随机过程与排队论

信息与软件工程学院

顾小丰

Email：guxf@uestc.edu.cn

2020年9月27日星期日

上一讲内容回顾

- **二阶段循环排队系统**
  - 问题的引入
  - I号台的队长
  - 车辆在I号台的等待时间
- **一般服务的M/G/1/∞排队系统**
  - 嵌入马尔可夫链

本讲主要内容

- 一般服务的M/G/1/∞排队系统
  - 对长
  - 等待时间与逗留时间
  - 忙期
  - 输出过程
- 隐马尔科夫模型简介

第七章 一般服务的M/G/1/∞排队系统

前面内容着重讨论了按泊松流到达与负指数服务时间的简单排队系统，它的主要特点是在任何时刻系统都具有较好的马尔可夫性，能比较容易地得到队长分布的平稳解，因此部分内容相对讲可以看作是初等的。

对于一般服务或一般到达的排队系统，并不是任何时刻系统都具有马尔可夫性，只是在某些特殊的随机时刻系统才具有这种性质，我们称这种随机时刻为再生点，即从这个时刻起，系统好像又重新开始一样。利用再生点，一般服务或一般到达的排队系统可化成马尔可夫链，用马尔可夫链的方法来解决，这种方法叫做嵌入马尔可夫链法。此方法的精髓在于找到再生点。

§ 7.1 嵌入马尔可夫链

1. M/G/1/∞排队系统的叙述

- 顾客按参数λ(λ＞0)的泊松流到达，即相继到达的间隔时间序列$\{\tau_n, n\geq 1\}$独立、服从参数为λ(λ＞0)的负指数分布$F(t)=1-e^{-\lambda t}, t\geq 0$；
- 顾客所需的服务时间序列$\{\chi_i, i\geq 1\}$独立、同一般分布$G(t), t\geq 0$，记平均服务时间为 $0<\frac{1}{\mu}=\int_0^{\infty} t\,dG(t)$；
- 系统中只有一个服务台，容量为无穷大；
- 顾客到达时，若服务台空闲就立即接受服务，否则就排队等待，并按先到先服务的顺序接受服务，而且到达过程与服务过程彼此独立。

2.嵌入马尔可夫链

假定N(t)表示在时刻t系统中的顾客数(队长)，对于M/G/1/∞ 排队系统，由于服务时间是一般分布，对任选的一个时刻t正在接受服务的顾客可能还没有服务完。从时刻t起的剩余服务时间分布可能不具有无记忆性，于是队长{N(t)，t≥0}不再具有马尔可夫性。但是，若令N_n^+表示第n个顾客服务完毕离开时留在系统中的顾客数，即留下的队长，n≥1，则下面定理表明$\{N_n^+, n\geq 1\}$是马尔可夫链，被称为队长过程{N(t)，t≥0}的嵌入马尔可夫链。

定理

$\{N_n^+, n\geq 1\}$为一不可约、非周期的齐次马尔可夫链，其一步转移概率为

$$p_{ij}=P\{N_{n+1}^+=j\mid N_n^+=i\}$$

$$=\begin{cases}\int_0^\infty \frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t), & i=0\\ \int_0^\infty \frac{(\lambda t)^{j-i+1}}{(j-i+1)!}e^{-\lambda t}dG(t), & i\geq 1, j\geq i-1\\ 0 & i\geq 1, j<i-1\end{cases}$$

证明

设v_n表示在第n个顾客的服务时间χ_n内到达的顾客数，则容易看出$\{v_n，n\geq 1\}$相互独立同分布

$$a_j = P\{v_n = j\} = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dG(t), \qquad j \geq 0$$

而且

$$N_{n+1}^+ = \begin{cases} N_n^+ - 1 + v_{n+1}, & N_n^+ > 0 \\ v_{n+1}, & N_n^+ = 0 \end{cases}, \qquad n \geq 1$$

由于$\{v_n，n\geq 1\}$相互独立同分布，所以令$v_n = v$，$n\geq 1$，有

$$N_{n+1}^+ = \begin{cases} N_n^+ - 1 + v, & N_n^+ > 0 \\ v, & N_n^+ = 0 \end{cases}, \qquad n \geq 1$$

证明(续1)

从上式可以看出，当已知N_n^+时，N_{n+1}^+只与到达过程有关，而与$N_1^+, N_2^+,\dots, N_{n-1}^+$无关，所以是马尔可夫链，其状态空间E＝{0,1,2,…}。

其一步转移概率为：

$$p_{ij} = P\{N_{n+1}^+ = j \mid N_n^+ = i\}$$

$$= \begin{cases} P\{v = j-i+1\}, & i \geq 1 \\ P\{v = j\}, & i = 0 \end{cases}$$

当i≥1时，

$$P\{v = j-i+1\} = \int_0^\infty \frac{(\lambda t)^{j-i+1}}{(j-i+1)!} e^{-\lambda t} dG(t),$$

证明(续2)

当i＝0时，

$$P\{v=j\}=\int_0^{\infty}\frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t),$$

从一步转移概率表达式容易看出，p_{ij}，$i,j=0,1,2,\ldots$ 与时间的起点无关，而且任意两个状态是互通的，$p_{ii}>0$，$\{N_n^+, n\geq 1\}$为一不可约、非周期的齐次马尔可夫链。

V_n的均值

v_n表示在第n个顾客的服务时间χ_n内到达的顾客数，v_n分布函数为

$$a_j = P\{v_n = j\} = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dG(t), \qquad j \geq 0$$

V_n均值为

$$E[v_n] = \sum_{j=0}^{\infty} j a_j = \sum_{j=0}^{\infty} j \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dG(t)$$

$$= \int_0^{\infty} e^{-\lambda t} \sum_{j=0}^{\infty} j \frac{(\lambda t)^j}{j!} dG(t) = \lambda \int_0^{\infty} t e^{-\lambda t} \sum_{j=1}^{\infty} \frac{(\lambda t)^{j-1}}{(j-1)!} dG(t)$$

$$= \lambda \int_0^{\infty} t e^{-\lambda t} e^{\lambda t} dG(t) = \lambda \int_0^{\infty} t dG(t) = \frac{\lambda}{\mu} = \rho$$

V_n的二阶矩

V_n的二阶矩为

$$E[v_n^2]=\sum_{j=0}^{\infty} j^2 a_j=\sum_{j=0}^{\infty} j^2\int_0^{\infty}\frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t)$$

$$=\int_0^{\infty}\sum_{j=0}^{\infty} j^2\frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t)=\int_0^{\infty}E[T_n^2]dG(t)$$

$$=\int_0^{\infty}\{D[T_n]+E^2[T_n]\}dG(t)=\int_0^{\infty}[\lambda t+(\lambda t)^2]dG(t)$$

$$=\lambda\int_0^{\infty}tdG(t)+\lambda^2\int_0^{\infty}t^2dG(t)=\frac{\lambda}{\mu}+\lambda^2E[\chi^2]$$

$$=\frac{\lambda}{\mu}+\lambda^2\{D[\chi]+E^2[\chi]\}=\rho+\lambda^2D[\chi]+\rho^2$$

