A Stochastic EM algorithm for construction of Mortality Tables

Luz Judith Rodriguez Esparza, F. Baltazar-Larios

MAM-9

Budapest, June 28 - 30, 2016

Aim

We propose to use the concept of physiological age to modelling the aging process by using phase-type distributions to calculate the probability of death.

For the estimation part, we will use the EM algorithm which in turn uses the Bisection method.

Outline

1. BACKGROUND:
   - MORTALITY MODELS
   - PHYSIOLOGICAL AGE
   - PHASE-TYPE DISTRIBUTIONS
2. MODEL AND DATA
3. ESTIMATION: STOCHASTIC EM ALGORITHM
4. APPLICATION: REAL DATA

Mortality risk

- A. de Moivre in **1725** proposed the first mathematical mortality model.
- In **1825** B. Gompertz proposed that a law of geometric progression pervades in mortality after a certain age. He obtained the following expression:

$$\mu_x = \alpha e^{\beta x}.$$

- In **1860**, W. M. Makeham extended the Gompertz model by adding a constant:

$$\mu_x = \sigma + \alpha e^{\beta x},$$

where σ represents all random factors with no willingness to death, for example accidents, epidemics, etc.

Motivation

- Several factors can alter the probability of death, the more considered factor is the age but there are other important characteristics such as sex, clinical history, smoking, etc.
- In most mortality models we cannot determine the distribution of the time of death explicitly.
- Lin and Liu (2008) used phase-type distributions to model human mortality.
- We consider a more general case, giving another interpretation of the states and using a stochastic EM algorithm for the estimation part.

PHYSIOLOGICAL AGE

- **Aging process:** “the progressive, and essentially irreversible diminution with the passage of time of the ability of an organism or one of its parts to adapt to its environment, manifested as diminution of its capacity....".
- To model the aging process, we consider the concept of **physiological age:** relative health index representing the degree of aging on the individual.
- It is natural assume that the time spent in each state has an exponential behavior.

PHASE-TYPE DISTRIBUTIONS

- Dimension n.
- Infinitesimal matrix:

$$\Lambda = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{r} \\ \boldsymbol{0} & 0 \end{pmatrix}. \tag{1}$$

- Let α be the initial distribution of the process.

MODEL AND DATA

Let consider a finite-state Markov jump process to model the hypothetical aging process. If a person has physiological age i for $i \in \{1, \dots, n-1\}$, we consider three possible cases:

1. Natural development of the aging process: the person eventually transits to the next physiological age $i+1$.
   Intensity of this transition: $\lambda_{i,i+1}$.

2. The aging process is affected by an unusual incident: the person transits to some physiological age j, with $j \in \{i+2, i+3, \dots, n\}$.
   Intensity of this transition: λ_{ij}.

3. The possibility of death for the person at that physiological status.
   Intensity of this transition: r_i.

- Thus, the sub-intensity matrix is given by

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_1 & \lambda_{12} & \lambda_{13} & \dots & \lambda_{1n} \\ 0 & -\lambda_2 & \lambda_{23} & \dots & \lambda_{2n} \\ 0 & 0 & -\lambda_3 & \dots & \lambda_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_n \end{pmatrix}.$$

- We denote by $q_i(t)$ the probability of death for a person at physiological age $i \in \{1, 2 \dots, n\}$, in the interval $[0, t]$, which is given by

$$q_i(t) = P(\tau_i \leq t) = 1 - \boldsymbol{e}_i \exp(t\boldsymbol{Q})\boldsymbol{e}, \qquad (2)$$

where $\tau_i \sim PH(\boldsymbol{e}_i, \boldsymbol{Q})$.

- Considering a population of size M and the time interval $[0, T]$, let $X^m = \{X_t^m\}_{t \geq 0}^T$, $m = 1, \ldots, M$, be independent Markov jump processes with the same finite state space E and the same intensity matrix.

- Each X^m represents the aging process for each person in the population.

- We will use a stochastic EM algorithm for finding maximum likelihood estimators of $(\alpha, \boldsymbol{Q})$ considering the two scenarios and using the Bisection method.

