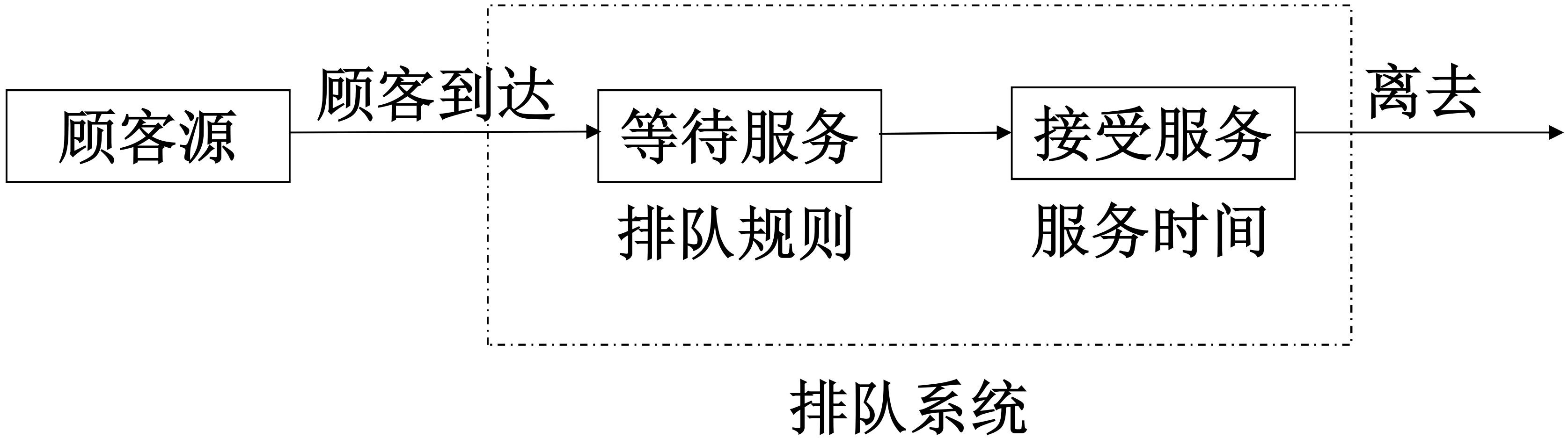
排队论

排队过程的描述：



现实社会中形形色色的排队系统

到达的顾客	要求服务的内容	服务机构
不能运转的机器	修理	修理技工
病人	诊断或东手术	医生
电话传呼	通话	交换台
文字稿	打字	打字员
到达机场上空的飞机	降落	跑道
驶入港口的货船	装卸货	装卸货码头

