GENERAL BIRTH-DEATH PROCESS AND SOME OF THEIR EM (EXPETATION-MAXIMATION) ALGORITHM

Islam Haliti, Lazim Kamberi, Ylldrita Seferi
State University of Tetovo, 1200, Tetovo Macedonia
islam.haliti@unite.edu.mk; lazim.kamberi@unite.edu.mk; ylldrita_mk@hotmail.com

ABSTRACT

Birth and death process is a continuous-time Marcov chain, that models a nonnegative integer of particles in a system. The aim of this paper is to evaluate finite-time transition probabilities for estimation of arbitrary birth and death rates in general BDPs. We first obtain expressions for Laplasce transforms of these transition probabilities. It is often necessary to evaluate finite-time transition probabilities. For estimation we use EM (expectation-maximization) algorithms including E-steps by giving four examples like: simple and generalized BDPs, linear BDPs via immigration, Moran model and Regression of count data. In the end we give control of numerical error of transition probabilities.

Key words: *General Birth-death process, transition probabilities, EM algorithms, E-Step , numerical error*

1. Introduction

The continuous time birth and death Marcov chain $X(t)$ has either a finite or infinite state space $\{0,1,2,....,N\}$ or $\{1,2,...\}$.

Assume the infinitesimal transition probabilities of a general birth and death process satisfy.

$$P_{k+j,k}(\Delta t) = P(\{\Delta X(t) = j \mid X(t) = k\}) = \begin{cases} \lambda_k \Delta t + o(\Delta t) & j = 1 \\ \mu_k \Delta t + o(\Delta t) & j = -1 \\ 1-(\lambda_k + \mu_k)\Delta t + o(\Delta t) & j = 0 \\ o(\Delta t) & j \neq 1,-1,0 \end{cases} \quad ...(1)$$

for Δ sufficiently small where $\lambda_k \geq 0, \mu_k \geq 0$ for $k = 1,2...$ and $\mu_0 = 0$ and $P_{kj}(0) = \delta_{kj}$. It is often the case that $\lambda_0 = 0$ also , except, for example when there is immigration.

If the state space is finite, then the initial transition matrix $P(0) = (P_{kj}(0)) = I$ (identity matrix).

In a small interval of time Δt, at most one change in state can occur, either a birth or a death. If the population size is k, and a birth occurs, then $k \rightarrow k+1$, but if a death occurs then $k \rightarrow k-1$.

We consider the transition probability $P_{jk}(t+\Delta t)$

$$P_{jk}(t+\Delta t) = P_{j-1,k}(t)[\lambda_{j-1}\Delta t + o(\Delta t)] + P_{j+1,k}(t)[\mu_{j+1}\Delta t + o(\Delta t)] + P_{jk}(t)[1-(\lambda_j + \mu_j)\Delta t + o(\Delta t)]$$

$$+ \sum_{i \neq -1,0,1} P_{j+i,k}(t) o(\Delta t) = P_{j-1,k}(t)\lambda_{j-1}\Delta t + P_{j+1,k}(t)\mu_{j+1}\Delta t + P_{jk}(t)[1-(\lambda_j + \mu_j)\Delta t + o(\Delta t)] \quad ...(2)$$

In the case of finite state where $j = N$ then :

$$P_{Nk}(t+\Delta t) = P_{N-1,k}(t)\lambda_{N-1}\Delta t + P_{Nk}(t)(1-\mu_N \Delta t) + o(\Delta t)$$ and then $j = 0$

$$P_{0k}(t+\Delta t) = P_{1k}\mu_1 \Delta t + P_{0k}(t)[1-\lambda_0 \Delta t] + o(\Delta t)$$

dividing by Δt and taking the limit as $\Delta t \to 0$ …(3)

Kolmogorov differential equations are obtained for the general birth and death process.

