松过程，平均到达率每分钟  $\lambda = 0.9$  (人)，服务(售票)时间服从负指数分布，平均服务率每分钟  $\mu = 0.4$  人。现设顾客到达后排成一队，依次向空闲的窗口购票如图12-14(a)，这就是一个M/M/c型的系统，其中  $c = 3$ ， $\frac{\lambda}{\mu} = 2.25$ ， $\rho = \frac{\lambda}{c\mu} = 2.253(<1)$  符合要求的条件，代入公式得：



(a)

## 3.1 标准的M/M/c模型

- ❖ (1) 整个售票处空闲概率

$$p_0 = \frac{1}{\frac{(2.25)^0}{0!} + \frac{(2.25)^1}{1!} + \frac{(2.25)^2}{2!} + \frac{(2.25)^3}{3!} \cdot \frac{1}{1 - 2.25/3}} = 0.0748$$

- ❖ (2) 平均队长

$$L_q = \frac{(2.25)^3 \cdot 3/4}{3!(1/4)^2} \times 0.0748 = 1.70$$

$$L_s = L_q + \lambda / \mu = 3.95$$

- ❖ (3) 平均等待时间和逗留时间

$$W_q = 1.70 / 0.9 = 1.89 \quad (\text{分钟})$$

$$W_s = 1.89 + 1 / 0.4 = 4.39 \quad (\text{分钟})$$

- ❖ 顾客到达后必须等待(即系统中顾客数已有3人即各服务台都没有空闲)的概率

$$P(n \geq 3) = \frac{(2.25)^3}{3!1/4} \times 0.0748 = 0.57$$

### 3.2 M/M/c型系统和c个M/M/1型系统的比较

例6. 某火车站售票处有三个窗口，同时售各车次的车票。顾客到达服从泊松分布，平均每分钟到达 $\lambda=0.9$ （人），服务时间服从负指数分布，平均服务率 $\mu=24$ （人/h），分两种情况：

1. 顾客排成一队，依次购票； ( $M/M/3/\infty/\infty$ )
2. 顾客在每个窗口排一队，不准串队。 ( $M/M/1/\infty/\infty$  三个系统并联)

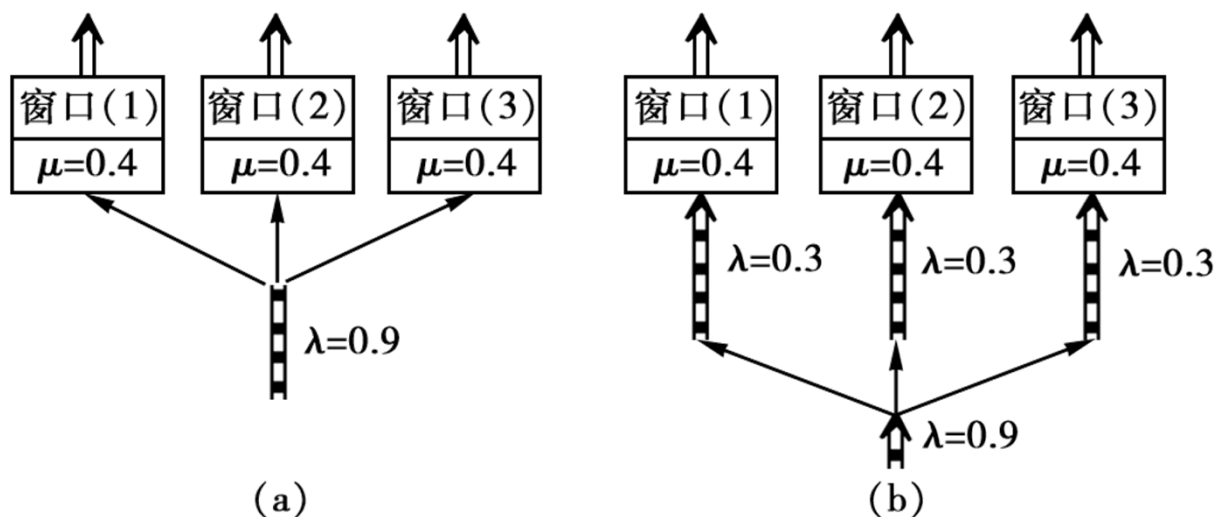
求：（1）售票处空闲的概率。  
（2）平均等待时间和逗留时间。  
（3）队长和队列长。