排队可能时人也可能是物，
为了一致，我们把要求得到服务的对象称为顾客
把提供服务的服务者成为服务机构

排队系统的基本组成部分： 输入过程、排队规则、
服务机构

输入过程：描述顾客的到达规律，包括顾客源的组成（有限、无限）、相继到达的顾客的间隔时间分布等。

排队规则：说明什么条件下允许（或顾客愿意）排队，排队的顾客按怎样的顺序接受服务。

损失制（服务台都被占用时，顾客自动离去）

等待制（服务台都被占用时，排队等候）

混合制（在一定条件下，等待；否则，离去）

服务顺序：先到先服务、后到先服务、按优先权、随机服务

服务机构：服务台数目、服务台之间的串并联结构、服务时间分布

排队模型的分类

记号　$X/Y/Z/A/B/C$

X：相继到达的顾客的间隔时间分布

M：负指数分布

G：一般分布

D：确定型（定长）分布

Y：服务时间分布

Z：服务台数目

A：系统容量限制（服务和等待的顾客数目）

B：顾客源数目

C：服务规则

注：X、Y、Z不能省；A、B、C缺省表示系统容量∞、顾客源∞、先到先服务

排队系统的数量指标

λ：顾客到达率，即单位 时间内到达的顾客平均 数

$\frac{1}{\lambda}$：到达时间平均间隔

μ：服务率，即单位时间 内平均服务的顾客数

$\frac{1}{\mu}$：平均服务时间，即平 均每位顾客的服务时间

L：平均队长，系统中顾客数的期望值

L_q：平均队列长，系统中等待服务的顾客数的期望值

W：平均逗留时间，顾客在系统中平均停留的时间

W_q：平均等待时间，顾客在系统中排队等待时间的期望值

ρ_e：平均服务强度，服务 台忙碌的概率

系统的状态：系统中的顾客数

计算排队系统数量指标的基础是表达系统的状态概率：

$P_n(t)$：时刻 t，系统中有 n 个顾客的概率。

$P_n(t)$ 为瞬态概率，不容易获得，即使获得也很难利用，实用中通常关心排队系统到达稳定或统计平衡状态时的各种数量指标，需要利用

稳态概率 $P_n = \lim_{t\to\infty} P_n(t)$

几个常用概率分布 设随机变量 T

定长分布 D： 概率函数 $\Pr(T=a)=1$

$\mathrm{E}(T)=a$ 。

负指数分布 M：

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

分布函数 $F(t)=\Pr(T\leq t)=\begin{cases}1-e^{-\lambda t}, & t\geq 0\\ 0, & t<0\end{cases}$

$\mathrm{E}(T)=\dfrac{1}{\lambda}$, $\mathrm{D}(T)=\dfrac{1}{\lambda^2}$

$\Pr(T>t)=e^{-\lambda t} \quad (t\geq 0)$

负指数分布具有无记忆性

$$\Pr(T > t + s \mid T > s) = \frac{\Pr(T > t + s,\ T > s)}{\Pr(T > s)}$$

$$= \frac{\Pr(T > t + s)}{\Pr(T > s)}$$

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

最简单流（普阿松流Possion）：

若用 $N(t)$ 表示在时间区间 $[0, t)$ 内到达的顾客数，

对于每个 t， $N(t)$ 是一个随机变量，

随机变量族 $\{N(t) \mid t \geq 0\}$ 称为一个随机过程。

若 $N(t)$ 满足以下三条件，则称为最简单流：

i) 平稳性：在 $[t, t+\Delta t)$ 内达到的顾客数与 t 无关，只与 Δt 有关。

设 $P_n(t) = \Pr(N(t) = n)$，$P_0(0) = 1$，$\sum_{n=0}^{\infty} P_n(t) = 1$，

则 $P_n(\Delta t)$ 表示在 $[0, \Delta t)$ 或 $[t, t+\Delta t)$ 内达到 n 个顾客的概率。

ii) 无后效性：在不重叠的时间区间内到达的顾客数相互独立，即 $[t, t+\Delta t)$ 内到达的顾客数与 t 之前到达的无关。

iii) 普通性：对于充分小的 Δt，在 $[t, t+\Delta t)$ 内到达两个或以上顾客的概率为 $o(\Delta t)$，即

$$\sum_{n=2}^{\infty} P_n(\Delta t) = o(\Delta t).$$

设 λ 表示单位时间内有一个 顾客到达的概率
（或单位时间内到达的 顾客平均数， $\lambda = P_1'(0)$
$P_1(\Delta t) = P_1(0) + P_1'(0)\Delta t + o(\Delta t) = P_1'(0)\Delta t + o(\Delta t)$ ）

