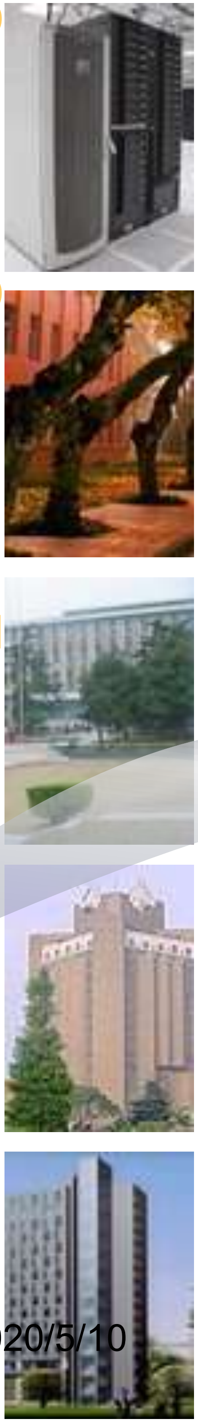
Computer Science and Technology



运 筹 学

讲者： 顾乃杰 教授、 黄章进 副教授

计算机科学与技术学院

排队论

Chap. 13 *Queueing Theory*

13. 排队论

- 13. 1 基本概念
- 13. 2 到达间隔的分布和服务时间的分布
- 13. 3 单服务台负指数分布排队系统的分析
- 13. 4 多服务台负指数分布排队系统的分析
- 13. 5 一般服务时间 M/G/1 模型
- 13. 6 经济分析—系统的最优化

13.1 基本概念

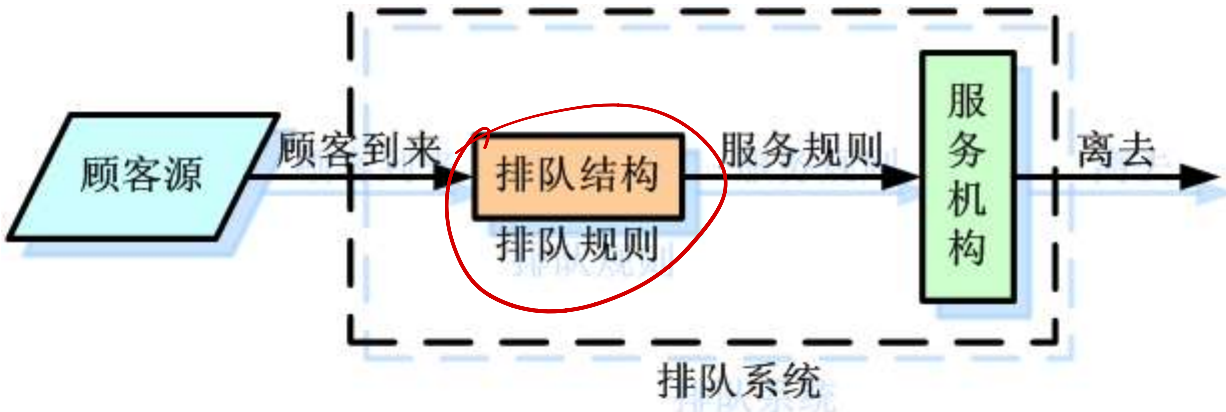
- 排队过程的一般表示
- 排队系统的组成和特征
- 排队模型的分类
- 排队问题的求解

13.1 基本概念

- 排队论（Queueing Theory）也称随机服务系统理论，它研究的内容有：
  - 性态问题：即研究各种排队系统的概率规律性，主要是研究队长分布、等待时间分布和忙期分布等，包括瞬态和稳态两种情形；
  - 最优化问题：又分静态最优和动态最优，前者指最优设计，后者指现有排队系统的最优运营；
  - 排队系统的统计推断：即判断一个给定的排队系统符合于哪种模型，以便根据排队理论进行分析研究。
- 排队过程的一般表示：
  - 有形排队现象：进餐馆就餐，到图书馆借书，车站等车，去医院看病，售票处售票，到工具房领物品等现象。