## 3.2 M/M/c型系统和c个M/M/1型系统的比较

- ❖ 现就例6说明，如果原题除排队方式外其他条件不变，但顾客到达后在每个窗口前各排一队，且进入队列后坚持不换，这就形成3个队列，见图12-14(b)而每个队列平均到达率为

$$\lambda_1 = \lambda_2 = \lambda_3 = 0.9 / 3 = 0.3 \quad (\text{每分钟})$$

- ❖ 这样，原来的系统就变成3个M/M/1型的子系统。  $\mu=0.4, \rho=\lambda/\mu=0.75$ ,  $P_0=1-\rho=0.25$



## 3.2 M/M/c型系统和c个M/M/1型系统的比较

- ❖ 现就例6说明，如果原题除排队方式外其他条件不变，但顾客到达后在每个窗口前各排一队，且进入队列后坚持不换，这就形成3个队列，见图12-14(b)而每个队列平均到达率为  $\lambda_1 = \lambda_2 = \lambda_3 = 0.9 / 3 = 0.3$ (每分钟)
- ❖ 这样，原来的系统就变成3个M/M/1型的子系统。现按M/M/1型解决这个问题，并与上面结果比较如下：

模型	(1) M/M/3型	(2) M/M/1型
指标		
服务台空闲的概率 $P_0$	0.0748	0.25(每个子系统)
顾客必须等待的概率	$P(n \geq 3) = 0.57$	0.75
平均队列长 $L_q$	1.70	2.25(每个子系统)
平均队长 $L_s$	3.95	9.00(整个系统)
平均逗留时间 $W_s$ (分钟)	4.39	10
平均等待时间 $W_q$ (分钟)	1.89	7.5

从表中各指标的对比可以看出：单队比三队有显著优越性。



### 3.3 系统的容量有限制的情形 (M/M/c/N/∞)

- ❖ 设系统的容量最大限制为 $N(\geq c)$ ，当系统中顾客数 $n$ 已达到 $N$ (即队列中顾客数已达 $N-c$ )时，再来的顾客即被拒绝，其他条件与标准的 $M/M/c$ 型相同。
- ❖ 这时系统的状态概率和运行指标如下：

$$\begin{cases} p_0 = \left[ \sum_{k=0}^{c-1} \frac{1}{k!} \left( \frac{\lambda}{\mu} \right)^k + \frac{1}{c!} \cdot \frac{1}{1-\rho} \cdot \left( \frac{\lambda}{\mu} \right)^c \right]^{-1} \\ p_n = \begin{cases} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n P_0 & (n \leq c) \\ \frac{1}{c! c^{n-c}} \left( \frac{\lambda}{\mu} \right)^n P_0 & (n > c) \end{cases} \end{cases} \quad (12-29)$$

- ❖ 其中  $\rho = \frac{\lambda}{c\mu}$ ，但现在已不必对  $\rho$  加以限制。

### 3.3 系统的容量有限制的情形 (M/M/c/N/∞)

$$\begin{cases} L_q = \frac{P_0 \rho (c\rho)^c}{c!(1-\rho)^2} [1 - \rho^N - c - (N-c)\rho^N - c(1-\rho)] \\ L_s = L_q + c\rho(1-P^N) \\ W_q = \frac{L_q}{\lambda(1-P^N)} \\ W_s = W_q + \frac{1}{\mu} \end{cases} \quad (12-32)$$

❖ 特别当  $N=c$  (即时制) 时

例如，停车场就不允许排队等待空位的情形！

$$\begin{cases} P_0 = \frac{1}{\sum_{k=0}^c \frac{(c\rho)^k}{k!}} \\ P_n = \frac{(c\rho)^n}{n!} P_0, \quad 0 \leq n \leq c \end{cases} \quad (12-33)$$

❖ 当  $N=c$  即关于  $P_c$  的公式被称为爱尔朗损失公式。

### 3.3 系统的容量有限制的情形 (M/M/c/N/∞)

❖ 这时的运行指标如下：

$$\begin{cases} L_q = 0, W_q = 0, W_s = \frac{1}{\mu} \\ L_s = \sum_{n=1}^c n P^n = \frac{c \rho \sum_{n=0}^{c-1} \frac{(c \rho)^{n-1}}{n!}}{\sum_{n=0}^c \frac{(c \rho)^n}{n!}} = c \rho (1 - P_c) \end{cases} \quad (12-34)$$