V_n的方差

V_n的方差为

$$D[v_n] = E[v_n^2] - E^2[v_n] = \rho + \lambda^2 D[\chi]$$

一步转移概率

嵌入马尔可夫链$\{N_n^+, n\geq 1\}$的一步转移概率矩阵为：

$$P=\begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & 0 & a_0 & a_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

其中：

$$a_j=\int_0^\infty \frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t), \qquad j=0,1,2,\cdots$$

引理

对于不可约非周期的马尔可夫链，令$\{p_{ij}$，$i,j = 0,1,2,\ldots\}$为一步转移概率，若不等式组

$$\sum_{j=0}^{\infty} p_{ij} y_j \le y_i - 1, \qquad i \ne 0$$

存在一个满足条件

$$\sum_{j=0}^{\infty} p_{0j} y_j < \infty$$

的非负解，则此马尔可夫链是正常返的。

定理

嵌入马尔可夫链$\{N_n^+，n \geq 1\}$为正常返的充分必要条件是$\rho = \lambda/\mu < 1$。

证明

充分性。设ρ＝λ/μ＜1，定义

$$y_j = \frac{j}{1-\rho} \geq 0, \qquad j = 0,1,2,\cdots$$

则

$$\sum_{j=0}^{\infty} p_{ij} y_j = \sum_{j=i-1}^{\infty} a_{j-i+1} \frac{j}{1-\rho} = \frac{i}{1-\rho} - 1 = y_i - 1, \qquad i \geq 1$$

$$\sum_{j=0}^{\infty} p_{0j} y_j = \sum_{j=0}^{\infty} a_j \frac{j}{1-\rho} = \frac{1}{1-\rho} E[v_n] = \frac{\rho}{1-\rho}$$

即$\{y_j，j\geq 0\}$满足引理的条件，因此嵌入马尔可夫链$\{N_n^+，n\geq 1\}$是正常返的。

证明(续1)

必要性。 设嵌入马尔可夫链$\{N_n^+, n\geq 1\}$是正常返的，则由极限定理知，它是遍历马氏链，且

$$\lim_{n\to\infty} P\{N_n^+ = j\} = p_j^+ > 0, \qquad j = 0,1,2,\cdots$$

存在，且$\{p_j^+, j\geq 0\}$是$\{N_n^+, n\geq 1\}$唯一的平稳分布，且满足

$$p_j^+ = \sum_{i=0}^{\infty} p_i^+ \cdot p_{ij}, \qquad j\geq 0$$

$$\sum_{j=0}^{\infty} p_j^+ = 1$$

绝对分布由初始分布和转移概率确定。而平稳分布的初始分布即为极限分布。

利用一步转移概率矩阵，有

$$p_j^+ = p_0^+ a_j + \sum_{i=1}^{j+1} p_i^+ \cdot a_{j-i+1}, \qquad j\geq 0$$

证明(续2)

引入母函数

$$P^+(z)=\sum_{j=0}^{\infty}p_j^+\cdot z^j,\quad A(z)=\sum_{j=0}^{\infty}a_j\cdot z^j,\quad |z|<1$$

得

$$P^+(z)=p_0^+\sum_{j=0}^{\infty}a_j\cdot z^j+\sum_{j=0}^{\infty}\sum_{i=1}^{j+1}p_i^+\cdot a_{j-i+1}\cdot z^j$$

$$=p_0^+A(z)+\frac{1}{z}\sum_{i=1}^{\infty}p_i^+z^i\left(\sum_{j=i-1}^{\infty}a_{j-i+1}z^{j-i+1}\right)$$

$$=p_0^+A(z)+\frac{1}{z}(P^+(z)-p_0^+)A(z)$$

于是

$$P^+(z)=\frac{p_0^+(1-z)A(z)}{A(z)-z}$$

证明(续3)

由于$P^+(1)=A(1)=1$，用求极限的洛必塔法则，得

$$1=\lim_{z\to 1^-}P^+(z)=\lim_{z\to 1^-}\frac{-p_0^+A(z)+p_0^+(1-z)A'(z)}{A'(z)-1}=-\frac{p_0^+}{A'(1)-1}$$

又

$$A'(z)=\sum_{j=0}^{\infty}ja_j=E[v_n]=\rho$$

所以

$$1=\frac{p_0^+}{1-\rho}$$

即

$$\rho=1-p_0^+<1$$

推论1

对于嵌入马尔可夫链$\{N_n^+, n\geq 1\}$，

1. 当$\rho=\lambda/\mu\geq 1$时，此马氏链是零常返或非常返的，n步转移概率的极限 $\lim\limits_{n\to\infty} p_{ij}(n)=0, \qquad i,j=0,1,2,\cdots$ 且不存在平稳分布。
2. 当$\rho=\lambda/\mu<1$时，此马氏链是正常返的，n步转移概率的极限存在，且 $\lim\limits_{n\to\infty} p_{ij}(n)=\lim\limits_{n\to\infty} P\{N_n^+=j\}=p_j^+>0$ 进一步，$\{p_j^+, j\geq 0\}$是唯一的平稳分布,有递推表达式

$$\begin{cases} p_0^+=1-\rho, \\ p_j^+=\dfrac{1}{a_0}\left(p_{j-1}^+-p_0^+a_{j-1}-\displaystyle\sum_{k=1}^{j-1}p_k^+a_{j-k}\right), \quad j\geq 1 \end{cases}$$

其中，$a_j=\displaystyle\int_0^\infty \frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t), \qquad j=0,1,2,\cdots$

推论2

对任意正整数m，有

$$\lim_{n\to\infty} P\{N_n^+ \le m\} = \begin{cases} \sum_{j=0}^{m} p_j^+, & \rho < 1 \\ 0, & \rho \ge 1 \end{cases}$$

证明：

$$P\{N_n^+ \le m\} = \sum_{j=0}^{m} P\{N_n^+ = j\} = \sum_{j=0}^{m}\sum_{i=0}^{\infty} P\{N(0) = i\}\cdot p_{ij}(n)$$

其中N(0)表示初始时刻t=0时系统中的顾客数，令n→∞，由推论1即得结论。

说明

推论1、2表明，当ρ＜1时，嵌入马尔可夫链的n步转移概率的极限总存在、为正、不依赖于初始状态，即为平稳分布；当ρ≥1时，不论怎样大的正整数m，第n个顾客服务完毕离开系统时留在系统中的顾客数≤m的概率总趋于0(n→∞)，这说明对长越来越长，系统达不到统计平衡。另外也可证明：若ρ＝1，则$\{N_n^+, n\geq 1\}$为零常返；若ρ＞1，则$\{N_n^+, n\geq 1\}$为非常返。两者的区别在于：当ρ＝1时，系统始终不空的概率为0；当ρ＞1时，系统始终不空的概率为正。

