文章编号：1001-2486（2002）03-0027-05

经马氏修正的 Poisson 过程的极大似然估计[※]

王春玲，李 兵，葛正坤
（国防科技大学理学院，湖南 长沙 410073）

摘 要：近年来，隐马氏模型成为研究相依随机变量的一个十分有用的工具。实际应用过程中的一个很重要的问题是如何对隐马氏模型的参数进行估计。将一类连续时间隐马氏模型的问题转化为离散时间隐马氏模型的问题，给出了具体的隐马氏模型——经马氏修正的 Poisson 过程的极大似然估计及其算法。此类过程被广泛用来对复杂电信网络的交通流进行建模。

关键词：隐马氏模型；前向算法；后向算法；经马氏修正的 Poisson 过程
中图分类号：O211.62；O24 **文献标识码**：A

A Maximum Likelihood Estimation for Markov-modulated Poisson Processes

WANG Chun-ling，LI Bing，GE Zheng-kun
（College of Science，National Univ. of Defense Technology，Changsha 410073，China）

Abstract：During the last decade，Hidden Markov Models（HMMs）have become a widespread tool for modeling sequence of dependent variables. Parameter estimation of HMMs is most important in actual application. By changing continued-time HMMs into discrete-time HMMs，we consider maximum likelihood estimation for a special HMMs which is called Markov-modulated Poisson processes. Such processes have been proposed for modeling traffic streams in complex telecommunication networks.

Key words：HMMs；forwards-algorithms；backwards-algorithms；Markov-modulated Poisson processes

HMM 是一种简单的数学模型，它的包容性大，内涵广。它也是一种不完全数据的统计模型。这种模型在应用中的弹性很大，因而有很大的适应性。Baum 和 Petrie 最先在 $\{Y_k\}$ 取值于有限集合的情况下对 HMMs 进行研究。近年来许多文献讨论了有关离散时间的 HMM 的具体问题，而有关连续时间 HMM 的文献不是很多。我们将一类连续时间隐马氏模型的问题转化为离散时间隐马氏模型的问题，给出了具体的隐马氏模型——经马氏修正的 Poisson 过程的极大似然估计及其算法。此类模型被广泛用来研究复杂通信网络的通信流模型建模。

1 背景

隐马氏模型（HMM）$\{X_k, Y_k\}$ 是一满足下列条件的离散时间的随机过程：

（1）$\{X_k\}$ 是一有限状态的马氏链；

（2）在 $\{X_k\}$ 的条件下，$\{Y_k\}$ 是条件独立的随机过程，Y_k 的条件分布只依赖于 X_k。

令 $\{X_t\}$ 是一连续时间状态取值于空间 $S=(1,\cdots,K)$ 的马氏过程，且有无穷小转移概率矩阵 $Q=\{q_{ij}\}$。令 $\lambda=(\lambda_1,\cdots,\lambda_K)$（其中 $\lambda_i\geqslant 0$，为记号方便我们也记 $\lambda_i=\lambda(i)$，$i=1,2,\cdots,K$），$\{N(t)\}$ 为非齐次 Poisson 过程，有随机强度函数 $\{\lambda(X(t))\}$，即给定 $\{\lambda(X(t))\}$，$\{N(t)\}$ 为条件独立增量且 $N(t+s)-N(t)$ 是均值为 $\int_t^{t+s}\lambda(X(u))\mathrm{d}u$ 的 Poisson 分布，这样的过程称为经马氏修正的 Poisson 过程（Markov-modulated Poisson processes）。模型参数是 Q，λ。