### 3.3 系统的容量有限制的情形 ( $M/M/c/N/\infty$ )

❖ **例7** 在某风景区准备建造旅馆，顾客到达为泊松流，每天平均到( $\lambda=$ )6人，顾客平均逗留时间( $1/\mu$ )为2天，试就该旅馆在具有( $c$ )1, 2, 3, ..., 8个房间的条件下，分别计算每天客房平均占用数 $L_s$  及满员概率  $P_c$ 。

❖ 这是即时式，因为在客房满员条件下，旅客显然不能排队等待。计算过程通过表12-12进行

(  $\lambda = 6, 1/\mu = 2, c\rho = \lambda/\mu = 12$  )。

### 3.3 系统的容量有限制的情形 (M/M/c/N/∞)

(1) $n$	(2) $(c\rho)^n = 12^n$	(3) $n!$	(4) $(c\rho)^n / n!$	(5) $\sum_{n=0}^c (c\rho)^n / n!$	(6) $P_c$ (答)	(7)	(8) $L_s$ (答)
0	1	1	1	1	1	—	—
1	$1.2 \times 10$	1	12	13	0.92	0.08	0.92
2	$1.44 \times 10^2$	2	72	85	0.85	0.15	1.83
3	$1.73 \times 10^3$	6	288	373	0.77	0.23	2.74
4	$2.07 \times 10^4$	24	864	$1.24 \times 10^3$	0.70	0.30	3.62
5	$2.49 \times 10^5$	120	$2.07 \times 10^3$	$3.31 \times 10^3$	0.63	0.37	4.48
6	$2.99 \times 10^6$	720	$4.15 \times 10^3$	$7.46 \times 10^3$	0.56	0.44	5.33
7	$3.58 \times 10^7$	$5.04 \times 10^3$	$7.11 \times 10^3$	$1.45 \times 10^4$	0.49	0.51	6.14
8	$4.30 \times 10^8$	$4.03 \times 10^4$	$1.07 \times 10^4$	$2.52 \times 10^5$	0.42	0.58	6.93

第(4)栏  
= (2)/(3)

第(5)栏: 第(4)栏各数累加

(4)/(5)  
得满员  
概率

用第(5)栏同行去除上一行结果  
(7) × 1  
2得, 为  
每天客房  
平均占用  
数

### 3.4 顾客源为有限的情形 ( $M/M/c/\infty /m$ )

- ❖ 设顾客总体(顾客源)为有限数 $m$ ，且 $m > c$ ，顾客到达率 $\lambda$ 是按每个顾客来考虑的。
  - 在机器管理问题中，就是共有 $m$ 台机器，有 $c$ 个修理工人，顾客到达就是机器出了故障
  - 每个顾客的到达率 $\lambda$ 是指每台机器每单位运转时间出故障的期望次数
  - 系统中顾客数 $n$ 就是出故障的机器台数
    - 当 $n \leq c$ 时，所有的故障机器都在被修理，有 $(c-n)$ 个修理工人在空闲
    - 当 $c < n \leq m$ 时，有 $(n-c)$ 台机器在停机等待修理，而修理工人都在繁忙状态
  - 假定这 $c$ 个工人修理技术相同，修理(服务)时间都服从参数为 $\mu$ 的负指数分布，并假定故障的修复时间和正在生产的机器是否发生故障是相互独立的。

### 3.4 顾客源为有限的情形 (M/M/c/∞ /m)

❖ (1)

$$p_0 = \frac{1}{m!} \cdot \frac{1}{\sum_{k=0}^c \frac{1}{k!(m-k)!} \left(\frac{c\rho}{m}\right)^k + \frac{c^c}{c!} \sum_{k=c+1}^m \frac{1}{(m-k)!} \left(\frac{\rho}{m}\right)^k}$$

其中

$$\rho = \frac{m\lambda}{c\mu}$$

$$p_n = \begin{cases} \frac{m!}{(m-n)!n!} \left(\frac{\lambda}{\mu}\right)^n P_0 & (0 \leq n \leq c) \\ \frac{m!}{(m-n)!c!c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n P_0 & (c+1 \leq n \leq m) \end{cases} \quad (12-35)$$

