· 研究简报 ·

马尔科夫排队过程

杨春巍
(重庆建筑大学基础科学系　400045)

摘　要　马尔科夫排队过程是马尔科夫决策规划的一个重要应用。本文研究了将一般的排队系统转化为马尔科夫排队过程，因而可以利用马尔科夫决策规划的求值运算来求解。本文着重介绍了顾客逐一的接受服务和顾客成批的接受服务两种最主要类型，并计算给出相应的结果。

关键词　马尔科夫排队过程，马尔科夫决策规划，队长分布，等待时间分布
中图法分类号　O211.62

对随机地发生的需求提供服务的系统，无论类型各异，总可以建立起数学模型：一个“顾客”来到“服务台”要求服务，若服务员正忙于接待着一位顾客，新到的顾客必须一直等到服务员为上一位顾客服务完之后，才能得到服务。在这段时间中，可能又有其他顾客来到服务台要求服务，如果有很多顾客在服务员忙碌的时候到来，就必须排队等候，直到轮到他接受服务为止。为了研究排队系统，必须考察系统的结构：

1)　输入过程：表示“顾客”来到服务台的概率规律，假定顾客们于时刻 $t_1, t_2, \cdots\cdots t_n$ $(0 < t_1 < t_2 < \cdots\cdots < t_n < \infty)$ 来到服务台，并令 $\tau_n = t_{n+1} - t_n$ 表示第 $n+1$ 个与第 n 个顾客来到的时刻差。输入过程由“来到时刻序列”$\{t_n\}$ 与“间隔时刻序列”$\{\tau_n\}$ 所遵循的概率规律所确定。一般实际问题中，输入的来到时刻遵循具有一个参数 λ 的 Poisson 过程，即遵循

$$P\{\xi = k\} = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}$$

此时，相隔时间 τ_n 遵循负指数分布，即

$$A(\tau) = \begin{cases} 1 - e^{-\lambda\tau} & \tau \geq 0 \\ 0 & \tau < 0 \end{cases}$$

2)　排队规则：确定为“先到先服务”

3)　服务机制：令　RV. ξ 表示第 n 个顾客所需的服务时间，即“服务时间序列”$\{\xi_n\}$ 所遵循的概率规律表示了排队系统的服务机制。相继的服务时间 $\xi_1, \xi_2 \cdots\cdots \xi_n \cdots\cdots$ 是相互独立且具有相同的分布函数 $B(\xi)$，$0 < \xi < \infty$，实际问题中，特别重要的分布函数是：

$$B(\xi) = \begin{cases} 1 & \xi > \dfrac{1}{\mu} \\ 0 & \xi \leq \dfrac{1}{\mu} \end{cases} \tag{1}$$

称为确定性或正则的；

收稿日期：1997 - 03 - 24
杨春巍．男，1940 年生，教授

$$B(\xi)=\begin{cases}1-e^{-\lambda\xi} & \xi\geq 0\\ 0 & \xi<0\end{cases}\tag{2}$$

称为负指数的。

其中 $\mu=E[\xi]$ 是平均或期望服务时间。或更一般的分布函数是由 Erlang 引入的

$$\mathrm{d}B(\xi)=\frac{(\mu n)^n}{n!}\xi^{n-1}e^{-\mu n\xi}\mathrm{d}\xi\tag{3}$$

表示服务时间是自由度为 $2n$ 的 X^2 分布，此时，顾客必须通过 n 个服务"阶段"，当 $n=1$ 时，即是(1)，当 $n\to\infty$ 则是(2)

对排队系统引入符号来规定所研究的排队系统：

D：确定性的或正则的，

M：随机的或 Poisson 的，

E_n：具 $2n$ 自由度的 X^2 的，

G：对 $B(\xi)$ 未作特殊的假定，

GI：一般的独立输入，即 τ_n 是相互独立的，且具有相同的分布函数 $A(\tau)$。

用这些记号，就可以用一个标号来标明一种类型的排队系统，如 $M/G/1$，表示 Poisson 来到，对服务时间分布无特殊假定，一个服务员，即，左边记号表示输入过程，中间的记号标明服务时间分布，右边记号是服务员的数目。

