



天津大学

TIANJIN UNIVERSITY

精品课程运筹学

第六章 排队系统分析

(Queuing Systems Analysis)

6.1 排队的基本概念

6.2 到达与服务的规律

6.3 M/M/1排队模型

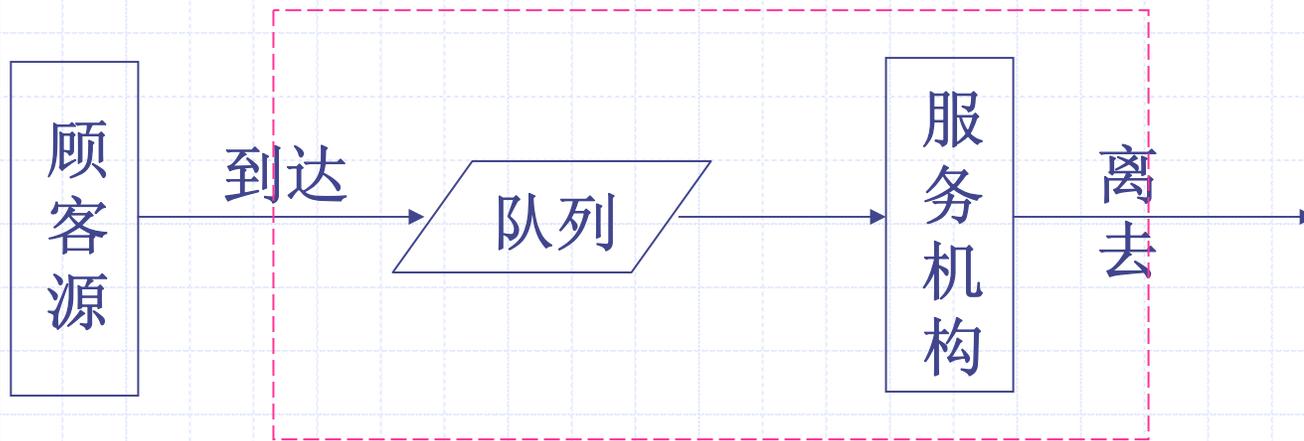
6.4 M/M/C排队模型

6.5 M/G/1排队模型

6.6 排队系统优化

6.1 排队的基本概念

一、排队系统的组成



现实系统中形形色色的排队系统

到达的顾客	服务内容	服务台
收费站排队的车辆	收费	收费站
病人	看病	医生
到达机场上空的飞机	降落	跑道
到达港口的货船	装货（卸货）	装卸码头
进入我方阵地的敌机	我方的防空火力	我方的防空导弹
需要加油的车辆	加油	加油站的加油机
电话呼唤	通话	交换台
出故障的机器	修理	修理技工

1、输入过程

- (1) 顾客源：分为
- 无限 ∞ (如电话呼唤)
 - 有限 m (如车间里待修理的机器)

(2) 到达规律：指到达间隔时间 T 的分布

分为

- 定长 D

- 负指数 M

- k 阶爱尔朗 E_k



2. 排队规则

(1) **损失制**: 指顾客到达时若所有服务设施均被占用, 则顾客自动离去。

(2) **等待制**: 指顾客到达时若所有服务实施均被占用, 则留下来等待, 直至被服务完离去。

等待的服务规则又可分为

- 先到先服务 (FCFS)
- 后到先服务 (LCFS)
- 带优先权服务 (PS)

(3) **混合制**: 分为

- 系统容量有限制
- 等待时间有限制

3. 服务机构

(1) 服务台个数 $C \begin{cases} = 1 \\ > 1 \end{cases}$ (并列多台)

(2) 服务规律: 指服务时间 v 的分布

分为 • 定长 D

• 负指数 M

• k 阶爱尔朗 E_k

• 一般分布 G



二、排队模型的表示

用记号 (X/Y/Z/A/B/C) 表示, 其中

X: 顾客到达时间间隔的分布

Y: 服务时间的分布

Z: 服务台个数

A: 系统容量 N

B: 顾客源数量 m

C: 服务规则

X ——表示顾客相继到达间隔时间分布，常用下列符号：

M ——表示到达过程为泊松过程或负指数分布；

D ——表示定长输入；

E_k ——表示 k 阶爱尔朗分布；

G ——表示一般相互独立的随机分布。

Y——表示服务时间分布，所用符号与表示顾客到达间隔时间分布相同。

M——表示服务过程为泊松过程或负指数分布；

D——表示定长分布；

E_k ——表示k阶爱尔朗分布；

G——表示一般相互独立的随机分布。

例1 $(M/M/1/\infty/\infty/FCFS)$ 表示:

到达间隔为负指数分布，服务时间也为负指数分布，1个服务台，顾客源无限，系统容量也无限，先到先服务。

三、排队问题的求解

主要是计算描述系统运行状态的指标:

1. 队长和排队长

队长: 系统中的顾客数; 其概率分布称状态概率, 记为 P_n , 表示系统中有 n 个顾客的概率; 队长的平均值记为 L_s 。

排队长: 系统中正在排队等待的顾客数, 记其均值为 L_q 。

2. 逗留时间和等待时间

逗留时间：一个顾客在系统中的停留时间，记为 W ，其均值记为 W_s 。

等待时间：一个顾客在系统中排队等待的时间，记其均值为 W_q 。

6.2 到达与服务规律

一、到达的规律

描述顾客到达规律可从两方面

- 到达间隔 (时间)
- 到达数 (人数)

