排队论

第十章

单服务台负指数分布排队系统的分析

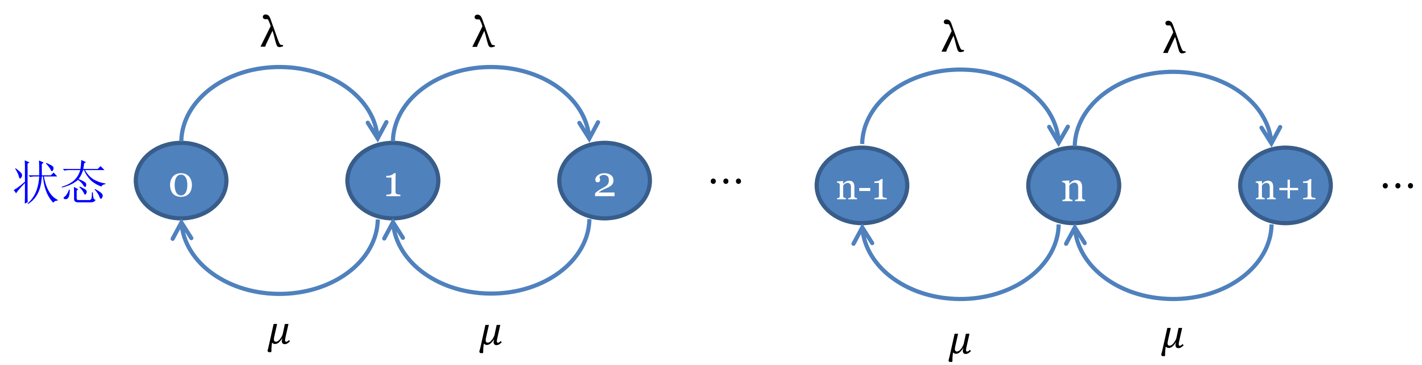
第4节

$M/M/1/\infty/\infty$模型

- 模型概述
  - 输入过程
    - 顾客源为无限，顾客单个到来，相互独立，一定时间内到达的顾客数服从泊松分布，到达过程是平稳的。
  - 排队规则
    - 单队，队长无限制，先到先服务。
  - 服务机构
    - 单服务台，各顾客的服务时间相互独立，服从相同的负指数分布。
  - 到达间隔时间和服务时间是相互独立的

$M/M/1/\infty/\infty$模型

- 模型求解
  - 系统的状态转移率图



  - 系统各状态的平衡方程

$$\begin{cases} \lambda P_{n-1} + \mu P_{n+1} = (\mu + \lambda) P_n \quad (n = 1,2, \ldots) \\ \mu P_1 = \lambda P_0 \end{cases}$$

$M/M/1/\infty/\infty$模 型

- 模型求解
  - 状态平衡方程的解：稳态概率

$$P_1=\left(\frac{\lambda}{\mu}\right)P_0,\ P_2=\left(\frac{\lambda}{\mu}\right)^2P_0,\ldots,\ P_n=\left(\frac{\lambda}{\mu}\right)^nP_0$$

设$\rho=\frac{\lambda}{\mu}<1$ (否则队列将排至无限远)，则由概率的性质可知

$$\sum_{n=0}^{\infty}P_n=1$$

从而有

$$P_0\sum_{n=0}^{\infty}\rho^n=P_0\frac{1}{1-\rho}=1$$

得到系统状态为n的概率(稳态概率)

$$\begin{cases}P_0=1-\rho \\ P_n=(1-\rho)\rho^n, \qquad (n=1,2,\ldots)\end{cases}$$

$M/M/1/\infty/\infty$模 型

- 模型求解
  - ρ的实际意义

$$\begin{cases} P_0 = 1-\rho \\ P_n = (1-\rho)\rho^n, \qquad (n = 1,2,\dots) \end{cases}$$

- 上式中，ρ的表达式不同，可以有不同的解释：
  - $\rho = \frac{\lambda}{\mu}$ ： 平均到达率与平均服务率之比；
  - $\rho = \frac{1/\mu}{1/\lambda}$：一个顾客的服务时间与到达间隔时间之比，称为服务强度（traffic intensity），或称为话务强度，这是早期排队论创立者爱尔朗等人在研究电话理论时用的术语，一直沿用至今；
  - $\rho = 1 - P_0$：系统中至少有一个顾客的概率，即服务机构的利用率，反映了服务机构的繁忙程度。