1　排队过程的马尔科夫过程化

1.1　Markov 链表示

要利用马尔科夫决策规划的成熟理论及求值运算解决排队过程，必须使一般的非马尔科夫型的排队过程马尔科过程化。以嵌入马尔科夫链的方法来解决：在时刻 t 时，排队系统的状态用 RV，$Y(t)$ 表示，于是在过程 $\{Y(t),t\geq 0\}$ 的任一实现中，系统的历史可用定义域 $(0,\infty)$ 的时间函数 $Y(\cdot)$ 表示，以集 Ωt 表示定义域为 $[0,t]$ 且值域与 $Y(\cdot)$ 相同的全体函数，对每个实现，令 T 为 $t\in[0,\infty]$ 的集，对任意 t，$Y(\cdot)$ 限于 $[0,t]$ 上是 θ 的一个元素。另外，定义 RV，$X(t)=f_t\{Y(\tau):\tau\leq t,t\in T\}$，其中 f_t 是一个特定的具定义域 θ 的泛函，对 $t\in[0,\infty]$ 使得 t 可顺序排列 $\cdots\cdots t_{n-1},t_n,t_{n+1},\cdots\cdots$，对任意 $t_n\in T$ 有 $X_n=X_n(t_n)$，则对一切 n 有 $P\{X_{n+1}=x_{n+1}\mid X_n=x_n,X_{n-1}=x_{n-1}\cdots\cdots\}=P\{X_{n+1}=x_{n+1}\mid X_n=x_n\}$

于是称过程 $\{X_n,n=0,1,2\cdots\cdots\}$ 为嵌入 Markov 链。嵌入 Markov 链的转移概率为

$$P_{ij}=P\{X_{n+1}=j\mid X_n=i\}$$

例如考察排队系统 $M/G/1$，设 i 与 j 表示两个相继离去的顾客后面的顾客数，令 $\mathrm{P}=(P_{ij})$ 表征嵌入 Markov 链的转移概率矩阵，于是

$$P=\left\{\begin{array}{lllll}k_0 & k_1 & k_2 & k_3 & \cdots\cdots\\ k_0 & k_1 & k_2 & k_3 & \cdots\cdots\\ 0 & k_0 & k_1 & k_2 & \cdots\cdots\\ 0 & 0 & k_0 & k_1 & \cdots\cdots\\ \cdots\cdots & & \cdots\cdots\cdots\cdots & & \end{array}\right.$$

其中 $k_r > 0 \quad (r = 0,1,2\cdots\cdots)$ 由

$$k_r = \frac{1}{\Gamma\,!}\int_0^\infty (\lambda\xi)^\Gamma e^{-\lambda\xi}\,\mathrm{d}\,B(\xi)$$

给出,其中 $\lambda > 0$ 是表征 Poisson 输入的参数。

1.2 Morkov 过程表示

令 RV, $X(t)$ 表时刻 t 时排队的顾客数,即在时刻 t 的队长,则随机过程 $\{X(t), t \geq 0\}$ 是可数状态的 Morkov 过程,令 $P(t) = [P_{ji}(t)]$ 表示过程 $\{X(t)\}$ 的转移矩阵,且满足 Kolmogorov 方程组 $\frac{\mathrm{d}\,P(t)}{\mathrm{d}\,t} = P(t)A(t)$ 且 $P(0) = I$ 为单位阵或表示为

$$\frac{\mathrm{d}\,P_{ij}(t)}{\mathrm{d}\,t} = \sum_{k=0}^{\infty} P_{ik}(t)\,a_{kj}(t) \qquad i, j = 0,1,2\cdots\cdots$$

且 $$P_{ij}(0) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

(δ_{ij}) 称 Kronecker 算子)

为解 Kolmogorov 方程,需规定函数 $a_{ij}(t)$,它们在排队系统中都是时间 t 的函数,并且是刻划时间间隔及服务时间分布函数的参数的函数。因而就可按通常的方法解 Kolmogorov 方程。以此为基础的排队系统可分为两个部份:(1) 非平衡理论:主要解决 $(P_x(t)) = P\{X(t) = x\}$ $x = 0, 1\cdots\cdots$,即寻找时刻 t 时排队长度为 x 的概率。