现实中有许多服务系统，其顾客的到达具有下述特征：

- (1) 无后效性：任一时段的到达数不受前一时段的影响；
- (2) 平稳性：顾客到达是均匀的；
- (3) 稀有性：瞬时内只可能有1个顾客到达。

称具有上述特征的输入为泊松流，其在 t 时段内到达 n 个顾客的概率为

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, n = 0, 1, \dots$$

即参数为 λt 的泊松分布。

由概率论知识可知，泊松分布的参数即其均值。因此， λ 的含义是单位时间到达系统的平均顾客数，即到达率。

下面考察，当顾客按泊松流到达时，其到达的间隔时间 T 是服从什么分布呢？

$$P(T \leq t) = 1 - e^{-\lambda t}, \quad \text{即 } F(t) = 1 - e^{-\lambda t}$$

而这正是负指数分布的分布函数，说明 T 服从负指数分布，且参数同为 λ 。

可证反之也成立。于是得到关于到达规律的重要性质：

到达数为泊松流 \iff 到达间隔服从负指数分布（同参数）。

由概率论知识可知，负指数分布的表达式（密度函数）为

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

参数 λ 即其均值的倒数。因此， $\frac{1}{\lambda}$ 的含义是平均间隔时间，这与 λ 为单位时间到达系统的平均顾客数的含义一致。

负指数分布有一个有趣的性质：无记忆性，即

$$P(T > t_0 + t | T > t_0) = P(T > t)$$

直观上看，在已知 $T > t_0$ 的条件下估计 $T > t$ 的概率，与无条件时估计 $T > t$ 的概率相同，把以前的 t_0 时间给忘了。

事实上，

$$\begin{aligned} P(T > t_0 + t | T > t_0) &= \frac{P(T > t_0 + t) \cap (T > t_0)}{P(T > t_0)} \\ &= \frac{P(T > t_0 + t)}{P(T > t_0)} = \frac{e^{-\lambda(t_0+t)}}{e^{-\lambda t_0}} = e^{-\lambda t} = P(T > t) \end{aligned}$$

假若 T 表示某种电子元件的寿命，则当元件已使用了 t_0 时间后估计它还能再使用 t 时间的概率，与刚开始用时的概率一样。说明这种元件是高度耐磨损的。

二、服务的规律

主要讨论服务时间 v 服从负指数分布的情形，参数为 μ ，即

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

由于 v 的均值为 $\frac{1}{\mu}$ ，即平均对每位顾客的服务时间为 $\frac{1}{\mu}$ ，可得

参数 μ 的含义：服务率，即单位时间平均服务完 μ 人。

注：负指数分布的一般化——爱尔朗分布，可用于描述由 k 道程序组成的 1 个服务台的服务时间的分布。

6.3 M/M/1排队模型

一、标准的M/M/1模型 (M/M/1/ ∞ / ∞)

1、问题的一般提法

设：泊松输入 / 负指服务 / 单服务台 / 系统无限制 / 顾客源无限制

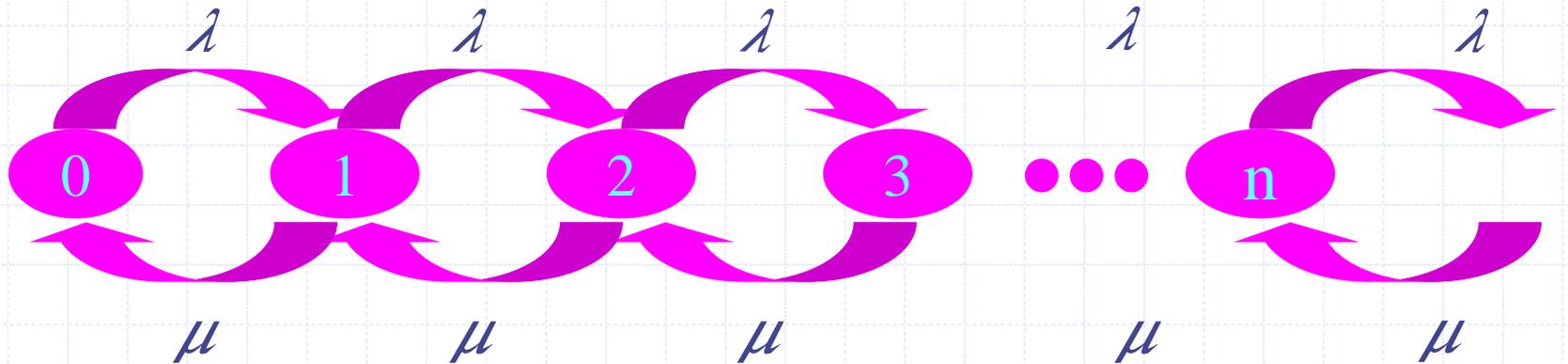
求：(1) 系统状态概率 P_n ;

(2) 系统运行指标 L_s , L_q , W_s , W_q 。

2. 系统状态概率

(1) 利用状态转移图列出平衡方程

状态转移图是处理稳态M/M/C系统的一种工具，设到达与服务率分别为 λ 和 μ ，则



⊕ (2) 由此列出平衡方程:

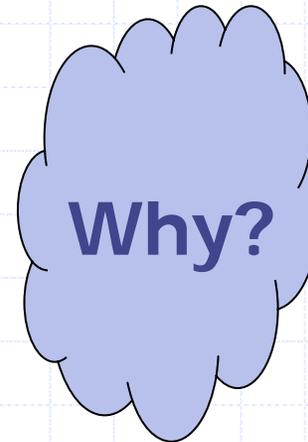
$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n, n \geq 1 \end{cases}$$

可解得状态概率:

$$\begin{cases} P_0 = 1 - \frac{\lambda}{\mu} \\ P_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right), n \geq 1 \end{cases}$$

记 $\frac{\lambda}{\mu} = \rho$ ，称为服务强度，规定 $\rho < 1$ ，则

$$\begin{cases} P_0 = 1 - \rho \\ P_n = \rho^n P_0 \end{cases}$$



3. 系统运行指标

(1) L_s 与 L_q

$\therefore L_s$ 表示系统中的平均顾客数，由期望定义，

$$\begin{aligned}\therefore L_s &= \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n\rho^n(1-\rho) = \rho(1-\rho)\sum_{n=1}^{\infty} n\rho^{n-1} \\ &= \rho(1-\rho)\sum_{n=1}^{\infty} \frac{d\rho^n}{d\rho} = \rho(1-\rho)\frac{d}{d\rho}\sum_{n=0}^{\infty} \rho^n \\ &= \rho(1-\rho)\frac{d}{d\rho}\frac{1}{1-\rho} = \rho(1-\rho)\frac{1}{(1-\rho)^2} \\ &= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}\end{aligned}$$

$$\begin{aligned} L_q &= \sum_{n=1}^{\infty} (n-1)P_n = \sum_{n=0}^{\infty} nP_n - \sum_{n=1}^{\infty} P_n = L_s - (1 - P_0) \\ &= L_s - \rho \end{aligned}$$

其中 $\rho = \frac{\lambda}{\mu} < 1$ 。问题：为什么 $L_s - L_q = \rho < 1$ (而不是 $= 1$) 呢？

——因为是均值。

(2) W_s 与 W_q

首先可证，逗留时间 W 服从参数为 $\mu - \lambda$ 的负指数分布，而负指数分布的均值等于其参数的倒数，故平均逗留时间

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

平均等待时间等于平均逗留时间减去平均服务时间，即

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

(3) 上述4个指标之间的关系——Little公式

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

一般的里特公式中 λ 应为 λ_e ，称有效到达率，即实际进入系统率。本模型中因系统容量无限制，故 $\lambda_e = \lambda$ 。

例2 某修理店只有一个修理工人，来修理的顾客到达数服从泊松分布，平均每小时4人；修理时间服从负指数分布，平均需6分钟。求：（1）修理店空闲的概率；（2）店内有3个顾客的概率；（3）店内至少有1个顾客的概率；（4）店内顾客的平均数；（5）顾客在店内的平均逗留时间；（6）等待服务的顾客平均数；（7）平均等待修理时间；（8）必须在店内消耗15分钟以上的概率。

解：此为标准的 M/M/1 模型， $\lambda = 4$ 人/小时， $\mu = \frac{1}{6}$ 人/分钟 = 10 人/小时，

$$\rho = \frac{\lambda}{\mu} = \frac{2}{5}。$$

$$(1) P_0 = 1 - \rho = \frac{3}{5};$$

$$(2) P_3 = \rho^3 (1 - \rho) = \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right) = 0.0384;$$

$$(3) 1 - P_0 = \frac{2}{5};$$

$$(4) L_s = \frac{\lambda}{\mu - \lambda} = \frac{4}{6} = \frac{2}{3} \text{ (人/小时)};$$

$$(5) W_s = \frac{1}{\lambda} L_s = \frac{1}{6} \text{ (小时/人)};$$

$$(6) L_q = L_s - \rho = \frac{2}{3} - \frac{2}{5} = \frac{4}{15} \text{ (人/小时)};$$

$$(7) W_q = W_s - \frac{1}{\mu} = \frac{1}{6} - \frac{1}{10} = \frac{1}{15} \text{ (小时/人)};$$

$$(8) P(W \geq \frac{1}{4}) = 1 - P(W < \frac{1}{4}) = 1 - F\left(\frac{1}{4}\right) = e^{-(10-4)\frac{1}{4}} = e^{-1.5} = 0.223。$$

二、系统容量有限的M/M/1模型 (M/M/1/N/∞)

1、与 (M/M/1/∞/∞) 的区别

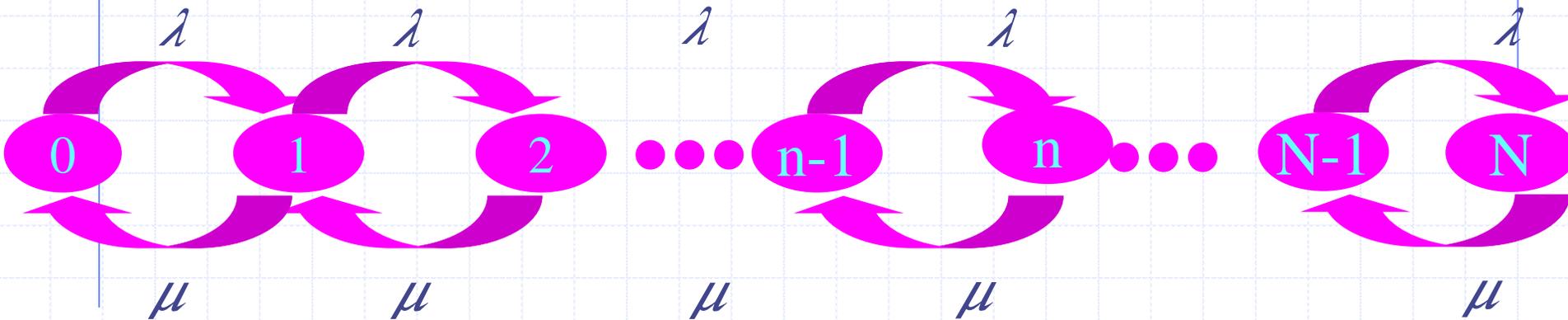
(1) 系统状态 $n = 0, 1, \dots, N$;