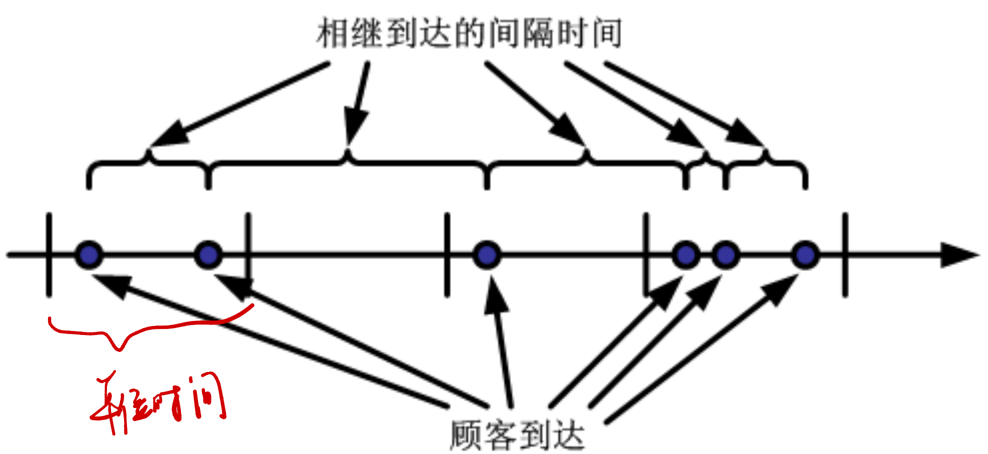
13.1 基本概念(cont.)

- 无形排队现象：如几个旅客同时打电话订车票；如果有一人正在通话，其他人只得在各自的电话机前等待，他们分散在不同的地方，形成一个无形的队列在等待通电话。
- 排队的不一定是人，也可以是物。如生产线上的原材料，半成品等待加工；因故障而停止运行的机器设备在等待修理；码头上的船只等待装货或卸货；要下降的飞机因跑道不空而在空中盘旋等。当然，进行服务的也不一定是人，可以是跑道，自动售货机，公共汽车等。



13.1 基本概念(cont.)

- 排队系统的组成和特征
  - 输入过程——即指顾客到达排队系统，可能有下列不同情况，这些情况并不彼此排斥：
    - 顾客的总体（称为顾客源）的组成可能是有限的，也可能是无限的。
    - 顾客到来的方式可能是一个一个的，也可能是成批的。
    - 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。



很好的刻画

随机型中的两种本质相同的表述：

单位时间内顾客到达数(期望)

相继到达间隔时间的概率分布

13.1 基本概念(cont.)

- 顾客的到达可以是相互独立的，即以前的到达情况对以后顾客的到来没有影响。或者是有关联的。
- 输入过程可以是平稳的，或称对时间是齐次的，指描述相继到达的间隔时间分布和所含参数（期望值、方差等）都是与时间无关的。否则就是非平稳的，非平稳情况的数学处理很难。

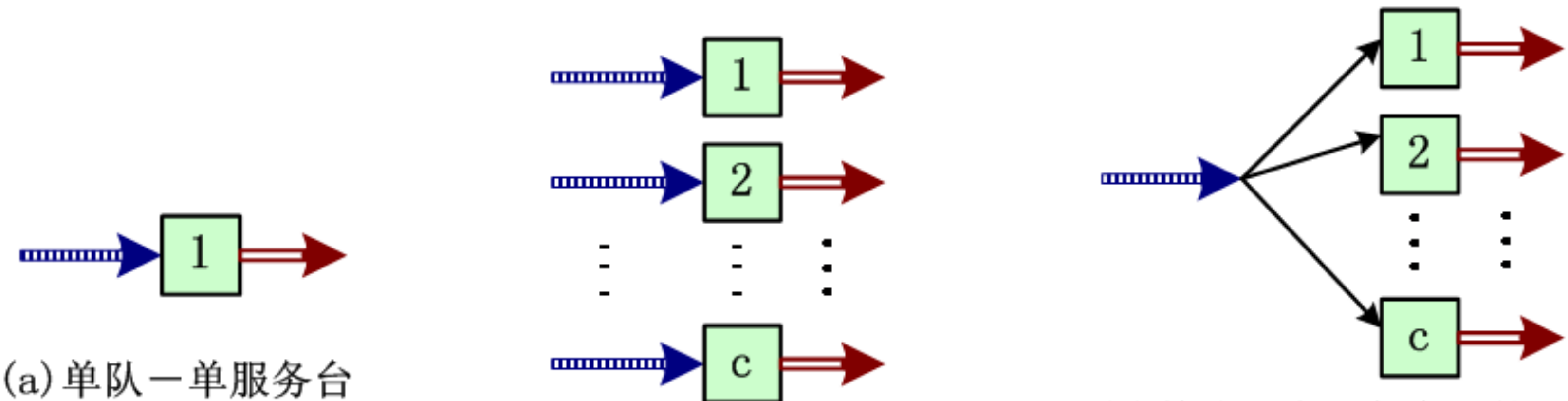
– 排队规则

- 顾客到达时，如果所有服务台都被占用，在这种情况下顾客可以随即离去，也可以排队等候。随即离去的称为即时制或损失制，排队等候的称为等待制。等待制可以采用下列规则：先到先服务、后到先服务、随机服务、有优先权的服务(VIP)。
- 从占有的空间看，队列可以排在具体的处所，也可以是抽象的。由于空间的限制或其它原因，有的系统要规定容量的最大限；有的没有限制。
- 从队列的数目看，可以是单列，也可以是多列。在多列情况下，各列间的顾客有的可以相互转移，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。

13.1 基本概念(cont.)