推论3

对M/G/1/∞排队系统，若$\rho=\lambda/\mu<1$，则平稳

分布$\{p_j^+,\ j\geq 0\}$的母函数为

$$P^+(z)=\frac{(1-\rho)(1-z)g(\lambda(1-z))}{g(\lambda(1-z))-z},\qquad |z|<1$$

其中

$$g(\lambda(1-z))=\int_0^\infty e^{-\lambda(1-z)t}dG(t)=\int_0^\infty e^{\lambda zt-\lambda t}dG(t)$$

$$=\int_0^\infty\sum_{j=0}^\infty z^j\frac{(\lambda t)^j}{j!}e^{-\lambda t}dG(t)=\sum_{j=0}^\infty a_j z^j=A(z)。$$

$\{N_n^+, n\geq 1\}$ 的平均对长

前面，我们讨论了队长过程$\{N(t), t\geq 0\}$的嵌入过程$\{N_n^+, n\geq 1\}$的平稳分布。

平均队长为

$$\overline{N}=\sum_{j=0}^{\infty} jp_j^+ = P^{+\prime}(1)=\frac{d}{dz}\left[\frac{(1-\rho)(1-z)A(z)}{A(z)-z}\right]\Bigg|_{z=1}$$

$$=(1-\rho)\left\{\frac{-A(z)\left[A(z)-z\right]+(1-z)\left[A(z)-A'(z)z\right]}{\left[A(z)-z\right]^2}\right\}\Bigg|_{z=1}$$

用求极限的洛必塔法则，得

$$\overline{N}=(1-\rho)\frac{-2\left[A'(1)\right]^2+2A'(1)+A''(1)}{2\left[A'(1)-1\right]^2}$$

$\{N_n^+, \ n\geq 1\}$ 的平均对长(续)

而

$$A'(1)=\sum_{j=0}^{\infty} ja_j = E[v_n]=\rho$$

$$A''(1)=\sum_{j=0}^{\infty} j(j-1)a_j=\sum_{j=0}^{\infty} j^2a_j-\sum_{j=0}^{\infty} ja_j = D[v_n]+E^2[v_n]-E[v_n]$$

$$=\rho+\lambda^2 D[\chi]+\rho^2-\rho=\lambda^2 D[\chi]+\rho^2$$

故

$$\overline{N}=\rho+\frac{\lambda^2 D[\chi]+\rho^2}{2(1-\rho)}=\rho+\frac{\lambda^2 E[\chi^2]}{2(1-\rho)}$$

上式称为**扑拉克—辛钦**(Pollaczek-Khinchin)**均值公式**。

扑拉克—辛钦均值公式

上式说明，平均对长只与ρ和D[χ]有关，即只与交通强度和服务分布的方差有关，而与服务分布的其它性质无关。当D[χ] ＝0时，则服务分布为定长分布，$\overline{N}$ 取最小值。一般情况下，平均对长$\overline{N}$ 是服务分布方差的线性函数。平均等待队长为

$$\overline{N}_q = \sum_{j=0}^{\infty} j p_{j+1}^{+} = \sum_{j=0}^{\infty} j p_j^{+} - \sum_{j=1}^{\infty} p_j^{+} = \overline{N} - (1 - p_0^{+}) = \frac{\lambda^2 E[\chi^2]}{2(1-\rho)}$$

由Little公式

平均等待时间 $\overline{W}_q = \frac{\overline{N}_q}{\lambda} = \frac{\lambda E[\chi^2]}{2(1-\rho)}$

平均逗留时间 $\overline{W} = \frac{\overline{N}}{\lambda} = \frac{1}{\mu} + \frac{\lambda E[\chi^2]}{2(1-\rho)}$

§7.2 对长

对于M/G/1/∞排队系统，令b表示从第一个顾客开始的系统忙期，且设

$$B(t)=P\{b\le t\},\ t\ge 0;\qquad b(s)=\int_0^{\infty}e^{-st}dB(t)$$

于是，我们可以先讨论系统的忙期分布，B(t)和b(s)也可由§7.4的内容给出。

下面我们在已知B(t)和b(s)的前提下，利用对长过程{N(t),t≥0}在忙期中的特性来研究对长的绝对分布和平稳分布。令

$$Q_j(t)=P\{b>t\ge 0;\ N(t)=j\},\ j\ge 1 \qquad (1)$$

表示在忙期b中系统的瞬时对长为j的概率，且在t＝0时只有一个顾客，忙期b刚开始，即

$$Q_1(0)=1,\ Q_j(0)=0,\quad j>1 \qquad (2)$$

定理1

令$q_j^*(s)=\int_0^\infty e^{-st}Q_j(t)dt$为$Q_j(t)$的拉普拉斯变换，则对s的实部R(s) ＞0和j≥1，有

$$q_j^*(s)=\frac{b(s)}{g(s+\lambda)}\int_0^\infty e^{-st}\Delta_{j-1}(t)dt$$

$$+\frac{1}{g(s+\lambda)}\sum_{k=1}^{j-1}\frac{q_{j-k}^*(s)}{b^k(s)}\left\{b(s)-\sum_{i=0}^{k}\int_0^\infty e^{-(s+\lambda)t}\frac{[\lambda b(s)t]^i}{i!}dG(t)\right\} \tag{3}$$

其中：$\Delta_j(t)=(1-G(t))\frac{(\lambda t)^j}{j!}e^{-\lambda t}=\overline{G}(t)\frac{(\lambda t)^j}{j!}e^{-\lambda t}$，$j\geq 0$;

当i≤0时 $\sum_{k=1}^{i}=0$;

$g(s+\lambda)=\int_0^\infty e^{-(s+\lambda)t}dG(t)$。

证明

设v表示在忙期b中第一个顾客的服务时间χ内到达的顾客数，显然有

$$P\{v=i\}=\int_0^\infty \frac{(\lambda t)^i}{i!}e^{-\lambda t}dG(t), \qquad i\geq 0 \qquad (4)$$

称在服务时间χ内到达的顾客为“第一类顾客”，在其后到达的顾客为“第二类顾客”。由于忙期长度与系统中的顾客数和服务顺序无关，所以我们可以重新安排服务顺序如下：

设在χ内到达的“第一类顾客”分别为$A_1,A_2,\ldots,A_v$。在服务完第一个顾客后，就服务A_1顾客并接着服务除$A_2,\ldots,A_v$外所有新到的“第二类

证明(续1)

顾客”，直到没有新到的“第二类顾客”时开始服务A_2，记这断时间为L_1，然后等服务完A_2后，又接着服务除$A_3,\ldots,A_v$外所有新到的“第二类顾客”，直到没有新到的“第二类顾客”时开始服务A_3，记这段时间为L_2，…,如此下去，直到最后服务A_v顾客及其后新到的所有“第二类顾客”，记这断时间为L_v，于是 $b=\chi+L_1+L_2+\ldots+L_v$