$M/M/1/\infty/\infty$模型

- 模型求解
  - 系统的运行指标
    - 在系统中的平均顾客数 (队长期望值)

$$L_s=\sum_{n=0}^{\infty} nP_n=\sum_{n=1}^{\infty} n(1-\rho)\rho^n$$
$$=(\rho+2\rho^2+3\rho^3+\cdots)-(\rho^2+2\rho^3+3\rho^4+\cdots)$$
$$=\rho+\rho^2+\rho^3+\cdots=\frac{\rho}{1-\rho},\quad 0<\rho<1$$

或

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

    - 在队列中等待的平均顾客数 (队列长期望值)

$$L_q=\sum_{n=1}^{\infty}(n-1)P_n=\sum_{n=1}^{\infty} nP_n-\sum_{n=1}^{\infty} P_n$$
$$=L_s-\rho=\frac{\rho^2}{1-\rho}=\frac{\rho\lambda}{\mu-\lambda},\quad 0<\rho<1$$

Haiyan Lu
Jiangnan University

$M/M/1/\infty/\infty$模型

- 模型求解
  - 系统的运行指标
    - 在系统中顾客逗留时间的期望值

$$W_s = E[W] = \frac{1}{\mu-\lambda}$$

    - 在队列中顾客等待时间的期望值

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

    - **备注**：
      - 可以证明：

在$M/M/1/\infty/\infty$情形下，顾客在系统中**逗留的时间**W服从参数为$\mu-\lambda$的负指数分布，其概率密度函数为$f_W(t) = (\mu-\lambda)e^{-(\mu-\lambda)t}$，从而有$E[W] = \frac{1}{\mu-\lambda}$.

      - 参见《运筹学》，第四版，365页。

Haiyan Lu
Jiangnan University

$M/M/1/\infty/\infty$模 型

- 模型求解
  - 系统的运行指标
    - 上述各式归纳如下

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

$$(3)\ W_s = \frac{1}{\mu-\lambda} \qquad (4)\ W_q = \frac{\rho}{\mu-\lambda}$$

    - 其相互关系如下（ **Little公式** ）

$$(1)\ L_s = \lambda W_s \qquad (2)\ L_q = \lambda W_q$$

$$(3)\ W_s = W_q + \frac{1}{\mu} \qquad (4)\ L_s = L_q + \frac{\lambda}{\mu}$$

    - 备注：
      - 下列公式称为**Little公式** (Little's queuing formula)。只要稳态概率分布存在，则该组公式对**任何**排队系统都适用：

$$L_s = \lambda W_s, \quad L_q = \lambda W_q$$

$M/M/1/\infty/\infty$模型

- 模型求解
  - 系统的运行指标
    - 在系统中顾客**逗留时间**W小于等于t的概率

$$P(W \le t) = \int_0^t (\mu - \lambda)e^{-(\mu-\lambda)t}dt = 1 - e^{-(\mu-\lambda)t}$$

    - 在系统中顾客**逗留时间**W大于t的概率

$$P(W > t) = 1 - P(W \le t) = e^{-(\mu-\lambda)t}$$

    - 在队列中顾客**等待时间**Q小于等于t的概率

$$P(Q \le t) = 1 - \rho e^{-(\mu-\lambda)t}, \qquad t \ge 0, \rho < 1$$

    - 在队列中顾客**等待时间**Q大于t的概率

$$P(Q > t) = \rho e^{-(\mu-\lambda)t}, \qquad t \ge 0, \rho < 1$$

$M/M/1/\infty/\infty$模型

- 模型求解
  - 系统的运行指标
    - **备注**：
      - 可以证明：在$M/M/1/\infty/\infty$情形下，顾客在队列中**排队等待的时间**Q的概率密度函数为

$$f_Q(t)=\begin{cases}1-\rho, & t=0\\ \rho(\mu-\lambda)e^{-(\mu-\lambda)t}, & t>0\end{cases},\quad 且有E[Q]=\frac{\rho}{\mu-\lambda}$$

      - 不同服务规则（先到先服务、后到先服务、随机服务）的不同点主要反映在等待时间的分布函数的不同，而一些期望值是相同的。上述讨论的各种指标都是期望值，因此其计算公式对三种服务规则都适用 (但对于具有优先权的服务规则不适用).