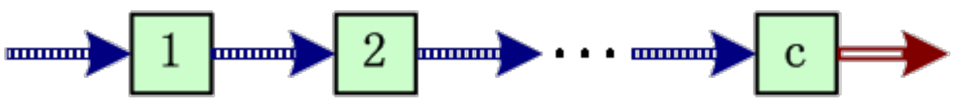
– 服务机构

- 服务机构可以没有服务员，也可以有一个或多个服务员。
- 在有多个服务台的情况下，它们可以是平行排列（并列的），可以是前后排列（串列的），也可以是混合的。

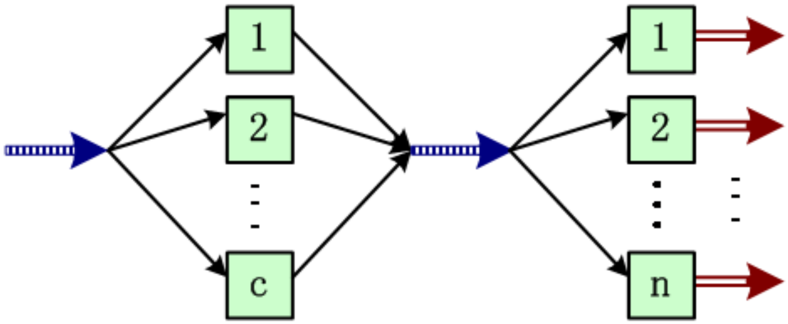


(a) 单队—单服务台　(b) 多队—多服务台（并列）　(c) 单队—多服务台（并列）

13.1 基本概念(cont.)



(d) 多服务台（串列）



(e) 多服务台（混合）

– 服务规则

- 服务方式可以对单个顾客进行，也可以对成批顾客进行。
- 和输入过程一样，服务时间也分确定型和随机型，对于随机型的服务时间，需要知道其概率分布。
- 和输入过程一样，服务时间的分布可以是平稳的，或非平稳的。

P8 有解释

13.1 基本概念(cont.)

- 本书排队论研究中的几点假设：
  - 假设顾客是单个到来的，即不研究成批到来情形；
  - 假设顾客的到达是相对独立的，即以前的到达情况对以后顾客的到来没有影响；
  - 假设输入过程和服务时间的分布是平稳的，即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；
  - 假设并列的各列间顾客不能相互转移、不能中途退出；
  - 假设服务方式只对单个顾客进行，即不研究对成批顾客服务的情形；
  - 假设排队规则采用先到先服务规则； →即假定为一队
  - 假设输入过程和服务时间至少有一个是随机型的情况，因为若二者都是确定型的则问题太过于简单。

13.1 基本概念(cont.)

- 排队模型的分类
  - 1953年，D.G.Kendall根据上面特征中最主要的、影响最大的三个，提出了Kendall记号，该记号只针对并列服务台的情形，表示为“X/Y/Z”，分别表示：
    - 相继顾客到达间隔时间分布；X
    - 服务时间的分布；Y
    - 服务台个数。Z
  - 表示相继到达间隔时间和服务时间的各种分布的符号是:
    - M——负指数分布（M是Markov的字头，因为负指数具有无记忆性，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。

13.1 基本概念(cont.)

– 例：
  - *M/M/1* 表示到达间隔时间为负指数分布、服务时间为负指数分布、单服务台的模型。
  - *D/M/c* 表示确定的到达间隔、服务时间为负指数分布、c个并行服务台（但顾客为一队）模型。

– 1971年关于排队论符号标准化会议上将Kendall符号扩充为：

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

X处填写顾客相继到达间隔时间的分布；

Y处填写表示服务时间的分布；

Z处填写并列的服务台数目；

A处填写系统容量限制N;

B处填写顾客源数目m；

C处填写服务规则，如先到先服务*FCFS*，后到先服务*LCFS*。

– 约定，如略去后三项，即指$X/Y/Z/\infty/\infty/FCFS$的情形。

13.1 基本概念(cont.)

- 排队问题的求解
  - 在求解排队问题时需要研究它属于哪个模型，其中顾客到达的间隔时间分布和服务时间分布需要实测的数据来确定。
  - 求解排队问题的目的是研究排队系统运行效率、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。
  - 解排队问题首先要求出用以判断系统运行优劣的基本数量指标的概率分布或特征数，常用的指标有：

正在服务的
正在排队的

    - 队长：指在系统中的顾客数，它的期望值记作 L_s；
    - 排队长（队列长）：指在系统中排队等待服务的顾客数，它的期望值记作L_q；

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

一般情形，L_s（或L_q）越大，说明服务效率越低。

$W_q = L_q/\lambda_e$

13.1 基本概念(cont.)