$$\frac{dp_{jk}(t)}{dt} = \lambda_{j-1} P_{j-1,k}(t) - (\lambda_j + \mu_j)P_{jk}(t) + \mu_{j+1} P_{j+1,k}(t)$$

$$\frac{dp_{ok}(t)}{dt} = -\lambda_0 P_{0k}(t) + \mu_1 P_{1k}(t) \quad \text{…(4)}$$

$$\frac{dp_{Nk}(t)}{dt} = \lambda_{N-1} P_{N-1,k}(t) - P_{Nk}(t)\mu_N$$

In general those equations satisfy $\frac{dp}{dt} = QP$ where Q is generator matrix when the state is finite :

$$Q = \begin{bmatrix} -\lambda_0 & \mu_1 & 0 & \dots & 0 \\ \lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & \mu_N \\ 0 & 0 & 0 & \dots & -\mu_N \end{bmatrix}$$

and when the state is infinite:

$$Q = \begin{bmatrix} -\lambda_0 & \mu_1 & 0 & 0\dots \\ \lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & 0\dots \\ 0 \quad \lambda_1 & -\lambda_2 - \mu_2 & \mu_3 \dots & \\ \dots\dots\dots\dots\dots\dots & & & \end{bmatrix}$$

For birth and death chains , there is an iterative procedure for computing the stationary probability distribution.

Karlin and McGregor (1957) show that for arbitrary starting state j , transition probabilities can be represented in the form: $P_{jk}(t) = \pi_k \int_0^{\infty} e^{-xt} Q_j(x)Q_k(x)\psi(dx)$ …(5)

and Laplace transform of (4) is:

$$f_{jk}(s) = L[P_{jk}(t)](s) = \int_0^{\infty} e^{-st} P_{jk}(t)dt \quad \text{…(6)}$$

Applying the Laplace transform to (3), with $j = 0$:

$$sf_{0,0}(s)-P_{0,0}(0)=\mu_1 f_{01}(s)-\lambda_0 f_{0,0}(s)$$
$$sf_{0,k}(s)-P_{0,k}(0)=\lambda_{k-1} f_{0,1}(s)+\mu_{k+1} f_{0,k+1}(s)-(\lambda_k+\mu_k)f_{0k}(s)$$
$$f_{oo}(s)=\frac{1}{s+\lambda_0+\mu_1(\frac{f_{01}(x)}{f_{oo(x)}})}$$

and

$$\frac{f_{ok}(s)}{f_{0,k-1}(s)}=\frac{\lambda_{k-1}}{s+\mu_k+\lambda_k-\mu_{k+1}(\frac{f_{0,k+1}(s)}{f_{0,k}(s)})}\Rightarrow f_{0,0}(s)=\cfrac{1}{s+\lambda_0-\cfrac{\lambda_0\mu_1}{s+\lambda_1+\mu_1-\cfrac{\lambda_1\mu_2}{s+\lambda_2+\mu_2-\dots}}}$$

Let $a_1=1, a_n=-\lambda_{n-2}\mu_{n-1}$ and $b_1=s+\lambda_0, b_n=s+\lambda_{n-1}+\mu_{n-1}$ for $i\geq 2$, then:

$$f_{0,0}(s)=\cfrac{a_1}{b_1+\cfrac{a_2}{b_2+\cfrac{a_3}{b_3+\dots}}}$$ or $$f_{0,0}(s)=\frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\dots$$ and kth convergent of

$$f_{00}(s) \text{ is } f_{oo}^{(m)}(s)=\frac{a_1}{b_1+}\frac{a_2}{b_2+}\dots\frac{a_m}{b_m+}=\frac{A_m(s)}{B_m(s)} \quad \dots(7)$$

2. Estimation via EM algorithm

We now proceed with estimation of the parameters of general BDP using discrete observations.

Let $Y=\{X(0)=i, X(t)=j\}$ the form of one or more independent observations from a general BDP. For the single realization of the process starting at $X(0)=i$ and ending at $X(t)=j$, let T_k be the total time spent in state k, U_k be number of "up" steps (birth) from state k and let D_k be number of "down" steps (deaths) from the state k.

Let the number of total up and down steps in a realization of the process be denoted by $U=\sum_{k=0}^{\infty}U_k$ and $D=\sum_{k=0}^{\infty}D_k$, and total particle time: $T_p=\int_0^t X(x)dx=\sum_{k=0}^{\infty}kT_k$

Wolff (1965) show the log-likelihood for a continuously observed process:

$$\ell(\theta)=\sum_{k=0}^{\infty}\left[U_k\log\lambda_k(\theta)+D_k\log\mu_k(\theta)-(\lambda_k(\theta)+\mu_k(\theta))T_k\right] \quad \dots(8)$$

When BDP is sampled discretely such that only $X(0)=i$ and $X(t)=j$ are observed, then U_k, D_k, T_k are unknown for every k and we cannot maximize the (8) without them.