ESTIMATION

Cases:

1. Continuous time information of the aging process of the population. (See [2]).
2. There are reports of the development process only at determined moments, i.e., there are only discrete time observations of a Markov jump process. (Discrete case).

Continuous case

Considering that the X^m's have been observed continuously in the time interval $[0, T]$.

- Complete likelihood function:

$$L_T^c(\theta) = \prod_{i=1}^{n} \alpha_i^{B_i} \prod_{i=1}^{n} \prod_{j \neq i}^{n} \lambda_{ij}^{N_{ij}(T)} e^{-\lambda_i Z_i(T)} \prod_{i=1}^{n} r_i^{N_i(T)}, \tag{3}$$

- Log-likelihood function:

$$\begin{aligned} \ell_T^c(\boldsymbol{\theta}) &= \sum_{i=1}^{n} B_i \log(\alpha_i) + \sum_{i=1}^{n} \sum_{j \neq i}^{n} N_{ij}(T) \log(\lambda_{ij}) \\ &- \sum_{i=1}^{n} \sum_{j \neq i}^{n} \lambda_{ij} Z_i(T) + \sum_{i=1}^{n} N_i(T) \log(r_i) - \sum_{i=1}^{n} r_i Z_i(T). \end{aligned} \tag{4}$$

Discrete Case

- We are interested in the inference about the intensity matrix Λ based on samples of observations of X^m's at discrete times points.

- Suppose that all processes have been observed only at K time points $0 = t_1 < \ldots < t_K = T$ denoted by $Y_k^m = X_{t_k}^m$.$t_{k+1} - t_k = \Delta$, then $Y^m = \{Y_k^m : k = 1, \ldots, K\}$ is the discrete time Markov chain associated with X^m with transition matrix

$$P(\Delta) = \exp(\Delta\Lambda).$$

- Observed values by $\boldsymbol{y} = \{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M\}$ where

$$\boldsymbol{y}^m = \{Y_1^m = y_1^m, \ldots, Y_K^m = y_K^m\}.$$

EM algorithm

Let $\theta_0 = (\alpha_0, Q_0)$ denote any initial value of parameters.

1. (E-step) Calculate the function
$$h(\theta) = E_{\theta_0}(\ell_T^c(\theta) | Y = y);$$
2. (M-step)
$$\theta_0 = argmax_{\theta} h(\theta);$$
3. Go to 1

E-step

$$
\begin{aligned}
E_{\theta_0}(\ell_T^c(\theta)|Y=y) \quad &= \quad \sum_{i=1}^{n} \log(\alpha_i) E_{\theta_0}(B_i|Y=y) \\
&+ \sum_{i=1}^{n}\sum_{j\neq i}^{n} \log(\lambda_{ij}) E_{\theta_0}(N_{ij}(T)|Y=y) - \sum_{i=1}^{n}\sum_{j\neq i}^{n} \lambda_{ij} E_{\theta_0}(Z_i(T)|Y=y) \\
&+ \sum_{i=1}^{n} \log(r_i) E_{\theta_0}(N_i(T)|Y=y) - \sum_{i=1}^{n} r_i E_{\theta_0}(Z_i(T)|Y=y). \qquad (5)
\end{aligned}
$$

Conditional expectations of the statistics:

$$E_{\theta_0}(B_i|Y=y)=\sum_{m=1}^{M}1_{\{Y_1^m=i\}};\quad i=1,\ldots,n, \tag{6}$$

where $1_{\{\cdot\}}$ is the indicator function,

$$E_{\theta_0}(N_{ij}(T)|Y=y)=\sum_{m=1}^{M}\sum_{k=2}^{K}\tilde{N}^{mij}_{y^m_{k-1}y^m_k}(t_k-t_{k-1}), \tag{7}$$

$$E_{\theta_0}(Z_i(T)|Y=y)=\sum_{m=1}^{M}\sum_{k=2}^{K}\tilde{Z}^{mi}_{y^m_{k-1}y^m_k}(t_k-t_{k-1}), \tag{8}$$

and

$$E_{\theta_0}(N_i(T)|Y=y)=\sum_{m=1}^{M}\sum_{k=2}^{K}\tilde{N}^{mi}_{y^m_{k-1}}(t_k-t_{k-1}). \tag{9}$$

Markov Bridges

- To calculate these expectations we propose to generate L sample paths of the Markov bridge $X^m(r, s_1, s, s_2)$ using the parameter value $\theta_0 = (\alpha_0, Q_0)$.
- I.e., a stochastic process defined on $[s_1, s_2]$ and having the same distribution of the Markov jump process $\{X_t^m\}_{t \in [s_1, s_2]}$ conditioned on $X_{s_1}^m = r$ and $X_{s_2}^m = s$ for $m = 1, \ldots, M$.