(2) 进入系统的速率
$$\begin{cases} \lambda, & \text{当 } n < N \\ 0, & \text{当 } n \geq N \end{cases}$$

故平均到达率 $\lambda_e = \lambda(1 - P_N) + 0P_N = \lambda(1 - P_N)$

注：由于系统稳态时应达到统计平衡，即进入速率应等于离去速率，故 $\lambda(1 - P_N) = \mu(1 - P_0)$ 。

2、状态概率



由此列出平衡方程:

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n, n = 1, \dots, N-1 \\ \lambda P_{N-1} = \mu P_N \end{cases}$$

先解得 $P_n = \rho^n P_0$,

再由 $\sum_{n=0}^N P_n = P_0 + \rho P_0 + \cdots + \rho^N P_0 = P_0 \frac{1 - \rho^{N+1}}{1 - \rho} = 1$ 可解得 P_0 ,

$$\text{故} \begin{cases} P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} \\ P_N = \rho^N P_0 \end{cases}$$

3. 系统运行指标

$$L_s = \sum_{n=0}^N nP_n = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}},$$

$$L_q = L_s - (1 - P_0),$$

$$W_s = \frac{L_s}{\lambda_e},$$

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

其中 $\lambda_e = \lambda(1 - P_N)$ 为有效到达率。

例3 某修理站只有1个修理工，且站内最多只能停放3台待修理的机器。设待修理的机器按泊松流到达，平均每小时到达1台；修理时间服从负指数分布，平均每1.25小时可修理1台。试求：（1）站内空闲率；（2）顾客损失率；（3）有效到达率；（4）站内平均队长；（5）机器为修理而需等待的平均时间。

解：此为 $(M / M / 1 / 4 / \infty)$ 排队系统， $\lambda = 1$ ， $\mu = 0.8$ ， $\rho = \frac{1}{0.8} = 1.25$ 。

$$(1) P_0 = \frac{1-\rho}{1-\rho^{4+1}} = \frac{1-1.25}{1-1.25^5} = 0.122;$$

$$(2) P_4 = \rho^4 P_0 = 1.25^4 \times 0.122 = 0.298;$$

$$(3) \lambda_e = \lambda(1-P_4) = 1 \times (1-0.298) = 0.702;$$

$$(4) L_s = \frac{\rho}{1-\rho} - \frac{(4+1)\rho^5}{1-\rho^5} = \frac{1.25}{1-1.25} - \frac{5 \times 1.25^5}{1-1.25^5} = 2.44(\text{台});$$

$$(5) W_q = \frac{L_q}{\lambda_e} = \frac{L_s - (1-P_0)}{\lambda_e} = \frac{2.44 - (1-0.122)}{0.702} = 2.23(\text{小时}).$$

例4: 为开办一个小型汽车冲洗站, 必须决定提供等待汽车使用的场地大小。设要冲洗的汽车到达服从泊松分布, 平均每4分钟1辆, 冲洗的时间服从负指数分布, 平均每3分钟洗1辆。试计算当所提供的场地仅能容纳 (a) 1辆; (b) 3辆; (c) 5辆 (包括正在被冲洗的1辆) 时, 由于等待场地不足而转向其它冲洗站的汽车的比例。

解： $\frac{1}{\lambda} = 4$, $\lambda = \frac{1}{4}$ 辆/分钟, $\mu = \frac{1}{3}$ 辆/分钟, $\rho = \frac{3}{4}$ 。

$$(a) \quad N = 1, \quad P_0 = \frac{1 - \rho}{1 - \rho^2} = \frac{1 - \frac{3}{4}}{1 - \frac{9}{16}} = \frac{4}{7}, \quad P_1 = \rho P_0 = \frac{3}{7} = 0.428$$

$$(b) \quad N = 3, \quad P_0 = \frac{1 - \rho}{1 - \rho^4} = \frac{1 - \frac{3}{4}}{1 - \frac{81}{256}} = \frac{64}{175}, \quad P_3 = \rho^3 P_0 = \frac{27}{175} = 0.154$$

$$(c) \quad N = 5, \quad P_0 = \frac{1 - \rho}{1 - \rho^6} = \frac{1 - \frac{3}{4}}{1 - 0.178} = 0.304, \\ P_5 = \rho^5 P_0 = 0.237 \times 0.304 = 0.072$$

三、顾客源有限的M/M/1模型 ($M/M/1/\infty/m$)

1、与 ($M/M/1/\infty/\infty$) 的区别

(1) 系统状态 $n = 0, 1, \dots, m$;

(2) λ 的含义

{	(∞/∞) : 表全体顾客的平均到达率, 即单位时间平均到 λ 人 (每人率不同);
	(∞/m) : 表每个顾客的平均到达率, 即单位时间每人到 λ 次 (每人率同 λ)。