可以推出

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

称 $N(t)$ 服从普阿松分布。

可证： $\mathrm{E}(N(t)) = \lambda t$ ， $\mathrm{D}(N(t)) = \lambda t$

最简单流的到达时间间隔服从负指数分布；
到达时间间隔相互独立且服从同一负指数分布的顾客流是最简单流。

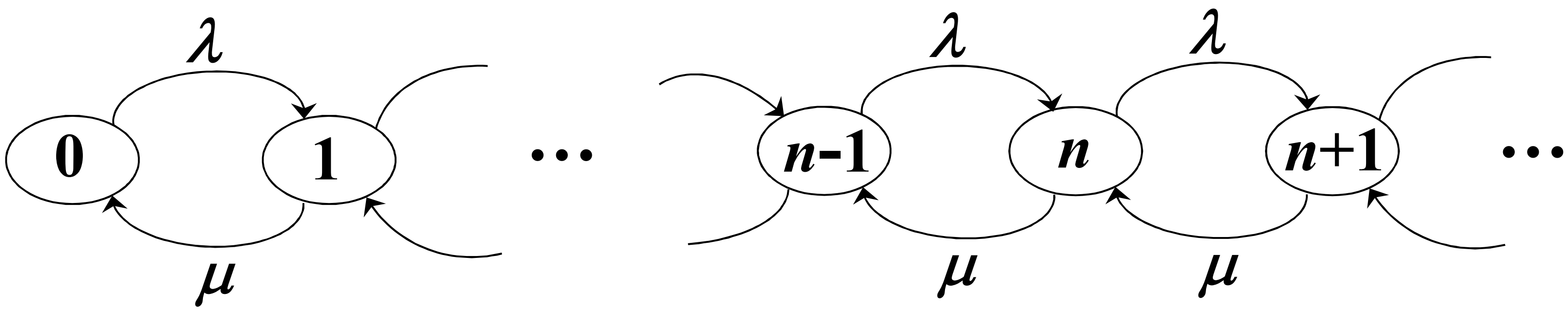
1. $M/M/1$ 模型

顾客流是参数 λ 的最简单流　　单位时间来的顾客数

服务时间服从参数 μ 的负指数分布 单位时间服务的顾客数

P_n：稳定状态下系统中有 n 个顾客的概率

各状态间的转移关系:



$$\lambda P_0 = \mu P_1$$

$$\lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n \,, \quad n \geq 1$$

$$P_1 = \frac{\lambda}{\mu} P_0$$

$$P_{n+1} - P_n = \frac{\lambda}{\mu}(P_n - P_{n-1}) = \cdots = \left(\frac{\lambda}{\mu}\right)^n (P_1 - P_0)$$

记 $\rho = \lambda / \mu$，ρ 表示单位时间内到达的 顾客数与服务完的顾客数之比，称为服务 强度，刻画了服务台的 忙碌程度。

若 $\rho \geq 1$，即单位时间内到达的 顾客数不少于服务完的 顾客数，系统中的顾客将 无限增多，系统不可能 达到稳定。

$\rho < 1$ 时，

$$P_{n+1} - P_0 = P_{n+1} - P_n + P_n - P_{n-1} + \cdots + P_1 - P_0$$
$$= (\rho^n + \rho^{n-1} + \cdots + \rho^0)(P_1 - P_0)$$

$$P_{n+1}=P_0+\frac{1-\rho^{n+1}}{1-\rho}(P_1-P_0)=P_0-(1-\rho^{n+1})P_0=\rho^{n+1}P_0$$

由 $\sum\limits_{n=0}^{\infty}P_n=\sum\limits_{n=0}^{\infty}\rho^nP_0=\frac{P_0}{1-\rho}=1$，得

$$P_0=1-\rho\,,\quad P_n=(1-\rho)\rho^n$$

$P_0=1-\rho$ 表示系统中无顾客，即 服务台空闲的概率，而 ρ 恰好是服务台忙碌的概 率。

平均队长 $L=\sum_{n=0}^{\infty}nP_n=\sum_{n=0}^{\infty}n(1-\rho)\rho^n$

$$=(1-\rho)\rho\sum_{n=0}^{\infty}n\rho^{n-1}$$

$$L=(1-\rho)\rho\left(\sum_{n=1}^{\infty}\rho^{n}\right)'=(1-\rho)\rho\left(\frac{\rho}{1-\rho}\right)'$$

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

平均队列长 $L_q=\sum_{n=1}^{\infty}(n-1)P_n$

$$=\sum_{n=1}^{\infty}nP_n-\sum_{n=1}^{\infty}P_n$$

$$=L-(1-P_0)$$

$$=L-\rho=\frac{\rho^{2}}{1-\rho}$$

平均等待时间 W_q ：

当一个顾客进入系统时，若系统中已有 n 个顾客，则其平均等待时间为 $n \cdot \frac{1}{\mu}$，而系统中有 n 个顾客的概率为 P_n，所以

$$W_q = \sum_{n=0}^{\infty} \frac{n}{\mu} P_n = \frac{1}{\mu} \sum_{n=0}^{\infty} n P_n = \frac{L}{\mu} = \frac{\rho}{\mu - \lambda}$$

