



排队论模型

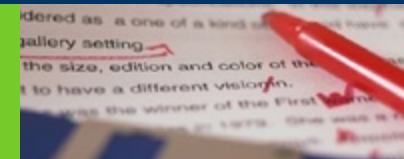
主讲人：泰山教育 小石老师

模型背景

排队论发源于上世纪初。当时美国贝尔电话公司发明了自动电话，以适应日益繁忙的工商业电话通讯需要。这个新发明带来了一个新问题，即通话线路与电话用户呼叫的数量关系应如何妥善解决，这个问题久久未能解决。

1909年，丹麦的哥本哈根电话公司A. K. 埃尔浪(Erlang)在热力学统计平衡概念的启发下解决了这个问题。

模型背景



排队现象

到达顾客

服务内容

服务机构

病人

诊断/手术

医生/手术台

进港的货船

装货/卸货

码头泊位

到港的飞机

降落

机场跑道

电话拨号

通话

交换台

故障机器

修理

修理技工

修理技工

领取修配零件

仓库管理员

上游河水

入库

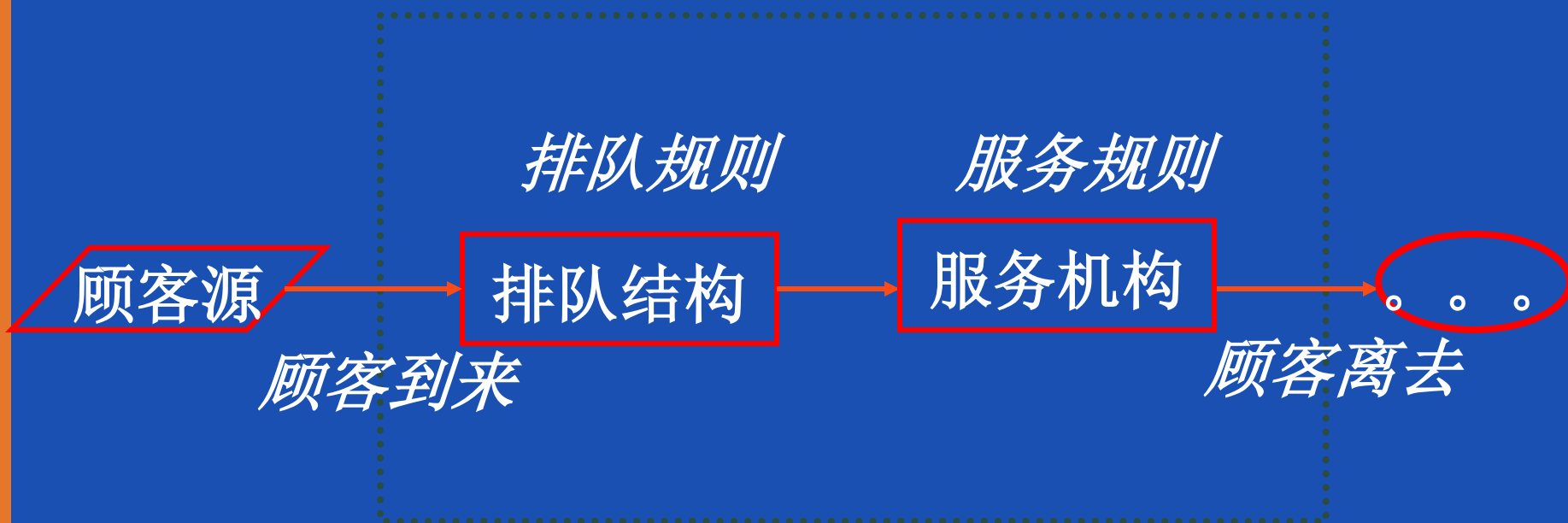