Haiyan Lu 5/27/2020
Jiangnan University

$M/M/1/\infty/\infty$模 型

- 例1 某医院手术室根据病人来诊和完成手术时间的记录，任意抽查了100个工作小时，每小时来就诊的病人数n的出现次数如下表所示。又任意抽查了100个完成手术的病例，所用时间v(小时)出现的次数亦见下表。

到达的病人数n	出现次数f_n	为病人完成手术时间v(小时)	出现次数f_v
0	10	0.0~0.2	38
1	28	0.2~0.4	25
2	29	0.4~0.6	17
3	16	0.6~0.8	9
4	10	0.8~1.0	6
5	6	1.0~1.2	5
6以上	1	1.2以上	0
合计	100	合计	100

Haiyan Lu
Jiangnan University

$M/M/1/\infty/\infty$模型

- 解：
  - 每小时平均到达率 $= \frac{\sum n f_n}{100} = 2.1$(人/小时)
  - 每次手术平均时间 $= \frac{\sum v f_v}{100} = 0.4$(小时/人)
  - 每小时完成手术人数(平均服务率) $= \frac{1}{0.4} = 2.5$(人/小时)
  - 取$\lambda = 2.1, \mu = 2.5$，可以通过**统计检验**的方法，认为病人到达数服从参数为$\lambda = 2.1$的**泊松分布**，手术时间服从参数为$\mu = 2.5$的**负指数分布**。
  - $\rho = \frac{\lambda}{\mu} = \frac{2.1}{2.5} = 0.84$，它说明服务机构（手术室）有84%的时间是繁忙的（被利用），有16%的时间是空闲的。
  - 在病房中病人数的期望值 $L_s = \frac{\lambda}{\mu-\lambda} = \frac{2.1}{2.5-2.1} = 5.25$(人)

    排队等待的病人数的期望值 $L_q = \rho L_s = 0.84 \times 5.25 = 4.41$(人)

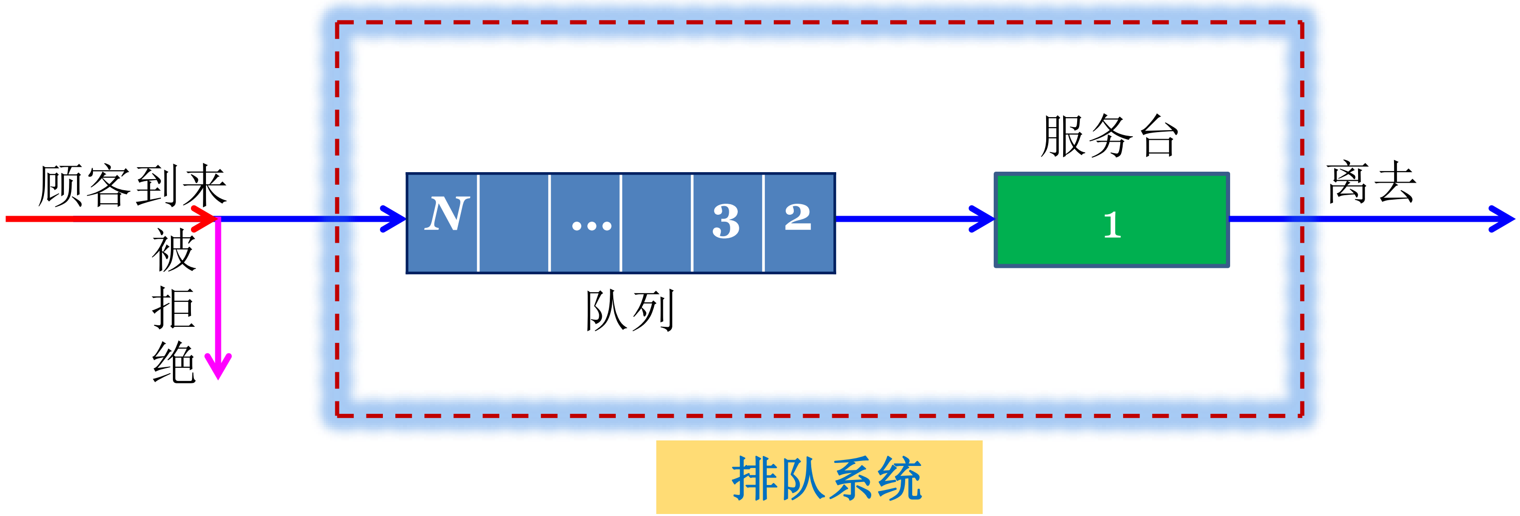
    病人在病房中逗留时间的期望值 $W_s = \frac{1}{\mu-\lambda} = \frac{1}{2.5-2.1} = 2.5$(小时)

