排队论

谭　忠

厦门大学数学科学学院

目录

案例引入

某大学工程技术学院每学期都有超过 100 人上课，期末时有近 500 份成绩上报.

核对登记成绩是一项单调和乏味的工作.

从人员设置的角度来考虑，

教务员数量多，平时人员冗余过多，经济性不好；

教务员数量少，忙期工作强度过大，易出错，效率低.

根据教务员平时的工作状态及教学运行的实际情况，

特以学期末的工作状态为考察对象，来度量决定教务员的合理数量.

根据统计数据观察，

教师相继提交成绩单的时间间隔服从负指数分布，

平均间隔时间为10分钟.

教务员登记成绩单所需时间也服从负指数分布，

且平均一个教务员登记一份成绩单耗时15分钟.

已知该教学科已有一名教务员，
现考虑是否需要再增加教务员，
增加几个合适？

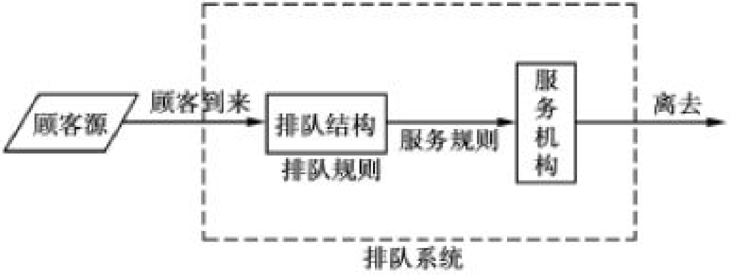
1 基本概念

排队是日常生活中常见的一种现象，
请求服务的人或物
(如候诊的病人，请求着陆的飞机等，称为“顾客”)
超过了服务机构
(为顾客提供服务的人或物，如医生，飞机跑道等)的
容量，就出现了排队现象.

基本概念

若增添服务设备，就要增加投资或发生空闲浪费；
但设备太少，排队现象会更加严重，
对顾客个人和对社会都带来不利影响.
因此，管理人员必须考虑在这两者之间取得平衡，
研究改进对策，以提高服务质量，降低成本.
排队论：为解决上述问题而发展的一门学科，
也称“随机服务系统理论”

基本概念



基本概念

排队论主要研究的内容有三个方面：

(1)性态问题：

研究排队系统的概率分布规律，

主要研究队长分布，

等待时间分布以及忙期分布等

基本概念

(2)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，

以便于根据排队理论进行分析研究

(3)最优化问题：

分为静态最优化和动态最优化，

即为系统的最优设计

和系统的最优运营问题

基本概念

一般而言，排队系统主要有三个基本部分组成：

输入过程、排队规则和服务过程.

1、输入过程：

即顾客到达排队系统的过程，其具有如下特征：

(1)顾客源的组成可能是有限，也有可能是无限的；

(2)顾客到达方式可能是一个一个的，也有可能是成批的；

基本概念

(3)顾客的到达是相互独立的；

(4)输入过程是平稳的，即相继到达的间隔时间分布及其数学期望、方差等数字特征都与时间无关.

2、排队规则：即到达排队系统的顾客按怎样的规则排队等待，可分为损失制，等待制和混合制三种.

(1)损失制：当顾客到达时，所有的服务台均被占用，顾客随即离去.

基本概念

(2)等待制：当顾客到达时，所有的服务台均被占用，顾客就排队等待，直到接受完服务才离去. 例如出故障的机器排队等待维修就是这种情况.

(3)混合制：介于损失制和等待制之间，即既有等待又有损失. 有队列长度有限和排队等待时间有限两种情况，在限度以内就排队等待，超过一定限度就离去.

基本概念

3、服务过程：

(1)服务机构：主要有以下几种类型：

单服务台，多服务台并联（每个服务台同时为不同顾客服务），多服务台串联（多服务台依次为同一顾客服务），混合型.

(2)服务规则：①先到先服务，这是通常的情形.

②后到先服务，如情报系统中，最后到的情报信息

基本概念

往往最有价值，因而常被优先处理.

③随机服务，服务台从等待的顾客中随机地取其一进行服务，而不管到达的先后.

④优先服务，如医疗系统对病情严重的病人给予优先治疗.

1.1 排队论模型的符号表示

目前广泛采用的服务系统符号表示为:

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

X 为顾客相继到达时间间隔的分布

Y 为服务时间的分布

Z 为并列服务台的个数

A 为排队系统的容量

基本概念

B 为顾客源数

C 服务规则

表示相继到达间隔时间和服务时间分布的符号为：

M——负指数分布

(Markov，负指数分布具有无记忆性)

D——定长分布(Deterministic)

E_k——k阶厄朗分布

基本概念

GI——一般相互独立的时间间隔分布

G——一般服务时间随机分布

例如，$M/M/1$ 表示相继到达间隔服从负指数分布、服务时间服从负指数分布、单服务台的模型. 一般约定，在服务规则为先到先服务，顾客源为 ∞，系统容量为 ∞，这可以省略不写.