In the EM algorithm we define a surrogate objective function Q by taking the expectation of (8), conditional of the observed data Y and the parameter values $\theta^{(m)}$. Taking the expectation of (8) conditional of Y and $\theta^{(m)}$ we form Q:

$$Q(\theta,\theta^{(m)}) = E\left(\ell(\theta)\mid Y,\theta^{(m)}\right) =$$
$$= \sum_{k=0}^{\infty}\left[E(U_k\mid Y)\log(\lambda_k(\theta)) + E(D_k\mid Y)\log(\mu_k(\theta)) - E(T_k\mid Y)(\lambda_k(\theta)+\mu_k(\theta))\right] \quad \ldots(9)$$

Where for clarity we have omitted the dependence of the expectation of parameter value $\theta^{(m)}$ from the m-th iterate.

3. Computing expectations of the E-step

In addition the state-space of BDP is generally infinite, it is tempting to approximate an infinite BDP as a similar process on the finite state-space $\{0,1,\ldots,N\}$ where N is chosen so that the probability of the process visiting state greater then N is small.

$$P(X(s) > N \mid X(0)=a, X(t)=b, 0<s<t) < \varepsilon \quad \forall\varepsilon > 0$$

Using transition probabilities we have:

$$E(U_k\mid Y) = \frac{\int_0^t P_{ak}(x)\lambda_k P_{k+1,b}(t-x)\,dx}{P_{ab}(t)}$$

$$E(D_k\mid Y) = \frac{\int_0^t P_{ak}(x)\mu_k P_{k-1,b}(t-x)\,dx}{P_{ab}(t)}$$

$$E(T_k\mid Y) = \frac{\int_0^t P_{ak}(x)P_{kb}(t-x)\,dx}{P_{ab}(t)} \quad \ldots(10)$$

Using Laplace convolution property, we arrive at the representations:

$$E(U_k\mid Y) = \lambda_k\frac{L^{-1}\left[f_{ak}(s)f_{k+1,b}(s)\right](t)}{P_{ab}(t)}$$

$$E(D_k\mid Y) = \mu_k\frac{L^{-1}\left[f_{ak}(s)f_{k-1,b}(s)\right](t)}{P_{ab}(t)}$$

$$E(T_k\mid Y) = \frac{L^{-1}\left[f_{ak}(s)f_{kb}(s)\right](t)}{P_{ab}(t)} \quad \ldots(11)$$

Where L^{-1} denote inverse Laplace transformation.

4. Simple birth and death process

For Δ sufficiently small, the transition probabilities for a simple birth and death process satisfy:

$$p_{k+j,k}(\Delta t) = P(\Delta X(t) = j \mid X(t) = k) = \begin{cases} k\mu\,\Delta t + o(\Delta t) & j = -1 \\ k\lambda\,\Delta t + o(\Delta t) & j = 1 \\ 1-(\lambda+\mu)k\,\Delta t + o(\Delta t) & j = 0 \\ o(\Delta t) & j \neq -1,0,1 \end{cases} \quad \ldots(12)$$

Kolmogorov differential equations:

$$\frac{dp_k(t)}{dt} = \lambda(k-1)p_{k-1}(t) + \mu(k+1)p_{k+1}(t) - (\lambda+\mu)k\,p_k(t)$$

$$\frac{dp_0(t)}{dt} = \mu p_1(t) \quad \text{for } i = 1,2,\ldots \text{ ,with } p_k(0) = \delta_{kN} \quad \ldots(13)$$

In this process birth and death happen at constant per-capital rates, so $\lambda_k = k\lambda$ and $\mu_k = k\mu$ the $\theta = (\lambda, \mu)$ and the surrogate function becomes:

$$Q(\theta) = \sum_{k=0}^{\infty} E(U_k \mid Y)\log(\lambda) + E(D_k \mid Y)\log(\mu) - E(T_k \mid Y)k(\lambda+\mu) \quad \ldots(14)$$

With derivation of (14) setting the result to 0 and solving for λ and μ we have:

$$\lambda^{(n+1)} = \frac{E(U \mid Y)}{E(T_p \mid Y)} \qquad \mu^{(n+1)} = \frac{E(D \mid Y)}{E(T_p \mid Y)} \quad \ldots(15)$$

5. Linear BDPs via immigration and emigration

Sometimes population are not closed and new individuals can enter, we call this action "immigration", where $\lambda_k = k\lambda + \nu$ and $\mu_k = k\mu$

Suppose immigration is include in the simple birth and death process at a constant rate ν.