若初始状态 x_0 已知,则 $\{P_{x_0 x}(t)\}$ 是时刻 t 的绝对概率分布。(2) 平衡理论:主要解决概率 $\pi_x = \lim_{t\to\infty} P_x(t)$,即要寻求状态 x 的极限或平稳概率分布 $\{\pi_x\}$。寻求 π_x 是排队论中一个重要问题。

2 顾客逐一接受服务的排队系统

考察 $M/M/1$ 型的排队系统,即系统属于单个服务员,具有负指数的间隔时间与服务时间的分布函数。这种类型有:

1) 在区间 $(t, t+\Delta t)$ 内,有一个顾客来到服务台的概率是 $\lambda\Delta t + 0(\Delta t)$, $\lambda > 0$;
2) 在区间 $(t, t+\Delta t)$ 内,有一个顾客接受服务完毕后离去的概率是 $\mu\Delta t + 0(\Delta t)$, $\mu > 0$;
3) 在区间 $(t, t+\Delta t)$ 内,既无顾客到来也无顾客离去的概率是 $0(\Delta t)$。

则得到差分微分方程

$$\begin{cases} \dfrac{\mathrm{d}\,P_x(t)}{\mathrm{d}\,t} = \lambda P_{x-1}(t) - (\lambda + \mu) P_x(t) + \mu P_{x+1}(t) \qquad x = 0,1\cdots\cdots \\ \dfrac{\mathrm{d}\,P_0(t)}{\mathrm{d}\,t} = -\lambda P_0(t) + \mu P_1(t) \end{cases} \tag{4}$$

为解这方程组,用母函数与 Laplace 变换法,令

$$F(s,t) = \sum_{x=0}^{\infty} P_x(t)\,S^x, \quad |S| \leq 1$$

表示概率 $P_x(t)$ 的母函数,所以母函数方程为

$$S\frac{\partial F(s,t)}{\partial t}=(1-S)\{(\mu-\lambda s)F(s,t)-\mu P_0(t)\} \tag{5}$$

如果在时刻 $t=0$,排队中的顾客数是 x_0 即 $X(0)=x_0$,则方程应在初始条件 $F(s,0)=S^{x_0}$ 下求解。令 $f(s,z)=\mathrm{L}\{F(s,t)\}$是 $F(s,t)$的 Laplace 变换,从(4)与(5)得:

$$f(s,z)=\frac{S^{x_0+1}-\mu(1-S)P_0(z)}{zS-(1-S)(\mu-\lambda S)} \tag{6}$$

其中 $P_0(z)=L\{P_0(t)\}$,由 Laplace 变换定义,$f(s,z)$必须在单位圆$|S|=1$内处处收敛,在此区域中,$f(s,z)$的表达式的分子与分母的零点必须重合,分母零点是

$$\xi_{1,2}(z)=\left\{\frac{(\lambda+\mu+z)\mp\left[(\lambda+\mu+z)^2-4\lambda\mu\right]^{1/2}}{2\lambda}\right\}$$

于是
$$\begin{cases}\xi_1+\xi_2=\dfrac{\lambda+\mu+z}{\lambda}, & \xi_1\xi_2=\dfrac{\mu}{\lambda}\\ z=\lambda(1-\xi_1)(1-\xi_2)\end{cases}$$

则 $\xi_1(z)$是单位圆内的唯一零点,故而有

$$P_0(z)=\frac{\xi_1^{x_0+1}}{\mu(1-\xi_1)}$$

把此式代入(6),以$(s-\xi_1)$除两边,再展成 s 的幂级数,即可将(6)改写为:

$$\begin{aligned}f(s,z)&=\frac{S^{x_0+1}-(1-S)\xi_1^{x_0+1}/(1-\xi_1)}{-\lambda(S-\xi_1)(S-\xi_2)}\\
&=\frac{1}{\lambda\xi_2}\sum_{i=0}^{x_0-1}\left\{\frac{(\xi_1/\xi_2)^i-1}{(\xi_1/\xi_2)-1}+\frac{(\xi_1/\xi_2)^i}{1-\xi_1}\right\}\xi_1^{x_0-i}s^i\\
&\quad+\frac{1}{\lambda\xi_2}\left\{\frac{(\xi_1/\xi_2)^{x_0}-1}{(\xi_1/\xi_2)-1}+\frac{(\xi_1/\xi_2)^{x_0}}{1-\xi_1}\right\}\sum_{i=x_0}^{\infty}\xi_2^{x_0-i}s^i\end{aligned} \tag{7}$$