1.2 排队系统的运行指标

(1)队长:

指系统内顾客数

(包括正被服务的顾客与排队等待服务的顾客)

的数学期望,

记作 $\boldsymbol{L_s}$.

(2)队列长：

指系统内等待服务的顾客数的数学期望，

记作 L_q .

队长和队列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(即平均队长和平均队列长)及有关的矩. 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.

(3)逗留时间：

指顾客在系统内逗留时间(包括排队等待的时间和接受服务的时间) 的数学期望，记作 W_s.

(4)等待时间：

指一个顾客在排队系统中排队等待时间的数学期望，记作 W_q .

(5)忙期：

指服务机构连续繁忙时间(顾客到达空闲服务机构起，到服务机构再次空闲止的时间)长度的数学期望，记为 T_b

1.3 Little公式

在排队论中,

有两个重要的公式,

适用于存在稳态分布的任何排队系统.

（1）$L_s = \lambda W_s$

此公式揭示了 L_s（队长）与 W_s（逗留时间）之间的数量关系.即：平均逗留队长等于单位时间内到达并进入系统的平均顾客数与平均逗留时间的乘积.

λ为平均强度，单位时间内到达并进入系统的平均顾客.

（2）$L_q = \lambda W_q$

此公式揭示了 L_q（队列长）与 W_q（等待时间）之间的数量关系.

即：平均等待队长等于单位时间内到达并进入系统的平均顾客数与平均等待时间的乘积.

2 到达间隔时间分布和服务时间分布

系统状态是求运行指标的基础,

所谓系统状态是指系统中顾客的数量.

如果系统中有 n 个顾客,

则说系统的状态是 n,

即 n 可能取值为：

(1)当队长无限制时，

则 $n = 0, 1, 2, \ldots$

(2)当队长有限制，且最大值为 N 时，

则 $n = 0, 1, 2, \ldots, N$

(3)当服务台个数为 c，且服务为即时制时，

则 $n = 0, 1, 2, \ldots, c$

到达间隔时间分布和服务时间分布

一般来说状态取值与时间 t 有关,

因此在时刻 t 系统状态取值为 n 的概率记为 $Pn(t)$,

若 $Pn(t) \to Pn$ 则称为稳态.

实际上多数平衡问题都是属于稳态的情况,

但不是真正的 $t \to \infty$,

即过一段时间以后就有 $Pn(t) \to Pn$.

2.1 定长分布

定长分布是最简单的一种常用概率分布.

它既能代表某些到达间隔的分布,

也能代表某些服务时间的分布.

到达间隔时间分布和服务时间分布

在定长输入中，

顾客等间隔地有规律到达，

即每隔固定时间 $c_1(c_1 > 0)$ 到达一位，

有间隔 $\tau_k = c_1$，

即 $P(\tau_k = c_1) = 1$,

到达间隔时间分布和服务时间分布

某些自动装配生产线上的装配件就是如此.

顾客相继到达的时间间隔分布函数为

$$A(t)=P(\tau_k\leq t)=1, t\geq c_1$$

类似地，当服务时间 v_k 服从定长分布时，若每个顾客的服务时间为常数 $c_2(c_2>0)$，则服务时间的分布函数为

$$B(t)=P(v_k\leq t)=1, t\geq c_2$$

2.2 泊松流

设 $\boldsymbol{N(t)}$ 表示在时间段 $[\mathbf{0}, \boldsymbol{t})$内到达的顾客数，

$\boldsymbol{P_n(t_1, t_2)}$ 表示在时间段 $[\boldsymbol{t_1}, \boldsymbol{t_2})$

内有 $\boldsymbol{n}(\boldsymbol{n} \geq \mathbf{0})$ 个顾客到达的概率，即

$$\boldsymbol{P_n(t_1, t_2) = P\{N(t_2) - N(t_1) = n\}}$$

到达间隔时间分布和服务时间分布

当 $P_n(t_1, t_2)$ 满足以下三个条件时，

则称顾客的到达形成泊松流：

(1)无后效性：

在不相交时间区间内顾客到达数是相互独立的，

即在时间段 $[t, t+\Delta t]$ 内

到达 k 个顾客的概率与 t 之前到达多少顾客无关.

到达间隔时间分布和服务时间分布

(2)平稳性：

对于充分小的 Δt，在时间段 $[t, t+\Delta t]$ 内到达 1 个顾客的概率只与时间长度 Δt 有关，而与起始时刻 t 无关，且

$$P_1(t, t+\Delta t) = \lambda\Delta t + o(\Delta t),$$

其中 $\lambda > 0$ 称为概率强度，

即表示单位时间内有一个顾客到达的概率.