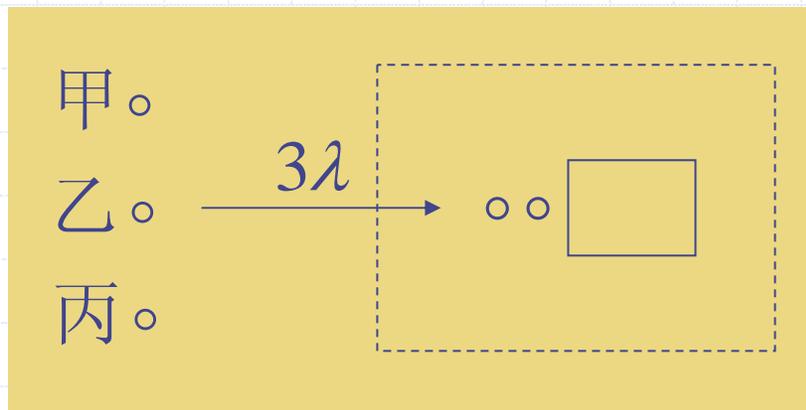
(3) 实际进入率

$$\left\{ \begin{array}{l} (\infty / \infty) : \lambda \text{ (与状态无关)} ; \\ (\infty / m) : \begin{cases} (m-n) \lambda \text{ (为什么?)}, n < m, \\ 0, n = m, \end{cases} \end{array} \right.$$

故平均到达率 $\lambda_e = E[(m-n)\lambda]$
 $= E(m\lambda) - E(n\lambda) = m\lambda - \lambda E(n)$
 $= \lambda(m - L_s).$

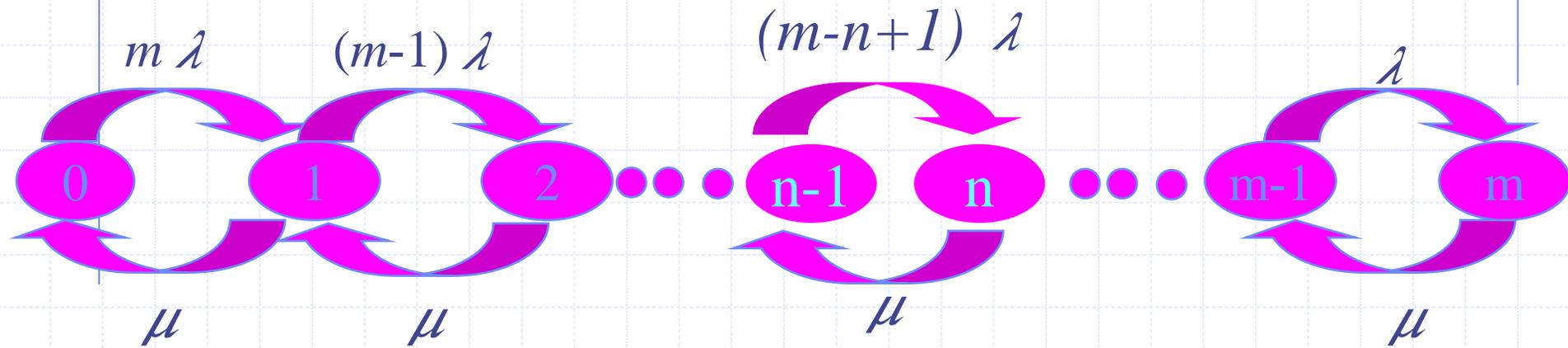
说明（进入率与状态有关）：

如 $m=5$ ， $n=3$ ，如下图所示



进入的或甲或乙或丙，故 3λ

2、状态概率



由此列出平衡方程：

$$\begin{cases} m\lambda P_0 = \mu P_1 \\ (m-n+1)\lambda P_{n-1} + \mu P_{n+1} = [(m-n)\lambda + \mu]P_n, n = 1, \dots, m-1 \\ \lambda P_m = \mu P_{m-1} \end{cases}$$

解得

$$\begin{cases} P_0 = \frac{1}{\sum_{i=0}^m \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu}\right)^i} \\ P_n = \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu}\right)^n P_0, n = 1, \dots, m \end{cases}$$

3、系统运行指标

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

$$\therefore \text{由里特公式, } L_s - L_q = \frac{\lambda_e}{\mu} = \frac{\lambda(m - L_s)}{\mu};$$

$$\text{另一方面, } L_s - L_q = 1 - P_0,$$

$$\text{解得 } L_s = m - \frac{\mu}{\lambda}(1 - P_0).$$

问题： L_s 的直观意义为何？

例5: 某车间有5台机器, 每台机器的连续运转时间服从负指数分布, 平均连续运转时间为15分钟。有1个修理工, 每次修理时间服从负指数分布, 平均每次需12分钟。

求 (1) 修理工空闲的概率; (2) 5台机器都出故障的概率; (3) 出故障机器的平均台数; (4) 等待修理机器的平均台数; (5) 每台机器的平均停工时间; (6) 每台机器的平均等待修理时间。