    病人排队等待时间的期望值 $W_q = \frac{\rho}{\mu-\lambda} = \frac{0.84}{2.5-2.1} = 2.1$(小时)

Haiyan Lu 5/27/2020
Jiangnan University

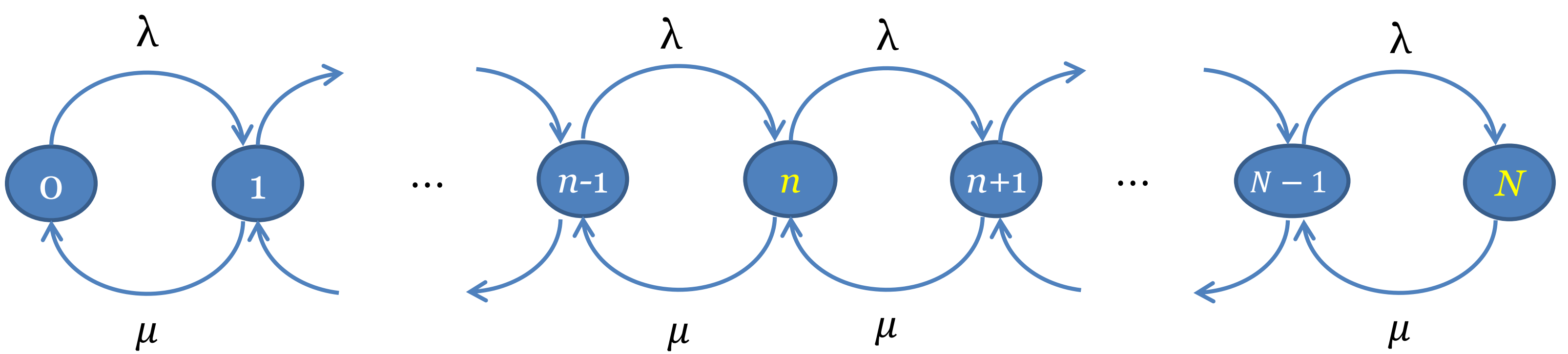
$M/M/1/N/\infty$模型

- 模型概述
  - 系统的最大容量为N，排队等待的顾客数最多为$N-1$.
  - 在某时刻一顾客到达时，若系统中已有N个顾客，则该顾客就被拒绝进入系统。
  - 当$N=1$时，即为即时制的情形；当$N\to\infty$时，即为容量无限制的情形($M/M/1/\infty/\infty$模型)



$M/M/1/N/\infty$模型

- 模型求解
  - 考虑稳态的情形，作出各状态间概率强度的转换关系图.



  - 根据上图，写出状态概率的平衡方程：

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

其中，$P_0 + P_1 + \cdots + P_N = 1$

$M/M/1/N/\infty$模 型

- 模型求解
  - 系统的各状态概率（稳态概率）

$$\begin{cases} P_0 = \frac{1-\rho}{1-\rho^{N+1}} \\ P_n = \frac{1-\rho}{1-\rho^{N+1}}\rho^n \quad (1 \le n \le N) \end{cases}$$

其中 $\rho = \frac{\lambda}{\mu}$， $\rho \neq 1.$

  - **备注**
    - 对于$M/\ M\ /1/N/\ \infty$模型, 若$\rho = 1$(即$\lambda = \mu$), 代入状态平衡方程求解可得

$$P_0 = P_1 = \cdots = P_N = \frac{1}{N+1}$$

从而 $L_s = \frac{N}{2}$.