而且易知$L_1,L_2,\ldots,L_v$相互独立，与b同分布B(t)，并独立于χ与v，且当$v=0$时，$L_1+L_2+\ldots+L_v=0$。

于是

证明(续2)

$$Q_j(t) = P\{\chi + L_1 + L_2 + \ldots + L_v > t \geq 0;\ N(t)=j\}$$

$$= P\{\chi > t \geq 0;\ N(t)=j\}$$

$$+ \sum_{i=0}^{\infty} \int_0^t \frac{(\lambda x)^i}{i!} e^{-\lambda x} \cdot P\{L_1 + L_2 + \cdots L_i > t - x; N(t-x) = j\} dG(t)$$

$$= \overline{G}(t) \frac{(\lambda t)^{j-1}}{(j-i)!} e^{-\lambda t}$$

$$+ \sum_{i=0}^{\infty} \int_0^t \frac{(\lambda x)^i}{i!} e^{-\lambda x} \cdot P\{L_1 + L_2 + \cdots L_i > t - x; N(t-x) = j\} dG(t)$$

$$j \geq 1 \qquad (5)$$

根据服务顺序的安排，无论时刻t-x落入哪个区间L_k（$1 \leq k \leq i$）内，都有在此时刻系统的顾客数等于“第一类顾客数”加上 “第二类顾客数”，而且L_k与忙期b有相同的概率特性，所以

证明(续3)

$P\{L_1+L_2+\ldots+L_i > t\text{-}x；\ N(t\text{-}x)=j\}$

$$=\sum_{k=1}^{i} P\{L_1+L_2+\cdots L_k > t-x;$$

$$L_1+L_2+\cdots L_{k-1} \le t-x; N(t-x)=j-(i-k)\}$$

$$=\sum_{k=1}^{i} Q_{j-(i-k)}(t-x) * B^{(k-1)}(t-x) \quad , \quad j\ge 1 \qquad (6)$$

其中，当$j\le 0$时，$Q_j(t)=0$； “*”号表示卷积运算。于是

$$Q_j(t)=\Delta_{j-1}(t)$$

$$+\sum_{i=0}^{\infty}\int_0^t \frac{(\lambda x)^i}{i!} e^{-\lambda x} \cdot \sum_{k=1}^{i} Q_{j-(i-k)}(t-x) * B^{(k-1)}(t-x) dG(t)$$

$$j\ge 1 \qquad (7)$$

证明(续4)

对R(s)>0，上式的拉普拉斯变换为

$$q_j^*(s)=\int_0^\infty e^{-st}\Delta_{j-1}(t)dt$$

$$+\sum_{i=0}^{\infty}\sum_{k=1}^{i}q_{j-i+k}^*(s)b^{k-1}(s)\int_0^t\frac{(\lambda t)^i}{i!}e^{-(s+\lambda)x}dG(t)$$

$$=\int_0^\infty e^{-st}\Delta_{j-1}(t)dt$$

$$+\sum_{k=1}^{\infty}b^{k-1}(s)\sum_{i=k}^{j+k-1}q_{j-i+k}^*(s)\int_0^t\frac{(\lambda t)^i}{i!}e^{-(s+\lambda)x}dG(t)$$

证明(续5)

$$
\begin{aligned}
&= \int_0^{\infty} e^{-st}\Delta_{j-1}(t)dt \\
&+ \frac{q_j^*(s)}{b(s)}\{g(s+\lambda-\lambda b(s)) - g(s+\lambda)\} \\
&+ \frac{q_{j-1}^*(s)}{b^2(s)}\left\{g(s+\lambda-\lambda b(s)) - \sum_{i=0}^{1}\int_0^{\infty} e^{-(s+\lambda)t}\frac{[\lambda b(s)t]^i}{i!}dG(t)\right\} \\
&+ \dots \\
&+ \frac{q_1^*(s)}{b^j(s)}\left\{g(s+\lambda-\lambda b(s)) - \sum_{i=0}^{j-1}\int_0^{\infty} e^{-(s+\lambda)t}\frac{[\lambda b(s)t]^i}{i!}dG(t)\right\}
\end{aligned}
$$

$$j \ge 1 \qquad (8)$$

并注意到§7.4中定理1，有b(s)＝g(s+λ-λ b(s))。于是本定理得证。

定理2

对t≥0，令

$$p_{ij}(t)=P\{N(t)=j\,|\,N(0)=i\}$$

$$p_{ij}^{*}(s)=\int_0^{\infty}e^{-st}p_{ij}(t)dt,\quad i,j\geq 0$$

对R(s) ＞0，i≥0，j≥1，有

$$p_{i0}^{*}(s)=\frac{b^{i}(s)}{s+\lambda-\lambda b(s)} \qquad (9)$$

$$p_{ij}^{*}(s)=\sum_{k=1}^{i}q_{j-i+k}^{*}(s)b^{k-1}(s)+\frac{\lambda q_{j}^{*}(s)b^{i}(s)}{s+\lambda-\lambda b(s)} \qquad (10)$$

其中：当i≤0时 $\sum_{k=1}^{i}=0$，q_j^* 由定理1给出；b(s)由§7.4中定理1确定。

证明

设$\hat{\tau}_j$与b_j分别表示从初始状态N(0)＝0出发，系统的第j个闲期与忙期长度，j≥1。由于到达是参数λ的泊松流，所以$\hat{\tau}_j$服从参数为λ的负指数分布，而且从任意初始状态出发，顾客离开后瞬时对长为n（n≥0）的那些时刻构成一个更新过程的更新时刻，于是$\{\hat{\tau}_j, j\geq 1\}$与$\{b_j, j\geq 1\}$相互独立，且各自为一个独立过程。

显然，在时刻t系统的对长等于0的充要条件是时刻t处于系统的闲期，所以

证明(续1)

$$P_{00}(t)=\sum_{k=1}^{\infty}P\left\{\sum_{j=1}^{k-1}(\hat{\tau}_j+b_j)\le t<\sum_{j=1}^{k-1}(\hat{\tau}_j+b_j)+\hat{\tau}_k\right\} \qquad (11)$$

$$=\sum_{k=1}^{\infty}\int_0^t[1-F(t-x)]d\left[F^{(k-1)}*B^{(k-1)}(x)\right]$$

$$P_{i0}(t)=\int_0^t p_{00}(t-x)dP\{b(i)\le x\},\qquad i\ge 1 \qquad (12)$$

其中$b^{(i)}$表示有i个顾客开始的忙期长度。由于到达是泊松流，所以类似前面服务顺序的安排，$b^{(i)}$可表示为

$$b^{(i)}=b_1+b_2+\ldots+b_i,\quad i\ge 1 \qquad (13)$$

而且$b_1,b_2,\ldots,b_i$相互独立,有相同分布B(t),所以

证明(续2)

$$P\{b^{(i)}\le t\}=B^{(i)}(t),\quad i\ge 1, t\ge 0 \qquad (14)$$

然后取拉普拉斯变换，经过整理即得式(9)。

同理，当$j\ge 1$时，有

$$P_{ij}(t)=P\{\text{时刻}t\text{处于忙期；}N(t)=j|N(0)=i\}$$

$$i\ge 0 \qquad (15)$$

于是，当$i=0$时，

$$P_{0j}(t)=\sum_{m=1}^{\infty}P\left\{\sum_{k=1}^{m-1}(\hat{\tau}_k+b_k)+\hat{\tau}_m\le t<\sum_{k=1}^{m}(\hat{\tau}_k+b_k);N(t)=j\right\}$$

$$=\sum_{m=1}^{\infty}\int_0^t P\{b_m>t-x;N(t-x)=j\}d\left[F^{(m)}*B^{(m-1)}(x)\right]$$

$$=\sum_{m=1}^{\infty}\int_0^t Q_j(t-x)d\left[F^{(m)}*B^{(m-1)}(x)\right],\qquad j\ge 1 \qquad (16)$$