解：此为 $(M/M/1/\infty/5)$ 排队系统， $\lambda = \frac{1}{15}$ ， $\mu = \frac{1}{12}$ ， $\rho = \frac{12}{15} = 0.8$ 。

$$(1) P_0 = \left[\frac{5!}{5!} (0.8)^0 + \frac{5!}{4!} (0.8)^1 + \frac{5!}{3!} (0.8)^2 + \frac{5!}{2!} (0.8)^3 + \frac{5!}{1!} (0.8)^4 + \frac{5!}{0!} (0.8)^5 \right]^{-1} = 0.0073$$

$$(2) P_5 = \frac{5!}{0!} (0.8)^5 P_0 = 0.287;$$

$$(3) L_s = 5 - \frac{1}{0.8} (1 - 0.0073) = 3.76 \text{ (台)};$$

$$\begin{cases} P_0 = \frac{1}{\sum_{i=0}^m \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu}\right)^i} \\ P_n = \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu}\right)^n P_0, n = 1, \dots, m \end{cases}$$

$$(4) L_q = 3.76 - (1 - 0.0073) = 2.77(\text{台});$$

$$(5) W_s = \frac{5}{\frac{1}{12}(1 - 0.0073)} - 15 = 46(\text{分钟})$$

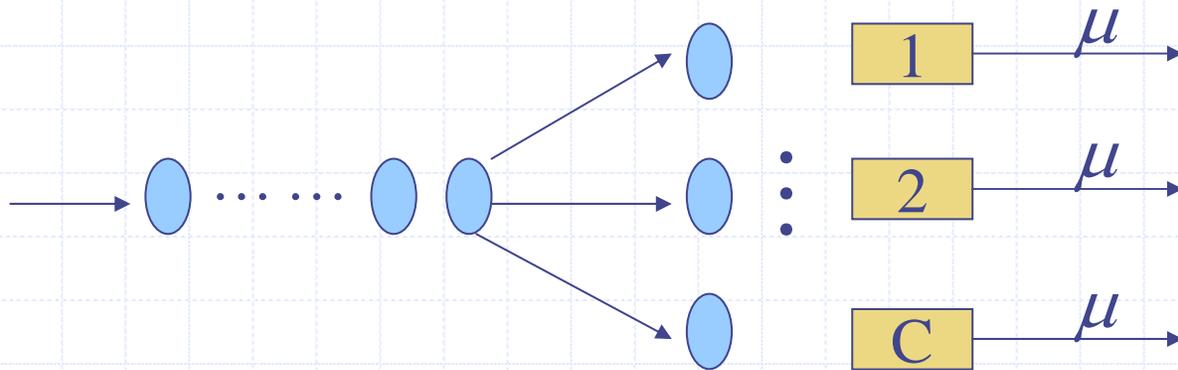
$$(6) W_q = 46 - 12 = 34(\text{分钟}).$$

由此可对该排队系统做何分析？

——机器停工时间过长，修理工几乎没有空闲时间应当提高服务率或增加修理工，或购置高效机器减少需修理率。

6.4 M/M/C排队模型

一、前提：单队、并列C台



模型仍可分为

$$\left\{ \begin{array}{l} (\infty / \infty / G) : \text{标准的} \\ (N / \infty / G) \\ (\infty / m / G) \end{array} \right.$$

我们仅讨论标准的M/M/C

二、(M/M/C): ($\infty / \infty / G$) 系统

1、与(M/M/1/ $\infty / \infty / G$)的区别

服务率与服务强度:

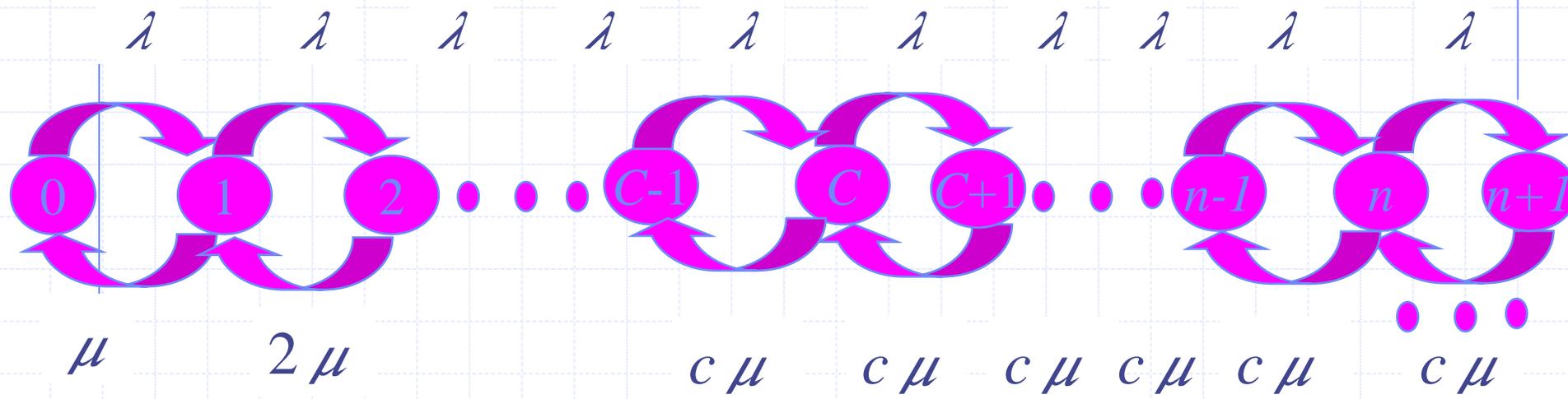
(M/M/1): 服务率与系统状态无关, 皆为 μ , 服务强度

$$\rho = \frac{\lambda}{\mu} < 1$$

(M/M/C): 服务率与系统状态有关, 为 $\begin{cases} n\mu, n < C \\ C\mu, n \geq C \end{cases}$

$\rho = \frac{\lambda}{C\mu}$, 表示每台单位时间内的平均负荷

2、状态概率



由此列出平衡方程：

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + (n+1)\mu P_{n+1} = (\lambda + n\mu)P_n, \quad n < C \\ \lambda P_{n-1} + C\mu P_{n+1} = (\lambda + C\mu)P_n, \quad n \geq C \end{cases}$$