- 逗留时间：指一个顾客在系统中的停留时间，它的期望值记作W_s；
- 等待时间：指一个顾客在系统中排队等待的时间，它的期望值记作W_q；

  [逗留时间]=[等待时间]+[服务时间]

- 忙期（Busy Period）：指从顾客到达空闲服务机构起，到服务机构再次为空闲为止这段时间长度，即服务机构连续繁忙的时间长度。该指标关系到服务员的工作强度。忙期和一个忙期中平均完成服务顾客数都是衡量服务机构效率的指标。
- 在即时制或排队有限制的情形，由于顾客被拒绝而使企业受到损失的损失率和服务强度都是衡量服务机构效率的指标。

– 计算上述指标的基础是表达系统状态的概率。系统的状态指系统中顾客数，如果系统中有n个顾客就说系统的状态是n，它的可能值是：
  - 队长没有限制：n=0,1,2,……
  - 队长有限制，最大数为N时，n=0,1,2,……,N；
  - 即时制，服务台个数是c时，n=0,1,2,……,c。

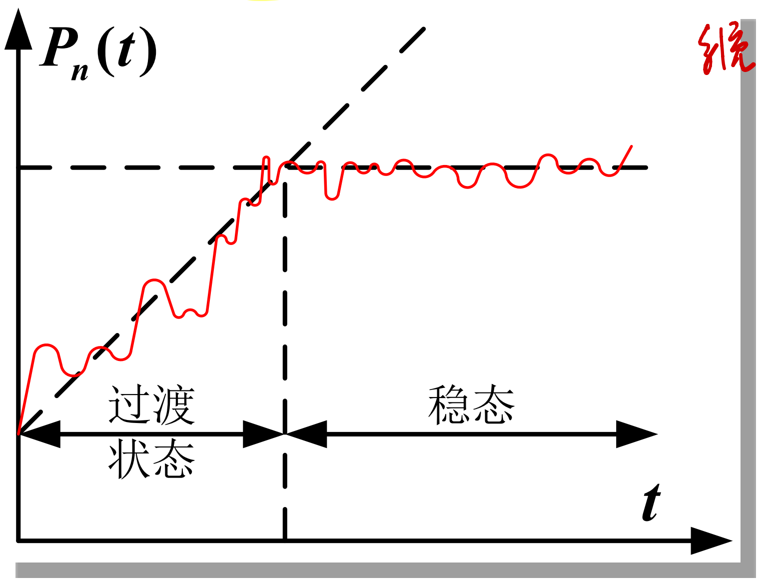
– 在上述第三种情况中，状态n又表示正在工作（繁忙）的服务台数。这些状态一般是随时刻t而变化，在时刻t、系统状态为n的概率用$P_n(t)$表示。

13.1 基本概念(cont.)

– 求状态概率$P_n(t)$的方法：

- 建立含$P_n(t)$的关系式（如图），因t是连续变量，而n只取非负整数，所以建立的$P_n(t)$关系式一般是微分差分方程（关于t的微分方程，关于n的差分方程）。
- 方程的解称为瞬态（或称过渡状态）（Transient State）解。
- 一般情况下，用极限 $\lim_{t\to\infty} P_n(t) = P_n$ 称为稳态（Steady State），或称统计平衡状态（Statistical Equilibrium State）的解。

系统一直运行下去的稳定状态



稳态的物理含义：

当系统运行了无限长的时间之后，初始（t=0）出发状态的概率分布（$P_n(t)$, $n \geqslant 0$）的影响将消失，而且系统的状态概率分布不会再随时间变化。

求稳态概率 P_n 时，不一定求 $t\to\infty$时 $P_n(t)$ 的极限，而只需令导数 $P_n'(t)=0$.

