排队论

第十章

内容提要

- 排队论的基本概念
- 到达时间间隔的分布和服务时间的分布
- 生灭过程
- 单服务台负指数分布排队系统的分析
- 多服务台负指数分布排队系统的分析
- 一般服务时间模型

排队论的基本概念

第1节

要求接受服务的对象统称为**顾客**。由于顾客到达和服务时间的**随机性**，可是说排队现象几乎是不可避免的。

排队论的基本概念

- 排队论概述
  - 日常生活或生产中随处可见有形或无形的排队现象，如
    - 顾客到商店购买物品
    - 医院电话预约挂号
    - 水库水量的存储调节
    - 船只进港待泊，等等
  - 排队论(Queuing Theory)，又称随机服务系统理论，研究随机服务系统中的排队现象，恰当解决**顾客排队时间**和**服务设施费用**这一对**矛盾**，做到既保证一定的服务质量，又使服务设施费用经济合理。其研究内容主要有三个部分：
    - **排队性态问题**：研究各种排队系统的概率规律性（队长分布、等待时间分布、忙期分布等）。
    - **最优化问题**：研究排队系统的最优设计以及最优运营。
    - **排队系统的统计推断**：判断一个给定的排队系统符合哪种模型，以便根据排队理论进行分析。

Haiyan Lu
Jiangnan University

排队论的研究历程

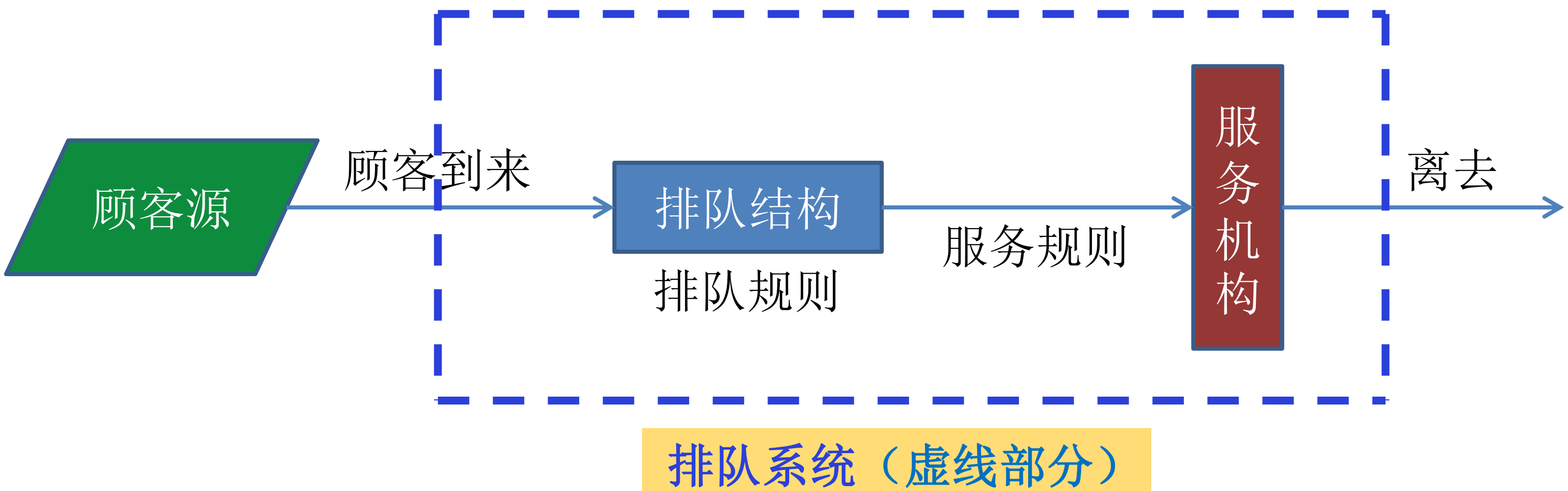
A.K. Erlang

- 理论起源及发展状况
  - 排队论起源于对电话系统的研究。
    - 1909年始，丹麦的电话工程师A. K. Erlang (1878–1929)等人对电话服务系统进行了长期的工作，取得了排队论最早的研究成果。
  - 随后，排队论陆续应用于陆空交通、机器管理、水库设计、可靠性理论等方面。
  - 20世纪60年代，随着计算机迅速发展的需要，排队论开始应用于计算机网络的最优设计。
  - 目前，无论在理论或应用上，排队论都有了飞速的发展，且由于计算机数字模拟技术的发展，排队论已成为工程设计和管理问题的有力工具。

Haiyan Lu 5/26/2020
Jiangnan University

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，服务完成后就离开。
  - **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。



排队系统的组成与特征

- 一般排队系统的构成
  - 输入过程(Input Process)
    - 又称到达过程(Arrival Process)
  - 输出过程(Output Process)
    - 又称服务过程(Service Process)
  - 排队规则(Queuing Discipline)
  - 服务机构(Server)

排队系统的组成与特征

- 输入过程
  - 指顾客到达服务系统的情况。可能有以下多种情况，且这些情况并非彼此排斥：
    - 顾客总体(称为顾客源)：
      - 有限顾客总体
      - 无限顾客总体
    - 顾客到来的方式：
      - 单个到来
      - 成批到来
    - 顾客相继到达的时间间隔：
      - 确定的到达时间间隔
      - 随机的到达时间间隔
    - 顾客到达的相互关系
      - 相互独立的到达(以前的到达情况对以后顾客的到达没有影响)
      - 有关联的到达
    - 顾客到达是否与时间相关
      - 平稳的输入过程(时间间隔的分布及其所含参数与时间无关)
      - 非平稳的输入过程