水闸管理员

模型介绍

- (1) 由于顾客到达和服务时间的随机性，现实中的排队现象几乎不可避免；
- (2) 排队过程，通常是一个随机过程，排队论又称“随机服务系统理论”；

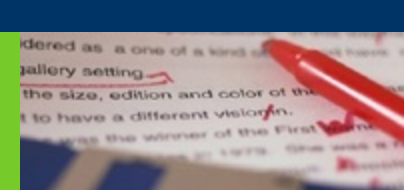
模型介绍

排队服务过程



排队系统

模型介绍



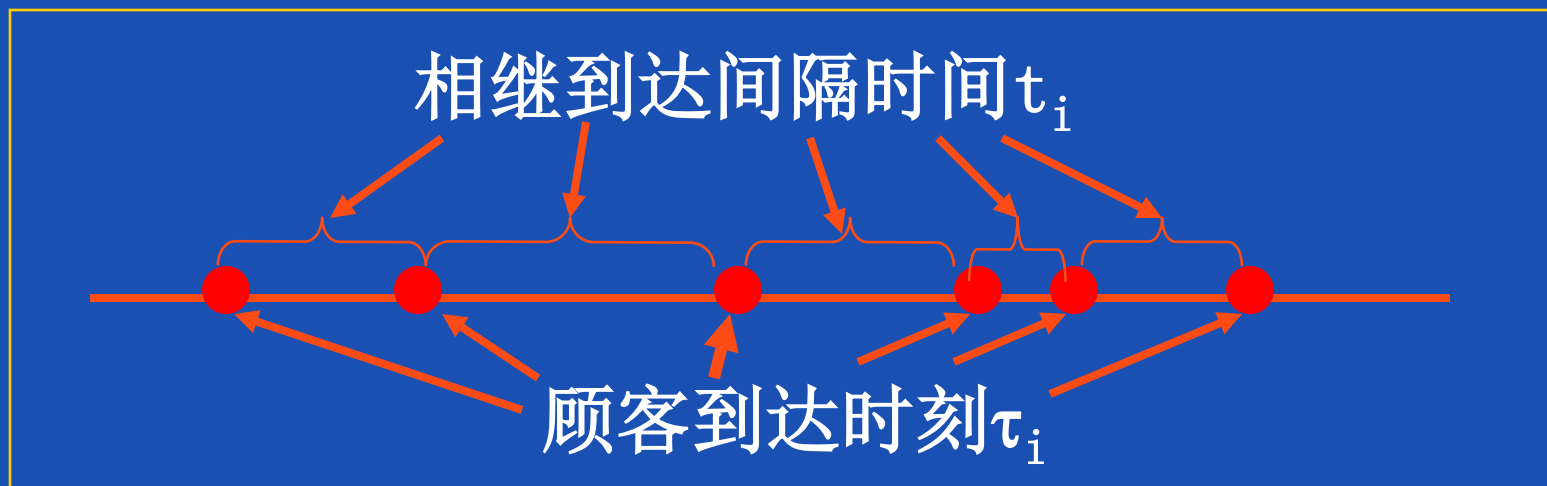
排队系统的要素

- (1) 顾客输入过程;
- (2) 排队结构与排队规则;
- (3) 服务机构与服务规则;

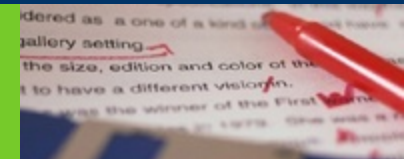
模型介绍

顾客输入过程

顾客源(总体): 有限/无限;
顾客到达方式: 逐个/逐批; (仅研究逐个情形)
顾客到达间隔: 随机型/确定型;
顾客前后到达是否独立: 相互独立/相互关联;
输入过程是否平稳: 平稳/非平稳; (仅研究平稳性)