13.2 到达间隔的分布和服务时间的分布

解决排队问题首先要根据原始资料作出顾客到达间隔和服务时间的经验分布，然后按照统计学的方法确定合于哪种理论分布，并估计它的参数值。

下面讨论到达间隔的分布和服务时间的分布。分种情形讨论：

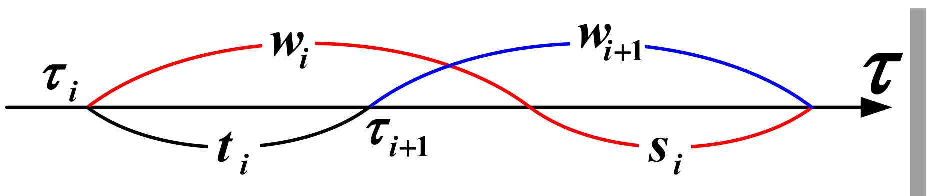
- 经验分布
- 泊松流
- 负指数分布
- 爱尔朗分布

13.2 到达间隔的分布和服务时间的分布

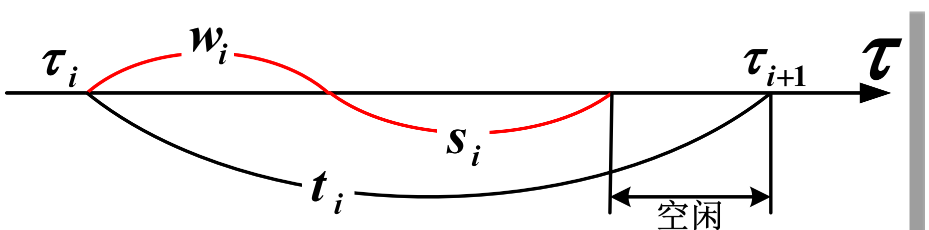
- 经验分布
  - 原始资料的整理，原始资料记录各顾客到达时刻和对各顾客的服务时间。

以τ_i表示第i号顾客到达的时间，以s_i表示对它的服务时间，

可算出相继到达的间隔时间$t_i(t_i=\tau_{i+1}-\tau_i)$和排队等待时间$w_i$。



$(a)w_i+s_i-t_i>0$



$(b)w_i+s_i-t_i<0$

如图所示：

间隔

$t_i=\tau_{i+1}-\tau_i$

等待时间

$$w_{i+1}=\begin{cases} w_i+s_i-t_i, & \text{当}w_i+s_i-t_i>0 \\ 0, & \text{当}w_i+s_i-t_i<0 \end{cases}$$

13.2 到达间隔的分布和服务时间的分布

– 例1：大连某港区2011年载货500吨以上船舶到达(不包括定期到达的船舶)逐日记录表：每月1至16日记录

月＼日	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	8	10	4	1	0	1	4	4	7	0	3	2	4	0	2	2
2	2	5	5	2	7	3	6	2	3	2	2	2	2	3	4	6
3	3	3	5	2	4	7	4	3	3	4	3	2	4	5	3	2
4	6	2	1	5	4	3	4	3	7	5	5	1	1	3	7	4
5	7	4	1	2	2	3	4	3	5	3	2	7	5	4	3	7
6	1	4	2	6	5	2	7	4	3	3	1	3	4	5	5	3
7	4	1	1	3	3	4	8	4	4	4	3	5	3	0	7	4
8	4	3	4	4	3	1	2	5	5	3	5	4	3	6	1	3
9	3	4	4	1	7	7	0	7	2	3	0	6	6	2	2	3
10	5	2	1	1	6	6	4	6	2	3	6	7	3	5	3	2
11	5	4	3	5	2	3	4	4	3	1	3	3	3	4	5	0
12	6	5	5	2	1	5	4	2	2	2	3	1	7	5	1	4

13.2 到达间隔的分布和服务时间的分布

– 例1：大连某港区**2011**年载货**500**吨以上船舶到达(不包括定期到达的船舶)逐日记录表：每月**17**至**31**日记录

月\日	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	5	4	2	1	1	3	6	2	3	4	4	2	0	3	1
2	2	2	6	5	2	1	5	3	5	5	2	2			
3	1	2	3	8	2	3	3	2	5	4	7	2	4	1	3
4	5	4	4	4	6	5	1	4	4	1	4	3	6	4	
5	2	1	6	3	2	5	2	5	4	2	4	4	4	4	2
6	5	2	4	3	3	3	6	3	5	5	6	2	1	4	
7	1	3	4	4	2	3	5	5	1	2	4	3	4	6	3
8	2	0	6	3	4	6	4	4	5	1	2	8	4	5	1
9	4	4	5	3	1	4	6	1	2	3	5	0	2	4	
10	6	2	5	1	0	7	9	3	2	5	1	7	3	5	3
11	2	1	0	3	6	6	3	5	5	1	2	2	2	4	
12	6	7	3	1	2	1	3	4	1	3	9	3	3	1	4

13.2 到达间隔的分布和服务时间的分布

– 原始资料的整理

船舶到达数 n	频　数 (天数)	频　率/%
0	12	0.033
1	43	0.118
2	64	0.175
3	74	0.203
4	71	0.195
5	49	0.134
6	26	0.071
7	19	0.052
8	4	0.011
9	2	0.005
10 以上	1	0.003
合　计	365	1.000