※ 收稿日期：2001-11-29
基金项目：国防科技大学预研项目（JC00-02-011）
作者简介：王春玲（1977—），女，助教，硕士。

为了与离散时间的隐马氏模型联系起来，令 $\tau_0=0$，τ_k 为 $\{N(t)\}$ 中第 k 个事件发生的时刻，$Y_k=\tau_k-\tau_{k-1}$ 且 $X_k=X(\tau_k)$，则不妨认为 $\{(X_k,Y_k)\}$ 为隐马氏链模型（事实上，给定 $\{X_k\}$ 时 Y_n 的分布可能依赖 X_{n-1} 和 X_n，此时用 $\{X'_k\}=\{(X_{k-1},X_k)\}$ 代替 $\{X_k\}$，则 $\{(X'_k,Y_k)\}$ 是隐马氏模型）。令参数

$$\phi=(\lambda_1,\lambda_2,\cdots,\lambda_K,q_{11},\cdots,q_{1K},\cdots,q_{K1},\cdots,q_{KK})$$

$$m=1,2,\cdots,K$$

本文讨论的 $\{X_t\}$ 是状态 i 全为逗留态（即 $0<q_i<\infty$，$q_i=-q_{ii}$），轨道右连续的马氏过程。由文献［1］可知，$\{X_t\}$ 几乎全部轨道是阶梯函数，在每个跳点上，由 m 出发的条件下，跳到 k 去的概率为 q_{mk}/q_m。不妨假设时刻 t_i 也为过程 $\{X_t\}$ 的跳点，且 $\{X_t\}$ 在 τ_i 和 τ_{i-1} 之间没有跳点。在没有特别指明的情况下，下面出现的 Q 及 q_i 与本段中的意义一样。因为 $\{X_t\}$ 为轨道右连续，所以其转移概率矩阵 $P(t)$ 为标准的[2]，即 $\lim\limits_{t\to 0}P(t)=I$，$I$ 是单位阵。

引理[2] 对有限状态马氏过程，若其转移阵为 P 标准，则它的 Q—矩阵保守。

注：Q—矩阵保守指的是 $\sum\limits_{j\neq i}q_{ij}=q_i<\infty$。

定理1 设 $X=\{X_t;t\in R^+\}$ 是 Q 保守的，无吸收态的轨道右连续的马氏链。令

$$\tau_0=0 \text{ 与 } \tau_k=\inf\{t>\tau_{k-1};X_t\neq X_{\tau_{k-1}}\}$$

再令

$$\eta_n=X_{\tau_n}$$

则 $\eta_n=X_{\tau_n}$ 是马氏链（称为 X 的嵌入链），而且它的转移概率矩阵是

$$P=(p_{ij})=((1-\delta_{ij})q_{ij}/q_i)$$

其中 $\delta_{ij}=\begin{cases}1 & i=j\\0 & i\neq j\end{cases}$，$Q=(q_{ij})$ 是 $X=\{X_t;t\in R^+\}$ 的 Q—矩阵。

证明略[2]。

定理2[2] 设马氏链 $X=\{X_t;t\in R^+\}$ 的轨道右连续（也就是 $P(\omega:\lim\limits_{t\to S}x_t(\omega)=x_S(\omega),\forall S)=1$），且 $0\leqslant q_i<\infty$，则有 $P(\tau>t\mid X_0=i)=e^{-q_it}$（$\{X_t\}$ 在状态 i 的逗留时间 τ 服从指数分布）。

证明略。

根据定理2可知 $\lambda_i=\lambda(i)\equiv q_i=-q_{ii}$ $(1\leqslant i\leqslant K)$，这样原参数组

$$\phi=(\lambda_1,\lambda_2,\cdots,\lambda_K,q_{11},\cdots,q_{1K},\cdots,q_{K1},\cdots,q_{KK})$$

中的 $\lambda_1,\cdots,\lambda_K$ 可省去为 $\phi=(q_{11},\cdots,q_{1K},\cdots,q_{K1},\cdots,q_{KK})$。

由文献［1］可知，在给定 X_k 的条件下 Y_k 的密度函数为：

$$P(Y_n=y_n\mid x(\tau_n))=\lambda(x(\tau_n))\exp\{-\lambda(x(\tau_n))y_n\}$$

且知在给定 $\{X_k\}_{k=1}^K$ 的条件下，$\{Y_k\}_{k=1}^K$ 的样本函数密度为：

$$P(Y=y\mid X=x,\phi)=\prod_{i=1}^n\lambda(x(\tau_i)\exp\{-\lambda(x(\tau_i))y_i\}$$