❖ (2) 平均顾客数(即平均故障台数)

$$L_s = \sum_{n=1}^m np_n$$

$$L_q = \sum_{n=c+1}^m (n-c)P_n$$

### 3.4 顾客源为有限的情形 (M/M/c/∞ /m)

- 有效的到达率  $\lambda_e$  应等于每个顾客的到达率  $\lambda$  乘以在系统外(即正常生产的)机器的期望数:

$$\lambda_e = \lambda(m - L_s)$$

- 在机器故障问题中, 它是每单位时间  $m$  台机器平均出现故障的次数。

❖ (3) 可以证明

$$\begin{cases} L_s = L_q + \frac{\lambda_e}{\mu} = L_q + \frac{\lambda}{\mu}(m - L_s) \\ M_s = L_s / \lambda_e \\ W_q = L_q / \lambda_e \end{cases} \quad (12-36)$$



### 3.4 顾客源为有限的情形 ( $M/M/c/\infty/m$ )

- ❖ 例8 设有两个修理工人，负责5台机器的正常运行，每台机器平均损坏率为每运转小时1次，两工人能以相同的平均修复率4(次/小时)修好机器。求：
- ① 等待修理的机器平均数；
  - ② 需要修理的机器平均数；
  - ③ 有效损坏率；
  - ④ 等待修理时间；
  - ⑤ 停工时间。

解：  $m = 5$ ,  $\lambda = 1$  (次/小时),  $\mu = 4$  (台/小时),  $c = 2$ ,  $c\rho/m = \lambda/\mu = 1/4$

$$P_0 = \frac{1}{5!} \left[ \frac{1}{5!} \left( \frac{1}{4} \right)^0 + \frac{1}{4!} \left( \frac{1}{4} \right)^1 + \frac{1}{2!3!} \left( \frac{1}{4} \right)^2 + \frac{2^2}{2!} \frac{1}{2} \left( \frac{1}{8} \right)^3 + \left( \frac{1}{8} \right)^4 + \left( \frac{1}{8} \right)^5 \right]^{-1}$$
$$= 0.3149$$

$$P_1 = 0.394, P_2 = 0.197, P_3 = 0.074, P_4 = 0.018, p_5 = 0.002$$

### 3.4 顾客源为有限的情形 (M/M/c/∞ /m)

- ❖ ①  $L_q = P_3 + 2P_4 + 3P_5 = 0.118$
- ❖ ②  $L_s = \sum_{n=1}^m nP_n = L_q + c - 2P_0 - P_1 = 1.094$
- ❖ ③  $\lambda_e = 1 \times (5 - 1.094) = 3.906$
- ❖ ④  $W_q = 0.118 / 3.906 = 0.03$  (小时)
- ❖ ⑤  $W_s = 1.094 / 3.906 = 0.28$  (小时)