– 表中列出了一年（365天）每天的船只到达数，然后分别统计了到达0、1、……9以及10以上船只数的出现次数，并计算对应的频率以及每天平均到达率（船只数）。这里的计算结果，就可以作为经验为第二年提供参考。

$$\text{平均到达率}=\frac{\text{到达总数}}{\text{总天数}}=\frac{1\,271}{365}=3.48(\text{艘/天})$$

13.2 到达间隔的分布和服务时间的分布

例2：某服务机构是单服务台，先到先服务，对41个顾客记录到达时刻和服务时间(单位为分钟)，在表中以第1号顾客到达时刻为0。全部服务时间为127分钟。

各栏含义：

(1)顾客编号;

(2)到达时刻;

(3)服务时间;

(4)到达间隔;

(5)排队等待时间;

前三栏是原始记录；后两栏通过公式(13-1)计算得到。

(1)i	(2)τ_i	(3)s_i	(4)t_i	(5)w_i	(1)i	(2)τ_i	(3)s_i	(4)t_i	(5)w_i
1	0	5	2	0	22	83	3	3	2
2	2	7	4	3	23	86	6	2	2
3	6	1	5	6	24	88	5	4	6
4	11	9	1	2	25	92	1	3	7
5	12	2	7	10	26	95	3	6	5
6	19	4	3	5	27	101	2	4	2
7	22	3	4	6	28	105	2	1	0
8	26	3	10	5	29	106	1	3	1
9	36	1	2	0	30	109	2	5	0
10	38	2	7	0	31	114	1	2	0
11	45	5	2	0	32	116	8	1	0
12	47	4	2	3	33	117	4	4	7
13	49	1	3	5	34	121	2	6	7
14	52	2	9	3	35	127	1	2	3
15	61	1	1	0	36	129	6	1	2
16	62	2	3	0	37	130	3	3	7
17	65	1	5	0	38	133	5	2	7
18	70	3	2	0	39	135	2	4	10
19	72	4	8	1	40	139	4	3	8
20	80	3	1	0	41	142	1		9
21	81	2	2	2					

13.2 到达间隔的分布和服务时间的分布

例2：整理可得41个顾客到达间隔分布表和服务时间分布表

- 到达间隔分布表用于记录到达间隔分别为1、2、……、9以及10以上出现的频率，
- 服务时间分布表记录服务时间分别为1、2、……、8以及9以上出现的次数。
- 根据这两个表可以计算出平均到达时间间隔以及平均服务时间。

到达间隔/分钟	次　数
1	6
2	10
3	8
4	6
5	3
6	2
7	2
8	1
9	1
10 以上	1
合　计	40

服务时间/分钟	次　数
1	10
2	10
3	7
4	5
5	4
6	2
7	1
8	1
9 以上	1
合　计	41

平均间隔时间＝142/40＝3.55(分钟/人)

平均到达率＝41/142＝0.28(人/分钟)

平均服务时间＝127/41＝3.12(分钟/人)

平均服务率＝41/127＝0.32(人/分钟)

13.2 到达间隔的分布和服务时间的分布

- 泊松流（以顾客到达说明）
  - 设$N(t)$表示在时间区间$[0,t)$内到达的顾客数$(t>0)$，令$P_n(t_1,t_2)$表示在时间区间$[t_1,t_2)(t_2>t_1)$内有$n(\geq 0)$个顾客到达（随机事件）的概率，即
    $P_n(t_1,t_2)=P\{N(t_2)-N(t_1)=n\}(t_2>t_1,n\geq 0)$
    当$P_n(t_1,t_2)$合于下列三个条件时，就称顾客的到达形成泊松流。

(1)在不相重叠的时间区间内顾客到达数是相互独立的，称该性质为无后效性；

(2)对充分小的Δt，在时间区间$[t,t+\Delta t)$内有1个顾客到达的概率与t无关，而约与区间长Δt成正比，即$P_1(t,\ t+\Delta t)=\lambda\Delta t+o(\Delta t)$，其中$o(\Delta t)$，当$\Delta t\to 0$时，是关于$\Delta t$的高阶无穷小。$\lambda>0$是常数，表示单位时间有一个顾客到达的概率，称为概率强度。

? 此与后面λ的理解不太一致?
(后4页PPT)

(3)对于充分小的Δt，在时间区间$[t,t+\Delta t)$内有2个或2个以上顾客到达的概率极小，以致可以忽略，即$\sum_{n=2}^{\infty}P_n(t,\ t+\Delta t)=o(\Delta t)$