设 $x=(x_{\tau_i},\cdots,x_{\tau_K})$，$y=(y_1,\cdots,y_K)$，按 HMCM 的意义，易知，

$$P(X=x\mid\phi)=\pi_{x_{\tau_1}}p_{x_{\tau_1}x_{\tau_2}}\cdots p_{x_{\tau_{n-1}}x_{\tau_n}}$$

$$P(Y=y\mid X=x,\phi)=\prod_{i=1}^n\lambda(x(\tau_i))\exp\{-\lambda(x(\tau_i))y_i\}$$

$$L(Y=y\mid\phi)=P(Y=y\mid\phi)=\sum_x\pi_{x_{\tau_1}}\prod_{i=2}^n\lambda(x(\tau_i)p_{x_{\tau_{i-1}}x_{\tau_i}})\exp\{-\lambda(x(\tau_i))y_i\}$$

由引理[2]及定理1知：$p_{x_{\tau_{i-1}}x_{\tau_i}}=P(X_{\tau_i}=x_{\tau_i}\mid X_{\tau_{i-1}}=x_{\tau_{i-1}})=(1-\delta_{x_{\tau_{i-1}}x_{\tau_i}})q_{x_{\tau_{i-1}}x_{\tau_i}}/q_{x_{\tau_i}}$，$x_{\tau_i}\in S$，$\pi_{x_{\tau_1}}\triangleq(p(X_{\tau_1}=x_{\tau_1}))$。

2 新模型与原模型的关系

原模型中的问题是根据观测序列 $\{N(t)\}$ 估计隐藏的时间连续的序列 $\{X_t\}$，本文假设 $\{X_t\}$

是一轨道右连续的过程，其几乎全部轨道是阶梯函数。在新模型中估计的是 $\{X_t\}$ 的跳点处的状态，由此可以估计出过程 $\{X_t\}$。

3 参数估计

根据似然估计的原则用归纳重估计的方法来得到一个较好的有效估计。实现它的主要程序是：首先给定一个初始参数 ϕ 且读入样本的观测序列 $y=\{y_1,\cdots,y_N\}$，在此基础上得到新参数 ϕ'，使得 $L(Y=y\mid\phi')\geqslant L(Y=y\mid\phi)$，当且仅当 ϕ 为 L 的临界点时等号成立；再在当前参数下重新估计参数，如此循环直至满足某个特殊的收敛条件为止。本文中模型的观测样本是过程 $N(t)$ 点事件的发生时间。利用将观测序列分为容量固定且为 N 的样本组，将这些样本组看做是独立的，然后利用这些样本组来对参数进行修正。本文中的 N 是固定的正整数。

重估计方法可以付诸实施的原因是可以用前向或后向递推公式来计算 L。

令 $\alpha_k(i)=P(Y_1=y_1,Y_2=y_2,\cdots,Y_k=y_k,X_{\tau_k}=i\mid\phi)$

$$\begin{aligned}\alpha_{k+1}(i)&=P(Y_1=y_1,Y_2=y_2,\cdots,Y_{k+1}=y_{k+1},X_{\tau_{k+1}}=i\mid\phi)\\&=\sum_j P(Y_1=y_1,Y_2=y_2,\cdots,Y_{k+1}=y_{k+1},X_{\tau_k}=j,X_{\tau_{k+1}}=i\mid\phi)\\&=\sum_j\alpha_k(j)a_{ji}b_i(y_{k+1})\end{aligned}$$

$$\beta_k(i)=P(Y_{k+1}=y_{k+1},Y_{k+2}=y_{k+2},\cdots,Y_n=y_n\mid X_{\tau_k}=i,\phi)$$

$$\begin{aligned}\beta_k(i)&=\sum_j P(Y_{k+1}=y_{k+1},Y_{k+2}=y_{k+2},\cdots,Y_n=y_n,X_{\tau_{k+1}}=j\mid X_{\tau_k}=i,\phi)\\&=\sum_j\beta_{k+1}(j)a_{ij}b_j(y_{k+1})\end{aligned}$$