    - 在$M/\ M\ /1/\ \infty\ /\ \infty$模型中，有$\rho < 1$这个要求；而在$M/\ M\ /1/N/\ \infty$模型中，则无此要求，但当$\rho > 1$时，**损失率$P_N$** (或被拒绝排队的顾客平均数$\lambda P_N$)将会很大。

$M/M/1/N/\infty$模型

- 模型求解
  - 系统的运行指标
    - 队长期望值

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

    - 队列长期望值

$$L_q = \sum_{n=1}^{N} (n-1)P_n = L_s - (1 - P_0)$$

    - 顾客逗留时间期望值

$$W_s = \frac{L_s}{\lambda_{\text{eff}}} = \frac{L_s}{\lambda(1-P_N)} = \frac{L_s}{\mu(1-P_0)}$$

    - 顾客等待时间期望值

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

Haiyan Lu 5/27/2020
Jiangnan University

$M/M/1/N/\infty$模型

- 模型求解
  - 系统的运行指标
    - **注意：** 在本模型中应用 Little公式时，需将**平均到达率** λ 换成**有效到达率** λ_{eff}。这里 $\lambda_{\text{eff}} = \lambda(1 - P_N)$，且有$1 - P_0 = \lambda_{\text{eff}}/\mu$.

$$(1)\ L_s = \lambda_{\text{eff}} W_s \qquad (2)\ L_q = \lambda_{\text{eff}} W_q$$

$$(3)\ W_s = W_q + \frac{1}{\mu} \qquad (4)\ L_s = L_q + \frac{\lambda_{\text{eff}}}{\mu}$$

$M/M/1/N/\infty$模型中系统运行指标的Little 公式

$M/M/1/N/\infty$模 型

- 例2 单人理发馆有6个椅子接待人们排队等待理发。当6个椅子都坐满时，后来到的顾客不进店就离开。顾客平均到达率为3人/小时，理发需时平均15分钟。求：
  - 某顾客一到达就能理发的概率；
  - 需要等待的顾客数的期望值；
  - 有效到达率；
  - 一顾客在理发馆内逗留时间的期望值；
  - 在可能到来的顾客中不等待就离开的概率。

$M/M/1/N/\infty$模 型

- 解：

此为$M/M/1/N/\infty$模型，其中$N=7$.

$\lambda=3$(人/小时)，$\mu=4$(人/小时)，$\rho=\lambda/\mu=3/4$.

- 某顾客一到达就能理发(相当于理发馆内没有顾客)的概率

$$P_0=\frac{1-\rho}{1-\rho^{N+1}}=\frac{1-3/4}{1-(3/4)^8}=0.2778$$

- 需要等待的顾客数的期望值

由$L_s=\frac{\rho}{1-\rho}-\frac{N+1}{1-\rho^{N+1}}\rho^{N+1}=\frac{3/4}{1-3/4}-\frac{7+1}{1-(3/4)^8}\times(3/4)^8=2.11$,

可得

$$L_q=L_s-(1-P_0)=2.11-(1-0.2778)=1.39$$

$M/M/1/N/\infty$模 型

- 解（续）：
  - 有效到达率

$$\lambda_{\text{eff}} = \mu(1 - P_0) = 4 \times (1 - 0.2778) = 2.89\ (人/小时)$$

  - 一顾客在理发馆内逗留时间的期望值

$$W_s = \frac{L_s}{\lambda_{\text{eff}}} = \frac{2.11}{2.89} = 0.73(小时) = 43.8(分钟)$$

  - 在可能到来的顾客中不等待就离开(即系统中有7个顾客)的概率(即顾客损失率)