# 提纲

---

- ❖ 基本概念
- ❖ M/M/1模型
- ❖ M/M/C模型
- ❖ 排队系统的最优化问题
- ❖ 分析排队系统的随机模拟法

## 第4节 经济分析——系统的最优化

- ❖ 4.1 排队系统的最优化问题
- ❖ 4.2 M/M/1模型中最优服务率 $\mu$
- ❖ 4.3 M/M/c模型中最优的服务台数 $c$

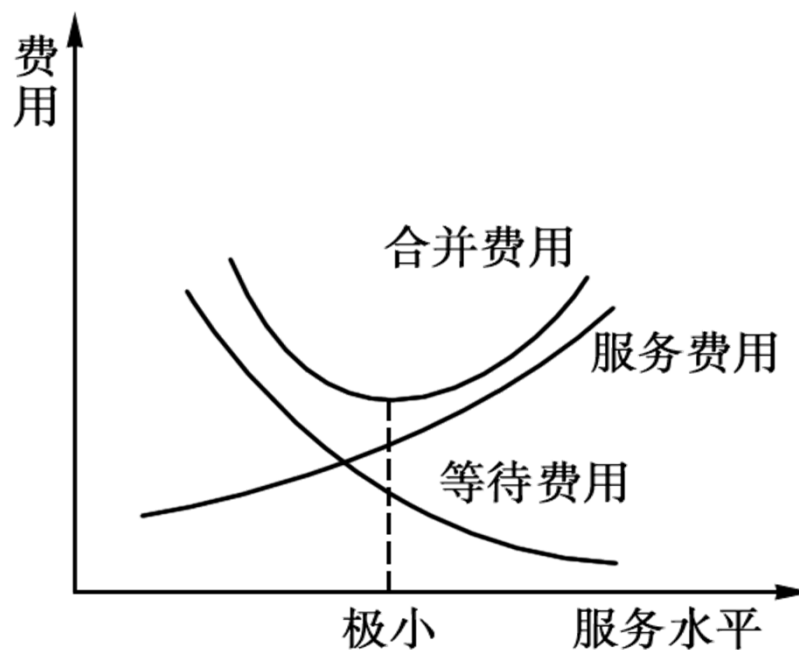
## 4.1 排队系统的最优化问题

- ❖ 以完全消除排队现象为研究目标是不现实的，那会造成服务人员和设施的严重浪费；
- ❖ 但是设施的不足和低水平的服务，又将引起太多的等待，从而导致生产和社会性损失。
- ❖ 从经济角度考虑，排队系统的费用应该包含以下两个方面：
  - 一个是服务费用，它是服务水平的递增函数；
  - 另一个是顾客等待的机会损失(费用)，它是服务水平的递减函数。
  - 两者的总和呈一条U形曲线。

## 4.1 排队系统的最优化问题

### ❖ 排队系统的最优化问题

- 一般情形下，提高服务水平(数量，质量)会降低顾客的等待费用(损失)，但却常常增加了服务机构的成本(如图)。
- 最优化的目标之一是使二者费用之和为最小，并确定达到这个目标的最优的服务水平。另一个常用目标是使纯收入或利润(服务收入与服务成本之差)为最大。



## 4.1 排队系统的最优化问题

- ❖ 排队系统的最优化问题包括如下两种：
  - 系统设计的最优化——静态问题，目的在于使设备达到最大效益，或者说，在一定的质量指标下要求机构最为经济。
  - 系统控制的最优化——动态问题，是指一个给定的系统，如何运营可使某个目标函数得到最优。

## 4.1 排队系统的最优化问题

- ❖ 归纳起来，排队系统常见的优化问题在于：
  - (1) 确定最优服务率 $\mu^*$ ；
  - (2) 确定最佳服务台数量 $c^*$ ；
  - (3) 选择最为合适的服务规则；
  - (4) 或是确定上述几个量的最优组合。
- ❖ 研究排队系统的根本目的在于以最少的设备得到最大的效益，或者说，在一定的服务质量的指标下要求机构最为经济。



## 4.1 排队系统的最优化问题

- ❖ 各种费用在稳态情形下，都是按单位时间来考虑的。
  - 服务费用(成本)是可以确切计算或估计的。
  - 等待费用有的可以确切估计的，有的只能根据统计的经验资料来估计。
  - 服务水平也可以由不同形式来表示，
    - 平均服务率 $\mu$ (代表服务机构的服务能力和经验等)
    - 服务设备，如服务台的个数 $c$ ，以及由队列所占空间大小所决定的队列最大限制数 $N$ 等
    - 服务强度 $\rho$
- ❖ 常用的求解方法
  - 对于离散变量常用边际分析法
  - 对于连续变量常用经典的微分法
  - 对于复杂问题可以用非线性规划或动态规划的方法

## 4.1 排队系统的最优化问题

- ❖ 由于系统动态最优控制问题涉及更多的数学知识，因此，只讨论系统静态的最优设计问题。这类问题一般可以借助于前面所得到的一些表达式来解决。
- ❖ 本节仅就 $\mu$ ,  $c$  这两个决策变量的分别单独优化，介绍两个较简单的模型，以便读者了解排队系统优化设计的基本思想。

## 第4节 经济分析——系统的最优化

- ❖ 4.1 排队系统的最优化问题
- ❖ 4.2 M/M/1模型中最优服务率 $\mu$
- ❖ 4.3 M/M/c模型中最优的服务台数 $c$
- ❖ 4.4 系统最优化应用举例

## 4.2 M/M/1模型中最优服务率 $\mu$

### 1. 标准的M/M/1模型

- 取目标函数 $z$ 为单位时间服务成本与顾客在系统逗留费用之和的期望值

$$z = c_s \mu + c_w L_s \quad (12-41)$$

- 其中  $c_s$  为当  $\mu=1$  时服务机构单位时间的费用;  $c_w$  为每个顾客在系统停留单位时间的费用。

- 将  $L_s = \lambda / (\mu - \lambda)$  代入  $L_s$ , 得  $z = c_s \mu + c_w \cdot \frac{\lambda}{\mu - \lambda}$
- 为了求极小值, 先求  $\frac{dz}{d\mu}$ , 然后令它为0,

$$\frac{dz}{d\mu} = c_s - c_w \lambda \cdot \frac{1}{(\mu - \lambda)^2}$$

$$c_s - c_w \lambda \cdot \frac{1}{(\mu - \lambda)^2} = 0$$

- 解出最优的

$$\mu^* = \lambda + \sqrt{\frac{c_w}{c_s} \lambda} \quad (12-42)$$

## 4.2 M/M/1模型中最优服务率 $\mu$

又由二阶条件  $\frac{d^2 z}{d\mu^2} = \frac{2c_w\lambda}{(\mu-\lambda)^3} > 0$  (因 $\mu > \lambda$ )