模型介绍



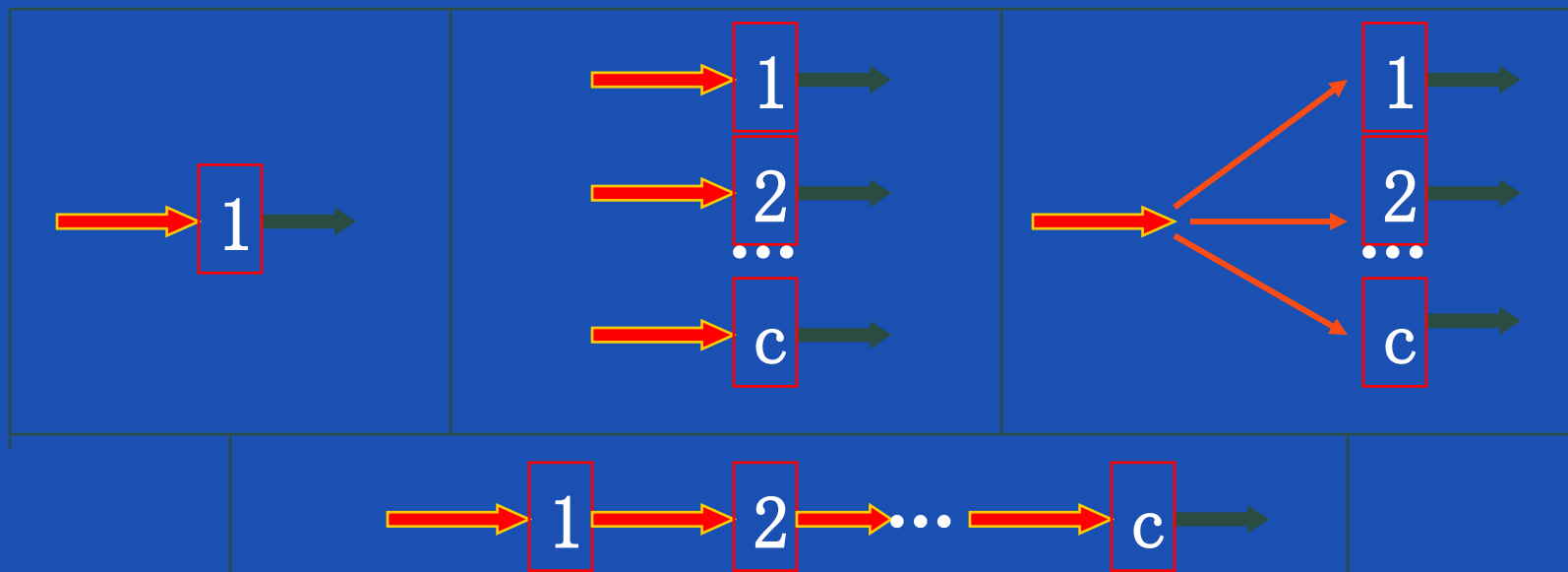
排队结构与排队规则

顾客排队方式：等待制/即时制(损失制)；
排队系统容量：有限制/无限制；
排队队列数目：单列/多列；
是否中途退出：允许/禁止；
是否列间转移：允许/禁止；
(仅研究禁止退出和转移的情形)

模型介绍

服务机构与服务规则

服务台(员)数目;单个/多个;
服务台(员)排列形式;并列/串列/混合;
服务台(员)服务方式;逐个/逐批;(研究逐个情形)
服务时间分布;随机型/确定型;
服务时间分布是否平稳:平稳/非平稳;(研究平稳情形)



模型介绍

服务台(员)为顾客服务的顺序:

- a) 先到先服务 (FCFS);
- b) 后到先服务 (LCFS);
- c) 随机服务;
- d) 优先服务;

模型介绍

到达间隔和服务时间典型分布

- | | | |
|-----|----------|---------|
| (1) | 泊松分布 | M ; |
| (2) | 负指数分布 | M ; |
| (3) | k阶爱尔朗分布 | E_k ; |
| (4) | 确定型分布 | D ; |
| (5) | 一般服务时间分布 | G ; |

模型介绍

排队模型示例

- $M/M/1$, $M/D/1$, $M/E_k/1$;
- $M/M/c$, $M/M/c/\infty/m$,
- $M/M/c/N/\infty$, . . .