（主要讨论平稳的情形，非平稳过程的数学处理比较困难。）

排队系统的组成与特征

- 输出过程
  - 指顾客从得到服务到离开服务机构的情况。
    - 服务方式：
      - 单个服务
      - 成批服务
    - 服务时间：
      - 确定的服务时间
      - 随机的服务时间
    - 备注
      - 若相继到达的间隔时间和服务时间都是确定型的，则问题就太简单。因此，排队论所讨论的是二者至少有一个为随机型。
      - 与输入过程一样，总假定服务时间的分布是平稳的，即分布的期望值、方差等参数都不受时间的影响。

排队系统的组成与特征

- 排队规则
  - 指确定顾客接受服务的次序的规则。
    - 损失制：
      - 又称即时制。顾客到达时，若所有服务设施都被占用，则顾客自动离去。
        - 举例：有些电话服务系统。
    - 等待制：
      - 顾客到达时，若所有服务设施都被占用，则顾客留下来排队等待服务，直到服务完毕才离去。
        - 先到先服务(First Come First Served, FCFS)
        - 后到先服务(Last Come First Served, LCFS)
        - 随机服务(Service In Random Order, SIRO)
        - 有优先权的服务(Priority, PR)
      - 队列的容量：有容量限制的队列，无容量限制的队列
      - 队列的数目：单列，多列（只讨论各队列间不能互相转移，也不能中途退出的情形）
    - 混合制

排队系统的组成与特征

- 服务机构
  - 指服务设施的个数、排列和服务方式。
    - 服务设施的个数：
      - 单站服务系统
      - 多站服务系统
    - 服务设施的排列形式
      - 并联服务系统
      - 串联服务系统
      - 串并联混合服务系统
    - 服务设施的服务方式
      - 单个服务
      - 成批服务

排队模型的分类

D.G.Kendall在提出排队模型分类方法，对分类方法影响最大的特征有三个：

- 排队论模型分类法
  - Kendall记号：X/Y/Z
    - X：相继顾客到达时间间隔的分布
    - Y：服务时间的分布
    - Z：并联的服务台的个数

相继到达时间间隔和服务时间的各种分布的表示符号为：

- M：负指数分布(Markov性，即无记忆性)
- D：确定型(deterministic)
- E_k：*k*阶爱尔朗(Erlang)分布
- GI：一般相互独立的(general independent)时间间隔的分布
- G：一般(general)时间间隔的分布

举例：

- M/M/1：相继到达时间间隔为负指数分布、服务时间为负指数分布、单服务台
- D/M/c：确定的到达间隔、服务时间为负指数分布、c个平行服务台(顾客为一队)

Haiyan Lu 5/26/2020
Jiangnan University

排队模型的分类

- 排队论模型分类法
  - 扩充的Kendall记号：X/Y/Z/A/B/C
    - X：顾客相继到达时间间隔的分布
    - Y：服务时间的分布
    - Z：并联的服务台的个数
    - A：系统容量限制N
    - B：顾客源数目m
    - C：服务规则

  约定：
    - 若略去后三项，即指X/Y/Z/∞/ ∞ /FCFS
    - 本章只讨论FCFS情形，故可略去第六项。

  备注：

  Kendall于1953年提出了Kendall记号，1971年国际排队符号标准会上，将Kendall记号进行了扩充。

排队问题的求解

- 排队问题的题设
  - 求解排队问题时，首先要研究它属于哪个模型，其中只有顾客到达的时间间隔和服务时间的分布需要实测的数据来确定，其他因素都是在问题提出时给定的。
- 排队问题的求解
  - 求解排队问题的目的，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。
  - 因此，必须确定用以衡量排队系统运行状况的基本数量指标。

Haiyan Lu
Jiangnan University

排队问题的求解

这里公式中的s指的是system，q指的是queue

- 排队系统的主要指标
  - **队长**：又称系统的**状态**，指在系统中的顾客数(包括正在接受服务的顾客数)，其期望值记作L_s.
    队列长：指在系统中排队等待服务的顾客数，其期望值记作L_q.
    队长 = 队列长 + 正被服务的顾客数
  - **逗留时间**：指一个顾客在系统中的停留时间，其期望值记作W_s.
    等待时间：指一个顾客在系统中排队等待的时间, 其期望值记作W_q.
    逗留时间 = 等待时间 + 服务时间
  - **忙期**：指从顾客到达空闲服务机构起到服务机构再次空闲为止这段时间长度，即服务机构连续繁忙的时间长度。
    服务机构工作强度：指服务机构累计的工作时间占全部时间的比例.

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。

Haiyan Lu 5/26/2020
Jiangnan University

排队问题的求解

- 系统的状态
  - 系统的**状态**(state)，指在系统中的顾客数(其期望值为L_s)。
  - 若系统中有n个顾客，就称系统的状态是n，它的可能值为：
    - 队长没有限制时，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
      - 对于**即时制**，状态为n又表示正在工作(繁忙)的服务台数。
  - **状态概率**$P_n(t)$：系统在时刻t的状态$N(t)$为n的概率。
  - **稳定状态**：当一个排队系统开始运转时，系统状态很大程度上取决于系统的初始状态和运转经历的时间。很多排队系统在运行了相当长一段时间后，系统的状态将独立于初始状态和经历的时间(即初始状态概率分布的影响将消失，且系统的状态概率分布不再随时间变化)，这时称系统处于稳定状态(steady state)。
    - 由于对系统的**瞬时状态**(transient state)研究分析起来比较困难，因此排队论中主要研究系统处于稳态时的工作情况，这时$P_n(t)$可写成P_n。

Thank you!

谢 谢！