若置
$$\xi_1=a^{1/2}\theta^{-1},\quad \xi_2=a^{1/2}\theta,\quad a=\mu/\lambda \tag{8}$$

则(7) 可写成为:

$$\begin{aligned}f(s,z)=\lambda^{-1}\sum_{i=0}^{\infty}\Bigg\{&\frac{1}{\theta^{x_0+i}(\theta-a)^{1/2}}-\frac{1}{\theta^{x_0+i+1}-\theta^{x_0+i-1}}\\
&+\frac{1}{\theta^{|x_0-i|+1}-\theta^{|x_0-i|-1}}\Bigg\}a^{1/2(x_0-i-1)}S^i\end{aligned} \tag{9}$$

为了简化,考察 $F(s,t)$对时间的导数的 Laplace 变换,由(5)得:

$$L\left\{\frac{\partial F(s,t)}{\partial t}\right\}=Zf(s,z)-s^{x_0} \tag{10}$$

将(9)代入(10)并用(8),即可看到 S^i 的系数是 $L\{\mathrm{d}Pt(t)/\mathrm{d}t\}$即 $\mathrm{d}Pt(t)/\mathrm{d}t$ 的 Laplace 变换。这个系数是六个项之和,每项都正比于形如

$$\frac{1}{\theta^{n+1}-\theta^{n-1}}\qquad n\geq 0$$

的表达式,它的 Laplace 变换是

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{e^{zt}}{\theta^{n+1}-\theta^{n-1}}\mathrm{d}z=(\lambda\mu)^{1/2}e^{(1/2)(\lambda+\mu)t}I_n[2(\lambda\mu)^{1/2}t]$$

其中 $I_n[2(\lambda\mu)^{1/2}t]$是 n 阶 Bessel 函数,因此,对任一的 i,有

$$\frac{\mathrm{d}P_i(t)}{\mathrm{d}t}=\left(\frac{\mu}{\lambda}\right)^{(1/2)(x_0-i)}e^{-(\lambda+\mu)t}\{-\lambda+\mu)I_{x_0-i}(\ \cdot\ )$$
$$+(\lambda\mu)^{1/2}I_{x_0-i-1}(\ \cdot\ )+(\lambda\mu)^{1/2}I_{x_0-i+1}(\ \cdot\ )$$
$$+I_{x_0+i+2}(\ \cdot\ )+2(\lambda+\mu)^{1/2}I_{x_0+i+1}(\ \cdot\ )+\mu I_{x_0+i}(\ \cdot\ )\} \qquad (11)$$

其中 $I_n(\ \cdot\ )=I_n[2(\lambda\mu)^{1/2}\ t]$,将(11)积分,就得到 $M/M/1$ 型系统的排队长度的概率分布。

这种排队模型中,队中的顾客是逐个被服务的,它在电话业务理论、在机器维修理论等领域中有着普遍的应用。

3 顾客成批接受服务的排队过程

这种排队过程可描述为:顾客们随机地来到服务台,然后加入排队,分批地在各个时间区间中得到服务,每一批接受服务的顾客数有一个固定的最大值。假设这个最大值为 m,即各批中受到服务的顾客数都不超过 m。因此,每次服务意味着从排队中至多减去 m 个顾客。假定:(1) 设在相继两次服务之间的时间间隔,长度为 ξ,是相互独立的,并且有相同的分布函数 $B(\xi)$, $0\leq\xi<\infty$; (2) 在任何给定的时间区间 ξ 内,新来到的顾客数 Γ 的分布是 Poisson 分布,即

$$\frac{(\lambda\xi)^{\Gamma}}{\Gamma\ !}e^{-\lambda\xi} \qquad \Gamma=0,1,2\cdots\cdots$$