Now, based on bridges we approximate the conditional expectations by

$$\tilde{N}^{mij}_{y^m_{k-1}y^m_k}(t_k - t_{k-1}) \approx \frac{1}{L}\sum_{l=1}^{L} N^{mij(l)}_{y^m_{k-1}y^m_k}(t_k - t_{k-1}), \tag{10}$$

$$\tilde{Z}^{mi}_{y^m_{k-1}y^m_k}(t_k - t_{k-1}) \approx \frac{1}{L}\sum_{l=1}^{L} Z^{mi(l)}_{y^m_{k-1}y^m_k}(t_k - t_{k-1}), \tag{11}$$

$$\tilde{N}^{mi}_{y^m_{k-1}}(t_k - t_{k-1}) \approx \frac{1}{L}\sum_{l=1}^{L} N^{mi(l)}_{y^m_{k-1}}(t_k - t_{k-1}), \tag{12}$$

respectively, for $i, j = 1, \ldots, n$; $m = 1 \ldots, M$; and $k = 1, \ldots, K$.

Stochastic EM algorithm

1. initial value of parameters: $\theta_0 = (\alpha_0, Q_0)$. Let $\theta = \theta_0$. Given M, K for $m = 1, \ldots, M$, $k = 2, \ldots, K$:
2. Generate L paths of the Markov bridge $X^m(y_{k-1}, t_{k-1}, y_k, t_k)$ using θ.
3. **E-step.**
   - Using (10), (11), and (12) calculate $\tilde{N}^{mij}_{y^m_{k-1} y^m_k}(t_k - t_{k-1})$, $\tilde{Z}^{mi}_{y^m_{k-1} y^m_k}(t_k - t_{k-1})$, and $\tilde{N}^{mi}_{y^m_{k-1}}(t_k - t_{k-1})$.
   - Using (6), (7), (8), and (9), calculate $E_\theta(B_i | Y = y)$, $E_\theta(N_{ij}(T) | Y = y)$, $E_\theta(Z_i(T) | Y = y)$, and $E_\theta(N_i(T) | Y = y)$.
4. **M-step** Calculate $\hat{\theta} = (\hat{\alpha}, \hat{\boldsymbol{Q}})$ by

$$\hat{\alpha}_i = \frac{E_\theta(B_i | Y = y)}{M}; \quad \hat{\lambda}_{ij} = \frac{E_\theta(N_{ij}(T) | Y = y)}{E_\theta(Z_i(T) | Y = y)}; \quad \hat{r}_i = \frac{E_\theta(N_i(T) | Y = y)}{E_\theta(Z_i(T) | Y = y)}.$$

5. $\theta = \hat{\theta}$ go to 2.

Bisection method

- To implement this algorithm, the main issue is how to sample Markov bridges.
- We use the bisection method proposed by S. Asmussen and A. Hobolth ([1]) because of its potential for variance reduction.
- Idea of this algorithm: formulate a recursive procedure where we finish off intervals with zero or one jump and keep bisecting intervals with two or more jumps.
- The recursion ends when no intervals with two or more jumps are presented.

Considering a Markov bridge $X(a, 0, b, T)$ and using the bisection algorithm we have two type of scenarios:

1. If $a = b$ and there are no jumps. In this case we are done: $X_t = a$ (a is not an absorbing state) for $0 \leq t \leq T$.
2. If $a \neq b$ and there is one jump we are done: $X_t = a$ for $t \in [0, \tau)$, and $X_t = b$ for $t \in [\tau, T]$. Here τ is the jumping time.

To determine τ we use the following lemma from [1].

