quantil

Use of Generalized Additive Models in the Estimation of Non-Homogeneous Discrete Markov Chains: Applications in Actuarial Multi-State Modeling

Sebastián Calcetero

sf.calcetero674@uniandes.edu.co

Researcher, Quantil SAS - Matemáticas aplicadas - quantil.co

Lecturer, Dpt. ingeniería industrial, Universidad de los Andes

Outline

Motivation

- A stochastic process $\{X_t, t \in \{0, 1 \ldots\}\}$ with countable state space $\mathcal{S}$, is a discrete-time Markov chain if it satisfies the *Markov property*:

$$P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \ldots)$$
$$= P(X_{t+1} = i_{t+1} | X_t = i_t) \forall i_s \in \mathcal{S}$$

- We define the one step conditional probability transition matrix $\mathbf{P}_t$ as :

$$[\mathbf{P}_t]_{(i,j)} := p_t^{i,j} := P(X_{t+1} = j | X_t = i)$$

- Similarly, we define the h-step probability transition matrix, ${}_h\mathbf{P}_t$, as:

$$[{}_h\mathbf{P}_t]_{(i,j)} :=_h p_t^{i,j} := P(X_{t+h} = j | X_t = i)$$

Motivation

- Using a conditional probability argument, one step at a time, it can be easily shown that:

$$ {}_h\mathbf{P}_t = \prod_{s=0}^{h-1} \mathbf{P}_{t+s} $$

- That is, the whole Markov chain can be completely specified by the one-step transition probabilities.
- The estimation is usually achieved using the maximum likelihood estimator:

$$ p_t^{ij} = \frac{n_t^{ij}}{\sum_{k \in \mathcal{S}} n_t^{ik}} $$

where n_t^{ij} denotes the number of transitions from state i to state j in the age t.

Motivation

- The main problem in the non-homogenous case, is the several amounts of parameters to estimate (one transition matrix per each age of the process). That is, the model is over-parametrized!
- Therefore the statistical complexity of the estimation is highly increased, and the overall quality of the estimation is reduced (credibility problems, overfitting, high variance in the estimation, etc).
- Moreover, if it is necessary to include profiling variables to take into account heterogeneity in the population, the complexity increases even more!

Objectives

In this paper we:

- Discuss the application of generalized additive models (GAMs) for the parsimonious and consistent estimation of the several transition matrices of a non-homogeneous discrete Markov chain.
- Include possibility of using profiling variables to take into account the heterogeneity of the population in the estimation.
- Discuss some advantages for modeling, including how to perform statistical inference for the parameters, *dynamic* modeling and longitudinal studies.
- Illustrate the methodology with real data in some actuarial models, including dynamic life tables.

Outline

Outline

1. Introduction
2. Applications of GAM's in Markov Chains Modeling
   - Generalized Additive Models (GAM's)
   - Estimation methodology and advantages
   - Consistency of the estimation
   - Longitudinal and Dynamical Approaches
   - Hypothesis Testing for Non-Homogeneity
3. Actuarial Applications
   - Dynamic Life Tables with Portfolio Experience
   - Multi-state Model for Health Status
4. Conclusions and Future Work

Generalized Additive Models (GAM's)

- The generalized additive models (GAM) are characterized by constituting the following components:

$$\begin{cases} Y \sim \text{Exponential Family} & \text{Random component} \\ \eta = \beta_0 + f_1(W_1) + \ldots + f_k(W_k) & \text{Systematic component} \\ g(\mu_Y) = \eta & \text{Link function} \end{cases}$$

  where the f_j are assumed to be *soft* functions, and g is a monotonous function.
- The estimation is achieved using the principle of *regularized maximum likelihood:*

$$L(\theta) + \lambda \left(\int (f(W))^2 \, dW \right)$$

- This estimator is asymptotically *consistent* and *efficient*.

Outline

Estimation methodology

- We propose the use of a generalized additive model with multinomial response to estimate the transition probabilities of a discrete non-homogeneous Markov Chain, If needed, profiling variables and dynamical behavior can be included.
- In order to take in count the non-homogeneity of the chain, we assume that the one step probabilities p_t^{ij} are soft functions f in t, the age of the process:

$$g(p_t^{ij}) = f^{ij}(t)$$

where g can be any link function appropriate for a probability parameter.