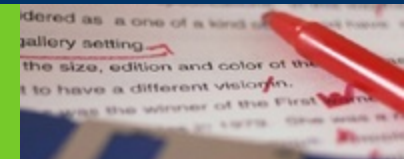
模型介绍

(一) 系统运行状态参数

系统状态 $N(t)$

——指排队系统在时刻 t 时的全部顾客数 $N(t)$ ，包括“排队顾客数”和“正被服务顾客数”；

模型介绍



系统状态概率：

(1) 瞬态概率 $P_n(t)$

——表示时刻 t 系统状态 $N(t)=n$ 的概率；

(2) 稳态概率 P_n

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

——一般排队系统运行了一定长的时间后，系统状态的概率分布不再随时间 t 变化，即初始时刻 ($t=0$) 系统状态的概率分布 ($P_n(0)$, $n \gg 0$) 的影响将消失。

模型介绍

(二) 系统运行指标参数

——评价排队系统的优劣。

1、队长与排队长

(1) 队长: 系统中的顾客数 (n) 期望值记为 L_s ;

(2) 排队长: 系统中排队等待服务的顾客数;
期望值记为 L_q

模型介绍

2、逗留时间与等待时间

(1) 逗留时间:

——指一个顾客在系统中的全部停留时间；
期望值，记为 W_s

(2) 等待时间:

——指一个顾客在系统中的排队等待时间；
期望值，记为 W_q

$$W_s = W_q + E[\text{服务时间}]$$

模型介绍

3、其他相关指标

(1) 忙期：指从顾客到达空闲服务机构起到服务机构再次空闲的时间长度；

(2) 忙期服务量：指一个忙期内系统平均完成服务的顾客数；

(3) 损失率：指顾客到达排队系统，未接受服务而离去的概率；

(4) 服务强度： $\rho = \lambda / s\mu$ ；

模型介绍

顾客到达时间间隔分布 泊松流与泊松分布

如果顾客到达满足如下条件,则称为泊松流:

- (1) 在不相互重叠的时间区间内, 到达顾客数相互独立(无后效性).
- (2) 对于充分小的时间间隔 $[t, t + \Delta t]$ 内, 到达1个顾客的概率与 t 无关, 仅与时间间隔成正比 (平稳性): $P_1(t, t + \Delta t) = \lambda \Delta t + o(\Delta t)$
- (3) 对于充分小的时间间隔 $[t, t + \Delta t]$, 2个及以上顾客到达的概率可忽略不计 (普通性)。

模型介绍

泊松流到达间隔服从负指数分布

- ❖ 若顾客到达间隔 T 的概率密度为

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

则称 T 服从负指数分布，分布函数如下：

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

- ❖ 若顾客流是泊松流时，顾客到达的时间间隔服从上述负指数分布
- ❖ $E[T] = 1/\lambda$; $\text{Var}[T] = 1/\lambda^2$; $\sigma[T] = 1/\lambda$

模型介绍

顾客服务时间分布 负指数分布

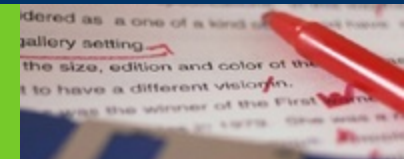
(1) 对一个顾客的服务时间 T_s ，等价于相邻两个顾客离开排队系统的时间间隔。若 T_s 服从负指数分布，其概率密度和分布函数分别为

$$f_{T_s}(t) = \begin{cases} \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad F_{T_s}(t) = \begin{cases} 1 - \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

则 $E[T_s] = 1/\mu$; $\text{Var}[T_s] = 1/\mu^2$; $\sigma[T_s] = 1/\mu$

(2) $E[T_s] = 1/\mu$: 每个顾客的平均(期望)服务时间;
 μ : 单位时间服务的顾客数, 平均(期望)服务率;