Lema

Considering an interval of length T, let $X_0 = a$, the probability that $X_T = b \neq a$ and there is only one single jump (from a to b) in the interval is given by

$$R_{ab}(T) = \lambda_{ab} \begin{cases} \frac{e^{-\lambda_a T} - e^{-\lambda_b T}}{\lambda_a - \lambda_b} & \lambda_a \neq \lambda_b \\ T e^{-\lambda_a T} & \lambda_a = \lambda_b. \end{cases}$$

The density of the time of state change is

$$f_{ab}(t; T) = \frac{\lambda_{ab} e^{-\lambda_b T}}{R_{ab}(T)} e^{-(\lambda_a - \lambda_b)t}; \quad 0 \leq t \leq T.$$

Example

$n = 4$, $M = 500$, $T = 100$, $\Delta = 5$, $K = 20$, $L = 50$, with arbitrary initial parameters.

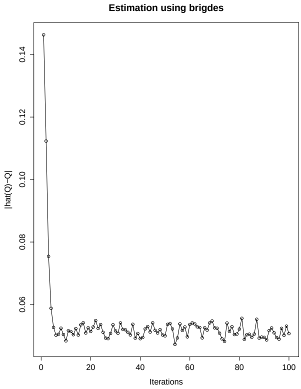

Figure: Norm-1 of $\hat{Q} - Q$ for 100 iterations, where the estimation was obtained using Bridges.

Table: Maximum likelihood estimators (MLEs) and Standard deviations (SDs)

Parameter	True value	MLE	SD
$\hat{\alpha}_1$	0.25	0.2680	0.0232
$\hat{\alpha}_2$	0.25	0.2820	0.0237
$\hat{\alpha}_3$	0.25	0.2480	0.0223
$\hat{\alpha}_4$	0.25	0.2020	0.0201
$\hat{\lambda}_{12}$	0.03125	0.0390	0.0064
$\hat{\lambda}_{13}$	0.03125	0.0320	0.0060
$\hat{\lambda}_{14}$	0.03125	0.0377	0.0060
$\hat{\lambda}_{23}$	0.03125	0.0252	0.0036
$\hat{\lambda}_{24}$	0.03125	0.0323	0.0041
$\hat{\lambda}_{34}$	0.03125	0.0345	0.0034
$\hat{r}_1$	0.03125	0.0210	0.0040
$\hat{r}_2$	0.03125	0.0354	0.0043
$\hat{r}_3$	0.03125	0.0325	0.0033
$\hat{r}_4$	0.03125	0.0334	0.0020

APPLICATION: REAL DATA

- Population data from the U.S.A. for the construction of hypothetical physiological ages.
- The database contains annual information on mortality and population from 1933 to 2013.
- In each year, the information is classified at 111 ages, from 0 to 110.
- Using a physiolgical age index we can obtain the probability that the person passes from age i to age j in one year ($i = 0, 1, \ldots, 109$; $j = 1, \ldots, 110$).

- We generate historical information of the aging process of the population of the United States observed at discrete times.
- We use the algorithm presented to estimate the corresponding infinitesimal generator and therefore the parameters of the phase type distribution used for building the mortality tables.
  $n = 111, , M = 1000, T = 50, \Delta = 1, K = 50, L = 20, I = 100$

Considering the year 2013, in figure 2 we plot the estimation of the mortality tables using the equation (2).

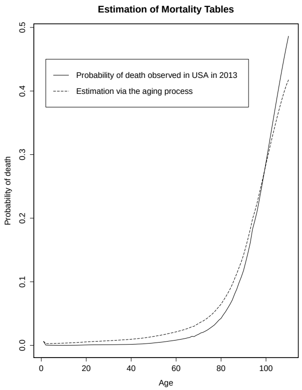

Figure: Estimation of Mortality Tables.

THANK YOU

QUESTIONS?

References

S. Asmussen and A. Hobolth.
Markov Bridges, Bisection and Variance Reduction.
Technical report, 2011.

S. Asmussen, O. Nerman, and M. Olsson.
Fitting phase-type distributions via the EM algorihtm.
Scandinavian Journal of Statistics, 23:419–441, 1996.