平均逗留时间 $W = W_q + \frac{1}{\mu} = \frac{\rho}{\mu - \lambda} + \frac{1}{\mu} = \frac{1}{\mu - \lambda}$

相关计算公式小结

λ：单位时间到达的人数　　记 $\rho = \lambda / \mu$，称为服务强度

μ：单位时间到达的人数

$P_0 = 1 - \rho$ 表示系统中无顾客，即 服务台空闲的概率

$$\boldsymbol{P_0 = 1 - \rho\ ,\quad P_n = (1 - \rho)\rho^n}$$

平均队长 $L = \dfrac{\rho}{1-\rho} = \dfrac{\lambda}{\mu - \lambda}$

平均队列长　$L_q = L - \rho = \dfrac{\rho^2}{1-\rho}$

平均等待时间 W_q：$= \dfrac{\rho}{\mu - \lambda}$

平均逗留时间　$W = W_q + \dfrac{1}{\mu} = \dfrac{1}{\mu - \lambda}$

相互关系（Little 公式）

$$L = \lambda W, \qquad L_q = \lambda W_q$$

$$L = L_q + \rho, \quad W = W_q + \frac{1}{\mu}$$

$L = \lambda W$ 之解释：

平均每个顾客逗留 W 个单位时间，每单位时 间有 λ 个顾客进入系统，因此 系统中平均有 λW 个顾客。

例 1. 某售票处有一个窗口，平均每分钟可服务两人，若平均每分钟有一位购票者到达，试研究售票窗口的排队情况。

解． $\lambda = 1$ 人/分，$\mu = 2$ 人/分， $\rho = \frac{\lambda}{\mu} = \frac{1}{2}$

$$L = \frac{\rho}{1-\rho} = 1 \text{ 人}, \quad L_q = \frac{\rho^2}{1-\rho} = \frac{1}{2} \text{ 人}$$

$$W = \frac{1}{\mu-\lambda} = 1 \text{ 分}, \quad W_q = \frac{\rho}{\mu-\lambda} = \frac{1}{2} \text{ 分}$$

不需等待的概率 $P_0 = 1-\rho = \frac{1}{2}$

等待顾客超过 2 人的概率

$$\sum_{n=3}^{\infty} P_n = \sum_{n=3}^{\infty} \rho^n P_0 = \frac{\rho^3}{1-\rho} P_0 = \frac{1}{8}$$

例 2. 某私人诊所平均每隔 20 分钟有一位病人前来就诊，医生平均 15 分钟可诊断一位病人，若希望有足够的座位给病人坐，即病人到达时没有座位的概率不超过 0.01，问至少应准备多少个座位？

解. 设有 k 个座位（包括医生旁的座位），则诊所中不超过 k 个病人的概率应不小于 0.99，即

$$\sum_{n=0}^{k} P_n = \sum_{n=0}^{k} \rho^n (1-\rho) = 1-\rho^{k+1} \geq 0.99$$

$$\rho^{k+1} \leq 0.01\,, \quad (k+1)\lg\rho \leq \lg 0.01 = -2$$

$$\rho = \frac{15}{20} = 0.75\,, \quad \lg 0.75 = -0.1249$$

$\therefore \quad k \geq \dfrac{2}{0.1249} - 1 \approx 15$，即至少应准备 15 个座位。

例 3. 某铁路公司为经常油漆车厢，考虑了两个方案。方案一是设置一个手工油漆车间，年总开支 20 万元，每节车厢的油漆时间为 6 小时。方案二是建立一个喷漆车间，年总开支 45 万元，每节车厢的油漆时间为 3 小时。设油漆车间常年开工（每年开工 365×24 = 8760 小时），平均每隔 8 小时有一节车厢要油漆，每节车厢闲置的损失为每小时 15 元，问应采用哪个方案？

