Finite Mixture Models for an underlying zero-one inflated Beta distribution

Cédric Noel*, Jang Schiltz**

*Université de Lorraine, IUT de Thionville-Yutz
Espace Cormontaigne, Impasse Alfred Kastler
57970 Yutz, France
University of Luxembourg, Department of Mathematic
E-mail: cedric.noel@univ-lorraine.fr
**Department of Finance, University of Luxembourg
6, rue Richard Coudenhove-Kalergi
L-1359 Luxembourg, Luxembourg
E-mail: jang.schiltz@uni.lu

Résumé. Les versions actuelles de modèles de mélanges finis avec des distributions bêta sous-jacentes ont comme problème que les données doivent vérifier $0 < y < 1$. Nous résolvons ce problème en utilisant la distribution bêta avec excès de zéro et un à la place et présentons notre package **R trajeR** qui permet de calibrer le modèle. Nous illustrons certaines des possibilités de **trajeR** au moyen d'un exemple avec des données simulées.

1 Introduction

Finite Mixture Models in the sense of Nagin (2005) are fuzzy logic cluster analysis models for time series. Starting from a sample of trajectories, the aim is to detect a number of subgroups of the sample, so that subjects in the same group exhibit quite similar data trajectories, whereas two subjects from two different groups have trajectories that differ in some sense. These models are part of a larger strand of models that analyze latent evolutions in longitudinal data.

Group-based trajectory models have no random effects to capture individual differences in a continuous way. All individual deviations from the typical trajectory of a group are treated as residual errors.

Initially, in finite mixture models, the error variance is supposed to remain constant over all groups and the complete time line. This implies that far fewer parameters need to be estimated and these kind of models remain useful for small samples. Schiltz (2015) introduced a generalization of this model in which the error variance remains constant inside a given group, but is allowed to differ across groups, which is a far more realistic assumption in a lot of applications.

While the conceptual aim of the analysis is to identify clusters of individuals with similar trajectories, the model's estimated parameters are not the result of a classical cluster analysis, but of maximum likelihood estimation Nagin (2005).

Finite mixture models for an underlying Beta distribution was introduced by Elmer et al. (2018) and a generalization of it to time-varying variances in Noel et Schiltz (2024). The family of Beta distributions has the big advantage of containing very diverse density shapes which makes it very interesting for modeling in a lot of real life applications. The two previous models present however the disadvantage that it cannot handle data with values of 0 or 1. We overcome that difficulty by presenting a model using a zero-one inflated Beta distribution instead and presents th **R** package **trajeR** which allows to calibrate the model. The package can be seen as an extension of the **SAS** procedure **Proc Traj** developed by Jones et al. (2001), Jones et Nagin (2005) and their **Stata** version (Jones et al., 2001) with the additional advantage that, as an **R** package, it is open source software and can be adapted to the needs of the users.

2 Finite mixture models

In group based trajectory modeling, we consider a population of size N divided into K latent classes. The assignment of the individuals into classes is based on the degree of similarity of the developmental trajectories.

More precisely, consider a time-varying variable of interest Y defined in a population Ω of size N. Let $Y_i = y_{i_1}, \cdots, y_{i_T}$ be T measures of the variable Y, taken at times $t_1, \cdots, t_T$ for subject number i.

The aim of the analysis is to divide the population into K sub-populations $G_1, ...G_K$, which are homogeneous in the sense that two subjects in the same group have similar trajectories for the variable of interest Y and two subjects in different groups have quite different trajectories for the variable of interest Y.

Let $P^k(Y_i)$ be the probability of Y_i given membership in group G_k and $P(Y_i)$ the unconditional probability of observing the realization Y_i of Y. Furthermore, for a given group G_k, we suppose conditional independence for the sequential realizations of the elements y_{i_t} over the T periods of measurements. Then,

$$P(Y_i) = \sum_{k=1}^{K} P\left(G_i = k\right) P^k(Y_i). \tag{1}$$

By definition of a finite mixture model (Nagin, 2005), the density f of Y is given by

$$f(y_i; \psi) = \sum_{k=1}^{K} \pi_k g_k(y_i; \Theta_k). \tag{2}$$

The role of the parameters Θ_k is to describe the shape of the trajectories in group k.

This models supposes no structural between-subject variability within a class, hence the error variance is assumed to be constant inside a given class. Moreover, the group size $\pi_k > 0$ denotes the probability of a given subject to belong to group number k and

thus

$$\sum_{k=1}^{K} \pi_k = 1.$$

Since in practice, it is difficult to constraint the π_k to be numbers between 0 and 1, we link the π_k to a set of parameters $\theta_1, ..., \theta_K$ such that

$$\pi_k = \frac{e^{\theta_k}}{\sum_{k=1}^{K} e^{\theta_k}}.$$

The model depends thus on the parameter set $\psi = (K, \theta_1, \cdots, \theta_{K-1}, \Theta_1, \cdots, \Theta_K)$.