(3)普通性：

对于充分小的 Δt，在时间段 $[t, t+\Delta t]$ 内有 2 个或 2 个以上的顾客概率极小，可以忽略不计，

即 $\sum\limits_{n=2}^{\infty} P_n(t, t+\Delta t) = o(\Delta t)$

定理:

若顾客的到达形成泊松流，则有

$$P_n(0,t) = P\{N(t) - N(0) = n\} = \frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

即在时间段 $[0,t)$ 内到达的顾客数 $N(t)$

服从参数为 λt 的泊松分布，且有

$$E[N(t)] = \lambda t, D[N(t)] = \lambda t$$

2.3 负指数分布

负指数分布是排队论中最常用、最重要的一种分布.

它既能代表某些到达时间的间隔的分布，也能代表某些服务时间的分布.

到达间隔时间分布和服务时间分布

若随机变量 T 的概率密度为

$f_T(t) = \lambda e^{-\lambda t}, t \geq 0$，其中常数 $\lambda > 0$，

则称 T 服从参数为 λ 的负指数分布.

对应的其分布函数为：

$$F_T(t) = 1 - e^{-\lambda t}, t \geq 0$$

其数学期望 $E(T) = \frac{1}{\lambda}$，方差 $D(T) = \frac{1}{\lambda^2}$.

负指数分布有下列性质：

(1)无后效性：

由条件概率公式易证

$$P\{T > t + s | T > s\} = P\{T > t\}$$

若 T 表示排队系统中顾客到达的间隔时间，无后效性说明一个顾客到来所需时间与过去一个顾客到来所需时间 s 无关.

(2)当输入过程是泊松流时，

那么顾客到达的间隔时间 T 必须服从负指数分布.

相继到达的间隔时间是独立同分布于负指数分布(密度函数为$\boldsymbol{\lambda e^{-\lambda t}, t \geq 0}$)，与输入过程为泊松流(参数为$\boldsymbol{\lambda}$)是等价的.

对于泊松流，$\boldsymbol{\lambda}$表示单位时间平均到达的顾客数，所以$\frac{1}{\lambda}$就表示顾客相继到达平均间隔时间，正与$\boldsymbol{E(T)}$意义

符合.

同理，对于服务时间v为负指数分布时，设其密度函数和分布函数为：

$$f_v(t) = \mu e^{-\mu t}, F_v(t) = 1 - e^{-\mu t}, t \geq 0$$

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，而$E(v) = \frac{1}{\mu}$表示一个顾客的平均服务时间.

2.4 厄朗分布

如果服务系统中有 k 个串联的服务台，

每个顾客必须经过每个服务台接受服务后才能离去.

设第 i 个服务台对每一个顾客的服务时间为 v_i 相互独立，都服从以 $k\mu$ 为参数的负指数分布，

则 k 个服务台的总服务时间是

$T=v_1+v_2+\ldots+v_k$，其中 T 是随机变量，

到达间隔时间分布和服务时间分布

总服务时间$T =$ 的分布密度为

$$b_k(t) = \frac{\mu k(\mu k t)^{k-1}}{(k-1)!} e^{-k\mu t}$$

这个分布称为 k 阶厄朗分布.

其数学期望和方差为

$$E(T) = \frac{1}{\mu}, D(T) = \frac{1}{k\mu^2}$$

爱尔朗分布随 k 而变化，当 $k = 1$ 时，成为负指数分布.

2.5 生灭过程

生灭过程是描述生物系统中生灭现象的一种随机过程.

在随机服务系统中，顾客的到达和离去就是一种生灭现象.

因此，可用生灭过程来解决某些随机服务系统问题.

到达间隔时间分布和服务时间分布

生灭过程的定义如下:

设一个系统具有可数个状态 $0, 1, 2, \ldots$

或有限个状态 $0, 1, 2, \ldots, m$

令系统在时刻 t 的状态记为 $N(t)$

在任一时刻 t，若系统处于 $N(t) = j$

则在时间区间$(t, t + \Delta t)$ 内，只能转移到相邻的状态或者保持状态不变，而不允许有两次或者更多的转移.

到达间隔时间分布和服务时间分布

其转移的概率如下：

(1)由状态 j 转移到 $(j+1)$ 的概率如下：

$$P\{N(t+\Delta t)=j+1/N(t)=j\}=\lambda_j\Delta t+o(\Delta t)$$

(2)由状态 j 转移到 $(j-1)$ 的概率为：

$$P\{N(t+\Delta t)=j-1/N(t)=j\}=\mu_j\Delta t+o(\Delta t)$$

(3)状态保持不变的概率为

$$P\{N(t+\Delta t)=j/N(t)=j\}=1-(\lambda_j+\mu_j)\Delta t+o(\Delta t)$$

λ_j 增长率，μ_j 消亡率，只与状态 j 有关的正常数.