其中 $b_j(y_{k+1})=P(Y_{k+1}=y_{k+1}\mid X_{k+1}=j)$，下文出现与此类似的符号不再一一解释。

前向算法：

(1) 初始化： $\alpha_1(i)=\pi_i b_i(y_1)\quad i=1,2,\cdots,K$

(2) 迭代： $$\alpha_{k+1}(i)=\sum_{i=1}^{K}\alpha_k(j)a_{ji}b_i(y_{k+1})\quad 1\leqslant i\leqslant K,k=1,2,\cdots,n\tag{1}$$

(3) 结果： $L(Y=y\mid\phi)=\sum_{i=1}^{K}\alpha_n(i)$

后向算法：

(1) 设置终值： $\beta_N(i)=1$

(2) 用递推公式： $$\beta_n(i)=\sum_j\beta_{n+1}(j)a_{ij}b_{jy_{n+1}}\tag{2}$$

(3) 得结论： $P(Y=y\mid\phi)=\sum_i\beta_1(i)\pi_i b_{iy_1}$

下面描述的参数重估计是以 $\alpha_k(j),\beta_k(j),(j=1,\cdots,K;k=1,\cdots,N)$ 为变量的简单函数。令 $Q(\phi,\phi')=\sum_x P(x,y\mid\phi)\log(P(x,y\mid\phi'))$，此式中 ϕ 是一常数，ϕ' 是变量。其中

$$\phi'=(q_{11}',\cdots,q_{1K}',\cdots,q_{K1}',\cdots,q_{KK}')$$

下文出现的与此类似的符号意义与此雷同。

容易验证

$$\begin{aligned}&Q(\phi,\phi')-Q(\phi',\phi)\\&=\sum_x P(x,y\mid\phi)\log\frac{P(x,y\mid\phi')}{P(x,y\mid\phi)}\\&\leqslant\sum_x P(x,y\mid\phi)\left(\frac{P(x,y\mid\phi')}{P(x,y\mid\phi)}-1\right)\\&=P(y\mid\phi')-P(y\mid\phi)\end{aligned}$$

因此，由 $Q(\phi,\phi')\geqslant Q(\phi,\phi)$，可以得到 $P(y\mid\phi')\geqslant P(y\mid\phi)$。以上事实说明，如果给定参数 ϕ，找出使

$Q(\phi,\phi')>Q(\phi,\phi)$的参数$\phi'$，就有$P(y|\phi')>P(y|\phi)$，即找到了更符合样本的参数，从而将对$P(y|\phi')$的优化问题转化为对$Q(\phi,\phi')$关于$\phi'$的约束优化问题。

对于本文中的模型，

$$P(x,y|\phi')=\pi'(x_{\tau_1})\sum_{i=2}^{N}p'_{x_{\tau_{i-1}}x_{\tau_i}}b'x_{\tau_i}(y_i)$$

$$\log P(x,y|\phi')=\log\pi'(x_{\tau_1})+\sum_{i=2}^{N}(\log p'_{x_{\tau_{i-1}}x_{\tau_i}}+\log b'x_{\tau_i}(y_i))$$

$$=\log\pi'(x_{\tau_1})+\sum_{i=2}^{n}(\log p'_{x_{\tau_{i-1}}x_{\tau_i}}+\log\lambda'(x(\tau_i))-(\lambda'(x(\tau_i))y_i))$$

$$Q(\phi,\phi')=\sum_x P(x,y|\phi)\ \{\log\pi'(x_{\tau_1})+\sum_{i=2}^{N}(\log p'_{x_{\tau_{i-1}}x_{\tau_i}}+\log\lambda'(x(\tau_i))-(\lambda'(x(\tau_i))y_i))\}$$

$$=\sum_x P(x,y|\phi)\ \{\log\pi'(x_{\tau_1})+\sum_{i=2}^{N}(\log q'_{x_{\tau_{i-1}}x_{\tau_i}}/q'_{x_{\tau_{i-1}}})+\log q'_{x(\tau_i)}-q'_{x(\tau_i)}y_i)\}$$