解． $\lambda = \frac{1}{8}$ 节 / 小时，

$$\mu_1 = \frac{1}{6} \text{节 / 小时}, \quad \mu_2 = \frac{1}{3} \text{节 / 小时}$$

$$W_1 = \frac{1}{\mu_1 - \lambda} = \frac{1}{\frac{1}{6} - \frac{1}{8}} = 24 \text{ 小时}$$

$$W_2 = \frac{1}{\mu_2 - \lambda} = \frac{1}{\frac{1}{3} - \frac{1}{8}} = \frac{24}{5} \text{ 小时}$$

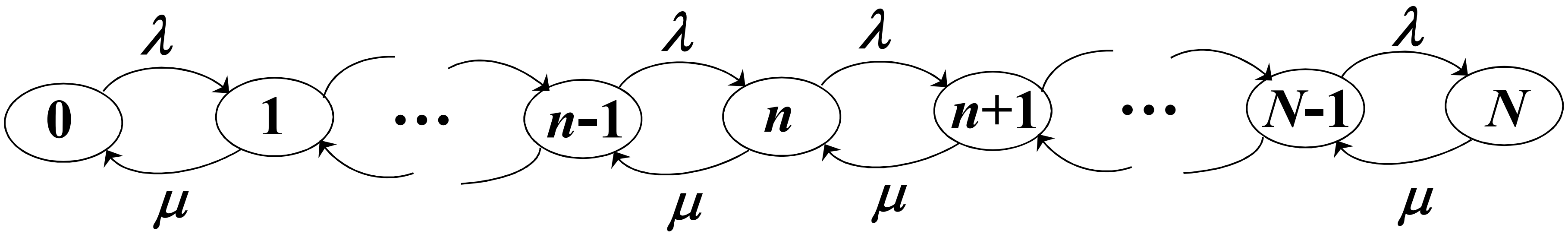
每年有 $8760 \div 8 = 1095$ 节车厢要油漆

方案一费用　$1095 \times 24 \times 15 + 200000 = 59.42$ 万元

方案二费用　$1095 \times \frac{24}{5} \times 15 + 450000 = 52.884$ 万元

2. $M/M/1/N$ 模型

系统容量为 N



$$\lambda P_0 = \mu P_1$$

$$\lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n \ , \quad n = 1, \cdots, N-1$$

$$\lambda P_{N-1} = \mu P_N$$

$$P_{n+1} - P_n = \frac{\lambda}{\mu}(P_n - P_{n-1}) = \cdots = \left(\frac{\lambda}{\mu}\right)^n (P_1 - P_0)$$

令 $\rho = \lambda / \mu$，则

$$P_{n+1} = P_0 + (\rho^n + \rho^{n-1} + \cdots + \rho^0)(P_1 - P_0)$$

$\rho \neq 1$ 时， $P_{n+1} = P_0 + \dfrac{1-\rho^{n+1}}{1-\rho}(\rho - 1)P_0 = \rho^{n+1} P_0$

由 $\sum_{n=0}^{N} P_n = \sum_{n=0}^{N} \rho^n P_0 = \dfrac{1-\rho^{N+1}}{1-\rho} P_0 = 1$，得

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}}$$

$$P_n = \rho^n P_0 = \frac{1-\rho}{1-\rho^{N+1}} \rho^n, \quad n = 0, 1, \cdots, N$$

$\rho = 1$ 时， $P_0 = P_1 = \cdots = P_N = \dfrac{1}{N+1}$

P_N 表示损失概率，

单位时间内平均损失的 顾客数为 λP_N ，

真正进入系统的顾客数 $\lambda_e = \lambda - \lambda P_N = (1 - P_N)\lambda$

平均服务强度 $\rho_e = \dfrac{\lambda_e}{\mu} = \dfrac{\lambda}{\mu}(1 - P_N) = 1 - P_0$

注：对于 $M/M/1/N$，平均服务强度不是 $\rho = \dfrac{\lambda}{\mu}$。

平均队长 $L = \sum_{n=0}^{N} nP_n$

$\rho = 1$ 时， $L = \dfrac{1}{N+1}\sum_{n=0}^{N} n = \dfrac{N}{2}$

$\rho \neq 1$ 时， $L = \sum_{n=0}^{N} n\rho^n \dfrac{1-\rho}{1-\rho^{N+1}}$

$$\therefore \quad L=\frac{(1-\rho)\rho}{1-\rho^{N+1}}\sum_{n=1}^{N} n\rho^{n-1}=\frac{(1-\rho)\rho}{1-\rho^{N+1}}\left(\sum_{n=1}^{N}\rho^{n}\right)'$$

$$=\frac{(1-\rho)\rho}{1-\rho^{N+1}}\left(\frac{(1-\rho^{N})\rho}{1-\rho}\right)'$$

$$=\frac{\rho}{1-\rho}-\frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}$$