先解得:

$$P_1 = \frac{\lambda}{\mu} P_0, \quad P_2 = \frac{\lambda}{2\mu} P_1, \quad P_3 = \frac{\lambda}{3\mu} P_2, \quad \dots$$

$$P_C = \frac{\lambda}{C\mu} P_{C-1} = \frac{1}{C!} \left(\frac{\lambda}{\mu}\right)^C P_0, \quad P_n = \frac{\lambda}{C\mu} P_{n-1} \quad (n > C)$$

再由 $\sum_{n=0}^{\infty} P_n = 1$ 解出 P_0 , 得:

$$\left\{ \begin{array}{l} P_0 = \left[\frac{(C\rho)^C}{C!} \frac{1}{1-\rho} + \sum_{n=0}^{C-1} \frac{(C\rho)^n}{n!} \right]^{-1}, \quad \rho = \frac{\lambda}{C\mu} < 1 \\ P_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n P_0, & n < C \\ \frac{1}{C! C^{n-C}} \left(\frac{\lambda}{\mu}\right)^n P_0, & n \geq C \end{cases} \end{array} \right.$$

3. 运行指标

$$L_s = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^C nP_n + \sum_{n=C+1}^{\infty} (n-C)P_n + \sum_{n=C+1}^{\infty} CP_n$$

$$= L_q + C - \sum_{n=0}^C (C-n)P_n = L_q + \frac{\lambda}{\mu}$$

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

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

注:

(1) 解释 $C - \sum_{n=0}^C (C-n) P_n = \frac{\lambda}{\mu}$ 的直观意义: 此式即 $\lambda = \mu [C - \sum_{n=0}^C (C-n) P_n]$,

其中 $\sum_{n=0}^C (C-n) P_n$ 为平均闲着的台数, $C - \sum_{n=0}^C (C-n) P_n$ 为平均忙着的台

数, μ 为每台的服务率, $\mu [C - \sum_{n=0}^C (C-n) P_n]$ 为系统的平均服务率,

由统计平衡, 它应等于平均到达率 λ 。

(2) $L_s - L_q = \frac{\lambda}{\mu}$, 这里与 $M/M/1$ 结果相同, 而与 C 无关, 从而

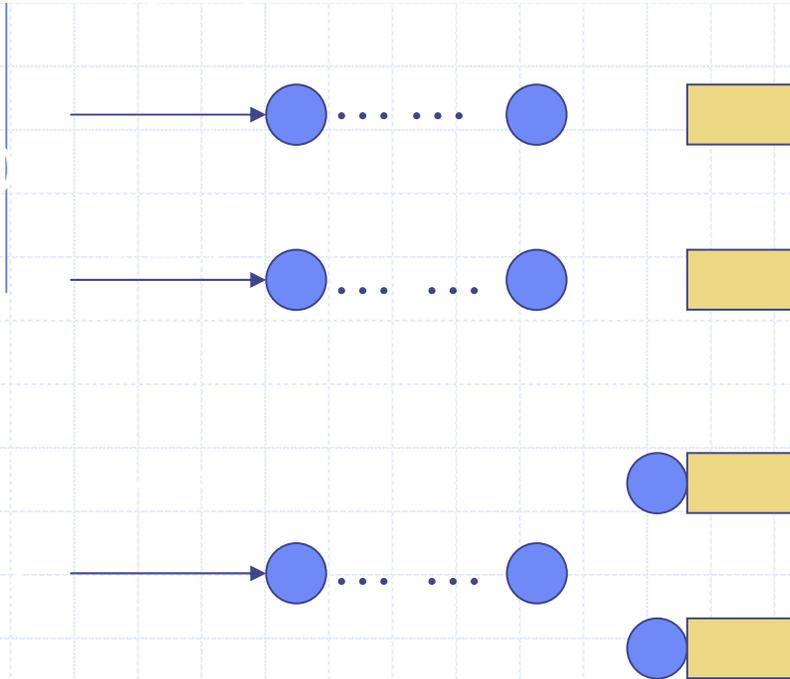
$$W_s - W_q = \frac{L_s}{\lambda} - \frac{L_q}{\lambda} = \frac{L_s - L_q + \frac{\lambda}{\mu}}{\lambda} = \frac{1}{\mu} \text{ 也与 } M/M/1 \text{ 相同}$$

(3) M/M/C指标有表可查:

$W_q \mu$ $\lambda / C \mu$	服务台数C			
	C=1	C=2	C=3	C=4
0.1	0.1111	0.0101	0.0014	0.0002
0.2	0.2500	0.0417	0.0103	0.0030
0.3	0.4286	0.0989	0.0333	0.0132
⋮	⋮	⋮	⋮	⋮

(4) 单队C台与C个单队单台系统比较

设 $C=2, \lambda=4, \mu=5$



$$\rho = \frac{\lambda}{\mu} = 0.8,$$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = 0.8$$

$$\rho = \frac{\lambda}{2\mu} = 0.8,$$

$$W_q = \frac{\lambda^{C-1} \rho P_0}{\mu^C C! (1 - \rho)^2} = 0.35$$

-----显然,单队C台效率高!