我们对$Q(\phi,\phi')$作为ϕ'的函数定义参数重估计，ϕ为目前最新估计。由于下面的定理，用此方法获得的重估计序列是单调上升的直到λ是似然函数的一个临界点。

定理3 参数ϕ是似然函数$P(Y|\phi)$的临界点的充分条件是目前的参数ϕ与参数重估计过程中新的参数ϕ'相同。

证明 若ϕ是似然函数$P(Y|\phi)$的临界点的定义为$\nabla_\phi P(Y|\phi))=0$，∇_ϕ表示梯度向量，而且下式成立，

$$\begin{aligned}\nabla_\phi P(Y|\phi)&=\nabla_\phi\sum_x P(Y,X=x|\phi)\\&=\sum_x\nabla_\phi P(Y,X=x|\phi)\\&=\sum_x P(Y,X=x|\phi)\nabla_\phi\log P(Y,X=x|\phi)\\&=\nabla_{\phi'}Q(\phi,\phi')|_{\phi'=\phi}\end{aligned}$$

即，$\nabla_\phi P(Y|\phi)=0$当且仅当$\nabla_{\phi'}Q(\phi,\phi')|_{\phi'=\phi}=0$。

我们对$Q(\phi,\phi')$作为ϕ'的函数定义参数重估计，ϕ为目前最新估计。由上面的定理2知，用下述方法获得的重估计序列是单调上升的直到ϕ是似然函数的一个临界点。由定理1知：$\sum_{k\neq m}(q'_{mk}/q'_m)=1$，此式在Lagrange乘子法中用到。

下面我们具体地给出重估计的公式：

(i)通过下面的公式求解$q'_{mk}/q'_m(m,k=1,\cdots,K,m\neq k)$

$$0=\frac{\partial}{\partial(q'_{mk}/q'_m)}\left\{Q(\phi,\phi')-\theta\Big(\sum_{k\neq m}(q'_{mk}/q'_m)-1\Big)\right\}\tag{$*$}$$

当$k\neq m$时，($*$)式为

$$0=\sum_x P(Y,X=x|\phi)\sum_{i\in I_{mk}(x)}\frac{1}{(q'_{mk}/q'_m)}-\theta\tag{3}$$

其中，θ是Lagrange乘子且$I_{mk}(x)$是集合$\{i:x_{\tau_{i-1}}=m,x_{\tau_i}=k\}$。

将(3)式中被加数的次序改变就得到下式，

$$(1/(q'_{mk}/q'_m))\sum_{i=1}^{N}\sum_{x\in X_{mk(i)}}P(Y,X=x|\phi)-\theta/q'_m=0\tag{4}$$

其中，$X_{mk}(i,\omega)=\{x=(x_{\tau_1}(\omega),\cdots,x_{\tau_N}(\omega)):x_{\tau_{i-1}}=m,x_{\tau_i}=k\}$，它是时刻$t_{i-1}$状态为$m$且时刻$t_i$状态为$k$的路径的集合。可以将(4)式简化为：

$$(1/(q'_{mk}/q'_m))\sum_{i=1}^{N}P(Y,x_{\tau_{i-1}}=m,x_{\tau_i}=k)-\theta=0\tag{5}$$

在(5)式的两边同乘以 q'_{mk}/q'_m，然后在所有不等于 m 的 k 处求和得到下式，

$$\theta=\sum_{i=1}^{N}P(Y,x_{\tau_{i-1}}=m\mid\phi) \tag{6}$$

则由(6)式可得

$$\theta=\sum_{i=1}^{N}P(Y,x_{\tau_{i-1}}=m\mid\phi)=\sum_{i=1}^{n}\alpha_{\tau_{i-1}}(m)\beta_{\tau_{i-1}}(m) \tag{7}$$