到达间隔时间分布和服务时间分布

当系统状态没有限制时，

设$\boldsymbol{\lambda}_j = \boldsymbol{\lambda}, \boldsymbol{\mu}_j = \boldsymbol{\mu}(j = 1, 2 \ldots)$

此时，令 $\frac{\lambda}{\mu} = \rho$，并设 $\rho < 1$，则：

$$P_0 = 1 - \rho \tag{1}$$

$$P_j = \rho^j P_0 = \rho^j (1 - \rho) \tag{2}$$

当系统状态有限时,

设 $\lambda_j = \lambda(j = 0, 1, 2, ..., m-1), \lambda_m = 0$

$\mu_j = \mu(j = 1, 2, ..., m)$, 令 $\frac{\lambda}{\mu} = \rho$,

系统状态到达 m 时，增长率为零，故不要求 $\rho < 1$,

则

$$P_0 = \frac{1-\rho}{1-\rho^{m+1}} \tag{3}$$

$$P_j = \frac{(1-\rho)\rho^j}{1-\rho^{m+1}} \qquad (1 \leq j \leq m) \tag{4}$$

3 单服务台排队模型

设系统输入过程服从泊松流，服务时间服从负指数分布，则单服务的排队系统有如下三种情况：

(1)标准型：$M/M/1/\infty/\infty$

(2)系统容量有限型：$M/M/1/N/\infty$

(3)顾客源有限型：$M/M/1/\infty/m$

3.1 标准型：$M/M/1/\infty/\infty$

排队模型 $M/M/1/\infty/\infty$ 表示顾客源无限，

顾客的到达相互独立，

到达规律服从参数为 λ 的泊松分布；

单服务台排队模型

各顾客的服务时间相互独立，且服从参数为 μ 的负指数分布；单服务台，队长无限，先到先服务.

因而在时间$[t, t+\Delta t)$内有：

(1)有一个顾客到达的概率是$\lambda\Delta t + o(\Delta t)$

没有顾客到达的概率为$1-\lambda\Delta t + o(\Delta t)$.

(2)有一个顾客被服务完而离去的概率是

$$\mu\Delta t + o(\Delta t)$$

未服务完留在系统中的概率是$1-\mu\Delta t + o(\Delta t)$.

(3)多于一个顾客到达或者离去的概率都是 $o(\Delta t)$ ，可以忽略不计.

因此 $N(t)$ 是一个生灭过程,

到达率为 λ 即增长率 λ_j，服务率 μ 即消亡率 μ_j.

所以稳态时状态概率为式(1)和(2).

其中 $\rho = \frac{\lambda}{\mu}$ 在服务系统中成为业务强度.

在单服务台系统中，必须保证 $\rho < 1$

理论证明当 $\rho < 1$，时，系统就一定能达到稳态。

否则顾客可能越来越多，系统达不到稳定状态.

下面计算各个运行指标：

(1)系统中的顾客平均数（队长期望值）

$$L_s = \sum_{j=0}^{\infty} jP_j = \sum_{j=0}^{\infty} j(1-\rho)\rho^j = \frac{\rho}{1-\rho}$$

或 $L_s = \dfrac{\lambda}{\mu - \lambda}$

(2)队列中等待顾客的平均数 L_q

$$L_q=\sum_{j=1}^{\infty}(j-1)P_j=\sum_{j=1}^{\infty}jP_j-\sum_{j=1}^{\infty}P_j=L_s-\rho$$

或

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

(3)系统中顾客平均逗留时间 W_s

根据 Little 公式有

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

(4)在队列中顾客平均等待时间 $\boldsymbol{W_q}$

根据 Little 公式有

$$\boldsymbol{W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)}}$$

例3.1 某公路收费入口处设有一个收费亭，
汽车进入路口必须向收费处交费.
收费亭的收费时间服从负指数分布，
平均每辆汽车的收费时间为 7.2s，
汽车的到达率为 400辆/h，服从泊松分布.
试求：

单服务台排队模型

(1)收费亭空闲的概率；

(2)收费亭前没有车辆排队的概率；

(3)收费亭前排队长度

超过100m(即车辆超过12辆)的概率；

(4)平均排队长度；

(5)车辆通过收费亭所花时间的平均值；

(6)车辆的平均排队时间.

解：这显然是一个 $M/M/1/\infty/\infty$ 问题，

收费亭是服务台，汽车是顾客，

汽车向收费亭交费便是接受服务.

$\boldsymbol{\lambda = 400}$辆$\boldsymbol{/h, u = \frac{1}{7.2}}$辆$\boldsymbol{/s = 500}$辆$\boldsymbol{/h;}$

服务强度$\boldsymbol{\rho = \frac{\lambda}{u} = \frac{400}{500} = 0.8 < 1,}$

故排队系统是稳定的.