Advantages

1. The assumption of a soft curve is not far from reality and reduces dramatically the complexity problem of usual approaches. That is, we have a *parsimonious* estimation of the probabilities.
2. Due to the smoothing, it is possible to have missing ages in the estimation, as long as there exist enough data for surrounding ages.
3. It is the possible to include profiling variables (W) to take into account the heterogeneity of the population:

$$g(p_t^{ij}) = f_1^{ij}(t) + f_2^{ij}(W)$$

4. Finally, all the estimation and the statistical inference procedures are already available in statistical software. Hence there is no need of doing extra coding

Outline

Consistency of the estimation

- We argue that the estimation approach is coherent within the context of Markov models, since the GAM estimator coincides with the maximum likelihood estimator, hence it inherits all of its properties:
    - The estimator is *consistent* and *asymptotically efficient*.
    - Statistical inference procedures for GAM's are valid for Markov Chains as well
- **We prove this by showing that the GAM likelihood is proportional to the one corresponding to a Markov chain with the same data set.**
- For simplicity, we illustrate the case where we have a discrete non-homogeneous Markov chain $\{X_t, t \in \{0, 1, \ldots\}\}$ with only two states. The general case is completely analogous.

Consistency of the estimation

Given a sample of N Markov chains $\{x_t^1\}_{t=0}^{T_1}$, $\{x_t^2\}_{t=0}^{T_2}$, $\ldots$, $\{x_t^N\}_{t=0}^{T_N}$, and denoting the vector of parameters as $\theta = \{p_t^{ij}\}, t \in \{0, 1, \ldots\}, i, j \in \{1, 2\}$, the *likelihood* function can be written as:

$$
\begin{aligned}
L(\theta) &= \prod_{k=1}^{N} P(X_0(\theta) = x_0^k, \ldots, X_{T_k}(\theta) = x_{T_k}^k) \\
&= \prod_{k=1}^{N} P(X_0(\theta) = x_0^k) \prod_{t=1}^{T_k} P(X_t(\theta) = x_t^k | X_{t-1}(\theta) = x_{t-1}^k) \\
&= \prod_{k=1}^{N} \pi_{x_0^k} \prod_{t=1}^{T_k} \prod_{i,j \in \mathcal{S}} \left(p_t^{ij}\right)^{1\{x_{t-1}^k = i, x_t^k = j\}}
\end{aligned}
$$

where $\pi_{x_0^k}$ is the long run probability of being in the state x_0^k, and $1\{.\}$ is the indicator function. The second line comes from conditioning and applying the Markov Property.

Consistency of the estimation

$$L(\theta) \propto \prod_{k=1}^{N} \prod_{t=1}^{T_k} \prod_{i,j \in \mathcal{S}} \left(p_t^{ij}\right)^{1\{x_{t-1}^k = i, x_t^k = j\}}$$

$$\propto \left(\prod_{k=1}^{N} \prod_{t=1}^{T_k} \prod_{j=1}^{2} \left(p_t^{1j}\right)^{1\{x_{t-1}^k = 1, x_t^k = j\}} \right) \left(\prod_{k=1}^{N} \prod_{t=1}^{T_k} \prod_{j=1}^{2} \left(p_t^{2j}\right)^{1\{x_{t-1}^k = 2, x_t^k = j\}} \right)$$

- The penultimate line shows that the Markov chain likelihood can be rewritten in a form which is proportional to a multinomial likelihood
- The last line shows that the likelihood can be factored in separate terms for every state in the chain, each one being a multinomial likelihood as well. Hence we can use separate estimations for each stage !

Outline

Longitudinal and Dynamical Approaches

- Since individuals are observed through time, we are in a *longitudinal study*. Here it is common to see the application of *mixed models* [3] which can be easily adapted for GAM's:

$$g(p_t^{ij}) = f^{ij}(t) + \alpha_\tau^{ij} + \nu_l^{ij}$$

  where α_τ^{ij} and ν_l^{ij} are fixed or random effects for being in the calendar time τ, and being the individual l, respectively.