模型介绍



单服务台负指数分布M/M/1排队系统

- ❖ 模型的条件是：
- ❖ 1、输入过程——顾客源是无限的，顾客到达完全是随机的，单个到来，到达过程服从普阿松分布，且是平稳的；
- ❖ 2、排队规则——单队，且队长没有限制，先到先服务；
- ❖ 3、服务机构——单服务台，服务时间的长短是随机的，服从相同的指数分布。

模型举例

对于M / M / 1模型有如下公式：

$$P_0 = 1 - \rho$$

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

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

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

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

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = W_s \rho$$

$$P(N > k) = \rho^{k+1} \quad N \text{ 表系统中的顾客数}$$

模型举例

例1 某医院急诊室同时只能诊治一个病人，诊治时间服从指数分布，每个病人平均需要15分钟。病人按泊松分布到达，平均每小时到达3人。试对此排队队系统进行分析。

解 对此排队队系统分析如下：

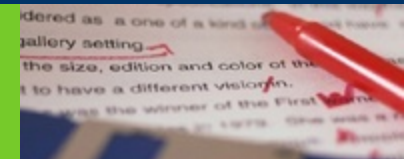
(1) 先确定参数值：这是单服务台系统，有：

$$\lambda = 3 \text{人} / h, \mu = \frac{60}{15} \text{人} / h = 4 \text{人} / h$$

故服务强度为：

$$\rho = \frac{\lambda}{\mu} = \frac{3}{4} = 0.75$$

模型举例



计算稳态概率：

$$P_0 = 1 - \rho = 1 - 0.75 = 0.25$$

这就是急诊室空闲的概率，也是病人不必等待立即就能就诊的概率。而病人需要等待的概率则为：

$$\rho = 1 - P_0 = 0.75$$

这也是急诊室繁忙的概率。

模型举例

(2) 计算系统主要工作指标。

急诊室内外的病人平均数:

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{3}{4 - 3} \text{人} = 3 \text{人}$$

急诊室外排队等待的病人平均数:

$$L_q = L_s \rho = 3 \times 0.75 \text{人} = 2.25 \text{人}$$

病人在急诊室内外平均逗留时间:

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{4 - 3} h = 1h = 60 \text{min}$$

病人平均等候时间:

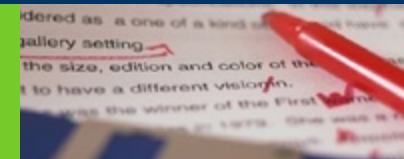
$$W_q = W_s \rho = 1 \times 0.75h = 0.75h = 45 \text{min}$$

模型举例

M/M/S模型

- ❖ 此模型与M/M/1模型不同之处在于有S个服务台，各服务台的工作相互独立，服务率相等，如果顾客到达时，S个服务台都忙着，则排成一队等待，先到先服务的单队模型。
- ❖ 整个系统的平均服务率为 $s\mu$ ， $\rho^* = \lambda / s\mu$ ， $(\rho^* < 1)$ 为该系统的服务强度。

模型举例



1 状态概率

$$P_0 = \left[\sum_{k=0}^{S-1} \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k + \frac{1}{s!} \frac{1}{1-\rho^*} \left(\frac{\lambda}{\mu} \right)^S \right]^{-1}$$

$$P_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0 & 0 < n \leq S \\ \frac{1}{S! S^{n-S}} \left(\frac{\lambda}{\mu} \right)^n P_0 & n \geq S \end{cases}$$