(1)收费亭空闲的概率：

即系统中没有车辆到达的概率 P_0，即

$$P_0 = 1 - \rho = 0.2$$

单服务台排队模型

(2)收费亭前没有车辆排队的概率：

当系统中没有车辆或者只有一辆车辆的时候

(这辆车正在被服务)，便没有车辆排队，即：

$$P(j \le 1) = P_0 + P_1 = 0.2 + 0.2 \times 0.8 = 0.36$$

(3)收费亭前排队长度超过 100m

(即车辆超过 12 辆)的概率：

即系统中车辆超过 13 辆的概率：

$$P(j>13)=1-P(j\le 13)=1-\sum_{i=0}^{13}(1-\rho)\rho^i$$

$$=1-(1-\rho)\times\frac{\rho^0\times(1-\rho^{14})}{(1-\rho)}=\rho^{14}=0.8^{14}=0.044$$

(4)平均排队长度:

$$L_q=\frac{\rho^2}{1-\rho}=\frac{0.8^2}{1-0.8}=3.2\text{辆}$$

(5)车辆通过收费亭所花时间的平均值:

$$W_s=\frac{L_s}{\lambda}=\frac{L_q+\rho}{\lambda}=\frac{3.2+0.8}{\frac{400}{3600}}=36s/\text{辆}$$

(6)车辆的平均排队时间:

$$W_q=\frac{L_q}{\lambda}=\frac{3.2}{\frac{400}{3600}}=28.8s/\text{辆}$$

4 多服务台并列排队模型

研究单队、并列的 n 个服务台的情形，

主要有以下三种情况：

(1)标准型：$M/M/n/\infty/\infty$

(2)系统容量有限型：$M/M/n/N/\infty$

(3)顾客源有限型：$M/M/n/\infty/m$

标准型：$M/M/n/\infty/\infty$

关于标准的 $M/M/n$ 模型各种特征的规定

与标准的 $M/M/1$ 模型的规定相似.

此外，我们规定：

顾客到达后排成一队，先到先服务.

多服务台并列排队模型

各服务台平均服务率相同，均为 μ

而且各台工作相互独立，

整个服务机构最大的平均服务率为 $n\mu$.

令 $\rho = \frac{\lambda}{n\mu}$，$\rho$ 称为此系统的业务强度.

只有当 $\rho < 1$ 时，队长才能稳定.

多服务台并列排队模型

(1)状态概率：多个服务台时，

正在接受服务的顾客数随系统状态而变，

$j \leq n$ 时，为 j 个，

而当 $j > n$ 时为 n 个，

故状态为 j 时的服务率为

$$\mu_j = \begin{cases} j\mu, j \leq n \\ n\mu, j > n \end{cases} \tag{1}$$

标准的$M/M/n$模型中，系统状态$N(t)$是一个生灭过程. 到达率λ即增长率，服务率μ_j即消亡率.

多服务台并列排队模型

稳态时的状态概率为：

$$P_0 = [\sum_{k=0}^{n-1} \frac{1}{k!}(\frac{\lambda}{\mu})^k + \frac{1}{n!}(\frac{\lambda}{\mu})^n \frac{1}{1-\rho}]^{-1}$$

$$P_j = \begin{cases} \frac{1}{j!}(\frac{\lambda}{\mu})^j P_0, 1 \le j < n \\ \frac{1}{n!n^{j-n}}(\frac{\lambda}{\mu})^j P_0, j \ge n \end{cases} \tag{2}$$

(2)平均队长 L_s 和平均队列长 L_q

平均队列长即排队等待的顾客人数，计算如下：

$$L_q = \sum_{j=n+1}^{\infty} (j-n)P_j = \frac{(n\rho)^n \rho P_0}{n!(1-\rho)^2}$$

多服务台并列排队模型

平均队长等于平均队列长加平均占用服务台数.

若单位时间内平均到达 $\boldsymbol{\lambda}$ 个顾客,

每台平均到达顾客为 $\frac{\boldsymbol{\lambda}}{\boldsymbol{n}}$.

所以每个服务台的利用率为 $\frac{\lambda}{n\mu} = \rho$.

即正在接受服务的顾客平均数.

所以，n 个服务台正在接受服务的顾客平均为 $n\rho$.

故队长为：

$$L_s = L_q + n\rho$$

多服务台并列排队模型

(3)平均等待时间 W_s 和逗留时间 W_q

利用 Little 公式可得：

$$W_q = \frac{L_q}{\lambda} = \frac{n\mu(\frac{\lambda}{\mu})^n}{n!}P_0 \cdot \frac{1}{(n\mu)^2(1-\frac{\lambda}{n\mu})^2}$$

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

例4.1 某售票所有三个窗口，顾客的到达数服从泊松分布，平均到达率为 0.9人/min，售票时间服从负指数分布，平均服务率为 0.4 人/min，现有顾客到达后排成一个队，以此向空闲的窗口购票.