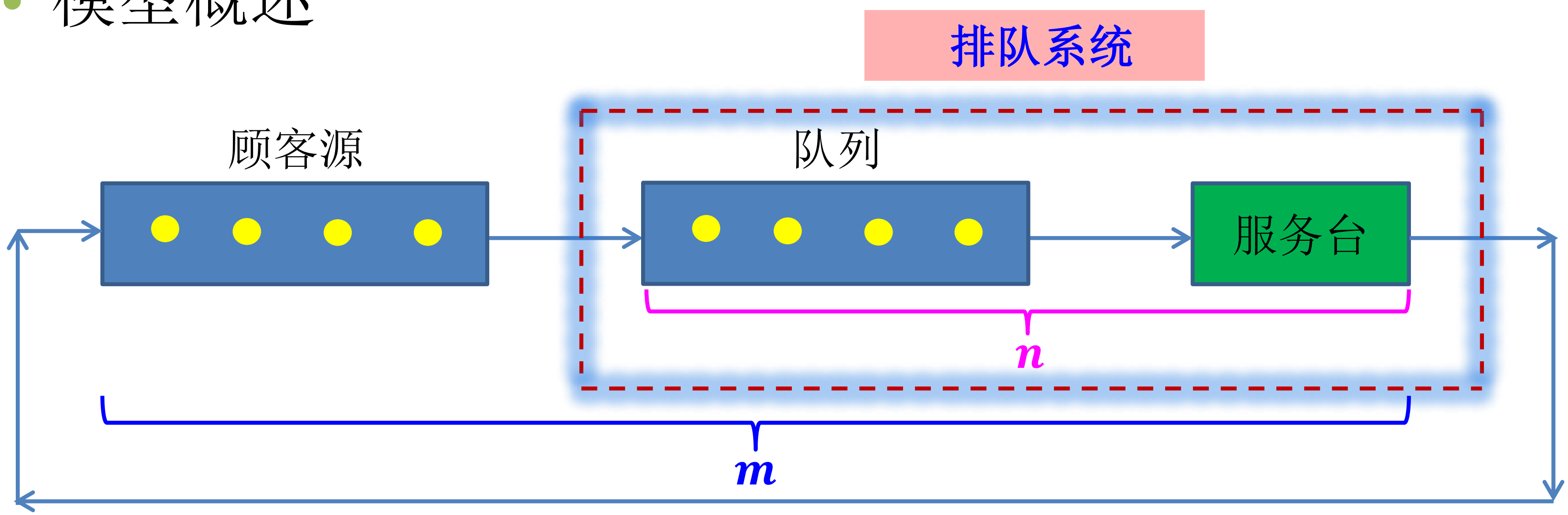
$$P_7 = \frac{1-\rho}{1-\rho^8}\rho^7 = \frac{1-3/4}{1-(3/4)^8} \times (3/4)^7 \approx 3.7\%$$

$M/M/1/\infty/m$ 模型

- 模型概述
  - 以机器因故障停机待修问题为例来说明：
    - 设共有m台机器（顾客总体由m台机器构成），机器因故障停机表示“到达”，待修的机器形成队列，修理工人为服务员，本模型考虑单服务员情形。
    - **类似例子：** m个打字员共用一台打字机；m个会计分析员共用一个计算机终端。
  - 顾客总体只有m个顾客，但每个顾客到来并经过服务后，仍回到原来总体，所有仍然可以到来。
    - 在机器故障问题中，同一台机器出了故障（到来）并经修好后（服务完成）仍可再出故障。
  - 模型符号$M/M/1/\infty/m$中第四项(系统容量限制)为∞，表示对系统的容量没有限制，但实际上它不会超过m，该模型符号又可写成$M/M/1/m/m$.