证明(续3)

当i≥1时，

$$P_{ij}(t) = P\{b^{(i)} > t \geq 0; N(t) = j\} + \int_0^t p_{0j}(t-x)dB^{(i)}(x) \quad (17)$$

类似$Q_j(t)$的讨论，有

$$P\{b^{(i)} > t \geq 0; N(t) = j\} = \sum_{k=1}^{i} Q_{j-(i-k)}(t) * B^{(k-1)}(t) \quad (18)$$

于是

$$P_{ij}(t) = \sum_{k=1}^{i} Q_{j-(i-k)}(t) * B^{(k-1)}(t) + \int_0^t p_{0j}(t-x)dB^{(i)}(x) \quad (19)$$

然后取拉普拉斯变换，经过整理即得式(10)。

定理3

令$p_j = \lim\limits_{t\to\infty} P\{N(t)=j\}$，$j\geqslant 0$，则对任意初始状态，有

1) 当$\rho=\lambda/\mu\geq 1$时，$p_j=0$，$=0,1,2,\ldots\ldots$

2) 当$\rho=\lambda/\mu<1$时，$\{p_j，j\geq 0\}$，且构成概率分布，有

$$p_0 = 1-\rho$$

$$p_j = \frac{1}{a_0}\left\{\lambda(1-\rho)\int_0^\infty \overline{G}(t)\frac{(\lambda t)^{j-1}}{(j-1)!}e^{-\lambda t}dt + \sum_{k=1}^{j-1} p_{j-k}[1-\sum_{i=1}^{k} a_i]\right\}$$

其中当$j\leq 0$时，有 $\sum\limits_{k=1}^{j} = 0$;

$$a_j = \int_0^\infty \frac{(\lambda t)^i}{i!}e^{-\lambda t}dG(t)$$

推论

1. 令$P(z)=\sum_{j=0}^{\infty}p_j z^j$，则当$\rho<1$时，

1) $P(z)=P^+(z)=\dfrac{(1-\rho)(1-z)g(\lambda(1-z))}{g(\lambda(1-z))-z}$，$|z|<1$

2) $\overline{N}=\rho+\dfrac{\lambda^2 E[\chi^2]}{2(1-\rho)}$

$$\overline{N}_q=\frac{\lambda^2 E[\chi^2]}{2(1-\rho)}$$

2. 在M/G/1/∞排队系统中，当$\rho<1$时，

$$p_j=p_j^+,\quad j=0,1,2,\cdots$$

§7.3 等待时间与逗留时间

1. 假定顾客是先到先服务。

令$W_q(t)$与$W(t)$在统计平衡下，顾客的等待时间分布函数和逗留时间分布函数，且它们的拉普拉斯—司梯阶变换为

$$w_q(s)=\int_0^\infty e^{-st}dW_q(t),\qquad w(s)=\int_0^\infty e^{-st}dW(t),$$

定理 对M/G/1/∞排队系统，在ρ<1下，有

$$1)\quad w(s)=\frac{s(1-\rho)g(s)}{s-\lambda[1-g(s)]},\qquad R(s)>0$$

$$2)\quad w_q(s)=\frac{s(1-\rho)}{s-\lambda[1-g(s)]},\qquad R(s)>0$$

其中2)称为扑拉克—辛钦公式； $g(s)=\int_0^\infty e^{-st}dG(t)$。

证明

显然，在统计平衡下，顾客服务完毕离开系统时留在系统中的顾客数应等于在该顾客的逗留时间内到达的顾客数，即

$$p_j^+ = P\{\text{在该顾客的逗留时间内到达}j\text{个}\}$$

$$= \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dW(t), \qquad j = 0,1,2,\cdots$$

于是 $$P^+(z) = \sum_{j=0}^{\infty} z^j p_j^+ = \int_0^\infty e^{-\lambda t} \sum_{j=0}^{\infty} \frac{(\lambda t z)^j}{j!} dW(t)$$

$$= \int_0^\infty e^{-\lambda(1-z)t} dW(t) = w(\lambda(1-z))$$

即 $$\frac{(1-\rho)(1-z)g(\lambda(1-z))}{g(\lambda(1-z)) - z} = w(\lambda(1-z))$$

证明(续)

令s＝λ(1-z)得

$$w(s)=\frac{s(1-\rho)g(s)}{s-\lambda[1-g(s)]}$$

因为逗留时间W等于等待时间W_q加上服务时间χ，即W＝W_q+χ，且W_q与χ相互独立，因此W的概率密度为W_q的概率密度和χ的概率密度的卷积，再利用卷积的拉普拉斯变换等于拉普拉斯变换的乘积，得

$$w(s)=w_q(s)\cdot g(s)$$

即

$$w_q(s)=\frac{s(1-\rho)}{s-\lambda[1-g(s)]}$$

推论1

令 $\hat{G}(t)=\mu\int_0^t[1-G(x)]dx$ 表示服务时间分布G(t)的平衡分布，则等待时间分布函数Wq(t)为

$$W_q(t)=\sum_{n=0}^{\infty}(1-\rho)\rho^n\hat{G}^{(n)}(t),\qquad t\geq 0$$

其中 $\hat{G}^{(n)}(t)$为$\hat{G}(t)$的n重卷积，$n\geq 1$，$\hat{G}^{(0)}(t)=1$，$t\geq 0$。

证明 令 $\hat{g}(s)=\int_0^t e^{-st}d\hat{G}(t)$，则 $\hat{g}(s)=\mu\dfrac{1-g(s)}{s}$，R(s)>0，于是

$$w_q(s)=\frac{1-\rho}{1-\rho\cdot\hat{g}(s)}$$

显然，当ρ＜1时，其模 $|\rho\cdot\hat{g}(s)|<1$ ，于是把$w_q(s)$展开，有

$$W_q(s)=\sum_{n=0}^{\infty}(1-\rho)\rho^n\left[\hat{g}(s)\right]^n$$

利用控制收敛定理，逐项求拉普拉斯－司梯阶的反演变换即得结论。

推论2、推论3

推论2 对M/G/1/∞排队系统，在ρ<1下，有

$$\overline{W}_q = \frac{\lambda E[\chi^2]}{2(1-\rho)}, \qquad \overline{W} = \frac{1}{\mu} + \frac{\lambda E[\chi^2]}{2(1-\rho)}$$