- A dynamical model can be estimated as well:

$$g(p_t^{ij}(\tau)) = f^{ij}(t) + \kappa^{ij}(\tau)$$

  where $p_t^{ij}(\tau)$ makes explicit the relation with calendar time, τ, and $\kappa^{ij}(\tau)$ is a smooth function of τ t.

Outline

Hypothesis Testing for Non-Homogeneity

- An important question that arises in this context, is if the transition probabilities do change according to the age. That is, if the chains is homogeneous:

$$H_0 : p_t^{ij} = p^{ij} \ \forall i, j \in \mathcal{S}, \ \forall t \in \{0, 1, \ldots\}$$
$$H_1 : \neg H_0$$

- We can use a *regularized likelihood ratio test* for comparison of models:

$$\text{Model 1: } g(p_t^{ij}) = \beta_0^{ij} \ \forall i, j \in \mathcal{S}, \ \forall t \in \{0, 1, \ldots\}$$
$$\text{Model 2: } g(p_t^{ij}) = \beta_0^{ij} + f^{ij}(t) \ \forall i, j \in \mathcal{S}, \ \forall t \in \{0, 1, \ldots\}$$

$$-2\left(ln(L(\theta_{\mathrm{M1}})) - ln(L(\theta_{\mathrm{M2}}))\right) \approx \chi^2_{\mathrm{dif}}$$

where $ln(L(\theta_{\mathrm{M}}))$ is the *loglikelihood* of the estimated model M, and dif is the difference between degrees of freedom

Outline

1. Introduction
2. Applications of GAM's in Markov Chains Modeling
   - Generalized Additive Models (GAM's)
   - Estimation methodology and advantages
   - Consistency of the estimation
   - Longitudinal and Dynamical Approaches
   - Hypothesis Testing for Non-Homogeneity
3. Actuarial Applications
   - Dynamic Life Tables with Portfolio Experience
   - Multi-state Model for Health Status
4. Conclusions and Future Work

Outline

Dynamic Life Tables with Portfolio Experience

- A life table describes a survival model in discrete time which is, in fact a two-state Markov chain with transition probability matrix given by:

$$\mathbf{P}_t = \begin{bmatrix} p_t & q_t \\ 0 & 1 \end{bmatrix}$$

  where $p_t (=: p_t^{11})$ is the one year survival probability for a person of age t, and $q_t (=: p_t^{12})$ is the one year death probability for a person of age t.
- Colombia's law declares that companies must use the RV08 life table, but it can be modified by the company if there is enough *experience in their portfolios*.

Dynamic Life Tables with Portfolio Experience

- The biggest problem for companies to use their own tables is related to the *credibility* of the estimation, since a lot of data is necessary.
- **Here we proceed to estimate a dynamic life table for the population of a confidential health company using GAM's**
- The data set contains the dates corresponding to birth, death (if the case), first ingress to the company and retirement of the company.
- In order to avoid the credibility problems, we use a "Bayesian alike.ᵃpproach to consider the RV08 as a baseline for the estimation, just as described in [4].

Dynamic Life Tables with Portfolio Experience

We consider a GAM with binary response, where the variable takes the value of 1 if the person survives the next year, and 0 otherwise:

$$\begin{aligned} log(-log(p_t(\tau))) = \beta_0 + f(t) + \beta_1 \mathrm{Gender} + \beta_2 log(-log(\widetilde{p}_t)) \\ + \beta_3 \left(\mathrm{Gender} * log(-log(\widetilde{p}_t))\right) + \kappa(\tau) \end{aligned}$$

where f is a soft function of the age, Gender is a dummy variable which takes value of 1 if the individual is male and 0 if female, $\widetilde{p}_t$ is the survival probability given by the RV08, and κ is a soft function of calendar time.