13.2 到达间隔的分布和服务时间的分布

– 顾客到达数n的概率分布

- 由条件(2)，取时间由0算起，简记$P_n(0,t)=P_n(t)$
- 由条件(2)(3)，在$[t,t+\Delta t)$区间内没有顾客到达的概率

$$P_0(t,t+\Delta t)=1-\lambda\Delta t+o(\Delta t)$$

- 区间$[0,t+\Delta t)$可以分成两个不重叠区间$[0,t)$和$[t,t+\Delta t)$，到达总数为n，分别出现在上面两个区间上，不外下表中的三种情况，见下表。

区间 \ 情况	$[0,t)$		$[t,t+\Delta t)$		$[0,t+\Delta t)$	
	个数	概率	个数	概率	个数	概率
(A)	n	$P_n(t)$	0	$1-\lambda\Delta t+o(\Delta t)$	n	$P_n(t)(1-\lambda\Delta t+o(\Delta t))$
(B)	$n-1$	$P_{n-1}(t)$	1	$\lambda\Delta t$	n	$P_{n-1}(t)\lambda\Delta t$
(C)	$n-2$	$P_{n-2}(t)$	2	$o(\Delta t)$	n	$o(\Delta t)$
	$n-3$	$P_{n-3}(t)$	3		n	
	$\vdots$		$\vdots$		$\vdots$	
	0		n		n	

13.2 到达间隔的分布和服务时间的分布

在$[0, t+\Delta t)$内到达n个顾客应该是表中三种情况之一，所以概率$P_n(t+\Delta t)$是表中三个概率之和：

$$P_n(t+\Delta t) = P_n(t)(1-\lambda\Delta t) + P_{n-1}(t)\lambda\Delta t + o(\Delta t)$$

$$\Rightarrow \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(\Delta t)}{\Delta t}$$

令$\Delta t \to 0$，得下列方程，并注意初始条件，有：

$$\begin{cases} \dfrac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t), & n \geq 1 \qquad (13.5) \\ P_n(0) = 0; \end{cases}$$

当$n = 0$时，没有$(B)(C)$两种情况，所以得：

$$\begin{cases} \dfrac{dP_0(t)}{dt} = -\lambda P_0(t) \\ P_0(0) = 1 \end{cases} \qquad (13.6)$$

泊松分布，λ：概率强度（前2页PPT）

解(13.5)和(13.6)即得：

$$\boxed{P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad t > 0, \quad n = 0,1,2,\cdots \qquad (13.7)}$$

13.2 到达间隔的分布和服务时间的分布

- $P_n(t)$表示长为t的时间区间内到达n个顾客的概率，由(13.7)式，随机变量$\{N(t)=N(s+t)-N(s)\}$服从泊松分布，其数学期望和方差是：
  $E[N(t)]=\lambda t$；$Var[N(t)]=\lambda t$ (t-取1)
- 期望值和方差相等是泊松分布的一个重要特征。

- 负指数分布
  - 对于负指数分布：

随机变量T的概率密度若是

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

则称服从负指数分布。其分布函数为

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

其数学期望$E[T]=\dfrac{1}{\lambda}$,方差$Var[T]=\dfrac{1}{\lambda^2}$,标准差$\sigma[T]=\dfrac{1}{\lambda}$

如果随机变量T表示顾客达到的间隔时间，则E[T]的值表明，λ是单位时间产生事件(到达)的速率。

∴在Kendall记号中均为M表示

- 负指数分布具有以下性质
  - 1）当输入过程是泊松流时，那么顾客相继到达的间隔时间T必须服从负指数分布。

    因为对于泊松流，在 [0,t) 区间内至少有一个顾客到达的概率是：

$$1-P_0(t)=1-\frac{(\lambda t)^0}{0!}e^{-\lambda t}=1-e^{-\lambda t},t>0$$

不一样吧？[0,t) 一人到达，那T不必≤t吗？

我觉得默认t=0时有顾客到来.

即 P{T≤t}（在≤t时间内有顾客相继到达，其时间间隔为T），有：

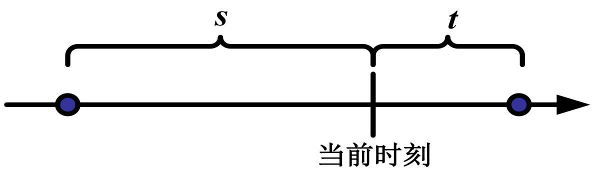
$$P(T\le t)=1-e^{-\lambda t}=F_T(t),t\ge 0$$

即间隔时间T服从负指数分布。

  - 2）具有Markov性（无记忆性）：若T表示排队系统中顾客到达的间隔时间，那么这个性质说明一个顾客到来还需的时间t与之前已经过去的时间s无关，即该情形下顾客到达是纯随机的。即：

$$P\{T>t+s\,|\,T>s\}=P\{T>t\}\qquad(13.11)$$

13.2 到达间隔的分布和服务时间的分布



- 证明： 根据条件概率有：

$$P\{T>t+s \mid T>t\}=\frac{P\{T>t+s, T>t\}}{P\{T>s\}}=\frac{P\{T>t+s\}}{P\{T>s\}}$$