For Δ sufficiently small, the transition probabilities for a simple birth and death process with immigration satisfy:

$$P_{k+j,k}(\Delta t) = P(\Delta X(t) = j \mid X(t) = k) = \begin{cases} k\mu\,\Delta t + o(\Delta t) & j = -1 \\ (\nu + k\lambda)\Delta t + o(\Delta t) & j = 1 \\ 1-(\nu + k(\lambda+\mu))\Delta t + o(\Delta t) & j = 0 \\ o(\Delta t) & j \neq -1,0,1 \end{cases} \quad \ldots(16)$$

Kolmogorov differential equations:

$$\frac{dp_k}{dt} = \lambda(k-1+\nu)p_{k-1} + \mu(k+1)p_{k+1} - ((\lambda+\mu)k+\nu)p_k$$

$$\frac{dp_0}{dt} = -\nu p_0 + \mu p_1 \qquad \text{...(17)}$$

The log-likelihood becomes :

$$\ell(\theta) = \sum_{k=0}^{\infty} U_k \log(k\lambda+\nu) + D_k \log(k\mu) - T_k(k(\lambda+\mu)+\nu) \qquad \text{...(18)}$$

$$Q(\theta) = \sum_{k=0}^{\infty} \mathrm{E}(U_k \mid Y)\log(k\lambda+\nu) + \mathrm{E}(D_k \mid Y)\log(k\mu) - \mathrm{E}(T_k \mid Y)(k(\lambda+\mu)+\nu) \qquad \text{...(19)}$$

Maximizing Q with respect to λ dhe μ yields the updates:

$$\lambda^{(n+1)} = \frac{\sum_{k=0}^{\infty} \mathrm{E}(U_k \mid Y)}{\mathrm{E}(S_p \mid Y)} \quad \text{and} \quad \nu^{(n+1)} = \frac{1}{t}\sum_{k=0}^{\infty} (1-p_k)\,\mathrm{E}(U_k \mid Y) \qquad \text{...(20)}$$

Likewise, consider a population model for the number of organisms in an area and suppose new emigrants leave at rate γ . In this case, we define the linear rate $\mu_k = k\mu + \gamma$. The process continuous like in the immigration case.

6. Moran model

Moran (1958) includes a model for the time-evolution of a biallelic locus when the population size is constant through time. This process keeps track of the number of alleles of a certain type at a biallelic locus in a haploid population of constant size $N < \infty$. Call the two alleles A and B, with fitness α and β respectively.

In the Moran model with mutation, A mutates B with probability u, and B mutates A with probability ν. If we let $X(t)$ be the number of A alleles in the population at time t, the rate of additions of new A is :

$$\lambda_n = \frac{N-n}{N}\left[\alpha \frac{n}{N}(1-u) + \frac{N-n}{N}\nu\right]$$ and the rate of removals of A is:

$$\mu_n = \frac{n}{N}\left[\frac{N-n}{N}(1-\nu) + \alpha \frac{n}{N}u\right]$$

where $\beta = 1$, $n = 0,1,\ldots,N$ and with $\lambda_n = 0$ when $n > N$

We can minimize the birth rate as:

$$\log(\lambda_n) \propto \log(n\alpha(1-u) + (N-n)v) \geq$$
$$\geq p_n^{(m)} \log\left(p_n^{(m)} n\alpha(1-u)\right) + (1-p_n^{(m)}) \log\left((1-p_n^{(m)})(N-n)v\right) \propto$$
$$\propto p_n^{(m)}(\log\alpha + \log(1-u)) + (1-p_n^{(m)})\log v$$

where: $$p_n^{(m)} = \frac{n\alpha^{(m)}(1-u^{(m)})}{n\alpha^{(m)}(1-u^{(m)}) + (N-n)v^{(m)}} \quad \text{...(21)}$$