Dynamic Life Tables with Portfolio Experience

Cuadro: Significance of GAM parameters for the life table

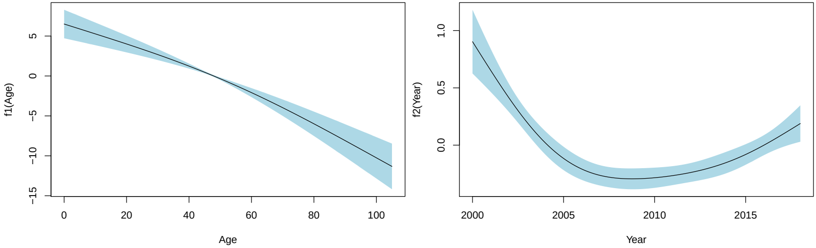
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	12.9110	1.5439	8.3628	< 0.0001
Gender	-0.4487	0.1373	-3.2679	0.0011
$log(-log(\widetilde{p}_t))$	-0.9161	0.2349	-3.9001	0.0001
Gender $* \, log(-log(\widetilde{p}_t))$	-0.0622	0.0230	-2.7043	0.0018
B. smooth terms	edf	Ref.df	F-value	p-value
$f(Age)$	1.9571	1.9982	63.2443	< 0.0001
$\kappa(Year)$	3.5903	3.9014	75.1478	< 0.0001

- All the terms are statistically significant, and Likelihood ratio tests for non-homogeneity and dynamic effects show that it is necessary to consider them.
- The *effective* number of terms used in the estimation is $1 + 1 + 1 + 1 + 1{,}96 + 3{,}59 = 9{,}55 \approx 10$.

Dynamic Life Tables with Portfolio Experience

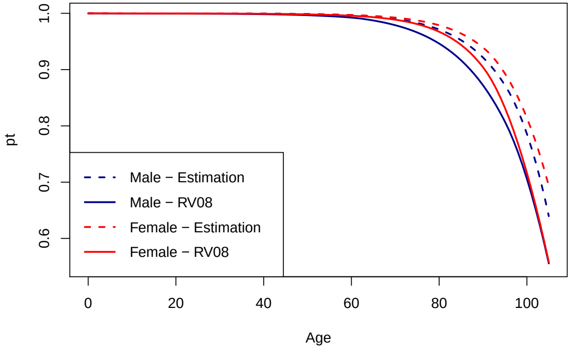
Figura: Estimation of GAM's smooth terms for p_t

(a) Estimation of $f(Age)$ (b) Estimation of $\kappa(Year)$



Dynamic Life Tables with Portfolio Experience

Figura: Forecast of survival probabilities p_t for 2019



Outline

1. Introduction
2. Applications of GAM's in Markov Chains Modeling
   - Generalized Additive Models (GAM's)
   - Estimation methodology and advantages
   - Consistency of the estimation
   - Longitudinal and Dynamical Approaches
   - Hypothesis Testing for Non-Homogeneity
3. Actuarial Applications
   - Dynamic Life Tables with Portfolio Experience
   - Multi-state Model for Health Status
4. Conclusions and Future Work

Multi-state Model for Health Status

- **Here we illustrate a Markov chain model to explain the evolution of the health status of individuals using real data from a confidential company.**
- The information consists of the diagnosis performed to the whole individuals according during a period of time on 10 years.
- Given the large number of possible diagnoses, we work only with the most common to be observed in the population.
- We develop separate models for each disease and the combine them in only one model.

Multi-state Model for Health Status

- Let $\{X_t^k, t \in \{0, 1, \ldots\}\}$ be DTMC that shows the presence of a disease k at the beginning of age t for a particular individual.