令 $h=\lambda E[\xi]$ 是每个时间间隔内来到的顾客的平均数, $E[\xi]$ 是 ξ 的期望值,因此,在一个随机的时间间隔内, Γ 的分布为:

$$k_{\Gamma}=\frac{1}{\Gamma\ !}\int_0^{\infty}(\lambda\xi)^{\Gamma}e^{-\lambda\Gamma}\mathrm{d}B(\xi) \qquad \Gamma=0,1,2\cdots\cdots$$

如果 $k(s)$ 表示任一区间内来到顾客数分布的母函数,则由上式可得:

$$k(s)=\sum_{\Gamma=0}^{\infty}k_{\Gamma}s^{\Gamma}=\phi[\lambda(1-s)] \quad |s|\leq 1 \qquad (12)$$

其中 $\phi(\ \cdot\ )$ 是 $B(\xi)$ 的 Laplace - Stieltjes 变换。

若以 RV. X 表示排队的长度,并以 X 与 X' 表示相继两批顾客已经接受服务完毕后的排队长度,在这情况下,由排队的长度所生成的嵌入 Markov 链的转移概率矩阵为:

$$P=(P_{ij})=\begin{bmatrix} k_0 & k_1 & k_2 & k_3 & k_4 & \cdots\cdots \\ k_0 & k_1 & k_2 & k_3 & k_4 & \cdots\cdots \\ & \cdots\cdots & & \cdots\cdots & & \\ & \cdots\cdots & & \cdots\cdots & & \\ k_0 & k_1 & k_2 & k_3 & k_4 & \cdots\cdots \\ 0 & k_0 & k_1 & k_2 & k_3 & \cdots\cdots \\ 0 & 0 & k_0 & k_1 & k_2 & \cdots\cdots \\ & \cdots\cdots & & \cdots\cdots & & \end{bmatrix}$$

其中前 $m+1$ 行完全相同,以成批服务排队的特征以及从转移矩阵的形式可知,在 $m=1$ 时(即顾客一个一个地接受服务),成批服务排队系统就化为 $M/G/1$。

3.1　排队长度的平衡分布

假定 Poisso 参数 $\lambda = 1$，则排队处于平衡的条件是繁忙强度 $\rho < 1$，在成批服务的情况下，$\rho = E[\xi]/m$，因此，为了平衡，必须有 $E[\xi] < m$。

由转移矩阵的构造，嵌入 Markov 链是不可约的、非周期的，因此，(1) 每个状态是滑过的或零常返的，并且对一切 i 与 j 有 $\lim\limits_{n\to\infty} P_{ij}^{(n)} = 0$，(2) 或者每个状态是各态历经的，并且对一切 i 与 j 有 $\lim\limits_{n\to\infty} P_{ij}^{(n)} = x_j^* > 0$，且适合 $\sum\limits_{j=1}^{\infty} x_j^* = 1$。

令 $\Pi^*(s)$ 表示极限概率 π_j^* 的母函数，由于 π_j^* 满足线性方程组。

$$\pi_j^* = \sum_{i=0}^{\infty} \pi_i^* P_{ij} \qquad j = 0, 1, 2 \cdots\cdots$$

因此 $\Pi^*(s)$ 为

$$\Pi^*(s) = \frac{\sum\limits_{j=0}^{m-1} \pi_j^* (s^m - s^i)}{s^m / [k(s) - 1]} \tag{13}$$

例如考察一个排队过程，其特征为：Poisson 来到，服务时间区间的分布是自由度为 $2n$ 的 X^2 分布，因此

$$\mathrm{d}B(\xi) = \frac{\mu^n}{\Gamma(n)} \xi^{n-1} e^{-\mu\xi} \mathrm{d}\xi$$