推论3 对M/G/1/∞排队系统，在ρ<1下，Little公式成立。

§7.4 忙期

当顾客到达空闲得服务台时，忙期就开始，直到服务台又一次得空闲时忙期才结束。令b表示从一个顾客开始的忙期长度，且令

$$B(t)=P\{b\le t\},\qquad b(s)=\int_0^{\infty} e^{-st}dB(t)$$

引理1

若 1) $R(s)\geq 0$，$|u|<1$；或

2) $R(s)>0$，$|u|\leq 1$；或

3) $R(s)\geq 0$，$|u|\leq 1$，$\rho=\lambda/\mu>1$，

则方程$z=g(s+\lambda(1-z))$在单位圆$|z|<1$内有唯一解$r(s,u)$，且$r(s,u)$可表示为

$$r(s,u)=\sum_{j=1}^{\infty}\frac{\lambda^{j-1}u^{j}}{j!}\int_0^t e^{-(s+\lambda)x}x^{j-1}dG^{(j)}(x)$$

而且

$$\lim_{u\to 1^-} r(0,u)=\begin{cases}1, & \rho\leq 1\\ \omega<1, & \rho>1\end{cases}$$

$$\lim_{s\to 0^+} r(s,1)=\begin{cases}1, & \rho\leq 1\\ \omega<1, & \rho>1\end{cases}$$

其中，$G^{(j)}(x)$为服务时间分布$G(x)$的j重卷积，$j\geq 1$；$g(s)$为$G(x)$的拉普拉斯-司梯阶变换；ω为方程$z=g(\lambda(1-z))$在(0,1)内的最小非负实根。

定理1

对M/G/1/∞排队系统，若R(s)＞0，则b(s)为方程

$$z = g(s + \lambda(1 - z))$$

在|z|＜1内的唯一解，且

$$B(t) = \sum_{j=1}^{\infty} \int_0^t \frac{(\lambda x)^{j-1}}{j!} e^{-\lambda x} dG^{(j)}(x)$$

若ρ≤1，则B(∞)＝1，此时B(t)为概率分布函数；若ρ＞1，则B(∞)＝ω＜1，此时B(t)不是概率分布函数，且忙期长度为无穷的概率等于1-ω。

定理2

令$D_j(t)$＝P{b≤t，且在忙期内b中服务j个顾客}，

$$d_j(s)=\int_0^\infty e^{-st}dD_j(t),\qquad j\geq 1$$

若R(s)＞0，|μ|＜1，则 $\sum_{j=0}^{\infty} d_j(s)\mu^j = r(s,\mu)$

其中r(s,μ)为方程 $z=ug(s+\lambda(1-z))$ 在|z|＜1内的唯一解，且

$$d_j(s)=\int_0^\infty \frac{(\lambda t)^{j-1}}{j!}e^{-(s+\lambda)t}dG^{(j)}(t)$$

反演，得

$$D_j(t)=\int_0^t \frac{(\lambda x)^{j-1}}{j!}e^{-\lambda x}dG^{(j)}(x)$$

推论

1) 忙期长度b的分布函数为

$$B(t)=\sum_{j=1}^{\infty}D_j(t)=\sum_{j=1}^{\infty}\int_0^t\frac{(\lambda x)^{j-1}}{j!}e^{-\lambda x}dG^{(j)}(x)$$

2) 忙期中服务j个顾客的概率为

$$D_j(\infty)=\int_0^{\infty}\frac{(\lambda x)^{j-1}}{j!}e^{-\lambda x}dG^{(j)}(x)$$

特别地，当$G(x)=1-e^{-\mu x}$时，有

$$D_j(\infty)=\frac{(2j-2)!\mu(\lambda\mu)^{j-1}}{j!(j-1)!(\lambda+\mu)^{2j-1}},\qquad j\geq 1$$

定理3

对M/G/1/∞排队系统，

1. 忙期b的平均长度为

$$E[b]=\begin{cases}\dfrac{1}{\mu-\lambda}, & \rho<1\\ \infty, & \rho\geq 1\end{cases}$$

2. 忙期b内服务完的顾客平均数为

$$E[N_b]=\begin{cases}\dfrac{1}{1-\rho}, & \rho<1\\ \infty, & \rho\geq 1\end{cases}$$

§5　输出过程

令T_n^+表示第n个顾客服务完毕的离去时刻，则T_{n+1}^+ - T_n^+表示离去的时间间隔，$n\geq 1$，于是，在统计平衡下有

$$\lim_{n\to\infty} P\{T_{n+1}^+ - T_n^+ \leq t\} = \begin{cases} (1-\rho)F(t)*G(t)+\rho G(t), & \rho<1 \\ G(t), & \rho\geq 1 \end{cases}$$

而且对M/G/1/∞排队系统，相继离去的间隔时间一般是不独立的，即使在统计平衡下也如此。

下面我们讨论在时间(0,t]内离去顾客的平均数以及其渐近展开问题。对$t\geq 0$，设

$A_i(t)$＝P{时刻t系统处于忙期|N(0)=i}，$i\geq 0$

引理1

令 $a_i^* = \int_0^\infty e^{-st} A_i(t)dt$， $i \geq 0$， $R(s) > 0$， 则

$$a_i^* = \frac{1}{s} - \frac{b^i(s)}{s+\lambda-\lambda b(s)}$$

且

$$\lim_{t\to\infty} A_i(t) = \lim_{s\to 0^+} s \cdot a_i^*(s) = \begin{cases} \rho, & \rho < 1 \\ 1, & \rho \geq 1 \end{cases}$$

定理1

我们用$M_i(t)$表示系统从初始状态N(0)=i出发，

在(0,t]内服务完的平均顾客数。

令 $m_i(s)=\int_0^\infty e^{-st}dM_i(t),\ i\ge 0,\ R(s)>0$，则

$$m_i(s)=\frac{g(s)}{1-g(s)}\left\{1-\frac{s\cdot b^i(s)}{s+\lambda-\lambda b(s)}\right\},\qquad i\ge 0$$

且

$$\lim_{t\to\infty}\frac{M_i(t)}{t}=\lim_{s\to 0^+}s\cdot m_i(s)=\begin{cases}\lambda, & \rho<1\\ \mu, & \rho\ge 1\end{cases}$$

其中$g(s)=\int_0^\infty e^{-st}dG(t)$,b(s)由§7.4中定理1确定.