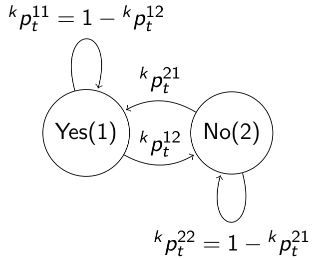


Figura: State graph of process X_t^k

Multi-state Model for Health Status

- Now, let $\{\mathbf{X}_t = (X_t^1, \ldots, X_t^n), t \in \{0, 1, \ldots\}\}$ be the aggregate process of health status of an individual at the beginning of age t. The state space is $\mathcal{S} = \mathcal{S}^1 \times \ldots \times \mathcal{S}^n$.
- The respective transition probabilities, under the assumption of independence, can be computed as:

$$
\begin{aligned}
{}_h p_t^{\mathbf{i},\mathbf{j}} &= P(\mathbf{X}_{t+h} = \mathbf{i} | \mathbf{X}_t = \mathbf{j}) \\
&= P(X_{t+h}^1 = j_1, \ldots, X_{t+h}^n = j_n | X_t^1 = i_1, \ldots, X_t^n = i_n) \\
&= \prod_{k=1}^{n} {}_h^k p_t^{i_k j_k}
\end{aligned}
$$

Or in matrix notation:

$$
{}_h\mathbf{P}_t = {}_h^1\mathbf{P}_t \otimes {}_h^2\mathbf{P}_t \otimes \ldots \otimes {}_h^n\mathbf{P}_t
$$

Multi-state Model for Health Status

To estimate the one step transition probabilities of each Markov chain, we consider two GAMs with binary response, where the variable takes the value of 1 if the contemplated transition occurs in the next step of the chain, and 0 otherwise.

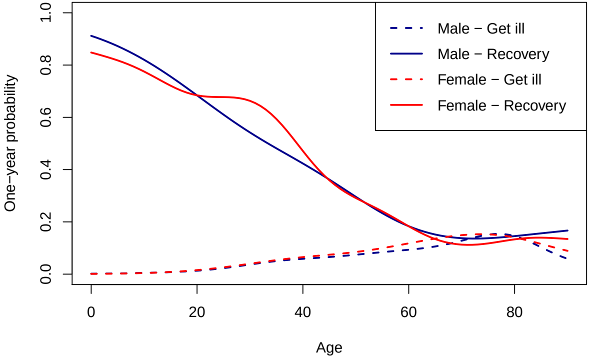
$$log(-log({}^{k}p_t^{12})) = {}^{k}\beta_0^1 + {}^{k}\beta_0^1 \mathrm{Gender} + {}^{k}f_1^1(t) + {}^{k}f_2^1(t) * \mathrm{Gender}$$

$$log(-log({}^{k}p_t^{21})) = {}^{k}\beta_0^2 + {}^{k}\beta_1^2 \mathrm{Gender} + {}^{k}f_1^2(t) + {}^{k}f_2^2(t) * \mathrm{Gender}$$

where the f are soft function of the age, Gender is a dummy variable that takes value of 1 if the individual is male and 0 if female, and * denotes an interaction term.

Multi-state Model for Health Status

Figura: Estimated one-year probabilities of getting ill and recovery for Heart affections (i.e ${}^{9}p_t^{12}$ and ${}^{9}p_t^{21}$)



Outline

Conclusions and Future Work

- These models are statistically consistent for the estimation of transition probabilities, and they are very flexible.
- It is easy to include complex relations, and the amount of effective parameters is much lower than those obtained with usual approaches.
- In the particular case of life tables, the aforementioned methodology solves naturally many problems that usually arises during the estimation (Credibility, Smoothing, Profiling, etc)
- There is great unexplored potential of this models that can be performed in actuarial sciences. Future work should study the case for higher order Markov chains, and other applications.

References

- S. Wood, Generalized Additive Models: An Introduction with R, Second Edition. CRC Press, 2017
- R. Jonsson, Tests of Markov Order and Homogeneity in a Markov Chain Model for Analyzing Rehabilitation. 2011
- G.Fitzmaurice & N Laird & J. WareApplied, Longitudinal Analysis, Second Edition. Wiley, 2004.
- A. Charpentier, Computational Actuarial Science with R CRC Press Book,2014.
- World Health Organization, International statistical classification of diseases and related health problems - 10th revision, Fifth edition, 2016.
- A. Zarruk & A. Villegas & F. Ortiz, Tablas de Mortalidad: Evolución en el Sector Asegurador Colombiano. Fasecolda,2013.
- E.Pitacco, Health Insurance. Springer, 2014.

GRACIAS