6.5 M/G/1排队模型

以上讨论了M/M/1和M/M/C系统，其前提均为泊松输入和负指数服务处理，这类系统的工具是生灭工程状态转移图。在实际中，有时到达仍为泊松过程，但服务时间并不服从负指数分布，即M/G/1系统这时不能用生灭过程处理，而主要依据布拉切克-钦辛公式(P-K公式)。

一、(M/G/1)

设：服务时间 ν 服从任意分布， $E(\nu)$ 与 $\sigma^2(\nu)$ 存在并已知，服务强度 $\rho = \lambda E(\nu) < 1$ 。其他条件同M / M / 1。
求：系统运行指标。

解：由布拉切克-钦辛 (P-K) 公式：

$$L_s = \rho + \frac{\rho^2 + \lambda^2 \sigma^2(\nu)}{2(1 - \rho)}$$

由里特公式：

$$W_s = \frac{L_s}{\lambda}, \quad W_q = W_s - E(\nu), \quad L_q = \lambda W_q$$

二、(M/D/1)系统 (定长服务时间)

若设: $E(v) = \frac{1}{\mu}$ 这时 $v \equiv E(v)$, $\sigma^2(v) = 0$

$$\text{则 } L_s = \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu(\mu - \lambda)}$$

$$\therefore L_q = L_s - \frac{\lambda}{\mu} = \frac{\lambda^2}{2\mu(\mu - \lambda)}, \quad W_q = \frac{L_q}{\lambda} = \frac{\lambda}{2\mu(\mu - \lambda)}$$

均为 M / M / 1 相应指标的一半。

可见，内部越有规律越省时间

三、 $(M/E_k/1)$ (k 阶爱尔郎服务时间)

由P-K公式:

$$L_s = \frac{\lambda}{\mu} + \frac{\lambda^2 \left(\frac{1}{\mu^2} + \frac{1}{k\mu^2} \right)}{2 \left(1 - \frac{\lambda}{\mu} \right)} = \rho + \frac{(k+1)\rho^2}{2k(1-\rho)},$$

由里特公式:

$$L_q = \frac{(k+1)\rho}{2k(1-\rho)}, \quad W_s = \frac{L_s}{\lambda}, \quad W_q = \frac{L_q}{\lambda}$$

可见, $k=1$ 时即 $(M/M/1)$, $k \rightarrow \infty$ 时即 $(M/D/1)$

注:对于到达与服务均为任意分布的情况,可采用随机模拟的方法求近似解。

6.6 排队系统最优化

排队系统优化

- 系统设计优化（静态）：如何设计一个系统（如何定 μ , C , 服务规则）使费用最经济（未必最小）。
- 系统控制优化（动态）：一个给定的系统如何运营使目标最优。

我们主要研究静态优化,
目标:费用(损失)最小。

待定指标 $\left\{ \begin{array}{l} \text{服务率 } \mu^* \\ \text{服务台数 } C^* \end{array} \right.$

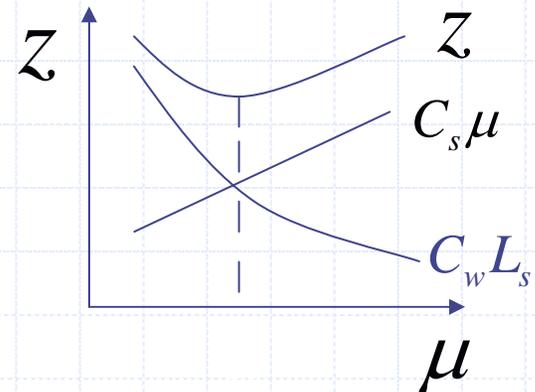
一、标准的M/M/1系统的最优服务率

设： C_s 为对每个顾客的单位时间服务费， C_w 为每个顾客在系统停留单位时间的损失费， z 为总费用。

目标： $Min z = C_s \mu + C_w L_s$, $L_s = \frac{\lambda}{\mu - \lambda}$, (单位时间费用最小)

解：令 $\frac{dz}{d\mu} = C_s - C_w \frac{\lambda}{(\mu - \lambda)^2} = 0$

得 $\mu^* = \lambda + \sqrt{\frac{C_w}{C_s} \lambda}$



注：若目标为服务利润最大，表达式为 $Max z = \mu(1 - P_0)G - C_s \mu$
其中 G 为单位时间对每位顾客服务的收入。

二.M/M/C系统的最佳服务台数

设： C_s 为每台单位时间服务成本， C_w 为每个顾客在系统停留单位时间的损失费， z 为总费用。

目标： $Min z = C_s C + C_w L_s$ ，（单位时间费用最小）

解： $z = z(C)$ 是 C 的离散函数，不能求导，采用边际分析法。

$\therefore C^*$ 是极小点

\therefore 应满足
$$\begin{cases} z(C^*) \leq z(C^* - 1) \\ z(C^*) \leq z(C^* + 1) \end{cases}$$

化简得： $L(C^*) - L(C^* + 1) \leq \frac{C_s}{C_w} \leq L(C^* - 1) - L(C^*)$