试计算该窗口的状态指标和运行指标.

多服务台并列排队模型

解：这显然是个 $M/M/3/\infty/\infty$ 系统，并且有：

$$n = 3, \lambda = 0.9\text{人}/\min, u = 0.4\text{人}/\min$$

$$\frac{\lambda}{u} = \frac{0.9}{0.4} = 2.25, \rho = \frac{\lambda}{nu} = \frac{0.9}{3 \times 0.4} = 0.75 < 1$$

故该排队系统稳定.

(1)整个售票所空闲的概率：

$$P_0 = \frac{1}{\sum_{k=0}^{n-1} \frac{1}{k!}(\frac{\lambda}{\mu})^k + \frac{1}{n!}(\frac{\lambda}{\mu})^n \frac{1}{1-\rho}}$$

$$= \frac{1}{\sum_{k=0}^{2} \frac{1}{k!}(2.25)^k + \frac{1}{3!}(2.25)^3 \frac{1}{1-0.75}}$$

$$= 0.0748$$

多服务台并列排队模型

(2)平均的排队长度：

$$
\begin{aligned}
L_q &= \frac{(n\rho)^n \rho P_0}{n!(1-\rho)^2} \\
&= \frac{2.25^3 \times 0.75 \times 0.0748}{3!(1-0.75)^2} \\
&= 1.70
\end{aligned}
$$

(3)系统中的平均顾客数量：

$$L_s = L_q + n\rho = 1.70 + 2.25 = 3.95$$

(4)每个顾客的平均等待时间：

$$W_q = \frac{L_q}{\lambda} = \frac{1.70}{0.9} = 1.89min$$

(5)每个顾客平均所花费的时间：

$$W_s = W_q + \frac{1}{u} = 1.89 + 2.5 = 4.39min$$

5 一般服务时间 M/G/1 模型

前面介绍的各种排队系统,

其输入过程都是泊松流,

服务时间都服从负指数分布,

故也统称泊松排队系统.

他们都属于生灭过程排队系统.

一般服务时间 M/G/1 模型

当一个排队系统的输入过程非泊松流，

或者其服务时间不服从负指数分布，

则称为非泊松排队系统.

本节仅以一般时间服务时间排队系统 $M/G/1$

为典型代表，重点介绍其性能指标的计算.

5.1 M/G/1一般服务时间系统

具体含义是:

(1)输入过程 $\{M(t), t \geq 0\}$

是强度为 λ 的泊松流,

设平均到达率 $\lambda > 0$.

一般服务时间 M/G/1 模型

(2)对每个顾客的服务时间 v_i 是相互独立且具有相同分布的随机变量，其期望值为 $E(v_i)$，方差为$D(v_i)$.

(3)单服务台，先到先服务.

(4)系统容量 ∞，为等待制系统，

有效平均到达率 $\boldsymbol{\lambda}_e = \boldsymbol{\lambda}$.

其状态集为可列状态集.

(5)顾客源 ∞.

(6)输入过程和服务过程相互独立.

采用嵌入马氏法可以

分析与求解 $M/G/1$ 系统.

下面直接引入$(P-K)$公式.

在 $M/G/1$ 系统中,

设顾客平均到达率 $\lambda > 0$,

一般服务时间 M/G/1 模型

一般服务时间 v_i 的

期望值为 $E(v_i)$，方差为 $D(v_i)$.

当 $\rho = \lambda E(v_i) < 1$ 时,

排队系统有稳态解，且有如下$(P-K)$公式

$$L_q = \frac{\rho^2 + \lambda^2 D(v_i)}{2(1-\rho)}$$

一般服务时间 M/G/1 模型

根据上述 P-K 公式以及 Little 公式等可以给出四个基本公式，

可以得到 $\boldsymbol{M/G/1}$ 系统的指标计算公式

一般服务时间 M/G/1 模型

$$L_q = \frac{\rho^2 + \lambda^2 D(v_i)}{2(1-\rho)}$$

$$L_s = L_q + \bar{S} = L_q + \rho$$

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

$$W_s = W_q + E(v_i)$$

显然，已知 $\lambda, E(v_i), D(v_i)$ 就可以求出 L_q，进而求出别的指标.而不管随机变量 v_i 有怎样的分布.

一般服务时间 M/G/1 模型

各项指标的值都与方差 $D(v_i)$ 有关，

只有方差为 0 时，

随机性的波动才不会影响各个指标的大小.

例5.1 有一个售票口，已知顾客按平均 2 分 30 秒的时间间隔的负指数分布到达，顾客在售票窗口前的服务时间平均为 2 分钟.

(1)若服务时间也服从负指数分布，

求顾客为购票所需的平均逗留时间和等待时间.