Likewise we minimize the death rate, where:

$$q_n^{(m)} = \frac{(N-n)(1-v^{(m)})}{(N-n)(1-v^{(m)}) + n\alpha^{(m)} u^{(m)}} \quad \text{...(22)}$$

and we form the minimizing function H as

$$H(\theta) = \sum_{k=0}^{\infty} \Big[U_k \left(p_k^{(m)}(\log\alpha + \log(1-u)) + (1-p_k^{(m)})\log v \right) +$$
$$+ D_k \left(q_k^{(m)} \log(1-v) + (1-q_k^{(m)})(\log\alpha + \log u) \right) -$$
$$- \frac{T_k}{N^2}\left[(N-k)k\alpha(1-u) + (N-k)^2 v + (N-k)k(1-v) + k^2\alpha \right] \Big] \quad \text{...(23)}$$

The update of α is

$$\alpha^{(m+1)} = \frac{\sum_{k=0}^{N} p_k^{(m)} N_k + (1-q_k^{(m)}) D_k}{\frac{1}{N^2}\sum_{k=0}^{N} T_k \left[(N-k)k(1-u^{(m)}) + k^2 u^{(m)} \right]} \quad \text{...(24)}$$

The solution of

$$\sum_{k=0}^{N} -uB_k p_k^{(m)} + (1-u) D_k (1-q_k^{(m)}) - u(1-u)\frac{T_k}{N^2}\left[k^2\alpha^{(m)} - (N-k)k\alpha^{(m)} \right] = 0$$ give the update of u

7. Regression of count data

Consider a collection of *n* independent BDPs, $X^i(t)$ with $\lambda_k^i = \exp(Z_i\beta)$ and $\mu_k = 0$

for $i = 1,2,...,n$; where Z_i is a $d\times 1$ vector of covariates and β is a covariate vector of corresponding dimension and $\mu_k = 0$ for all k. Letting $X^i(0) = 0$ and $X^i(t) = x_i$, then: $\ell(\beta) = \sum_{i=1}^{n} x_i Z_i\beta - \exp(Z_i\beta)$ is log-likelihood for classical Poisson regression.

The Newton-Raphson update for β is

$$\beta^{(n)} = \beta^{(n-1)} - \left(\nabla_\beta^2 \ell\right)^{-1} \nabla_\beta \ell \quad \text{...(25)}$$

where $\nabla_\beta Q = \sum_{i=1}^{n} \mathrm{E}\left(U^i \mid Y_i \right) Z_i - E\left(T_p^i \mid Y_i \right) Z_i \exp\left(Z_i \beta \right)$ and

$$d_\beta^2 Q = -\sum_{i=1}^{n} \mathrm{E}\left(T_p^i \mid Y_i \right) Z_i Z_i' \exp\left(Z_i \beta \right)$$

8. Transition probability error

To find the transition probability $P_{ai}(t)$ we set $G(s) = f_{ai}(s)$. We can evaluate the finite continued fraction $f_{ai}(s_j)$ the M_j-th convergent $f_{ai}^{(M_j)}(s_j)$

M_j is a positive integer chosen dynamically so that the error due truncation is

$$\left| f_{ai}(s_j) - f_{ai}^{(M_j)}(s_j) \right| \leq \varepsilon \text{ for every } \varepsilon > 0$$

$$\widetilde{P}_{ai}(t) = \frac{e^{\frac{A}{2}}}{2t} \mathrm{Re}\left(f_{ai}(s_0) \right) + \frac{e^{\frac{A}{2}}}{t} \sum_{j=1}^{\infty} (-1)^j \mathrm{Re}\left(f_{ai}(s_j) \right)$$

The discrete error is

$$\left| P_{ai}(t) - \widetilde{P}_{ai}(t) \right| \leq \sum_{j=1}^{\infty} e^{-jA} P\left((2j+1)t \right) \leq \frac{e^{-A}}{1-e^{-A}} \quad \ldots(26)$$

References:

1. Abate J. Whitt W.,*Computing Laplace Transforms of numerical inversion via continued fractions,* 1999, INFORMS J Comput, 11(4):394-405
2. Wolff R.W., *Problems of Statistical Inference for Birth and Death Queuing Models,* 1965, Operations Research 13,343-357
3. Lee E. T. and Wang J. W., *Statistical Methods for Survival Data Analysis, 3th ed.*, 2003, Wiley, New York
4. J. R. Norris, *Markov Chains,* 1998, Cambridge Series in Statistical and Probabilistic Mathematics
5. Prasanna Sahoo, *Probability and Mathematical Statistics,* 2013, University of Louisville
6. Athanasios Papoulis, *Probability Random Variables and Stochastic Processes 3th ed.* 1991, Mc Graw-Hill, inc, ISBN 0-07-048477-5