定理2

令

$$\hat{M}_i(t) = M_i(t) - (\mu t) * A_i(t), \quad i \geq 0,$$

若服务时间分布G(t)是非负的，且$E[\chi^2]<\infty$，则

$$\lim_{t\to\infty}\hat{M}_i(t) = \begin{cases} \rho\{\frac{\mu^2}{2}E[\chi^2]-1\}, & \rho<1 \\ \frac{\mu^2}{2}E[\chi^2]-1, & \rho\geq 1 \end{cases}$$

结论

上面的讨论思想，使我们更能清楚地了解到离去过程的独特结构：在忙期中嵌入一个更新过程，而且平均离去的顾客数与该更新过程的更新函数呈现一种卷积关系，且当t充分大时，可得便于计算$M_i(t)$的近似式：

$$M_i(t) \approx \begin{cases} \lambda t, & \rho < 1 \\ \mu t, & \rho \geq 1 \end{cases}$$

或

$$M_i(t) \approx \begin{cases} \lambda t + \rho\{\frac{\mu^2}{2}E[\chi^2] - 1\}, & \rho < 1 \\ \mu t + \frac{\mu^2}{2}E[\chi^2] - 1, & \rho \geq 1 \end{cases}$$

隐Markov模型

例　设某人在3个装有红白两种颜色的球的盒子中任取一个盒子，然后在此盒子中每次抽取1个球，连续地在同一盒子中按如下方式抽取m次，即各个盒子的内容与抽取方式分别为：

	红球数	白球数	每次抽取方式
盒1	90	10	随机抽取1球，记下颜色后不放回，而放进1个与它不同的球
盒2	50	50	随机抽取1球，记下颜色后放回
盒3	40	60	随机抽取1球，记下颜色后不放回，而放进1个红球

如果某人用上述方法得到一个记录（红，红，红，红，白）（即m＝5），但不告诉我们球出自哪个盒子，我们应如何推测他是从哪个盒子中抽取的观测样本呢？

例(续1)

令

$S_n^{(k)}$＝在第k个盒子（k=1,2,3）中第n次抽取完成后在各盒子中的红球数

那么，在k分别固定为1,2,3时，

$$\{S_n^{(k)}, \ n\geq 0\}$$

分别为马尔可夫链，且其转移概率分别为

例(续2)

$$p_{ij}^{(1)}=\begin{cases}\dfrac{i}{100}, & j=i-1\\ 1-\dfrac{i}{100}, & j=i+1\\ 0, & 其它\end{cases}$$

（逢红，红减1；逢白，红加1）

$$p_{ij}^{(2)}=\begin{cases}1, & j=i\\ 0 & j\neq i\end{cases}$$

（内容总是不变）

$$p_{ij}^{(3)}=\begin{cases}\dfrac{i}{100}, & j=i\\ 1-\dfrac{i}{100}, & j=i+1\\ 0, & 其它\end{cases}$$

（逢红不变；逢白，红加1）

例(续3)

而且初值分别为：$S_0^{(1)}=90$， $S_0^{(2)}=50$， $S_0^{(3)}=40$。于是这3个盒子就分别对应于3个不同的马尔可夫链模型，把这3个模型分别记为λ_1，λ_2，λ_3，并把某人观测到的样本序列中的第n个记为O_n。即令

O_n为抽到的记录列中第n个记录中的白球数

（只能为0或1）

从此例可以看出，在观测出自哪个盒子已知时，状态随机变量序列$\{S_n\}$与某人提供的观测随机变量序列$\{O_n\}$之间的条件概率计算的关系可以直观地写为：

$$S_0,O_1,S_1,\ldots, O_{m-1},S_{m-1}, O_m,S_m$$

其中在前面的一段随机变量序列取定值的条件下，继后的那个随机变量取值的条件概率就完全确定了。

例(续4)

在这3个模型下分别都有：

$P\{S_{n+1} \mid O_1, S_1, \ldots, O_n, S_n, O_{n+1}\} = P\{S_{n+1} \mid S_n\} = P\{S_2 \mid S_1\}$

$P\{O_{n+1} \mid O_1, S_1, \ldots, O_n, S_n\} = P\{O_{n+1} \mid S_n\} = P\{O_2 \mid S_1\}$

于是

$P\{S_0, O_1, S_1, \ldots, O_{m-1}, S_{m-1}, O_m, S_m\}$

$= P\{S_0\}P\{O_1|S_0\}P\{S_1|S_0\}P\{O_2|S_1\}\ldots P\{O_m|S_{m-1}\}P\{S_m|S_{m-1}\}$

具体地，我们有

在模型λ_1下（把取到的球换色）

$P\{(O_1, O_2, O_3, O_4, O_5) = (0, 0, 0, 0, 1) \mid \lambda_1\}$

$= 0.9 \times 0.89 \times 0.88 \times 0.87 \times 0.14 \approx 0.086$

例(续5)

在模型λ_2下（球的内容不变）

$$P\{(O_1, O_2, O_3, O_4, O_5) = (0, 0, 0, 0, 1) \mid \lambda_1\}$$
$$= (0.5)^5 \approx 0.031$$

在模型λ_3下（取红不变，取白换红）

$$P\{(O_1, O_2, O_3, O_4, O_5) = (0, 0, 0, 0, 1) \mid \lambda_1\}$$
$$= (0.4)^4 \times 0.6 \approx 0.015$$

再用贝叶斯公式分别得到(即取上面3个概率的归一化值)

$$P\{\lambda_1 \mid (O_1, O_2, O_3, O_4, O_5) = (0, 0, 0, 0, 1)\} \approx 0.64$$
$$P\{\lambda_2 \mid (O_1, O_2, O_3, O_4, O_5) = (0, 0, 0, 0, 1)\} \approx 0.24$$
$$P\{\lambda_3 \mid (O_1, O_2, O_3, O_4, O_5) = (0, 0, 0, 0, 1)\} \approx 0.12$$

可见从第一盒抽出样本（红，红，红，红，白）的概率要比从其它两盒中抽出该样本的概率要大得多。

隐马尔可夫模型的描述

定义1 设$\{X_n, n\geq 1\}$是取值于有限状态$\{1,2,\ldots,L\}$的随机变量序列，称为**状态链**，X_1的分布称为其初始分布，记为μ。假定X_n是齐次马尔可夫链，记

$$a_{ij} = P\{X_{n+1} = j \mid X_n = i\}, \quad A = (a_{ij})_{i,j\leq L}$$

为它的转移矩阵。假定X_n的取值、初始分布μ与转移矩阵A都不能测量得到。而能测量到的是另一个与它有联系的，且可以观测到的一个取值于有限集$\{v_1, v_2, \ldots, v_M\}$的随机变量序列$Y_n$，称为**观测链**，它们合起来还要满足如下的**隐马尔可夫条件（HMM条件）**

$$P\{Y=y, X=x\} = \mu_{i_1} b_{i_1 y_1} a_{i_1 i_2} \cdots b_{i_{N-1} y_{N-1}} a_{i_{N-1} i_N} b_{i_N y_N} \qquad (*)$$

隐马尔可夫模型的描述(续1)

其中N为样本观测的时间长度，而$X=(X_1,X_2,\ldots,X_N)$，$Y=(Y_1,Y_2,\ldots,Y_N)$，$x=(x_1,x_2,\ldots,x_N)$，$y=(y_1,y_2,\ldots,y_N)$，$y_n\in\{v_1,v_2,\ldots v_M\}$，$1\le i_n\le L$，$1\le n\le N$，初始分布为$\mu=(\mu_1,\mu_2,\ldots,\mu_L)$。由未知状态链与观测到的观测链一起（$X_n,Y_n$），就构成了隐马尔可夫模型，这里“隐”的含义是说状态链是隐藏起来的。