可知 $\mu^*$ 为 $(\lambda, \infty)$ 上的全局唯一最小点。将 $\mu^*$ 代入

$z = c_s\mu + c_wL_s$  (12-41) 中, 可得最小总平均费用

$$z^* = c_s\lambda + 2\sqrt{c_sc_w\lambda}$$

## 4.3 M/M/c模型中最优的服务台数c

标准的M/M/c模型，且在稳态情形下，这时单位时间全部费用(服务成本与等待费用之和)的期望值

$$z = c'_s \cdot c + c_w \cdot L \quad (12-43)$$

其中c是服务台数； $c'_s$  是每服务台单位时间的成本； $c_w$  为每个顾客在系统停留单位时间的费用；L是系统中顾客平均数 $L_s$  或队列中等待的顾客平均数 $L_q$ 。因为c只取整数， $z(c)$ 不是连续函数，故采用边际分析法。

根据 $z(c^*)$ 是最小的特点，有

$$\begin{cases} z(c^*) \leq z(c^* - 1) \\ z(c^*) \leq z(c^* + 1) \end{cases}$$

## 4.3 M/M/c模型中最优的服务台数c

- ❖ 将(12-43)式中z代入，得

$$\begin{cases} c'_s c^* + c_w L(c^*) \leq c'_s (c^* - 1) + c_w L(c^* - 1) \\ c'_s c^* + c_w L(c^*) \leq c'_s (c^* + 1) + c_w L(c^* + 1) \end{cases}$$

- ❖ 上式化简后，得

$$L(c^*) - L(c^* + 1) \leq c'_s / c_w \leq L(c^* - 1) - L(c^*)$$

- ❖ 依次求 L 的值，并作两相邻的L值之差，因  $c'_s / c_w$  是已知数，根据这个数落在哪个不等式的区间里就可定出c

## 4.3 M/M/c模型中最优的服务台数c

- ❖ **例12** 某检验中心为各工厂服务，要求作检验的工厂(顾客)的到来服从泊松流，平均到达率 $\lambda$ 为每天48次，每次来检验由于停工等原因损失为6元。服务(作检验)时间服从负指数分布，平均服务率 $\mu$ 为每天25次，每设置1个检验员服务成本(工资及设备损耗)为每天4元。其他条件适合标准的M/M/c模型，问应设几个检验员(及设备)才能使总费用的期望值为最小？
- ❖ **解：**  $c'_s = 4$  元/检验员， $c_w = 6$ 元/次， $\lambda = 48$   $\mu = 25$   $\lambda / \mu = 1.92$
- ❖ 设检验员数为 $c$ ，令 $c$ 依次为1, 2, 3, 4, 5，根据表12-11，求出 $L_s$ 。计算过程如下：

c	1	2	3	4	5
$\lambda/c\mu$	1.92	0.96	0.64	0.48	0.38
查表12-11 $W_q \cdot \mu$	—	10.2550	0.3961	0.0772	0.0170
$L_s = \lambda/\mu(W_q \cdot \mu + 1)$	—	21.610	2.680	2.068	1.952



## 4.3 M/M/c模型中最优的服务台数c

❖ 将  $L_s$  值代入(12.44)式得表12-13。

检验员数c	未检验顾客数 $L_s(c)$	$L(c) - L(c+1) \sim L(c) - L(c-1)$	总费用(每天) $z(c)$
1	$\infty$		$\infty$
2	21.610	18.930 $\sim\infty$	154.94
3	2.680	0.612 $\sim$ 18.930	27.87(*)
4	2.068	0.116 $\sim$ 0.612	28.38
5	1.952		31.71