模型举例

2 主要运行指标

$$L_q = \frac{(s\rho^*)^s \rho^*}{s! (1 - \rho^*)^2} P_0$$

$$L_s = L_q + s\rho^*$$

$$W_q = \frac{L_q}{\lambda}$$

$$W_s = \frac{L}{\lambda} = W_q + \frac{1}{\mu}$$

3 系统状态 $N \geq s$ 的概率

$$P(N \geq k) = \sum_{n=k}^{\infty} P_n = \frac{\rho^k}{k! (1 - \rho^*)} P_0$$

模型举例

例2 承接例1，假设医院增强急诊室的服务能力，使其同时能诊治两个病人，且平均服务率相同，试分析该系统工作情况。

解 这相当于增加了一个服务台，故有：

$$S=2, \lambda=3\text{人/h}, \mu=4\text{人/h}$$

$$\rho = \frac{\lambda}{\mu} = 0.75, \quad \rho^* = \frac{\lambda}{S\mu} = \frac{3}{2 \times 4} = 0.375$$

$$P_o = \left[1 + 0.75 + \frac{(0.75)^2}{2!(1-0.375)} \right]^{-1} = \frac{5}{11} = 0.45$$

模型举例

$$L_q = \frac{(0.75)^2 \times 0.375}{2!(1-0.375)^2} \times \frac{5}{11} = 0.27 \times \frac{5}{11} \text{ 人} \approx 0.12 \text{ 人}$$

$$L_s = L_q + \rho = (0.12 + 0.75) \text{ 人} = 0.87 \text{ 人}$$

$$W_s = \frac{L_s}{\lambda} = \frac{0.87}{3} = 0.29h = 17.4 \text{ min}$$

$$W_q = \frac{L_q}{\lambda} = \frac{0.12}{3} = 0.04h = 2.4 \text{ min}$$

病人必须等候的概率, 即系统状态 $N \geq 2$ 的概率:

$$P(N \geq 2) = \frac{(0.75)^2}{2!(1-0.375)} \times \frac{5}{11} \approx 0.20$$

模型举例

例三 某医院挂号室有三个窗口，就诊者的到达服从泊松分布，平均到达率为每分钟0.9人，挂号员服务时间服从指数分布，平均服务率每分钟0.4人，现假设就诊者到达后排成一队，依次向空闲的窗口挂号，显然系统的容量和顾客源是不限的，属于M/M/S型的排队服务模型。求：该系统的运行指标 $P_0, L_q, L, W_q, W, P(N \geq 3)$ 。

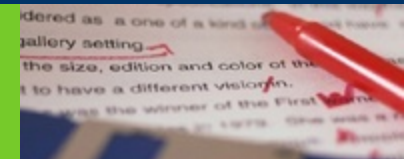
解

$$S=3, \rho = \frac{\lambda}{\mu} = \frac{0.9}{0.4} = 2.25, \quad \rho^* = \frac{\lambda}{S\mu} = \frac{2.25}{3} = \frac{3}{4} < 1$$

(1) 整个挂号间空闲的概率

$$P_0 = \left[\frac{(2.25)^0}{0!} + \frac{(2.25)^1}{1!} + \frac{(2.25)^2}{2!} + \frac{(2.25)^3}{3!} \frac{1}{1-2.25/3} \right]^{-1} = 0.0748$$

模型举例



(2) 等待挂号的平均人数或称队列长

$$L_q = \frac{(2.25)^3 \cdot 3/4}{3!} \times 0.0748 = 1.7 \text{ (人)}$$

(3) 挂号间平均逗留人数或称队长

$$L = L_q + \frac{\lambda}{\mu} = 1.7 + 2.25 = 3.95 \text{ (人)}$$

(4) 等候挂号的平均时间

$$W_q = \frac{1.7}{0.9} = 1.89 \text{ (分钟)}$$

(6) 就诊者到达后必须等待（即系统中就诊者不少于3人或各挂号员都没有空闲）的概率

(5) 在挂号间平均逗留时间

$$W = 1.89 + \frac{1}{0.4} = 4.39 \text{ (分钟)}$$

$$P(N \geq 3) = \frac{(2.25)^3}{3!} \times 0.0748 = 0.57$$



Thank You !