负指数分布定义有：

$$\frac{P\{T>t+s\}}{P\{T>s\}}=\frac{1-P\{T\le t+s\}}{1-P\{T\le s\}}=\frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}=e^{-\lambda t}$$

$$=1-P\{T\le t\}=P\{T>t\}$$

13.2 到达间隔的分布和服务时间的分布

对于泊松流，λ表示单位时间平均到达的顾客数，所以1/λ表示相继顾客到达平均间隔时间，与E[T]的意义相符。

- 服务时间v的分布即对一顾客的服务时间，也就是在忙期相继离开系统的两顾客的间隔时间，有时也服从负指数分布。这时设它的分布函数和密度分别是

$$F_v(t) = 1 - e^{-\mu t}, \quad f_v(t) = \mu e^{-\mu t} \tag{13.12}$$

其中，μ表示单位时间能被服务完成的顾客数，称为平均服务率，

$\frac{1}{\mu} = E(v)$表示一个顾客的平均服务时间，这里平均就是期望值。

有联系

– 负指数分布和泊松分布在Kendall记号中都用M表示

泊松分布是单位时间内独立事件发生次数的概率分布 指数分布是独立事件的时间间隔的概率分布

*一个关于泊松分布和指数分布的浅显说明：http://www.ruanyifeng.com/blog/2015/06/poisson-distribution.html

13.2 到达间隔的分布和服务时间的分布

- 爱尔朗分布

设 $v_1, v_2, \cdots, v_k$ 是 k 个相互独立的随机变量，服从相同参数 $k\mu$ 的负指数分布，那么 $T = v_1 + v_2 + \cdots + v_k$ 的概率密度是

$$b_k(t) = \frac{\mu k(\mu k t)^{k-1}}{(k-1)!} e^{-\mu k t}, \quad t > 0, \qquad (13.13)$$

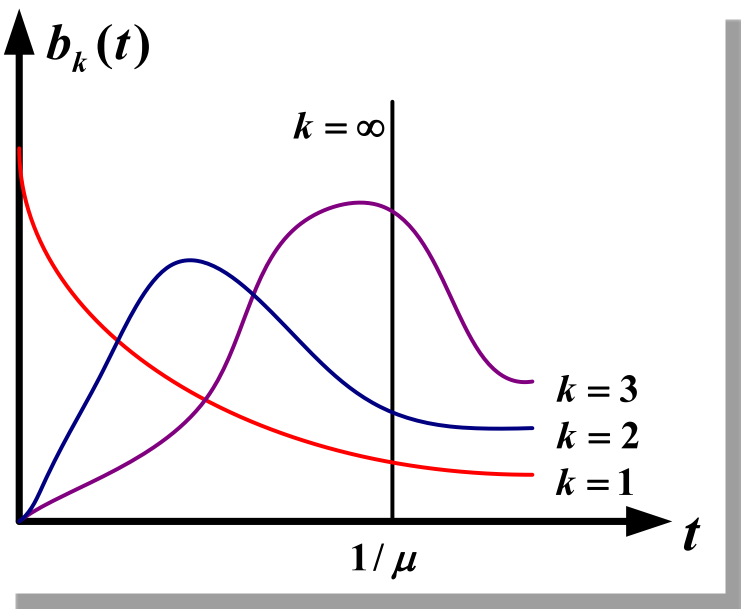
即 T 服从k阶爱尔朗分布。

- 其期望值和方差分别为：$E[T] = \frac{1}{\mu}$; $Var[T] = \frac{1}{k\mu^2}$ (13.14)
- 这是因为：$\because E(v_i) = \frac{1}{k\mu}, \quad i = 1,2,\cdots,k; \qquad \therefore E(T) = \sum_{i=1}^{k} E(v_i) = \frac{1}{\mu}$
- 例如串行的k个服务台，每台服务时间相互独立，服从相同参数的负指数分布，那么顾客走完这k个服务台总共需要的时间就服从上述的k阶爱尔朗分布。

爱尔朗分布提供更为广泛的模型类，比指数分布具有更大的适应性。

13.2 到达间隔的分布和服务时间的分布

爱朗分布提供更为广泛的模型类，能对现实世界提供更为广泛的适应性。



- 当k=1时，爱尔朗分布化为负指数分布，这可看成是完全随机的；
- 当k增大时，爱尔朗分布的图形逐渐变为对称的；
- 当k≥30时，爱尔朗分布近似于正态分布；
- 当k→∞时，Var[T]→0，因此这时爱尔朗分布化为确定型分布。