平均队列长 $L_q=\sum_{n=1}^{N}(n-1)P_n=L-(1-P_0)$

$\rho=1$时， $L_q=\dfrac{N}{2}-1+\dfrac{1}{N+1}=\dfrac{N(N-1)}{2(N+1)}$

$\rho\neq 1$时， $L_q=\dfrac{\rho}{1-\rho}-\dfrac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}+\dfrac{1-\rho}{1-\rho^{N+1}}-1$

$$\therefore \quad L_q = \frac{\rho}{1-\rho} - \rho - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} + \frac{(1-\rho)\rho^{N+1}}{1-\rho^{N+1}}$$

$$= \frac{\rho^2}{1-\rho} - \frac{(N+\rho)\rho^{N+1}}{1-\rho^{N+1}}$$

Little 公式：

$$L = \lambda_e W \text{ ，} \qquad L_q = \lambda_e W_q$$

$$L = L_q + \rho_e \text{ ，} \quad W = W_q + \frac{1}{\mu}$$

W、W_q 由 Little 公式可计算。

例 4．给定一个 $M/M/1/N$ 系统，$\lambda = 10$ 人/小时，$\mu = 30$ 人/小时，$N = 2$。管理者想改进服务机 构，方案甲是增加等待空间到 $N = 3$，方案乙是将平均服务率提高到 $\mu = 40$ 人/小时，问哪个方案获益 更大?