❖  $c'_s c_w = 0.666$ ，落在区间(0.612 $\sim$ 18.930)内，所以  $c^* = 3$ 。即以设3个检验员使总费用为最小，直接代入(12-43)式也可验证总费用为最小。

$$z(c^*) = z(3) = 27.87(\text{元})$$

## 4.4 系统最优化应用举例

例1. 兴建一座港口码头，只有一个装卸船只的泊位。要求设计装卸能力。装卸能力单位为（只/日）船数。已知：单位装卸能力的平均生产费用 $a=2$ 千元，船只逗留每日损失 $b=1.5$ 千元。船只到达服从泊松分布，平均速率 $\lambda=3$ 只/日。船只装卸时间服从负指数分布。目标是每日总支出最少。

## 4.4 系统最优化应用举例

解:  $\lambda=3$   $\mu$ 待定 模型  $M/M/1/\infty/\infty$

队长  $L_s = \lambda/(\mu - \lambda)$

总费用  $C = a\mu + bL_s = a\mu + b\lambda/(\mu - \lambda)$

求极值 (最小值)

求导  $dc/d\mu = a + (-b\lambda)/(\mu - \lambda)^2 \stackrel{\text{令}}{=} 0$

得:  $\mu - \lambda = \pm (b\lambda/a)^{1/2}$  (根据题意舍负)

所以  $\mu = \lambda + (b\lambda/a)^{1/2} = 3 + (2.25)^{1/2} = 4.5$  (只/日)

## 4.4 系统最优化应用举例

例2. 建造一口码头，要求设计装卸船只的泊位数。已知：预计到达 $\lambda=3$ 只/天，泊松流；装卸 $\mu=2$ 只/天，负指数分布。装卸费每泊位每天 $a=2$ 千元，停留损失费 $b=1.5$ 千元/日。目标是总费用最少。

解：模型  $M/M/c/\infty/\infty$   $c$  待定

❖ 总费用： $F=ac+bL_s(c)$  离散，无法用求导来解。

## 4.4 系统最优化应用举例

考虑:  $M/M/c/\infty/\infty$

要求:  $\rho = \lambda / (c\mu) < 1$  即  $c > (\lambda/\mu) = 1.5$

讨论  $c=2,3,4\dots$

$$P_0 = \left[ \sum_{n=0}^{c-1} \frac{\rho^n c^n}{n!} + \frac{c^c \rho^c}{c! (1-\rho)} \right]^{-1}$$

$$L_q = P_0 \frac{c^c \rho^{c+1}}{[c! (1-\rho)^2]}$$

$$L_s = L_q + \lambda/\mu$$

## 4.4 系统最优化应用举例

	$\rho = 3/4$	$\rho = 1/2$	$\rho = 3/8$
	$M/M/2/\infty/\infty$	$M/M/3/\infty/\infty$	$M/M/4/\infty/\infty$
$P_0$	$1/7=0.14286$	$4/19=0.21053$	$40/181=0.22099$
$Lq$	$27/14=1.92857$	$9/83=0.23684$	$81/1810=0.04475$
$Ls$	$24/7=3.42857$	$33/19=1.73684$	$1398/905=1.54475$
$F$	$64/7=9.14286$	$327/38=8.60526$	$9337/905=10.31713$

结论： $c=3$  即设计三个装卸泊位可使每天的总费用最少为8.60526千元。



# 提纲

---

- ❖ 基本概念
- ❖ M/M/1模型
- ❖ M/M/C模型
- ❖ 排队系统的最优化问题
- ❖ 分析排队系统的随机模拟法

## 第5节 分析排队系统的随机模拟法

在模拟一个带有随机因素的实际系统时,究竟用什么样的概率分布描述问题中的随机变量,是我们总是要碰到的一个问题,下面简单介绍确定分布的常用方法:

1° 根据一般知识和经验,可以假定其概率分布的形式,如顾客到达间隔服从指数分布  $\text{Exp}(\lambda)$ ; 产品需求量服从正态分布  $N(\mu, \sigma^2)$ ; 订票后但未能按时前往机场登机的人数服从二项分布  $B(n, p)$ 。然后由实际数据估计分布的参数  $\lambda, \mu, \sigma$  等, 参数估计可用极大似然估计、矩估计等方法。

2° 直接由大量的实际数据作直方图, 得到经验分布, 再通过假设检验, 拟合分布函数, 可用  $\chi^2$  检验等方法。

3° 既缺少先验知识, 又缺少数据时, 对区间  $(a, b)$  内变化的随机变量, 可选用 Beta 分布 (包括均匀分布)。先根据经验确定随机变量的均值  $\mu$  和频率最高时的数值 (即密