因此，每个服务区来到的顾客平均数 h 为

$$h = \frac{\lambda n}{\mu} = \frac{n}{\mu} \qquad (\because \lambda = 1)$$

由 (12) 有：

$$k(s) = \left[1 + \frac{h(1-s)}{h}\right]^{-n}$$

为使系统满足 $\sum\limits_{j=0}^{\infty} x_j^* = 1$，必须有 $\Pi^*(1-0) = 1$，此条件应用于 (12) 就得到：

$$\sum_{j=0}^{m-1} (m-j)\pi_j^* = m - h$$

即在一个服务时间区间内来到的顾客的平均数必须等于实际被服务的每批顾客的平均数。考察方程

$$s^m \left[1 + \frac{h(1-s)}{h}\right]^n - 1 = 0 \tag{14}$$

知道除了 $s = 1$ 这个零点外，恰好还有模大于 1 的 $m-1$ 个单零点，设为 $s = s_i$ ($i = 1, 2 \cdots\cdots m-1$)，因此总共有 $m+n$ 个零点，其中有 n 个具有小于或等于 1 的模。

利用 $\Pi^*(1) = 1$，可知母函数 $\Pi^*(s)$ 具有形式

$$\Pi^*(s) = \prod_{j=m}^{m+n-1} \left(\frac{s_j - 1}{s_j - s}\right) \tag{15}$$

利用累积量母函数 $G(u)$，可求得排队长度 X 的矩，从 (15) 有

$$G(u) = \sum_{j=m}^{m+n-1} \mathrm{Log}(s_j - 1) - \sum_{j=m}^{m+n-1} \mathrm{Log}(s_j - e^u)$$

因此,
$$E(X)=\sum_{j=m}^{m+n-1}\operatorname{Log}(s_j-1)^{-1}$$
及
$$D(X)=\sum_{j=m}^{m+n-1}\operatorname{Log}s_j(s_j-1)^{-2}$$

3.2 成批服务排队的等待时间分布

令 RV. W 表示等待时间。Downto[1]证明了等待时间的平衡分布的 Laplace 变换为

$$\gamma(t)=\frac{\Pi^*(1-t)}{E[\xi]^t}\left[\frac{1}{\phi(t)}-1\right] \tag{16}$$

因此, Laplace 变换可从排队长度的平衡分布的母函数求得, 由于这个排队系统的服务时间区间是 X^2 分布的,因此(16)为:

$$\gamma(t)=\frac{\mu\left[(1+ht/n)^n-1\right]}{nt}\prod_{j=m}^{m+n-1}\left(\frac{s_j-1}{s_j-1+t}\right)$$

其中 s_j 仍是(14)的模大于 1 的零点,给定 s_j 后, W 的平衡分布就可通过逆变换求得。

等待时间平均数与方差为: $E(W)=\sum\limits_{j=m}^{m+n-1}(s_j-1)^{-1}-\dfrac{n-1}{2\mu}$

$$D(W)=\sum_{j=m}^{m+n-1}(s_j-1)^{-2}+\frac{(n-1)(n-s)}{12\mu}$$

成批服务的排队过程在医院门诊部的研究以及运输过程(电梯、公共汽车等等)的研究中很有用。

参 考 文 献

1 董译清. 马尔科夫决策规划. 北京:中国科学院应用数学所,1985
2 Bailey,N. T. J. On Queueing Processes with Bulk Service,1964
3 Benes,V. E. On Queues with Poisson Arrivals,1967
4 E. Б.邓 肯. 马尔科夫过程论基础. 北京:科学出版社,1962
5 王梓坤. 随机过程论. 北京:科学出版社,1965

Markovion Queuing Process

Yang Chunwei
(Dept. of Fundamental Sciences, Chongqing Jianzhu University, 400045)

Abstract Markovion pueuing process is an important application to Markovion Decision Programming. This paper studies that any system of queue can be transformed into Markovion queuing process and therefore, the solution can be obtained by Markovion Decision Programming's value operation. This paper also attaches importance to the introduction to the service accepted by individual customer and the service accepted by a group of customer, both of which are the most principal types of receptive services and consequently the corresponding results can be calculated.

Key Words Markovion queuing process, Markovion decision programming, queue length distribution, queuing time distribution

(编辑:王秀玲)