解. 服务机构的收益与单位时间内实际进入系统的顾客数成正比。

$$\lambda_e = (1 - P_N)\lambda = \lambda \frac{1-\rho^N}{1-\rho^{N+1}}$$

方案甲： $N = 3$， $\lambda = 10$， $\mu = 30$， $\rho = \frac{1}{3}$

$$\lambda_e = 10 \cdot \frac{1-(\frac{1}{3})^3}{1-(\frac{1}{3})^4} = 9.75$$

方案乙： $N=2$， $\lambda=10$， $\mu=40$， $\rho=\frac{1}{4}$

$$\lambda_e=10\cdot\frac{1-(\frac{1}{4})^2}{1-(\frac{1}{4})^3}=9.52$$

方案甲获益大。

若 λ 增加到 30 人 / 小时，则

方案甲： $N=3$， $\lambda=\mu=30$， $\rho=1$

$$\lambda_e=(1-P_N)\lambda=\frac{\lambda N}{N+1}=\frac{30\times 3}{4}=22.5$$

方案乙： $N=2$， $\lambda=30$， $\mu=40$， $\rho=\frac{3}{4}$

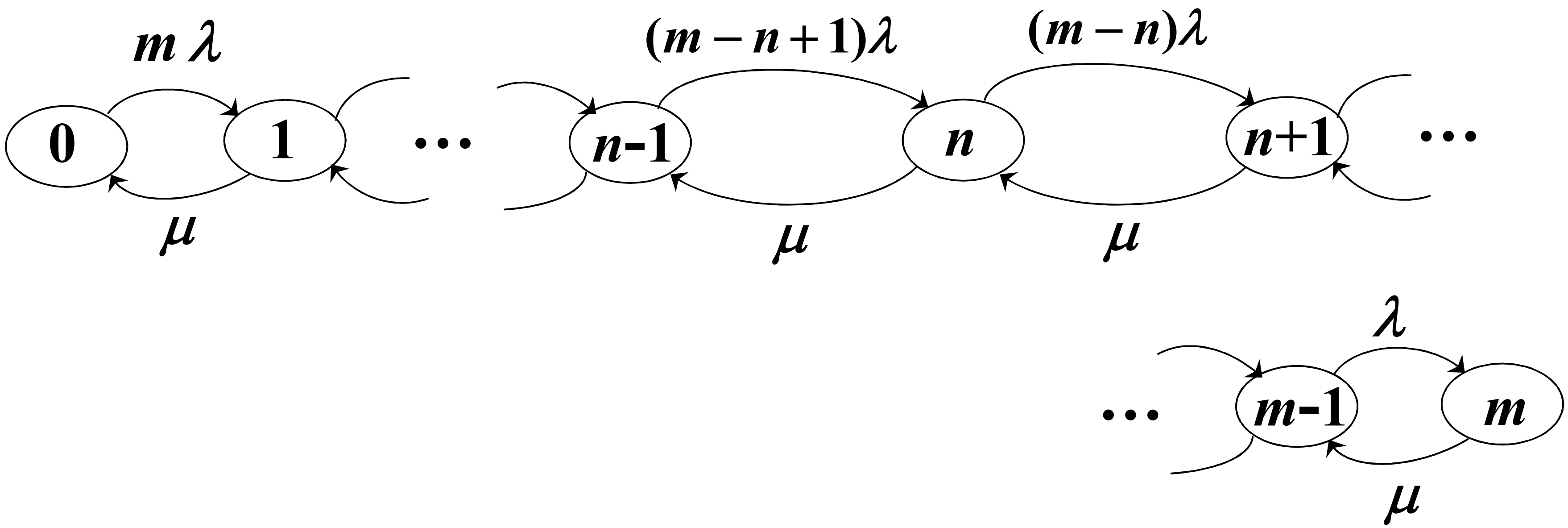
$$\lambda_e=30\cdot\frac{1-(\frac{3}{4})^2}{1-(\frac{3}{4})^3}=22.7$$

3. $M/M/1/\infty/m$ 模型

顾客源有限，与 $M/M/1/m/m$ 等价。

设λ 是单个顾客的到达率 (单位时间内要求服务的 次数).

若系统中已有 n 个顾客，则不在系统中 的顾客数为 $m-n$，对系统的有效到达率为 $\lambda_e=(m-n)\lambda=(m-L)\lambda$.



$$m\lambda P_0 = \mu P_1$$

$$(m-n+1)\lambda P_{n-1} + \mu P_{n+1} = ((m-n)\lambda + \mu)P_n$$

$$\lambda P_{m-1} = \mu P_m$$

可解出 $P_0 = \left(\sum_{i=0}^{m} \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu} \right)^i \right)^{-1}$

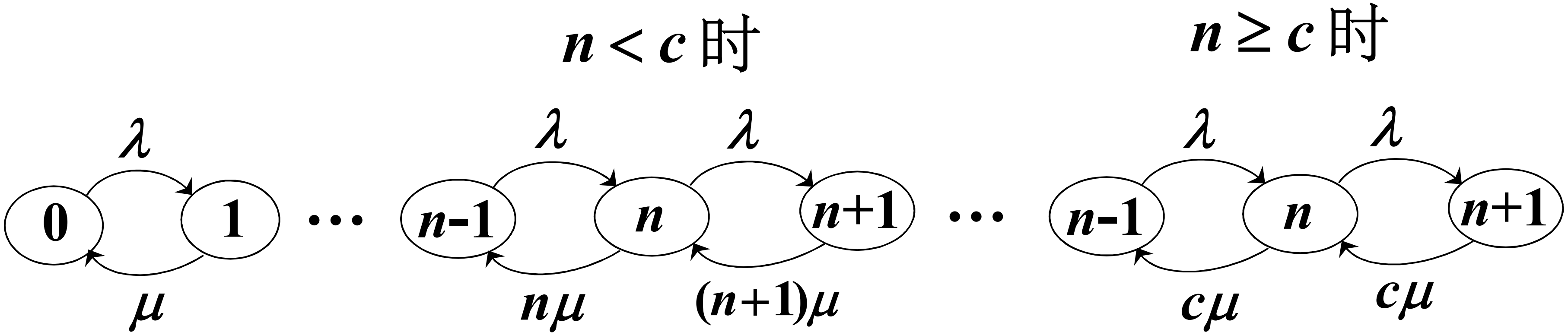
$$P_n = \frac{m!}{(m-n)!} \left(\frac{\lambda}{\mu} \right)^n P_0$$

$$L = \sum_{n=0}^{m} nP_n = m - \frac{\mu}{\lambda}(1-P_0), \quad L_q = L - (1-P_0)$$