隐马尔可夫模型的基本假定是：参数向量μ、参数矩阵A与$B=(b_{ik})_{L\times M}$($i\in\{1,2,\ldots,L\}$，$k\in\{v_1,v_2,\ldots v_M\}$)都是未知的，将它们合记为参数组$\lambda=(\mu,A,B)$。后者完全确定了状态链与观测链的联合统计规律。所以，我们通常用λ表示一个隐马尔可夫模型，并称之为隐马尔可夫模型λ（更确切地为隐马尔可夫链）。

隐马尔可夫模型的描述(续2)

在上述例子中，3个不同的模型就对应了3个不同的参数组。

只要令

$$X_n = S_n，\ Y_n = O_{n+1}$$

它们满足HMM条件，因而纳入了隐马尔可夫模型的框架。

(*)式是(X,Y)的联合分布通过参数表达的形式，它是计算各种边缘概率与条件概率的出发点。

而HMM的含义是：状态链与观测链的联合分布是由一系列简单转移与条件概率的乘积表达的。

隐马尔可夫模型的等价表述

HMM条件等价于：

对任意的$i \in \{1,2,\ldots,L\}$以及$k \in \{v_1,v_2,\ldots v_M\}$，有

$$P\{Y_n=v_k|X_n=i,Y_{n-1}=i_{n-1},X_{n-1}=v_{kn-1},\ldots,Y_1=i_1,X_1=v_{k1}\}$$

$$=P\{Y_n=v_k|X_n=i\}=b_{ik}$$

$$P\{X_{n+1}=j|X_n=i,Y_n=v_{kn},X_{n-1}=i_{n-1},Y_{n-1}=v_{kn-1},\ldots,X_1=i_1,Y_1=v_{k1}\}$$

$$=P\{X_{n+1}=j|X_n=i\}=a_{ij}$$

这两个等式只需要利用条件概率的定义就容易证明。

它们的直观含义就是：Y_n与X_{n+1}相对于历史条件的统计规律只与时间上最接近的X_n有关，而与其它更早的历史无关。

在应用中研究隐马尔可夫模型的主要方面

1. 从一段观测序列$\{Y_k, k \le m\}$及已知模型$\lambda = (\mu, A, B)$出发，估计Xn的最佳值，称为解码问题。这是状态估计问题。
2. 从一段观测序列出发，估计模型参数组$\lambda = (\mu, A, B)$，称为学习问题。就是参数估计问题。
3. 对于一个特定的观测链$\{Y_k, k \le m\}$，已知它可能是由已经学习好的若干模型之一所得的观测，要决定此究竟是得自其中那一个模型，这称为识别问题。就是分类问题。

解码问题

已知模型λ与观测Y＝y时，状态X的估计

1. 出现当前的观测的概率P{Y=y|λ}的计算

我们仍旧沿用记号

$X=(X_1,X_2,\ldots,X_N)$，$Y=(Y_1,Y_2,\ldots,Y_N)$，

$x=(x_1,x_2,\ldots,x_N)$，$y=(y_1,y_2,\ldots,y_N)$，

$y_n\in\{v_1,v_2,\ldots v_M\}$，$1\le i_n\le L$，$1\le n\le N$，

初始分布$\mu=(\mu_1,\mu_2,\ldots,\mu_L)$。

由(*)式，利用条件概率的性质易得

$$P\{Y=y|X=(i_1,i_2,\ldots,i_N),\lambda\}=b_{i_1y_1}b_{i_2y_2}\cdots b_{i_Ny_N}$$

$$P\{Y=y|X=x,\lambda\}=\mu_{i_1}b_{i_1y_1}a_{i_1i_2}\cdots b_{i_{N-1}y_{N-1}}a_{i_{N-1}i_N}b_{i_Ny_N}$$

$$P\{Y=y|\lambda\}=\sum_{i_1,\cdots i_N}\mu_{i_1}b_{i_1y_1}a_{i_1i_2}\cdots b_{i_{N-1}y_{N-1}}a_{i_{N-1}i_N}b_{i_Ny_N}$$

解码问题

对于$1\le n\le N$及观测样值$Y=y$，记（因为观测样值y是固定的，所以下面我们将在足标把它略去）

$\alpha_n(i)=P\{Y_1=y_1, Y_2=y_2, \ldots, Y_n=y_n, X_n=i|\lambda\}$ (依赖y)

则在模型λ给定下，关于观测资料$(y_1, y_2, \ldots, y_n)$的长度n，我们有递推公式(称为**向前递推公式**或**向前算法**)

$$\alpha_{n+1}(i)=\sum_j \alpha_n(j)a_{ji}b_{iy_{n+1}}$$

由此得到：

(1) 基于$\alpha_n(i)$的向前递推公式计算$P\{Y=y|\lambda\}$的步骤

计算初值 $\alpha_1(i)=\mu_i b_{iy_1}$

用递推公式 $\alpha_{n+1}(i)=\sum_j \alpha_n(j)a_{ji}b_{iy_{n+1}}$

最后得到结论 $P\{Y=y|\lambda\}=\sum_i \alpha_N(i)$

另一种途径

计算P{Y=y|λ}也可以通过另一种途径，为此记(观测样值y也是固定的，我们也把它在足标把它略去）

$$\beta_n(i)=P\{Y_{n+1}=y_{n+1},Y_{n+2}=y_{n+2},\ldots,Y_N=y_N|X_n=i,\lambda\}$$

则在模型λ给定下，关于观测资料$(y_1, y_2, \ldots, y_n)$的长度n，我们有递推公式(称为向后递推公式或向后算法)

$$\beta_n(i)=\sum_j \beta_{n+1}(j)a_{ij}b_{jy_{n+1}}$$

这就得到计算P{Y=y|λ}的另一个算法。

(2)基于$\beta_n(i)$的向后递推公式计算

$P\{Y=y|\lambda\}$的步骤

定义 $\beta_N(i)=1$

用递推公式 $\beta_n(i)=\sum_j \alpha_{n+1}(j)a_{ij}b_{jy_{n+1}}$

最后得到结论 $P\{Y=y|\lambda\}=\sum_i \beta_0(i)\mu_i b_{iy_1}$

2. 解码问题——已知模型λ与观测Y＝y时，状态X的估计

令

$$\gamma_N(i) = P\{X_n = i \mid Y = y, \gamma\}$$

那么

$$\gamma_N(i) = \frac{P\{Y = y, X_n = i \mid \gamma\}}{\sum_i P\{Y = y, X_n = i \mid \gamma\}} = \frac{\alpha_n(i)\beta_n(i)}{\sum_i \alpha_n(i)\beta_n(i)}$$

本讲主要内容

- 一般服务的M/G/1/∞排队系统
  - 对长
  - 等待时间与逗留时间
  - 忙期
  - 输出过程
- 隐马尔科夫模型简介

下一讲内容预告

- 例子
- 排队论部分习题解答
- 复习考试