(2)若经过调查，

一般服务时间 $M/G/1$ 模型

顾客在售票口前至少要占用 1 分钟，

且认为服务时间服从负指数分布是不恰当的，

而应该服从以下概率密度分布：

$$f_y = \begin{cases} e^{-y+1} & y \geq 1 \\ 0 & y < 1 \end{cases}$$

再求顾客的逗留时间和等待时间.

解：(1)$\boldsymbol{\lambda = 1/2.5 = 0.4}$

$$u = 1/2 = 0.5, \rho = \frac{\lambda}{u} = 0.8$$

$$W_s = \frac{1}{u-\lambda} = 10min$$

$$W_q = \frac{\rho}{u-\lambda} = 8min$$

(2)令 y 为服务时间，那么 $\boldsymbol{Y} = \mathbf{1} + \boldsymbol{X}$，

$\boldsymbol{X}$ 服从均值为 1 的负指数分布.于是：

$$\boldsymbol{E(Y) = 2}$$

$$\boldsymbol{D(Y) = D(1 + X) = D(X) = 1}$$

$$\boldsymbol{\rho = \lambda E(Y) = 0.8}$$

一般服务时间 M/G/1 模型

代入 P-K 公式，得：

$$L_q = \frac{\rho^2 + \lambda^2 D(Y)}{2(1-\rho)} = 2$$

$$L_s = L_q + \rho = 2.8$$

$$W_s = L_s/\lambda = 7min$$

$$W_q = L_q/\lambda = 5min$$

6 案例分析

案例一、排队论在收费站设计中的应用

某条道路上要设收费站，单向车流量为 800 辆/h.

假设工作人员平均能在 8s 内处理一辆汽车，符合负指数分布.

试分析收费亭单项至少需设多少通道，并对不同的系统进项评估.

问题分析:

在给定的服务水平条件下

针对不同的收费车道数量 n,

利用 $M/M/n$ 排队系统来进行试算.

(1)设单通道的 $M/M/1$ 系统

$$\lambda = 800, u = \frac{3600}{8} = 450$$

$$\rho = \frac{\lambda}{u} = \frac{800}{450} = \frac{16}{9} > 1$$

即排队系统不稳定,

队长会越来越长，排队得不到消散.

案例分析

(2)设双通道的 $M/M/2$ 系统

$$\lambda = 800, u = \frac{3600}{8} = 450, n = 2$$

$$\frac{\lambda}{u} = \frac{800}{450} = \frac{16}{9} > 1, \rho = \frac{\lambda}{nu} = \frac{8}{9} < 1$$

即系统是稳定的.

$$P_0 = [\sum_{k=0}^{n-1} \frac{1}{k!}(\frac{\lambda}{\mu})^k + \frac{1}{n!}(\frac{\lambda}{\mu})^n \frac{1}{1-\rho}]^{-1} = 0.0588$$

案例分析

$$L_q = \frac{(n\rho)^n \rho P_0}{n!(1-\rho)^2} = 6.7$$

$$L_s = L_q + n\rho = 8.5$$

$$W_q = \frac{L_q}{\lambda} = 30s$$

$$W_s = W_q + \frac{1}{\mu} = 38s$$

案例分析

(3)设三通道的 $M/M/3$ 系统

$$\lambda = 800, u = \frac{3600}{8} = 450, n = 3$$

$$\frac{\lambda}{u} = \frac{800}{450} = \frac{16}{9} > 1, \rho = \frac{\lambda}{nu} = \frac{16}{27} < 1$$

即系统是稳定的.

$$P_0 = [\sum_{k=0}^{n-1} \frac{1}{k!}(\frac{\lambda}{\mu})^k + \frac{1}{n!}(\frac{\lambda}{\mu})^n \frac{1}{1-\rho}]^{-1} = 0.1502$$

案例分析

$$L_q = \frac{(n\rho)^n \rho P_0}{n!(1-\rho)^2} = 0.5$$

$$L_s = L_q + n\rho = 2.3$$

$$W_q = \frac{L_q}{\lambda} = 2.25s$$

$$W_s = W_q + \frac{1}{\mu} = 10.25s$$

案例分析

问题分析：$M/M/1$ 系统的交通强度 $\rho > 1$，

说明该系统无法满足交通强度的要求，

车辆排队长度会越来越长，无法消散.

$M/M/2$ 系统和 $M/M/3$ 系统

均满足交通强度 $\rho < 1$ 的要求，

案例分析

但 $M/M/2$ 系统车辆排长队的概率高，
服务台劳动强度大，服务水平低；
而 $M/M/3$ 系统的各项指标
均要大大优于 $M/M/2$ 系统.
因此，在没有其他条件限制的情况下，
$M/M/3$ 系统应为该收费亭设计之首选.

案例分析

注意：

(1)系统的选择应综合分析各种限制条件，

通道多的系统车辆排队短，服务水平高，

但建设规模大，占地多，投资大，运营成本高，

因此作好交通量调查时计算分析的前提条件.