Little 公式: $L = \lambda_e W$， $L_q = \lambda_e W_q$

4. $M/M/c$ 模型

假设：各服务台的工作 相互独立 (不搞协作) 且平均服务率均为 μ，因此整个系统的平均 服务率为 $c\mu$ 。



$$\lambda P_0 = \mu P_1$$

$$\lambda P_{n-1} + (n+1)\mu P_{n+1} = (\lambda + n\mu)P_n\,, \quad n = 1, \cdots, c-1$$

$$\lambda P_{n-1} + c\mu P_{n+1} = (\lambda + c\mu)P_n\,, \quad n \geq c$$

5. $M/G/1$ 模型

各顾客的服务时间独立 同分布，均值 $\frac{1}{\mu}$，方差 δ^2 。

服务强度 $\rho = \frac{\lambda}{\mu}$ 。

$\rho < 1$ 时，系统能达到稳定， 且有 $P_0 = 1 - \rho$ 。

Little 公式： $L = \lambda W$ ， $L_q = \lambda W_q$

$$L = L_q + \rho, \quad W = W_q + \frac{1}{\mu}$$

Pollaczek-Khintchine（P-K）公式：

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

利用上述公式，可求出 L、W、W_q 。

6. $M/D/1$ 模型

各顾客的服务时间是确 定的常数，方差 $\delta = 0$ 。

Pollaczek-Khintchine（P-K）公式：

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

7. 排队系统的优化问题

(1) $M/M/1$ 系统的最优服务率

取目标函数 z 为单位时间的服务成本和顾客在系统的逗留费用之和

$$z = c_s\mu + c_w L = c_s\mu + \frac{c_w\lambda}{\mu-\lambda}$$

其中， c_s 为 $\mu=1$ 时单位时间的服务成本 ， c_w 为每个顾客单位逗留时间的费 用。

由 $\frac{dz}{d\mu} = c_s - \frac{c_w\lambda}{(\mu-\lambda)^2} = 0$，得

$$\mu^* = \lambda + \sqrt{\frac{c_w\lambda}{c_s}}$$ （为保证 $\rho<1$，根号前只能取 $+$）

(2) $M/M/1/N$ 系统的最优服务率

P_N：损失概率， $1-P_N$：能接受服务的概率

$\lambda(1-P_N)$：单位时间内实际进入系统的顾客数

设每服务一人收入 g 元，单位时间收入 $\lambda(1-P_N)g$ 元，

利润 $z=\lambda(1-P_N)g-c_s\mu$，（不考虑顾客逗留费用）

$$=\lambda g\frac{1-\rho^N}{1-\rho^{N+1}}-c_s\mu$$

$$=\lambda g\mu\frac{\mu^N-\lambda^N}{\mu^{N+1}-\lambda^{N+1}}-c_s\mu$$

由 $\frac{dz}{d\mu}=0$，得 μ^*满足 $\frac{N-(N+1)\rho+\rho^{N+1}}{(1-\rho^{N+1})^2}\rho^{N+1}=\frac{c_s}{g}$。

(3) $M/M/c$ 系统的最优服务台数目

取目标函数 $z = c_s' c + c_w L$

其中 c_s' 为每个服务台单位时间 的成本，c_w 为每个顾客单位逗留时间的费用。

c 只取整数，根据边际分析法，c^* 应满足

$$z(c^*) \le z(c^*+1)\,,\quad z(c^*) \le z(c^*-1)$$

即
$$c_s' c^* + c_w L(c^*) \le c_s'(c^*+1) + c_w L(c^*+1)$$
$$c_s' c^* + c_w L(c^*) \le c_s'(c^*-1) + c_w L(c^*-1)$$

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

依次求 $c = 1, 2, \cdots$ 时 L 的值和相邻两 L 的差，确定 $\dfrac{c_s'}{c_w}$ 落在哪个不等式里就可 定出 c^*。