将(7)式代入(5)式得

$$q'_{mk}/q'_m=\theta^{-1}\sum_{i=1}^{N}P(Y,x_{\tau_{i-1}}=m,x_{\tau_i}=k) \tag{8}$$

(ii) $q'_m,m=1,\cdots,K$ 为下式的解

$$0=\frac{\partial Q(\phi,\phi')}{\partial q'_m}=\sum_{x}P(Y,X=x\mid\phi)\sum_{i\in I_m(x)}\left\{\frac{1}{q'_m}-y_i\right\} \tag{9}$$

其中，$I_m(x)=\{i:x_{\tau_i}=m\}$。改变和式中被加数的次序可将(9)式简化为下式：

$$0=\frac{1}{q'_m}\sum_{i=1}^{N}\sum_{x\in X_m(i)}P(Y,X=x\mid\phi)-\sum_{i=1}^{N}y_i\sum_{x\in X_m(i)}P(Y,X=x\mid\phi) \tag{10}$$

$$\Rightarrow 0=\frac{1}{q'_m}\sum_{i=1}^{N}P(Y,x_{\tau_i}=m\mid\phi)-\sum_{i=1}^{N}y_iP(Y,x_{\tau_i}=m\mid\phi)$$

其中，$X_m(i)=\{x:x_{\tau_i}=m\}$，它是在时刻 τ_{i-1} 状态为 m 的路径的集合。由此可得，

$$q'_m=\left(\sum_{i=1}^{N}y_iP(Y,x_{\tau_1}=m\mid\phi)\right)^{-1}\sum_{i=1}^{N}P(Y,X_{\tau_1}=m\mid\phi) \tag{11}$$

其中，$P(Y,X_{\tau_i}=m\mid\phi)=\alpha_i(m)\beta_i(m)$。

到此为止，在重估计程序中根据目前参数 $q_m,m=1,\cdots,K$ 及 q_{mk}/q_m 给出了 $q'_m,m=1,\cdots,K$ 和 q'_{mk}/q'_m。由此可以根据目前 ϕ 参数组给出新的参数组 ϕ'。

4 小结

由上可知，重估计可由以下几个步骤来完成，其中可得到的信息是观测量 Y，ϕ' 为最新估计。

(1) 对于 $i=1$，…，N，$m=1$，…，K，α_i (m) 用前向递推公式 (1) 式来计算且将其储存在 $N\times K$ 维向量组中。

(2) 对于 $i=1$，…，N，$m=1$，…，K，β_i (m) 用后向递推公式 (2) 式来计算。

(3) q'_m，q'_{mk}/q'_m ($m\neq k$，$m=1$，2，…，K) 的值分别由 (11) 和 (8) 式累积得到。为了便利起见，β_i (m)，$m=1$，…，K 的 K 个值储存到后向递推公式的下一步中。

(4) 重复进行上述步骤，直到满足某一收敛条件为止。

参考文献：

[1] 邓永录，梁之舜．随机点过程及其应用［M］．北京：科学出版社，1992.

[2] 钱敏平，龚光鲁．随机过程论（第二版）［M］．北京：北京大学出版社，1997.

[3] Baum L E，Eagon J A. An Inequality with Applications to Statistical Estimation for Probabilistic Functions of Markov Processes and to a Model for Ecology [J]. Amer. Math. Soc. Bull.，1967，73：360-363.

[4] Baum L E. An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes [J]. Inequalities，1972，Ⅲ：1-8.

[5] Forney G D. The Viterbi Algorithm [J]．IEEE proc.，1973，61：266-278.

[6] Kullback S，Leibler R A. On Information and Sufficiency [J]．Ann.Math.Stat.，1951，22：79-86.

[7] Chang R W，Hancock T C. On Receiver Structures for Channels Having Memory [J]．IEEE Trans. Inform. Theory，1966，IT-12：464-468.