案例分析

(2)车辆在服务台停留时间的长短

对系统影响很大,

若采用先进的收费系统,

缩短服务所需的时间,

对减少运营成本,

提高服务水平，效果显著.

案例二、排队系统的优化

1、微分法:

一装卸队专门来到码头仓库的货车装卸货物,

设货车的到达服从泊松分布,

平均每 10 分钟一辆,

而装卸时间与装卸队的人数 x 成反比.

又设该装卸队每班(8小时)

的生产费用为 $(\mathbf{20+4}\boldsymbol{x})$ 元,

货车在码头装卸货物时每小时的损失是 15 元.

案例分析

若(1)装卸时间为常数，

一名装卸工人装卸一辆汽车需要 30 分钟；

(2)装卸时间为负指数分布，

一名装卸工人装卸一辆汽车需要 30 分钟；

试分别确定该装卸队应配备的装卸工人数.

解：计算一小时的费用，

该费用包括装卸队费用

和汽车在系统中逗留的损失，即

$$c=(\frac{20+4x}{8})+15L_s$$

(1)装卸时间为常数

$$\lambda = 6, u = 2x, \rho = \frac{3}{x}$$

代入 L_s 的表达式有：

$$L_s = \rho + \frac{\rho^2}{2(1-\rho)} = \frac{3}{x} + \frac{(\frac{3}{x})^2}{2(1-\frac{3}{x})} = \frac{3}{2}[\frac{2x-3}{x(x-3)}]$$

所以：$c=2.5+0.5x+\frac{45}{2}[\frac{2x-3}{x(x-3)}]$

令 $\frac{dc}{dx}=0.5+\frac{45}{2}[\frac{2}{x(x-3)}-\frac{(2x-3)^2}{x^2(x-3)^2}]=0$

可得：$0.5x^4-3x^3-40.5x^2+135x-202.5=0$

案例分析

经过试算，该方程在 11 和 12 之间有一个根，

分别比较二者所对应的费用值，

因有 $c(11) = 12.858, c(12) = 13.422$，

故装卸队应配备 11 名装卸工人.

(2)装卸时间为负指数分布

$$c=\frac{20+4x}{8}+15(\frac{\lambda}{u-\lambda})=2.5+0.5x+15(\frac{6}{2x-6})$$

令 $\frac{dc}{dx}=0.5-\frac{90}{2}\times\frac{1}{(x-3)^2}=0$

可得：$(x-3)^2=90$，从而 $x\approx 12.5$

比较 $\boldsymbol{c(12)}$ 和 $\boldsymbol{c(13)}$,

由于 $\boldsymbol{c(12) = c(13) = 13.5}$,

故装卸队配备 12 或 13 名装卸工人均可.

案例分析

2、边际分析法

仅讨论标准 $M/M/c$ 模型，确定最优服务台数

且在稳态情形下，单位时间全部费用

（服务成本与等待费用）的期望值z可以表示为：

案例分析

$$z = c_s' \cdot c + c_w \cdot L$$

其中 c 是服务台数，

c_s' 是每个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或者队列等待的顾客平均数 L_q

(它们随 c 值的不同而不同).

案例分析

因为 c_s' 和 c_w 都是给定的,

唯一可能变动的是服务台数 c,

所以 z 是 c 的函数 $z(c)$,

现在是求最优解 c^* 使 $z(c^*)$ 为最小.

案例分析

采用边际分析法，

根据 $z(c^*)$ 是最小的特点，我们有

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

案例分析

将 z 代入，得

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

案例分析

上式化简后，得

$$L(c^*) - L(c^*+1) \leq \frac{c_s^{'}}{c_w} \leq L(c^*-1) - L(c^*)$$

依次求 $c = 1, 2...$ 时 L 的值，

并作两相邻的 L 值之差，

因 $\frac{c_s^{'}}{c_w}$ 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^*.

案例分析

例：

某车间有一个工具维修部，要求维修的工具按泊松流到达，平均每小时 17.5 件；维修部工人每人每小时平均维修 10 件，服从负指数分布.已知每名工人每小时的工资为 6 元，因工具维修使机器停产的损失为每台每小时 30 元.要求确定该维修部的最佳工人数.

解：本例 $c_s^{'} = 6, c_w = 30,$

故 $\dfrac{c_s^{'}}{c_w} = 0.2.$

分别计算 $c = 1, 2..., N$ 时的 L 值，

并计算相邻两个 L 值的差，计算结果见下表：

案例分析

c	$L(c)$	$L(c-1)-L(c)$
1	∞	—
2	7.467	∞
3	2.217	5.25
4	1.842	0.375
5	1.769	0.073
6	1.754	0.015

因：$L(4)-L(5)$

$$=0.073<0.2<0.375=L(3)-L(4)$$

所以该维修部最佳应配备 4 名工人.

谢谢!