$M/M/1/\infty/m$ 模型

- 模型概述

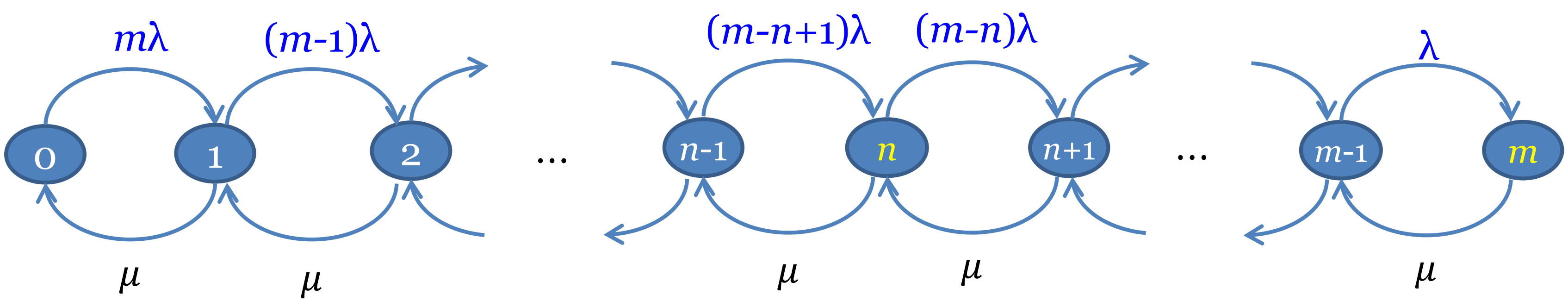


- 关于**平均到达率**
  - 当顾客总体为**无限**时，是按**全体顾客**来考虑。
  - 当顾客总体为**有限**时，则按**每个顾客**来考虑：
    - 设每个顾客的到达率(每台机器单位运转时间内发生故障的概率或平均次数)均为λ，这时在系统外的平均顾客数为$m-L_s$，故对系统的**有效到达率** λ_{eff}应为

$$\lambda_{\mathrm{eff}}=\lambda(m-L_s)=\mu(1-P_0)$$

$M/M/1/\infty/m$模型

- 模型求解
  - 考虑稳态的情形，作出各状态间概率强度的转换关系图.



  - 根据上图，写出状态概率的平衡方程：

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

其中， $P_0 + P_1 + \cdots + P_m = 1$（因而不要求$\rho = \lambda/\mu < 1$）

$M/M/1/\infty/m$模型

- 模型求解
  - 系统的各状态概率（稳态概率）

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

  - 系统的运行指标 (仍然满足**Little公式**，但需将λ换成 λ_{eff})

$$\begin{cases} L_s=m-\frac{\mu}{\lambda}(1-P_0) \\ L_q=m-\dfrac{(\lambda+\mu)(1-P_0)}{\lambda}=L_s-(1-P_0)=L_s-\dfrac{\lambda_{\text{eff}}}{\mu} \\ W_s=\dfrac{m}{\mu(1-P_0)}-\dfrac{1}{\lambda}=\dfrac{L_s}{\lambda_{\text{eff}}} \\ W_q=W_s-\dfrac{1}{\mu} \end{cases}$$

  - 在**机器故障问题**中，L_s即为平均故障机器台数，$m-L_s$为正常运转的平均机器台数： $m-L_s=\frac{\mu}{\lambda}(1-P_0)$

Haiyan Lu
Jiangnan University

$M/M/1/\infty/m$ 模型

- 例3 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15分钟，有一个修理工，每次修理时间服从负指数分布，平均每次12分钟。求：
  - 修理工空闲的概率；
  - 5台机器都出故障的概率；
  - 出故障的平均台数；
  - 等待修理的平均台数；
  - 平均停工时间；
  - 平均等待修理时间；
  - 评价这些结果。

$M/M/1/\infty/m$ 模型

- 解：

  此为$M/M/1/\infty/m$ 模型，$m=5$，$\lambda=\frac{1}{15}$（台/分钟），$\mu=\frac{1}{12}$（台/分钟），$\rho=\frac{\lambda}{\mu}=0.8$.

  - 修理工空闲的概率

$$P_0=\frac{1}{\sum_{i=0}^{m}\frac{m!}{(m-i)!}\left(\frac{\lambda}{\mu}\right)^i}=0.0073$$

  - 5台机器都出故障的概率

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

  - 出故障的平均台数

$$L_s=m-\frac{\mu}{\lambda}(1-P_0)=5-\frac{1}{0.8}(1-0.0073)=3.76\text{（台）}$$

$M/M/1/\infty/m$ 模型

- 解（续）：
  - 等待修理的平均台数

$$L_q = L_s - (1 - P_0) = 3.76 - (1 - 0.0073) = 2.77\text{（台）}$$

  - 平均停工时间

$$W_s = \frac{m}{\mu(1 - P_0)} - \frac{1}{\lambda} = \frac{5}{\frac{1}{12}(1 - 0.0073)} - 15 = 46\text{（分钟）}$$

  - 平均等待修理时间

$$W_q = W_s - \frac{1}{\mu} = 46 - 12 = 34\text{（分钟）}$$

  - 评价：机器停工时间过长，修理工几乎没有空闲时间，应当提高服务减少修理时间或增加修理工人。

Thank you!

谢 谢 !