## 第5节 分析排队系统的随机模拟法

度函数的最大值点)  $m$  , 则 Beta 分布中的参数  $\alpha_1, \alpha_2$  可由以下关系求出:

$$\mu = a + \frac{\alpha_1(b-a)}{\alpha_1 + \alpha_2}, \quad m = a + \frac{(\alpha_1 - 1)(b-a)}{\alpha_1 + \alpha_2 - 2}.$$

## 第5节 分析排队系统的随机模拟法

- ❖ 当排队系统的到达间隔时间和服务时间的概率分布很复杂时，或不能用公式给出时，就不能用以往的解析法求解。
- ❖ 这就需要用随机模拟法求解。

## 第5节 分析排队系统的随机模拟法

### ❖ 例13

- 设某仓库前有一卸货场，货车一般是夜间到达，白天卸货。每天只能卸货2车，若一天内到达数超过2车，那么就推迟到次日卸货。根据表12-14所示的经验货车到达数的概率分布(相对频率)平均为1.5车/天，求每天推迟卸货的平均车数。

到达车数	0	1	2	3	4	5	$\geq 6$
概率	0.23	0.30	0.30	0.10	0.05	0.02	0.00

- ❖ 可见，到达车数不服从普阿松分布，服务时间不服从负指数分布。
- ❖ 随机模拟法首先要求事件能按历史的概率分布规律出现。

## 第5节 分析排队系统的随机模拟法

到达车数	0	1	2	3	4	5	$\geq 6$
概率	0.23	0.30	0.30	0.10	0.05	0.02	0.00

### ❖ 卡片模拟

- 对例13的数据进行分析，取100张卡片，按表12-14的概率，取23张卡片填入0；取30张填1……。随机地一一取出，依次记录卡片上的数码，得到这一系列数据就是每天到达车数的模拟。
- 实际应用时可用随机数表(见书)

### ❖ MS Excel

- RAND() 函数产生随机数x，在 $0 < x < 1$ 之间，若需要整数两位的随机数，可以用ROUND(RAND()\*100, 0)产生随机数。
- 按到达车数的概率分配随机数

到达车数	概率	累积概率	对应的随机数
0	0.23	0.23	00~22
1	0.30	0.53	23~52
2	0.30	0.83	53~82
3	0.10	0.93	83~92
4	0.05	0.98	93~97
5	0.02	1.00	98~99

## 第5节 分析排队系统的随机模拟法

我们用 a1 表示产生的随机数，a2 表示到达的车数，a3 表示需要卸货车数，a4 表示实际卸货车数，a5 表示推迟卸货车数。编写程序如下：

```
clear
rand('state',sum(100*clock));
n=50000;
m=2
a1=rand(n,1);
a2=a1; %a2初始化
a2(find(a1<0.23))=0;
a2(find(0.23<=a1&a1<0.53))=1;
a2(find(0.53<=a1&a1<0.83))=2;
a2(find(0.83<=a1&a1<0.93),1)=3;
a2(find(0.93<=a1&a1<0.98),1)=4;
a2(find(a1>=0.98))=5;
a3=zeros(n,1);a4=zeros(n,1);a5=zeros(n,1); %a2初始化
a3(1)=a2(1);
if a3(1)<=m
    a4(1)=a3(1);a5(1)=0;
else
```

## 第5节 分析排队系统的随机模拟法

```
    a4(1)=m;a5(1)=a2(1)-m;
end
for i=2:n
    a3(i)=a2(i)+a5(i-1);
    if a3(i)<=m
        a4(i)=a3(i);a5(i)=0;
    else
        a4(i)=m;a5(i)=a3(i)-m;
    end
end
a=[a1,a2,a3,a4,a5];
sum(a)/n
```

## 第5节 分析排队系统的随机模拟法

### ❖ 模拟过程

(1)日期	(2)随机数	(3)到达数	(4)需要卸货车数	(5)卸货车数	(6)推迟卸货车数
x	97	4	4	2	2
x	02	0	2	2	0
x	80	2	2	2	0
1	66	2	2	2	0
2	96	4	4	2	2
3	55	2	4	2	2
4	50	1	3	2	1
...	...	...	...	...	...
48	55	2	2	2	0
49	54	2	2	2	0
50	23	1	1	1	0
总计		79			45
平均		1.58			0.90