There are two interesting generalizations of this basic finite mixture model. First, we can test if group membership of the individuals is influenced by a static set of R risk variables $X = (X_1 \cdots X_R)$, that are typically measured before the start of the trajectories Y. Second, we can investigate the relationship between the trajectories Y and a time-dependent covariate W which is independent of X. Let's point out, that this is basically regression analysis, so there is no automatic proof of causality here. We test if there is a significant relationship between Y and W and can only conclude that W influences the trajectories of Y if we get the causal status of the relationship by other means.
Combining these two extensions, the conditional density of Y given X and W is given by

$$f(y_i|x_i, w_i) = \sum_{k=1}^{K} \left(P(G_i = k|X_i = x_i) \prod_{t=1}^{T} P(Y_{i_t} = y_{i_t}|X_i = x_i, W_i = w_i, G_i = k) \right), \tag{3}$$

which can be written as

$$f(y_i|x_i, w_i) = \sum_{k=1}^{K} \left(\sum_{j=1}^{R} \frac{e^{x_i^j \theta_k^j}}{1 + e^{x_i^j \theta_k^j}} \prod_{t=1}^{T} P(Y_{i_t} = y_{i_t}|W_i = w_i, G_i = k) \right). \tag{4}$$

Nagin (2005) defined this model for underlying logit, (censored) normal and zero inflated Poisson distributions. In this paper, we extend the model to an underlying Beta distribution.

3 The Beta distribution

The Beta distribution is quite useful for modelizing percentages and proportions, since it takes values between 0 and 1. Its density depends on the two positive shape parameters α and β. The primary advantage of this distribution is the flexibility of the shape of its density, as can be seen in Figure 1.

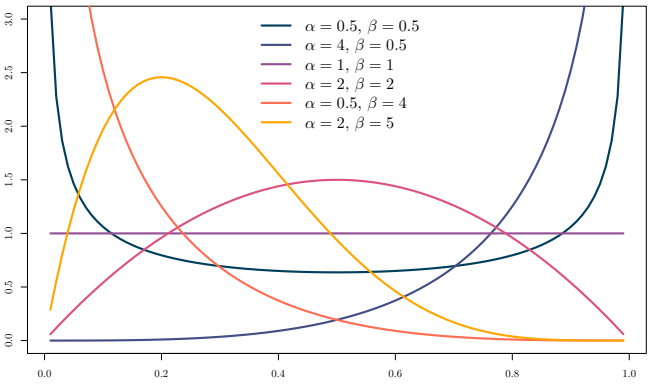


FIG. 1 – *Example of different shapes of the Beta density for some parameters.*

This flexibility allows to capture a lot of non normal distributions within the interval $[0; 1]$, which should help to resolve the problem that common Gaussian mixture models may extract too many latent groups due to non-normality of the outcome as highlighted by Bauer et Curran (2003).

In a regression context, another parametrization of the density is commonly used (Ferrari et Cribari-Neto, 2004). Let Y be a random variable following a Beta distribution with mean μ. Consider the parameter ϕ defined by

$$var(Y) = \frac{\mu(1-\mu)}{1+\phi}.$$

ϕ can be interpreted as a precision parameter, in the sense that a large value of ϕ implies a small variance of Y. The density Beta of Y can be written as

$$\text{Beta}(y; \mu; \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},$$

where $0 < \mu < 1$ and $\phi > 0$.

4 Finite mixture models for the zero-one inflated Beta distribution

As mentioned in the last paragraph, Beta distributions only have values between 0 and 1, which might seem quite reductive. But any continuous variable Y with bounds (a, b) can of course be transformed to the standard unit interval using $Y^* = (Y - a)/(b - a)$ and so the use of Beta distributions could actually be meaningful for most data sets.

For the finite mixture model with underlying zero-one inflated Beta distribution, called here the ZOIB model, we consider that

$$P\left(Y_{it}=y_{it}|W_i=wi,C_i=k\right)=\begin{cases}\rho_{ikt}^{01}\left(1-\rho_{ikt}^{1}\right) & \text{if } y_{it}=0\\ \left(1-\rho_{ikt}^{01}\right)\text{Beta}(\mu_{ikt},\phi_{ikt}) & \text{if } 0<y_{it}<1\\ \rho_{ikt}^{01}\rho_{ikt}^{1} & \text{if } y_{it}=1,\end{cases} \tag{5}$$

where ρ_{ikt}^{01} is the probability of the zero-one inflation, i.e. the probability that the variable Y takes the value 0 or 1, ρ_{ikt}^{1} is the conditional probability of inflation one, i.e. the probability that the variable takes the value 1 knowing that it is either 0 or 1.

As in the classical BETA model, we then establish the relationship between the trajectory and the time variable using the formulas

$$\mu_{ikt}=\frac{e^{\beta_k A_{it}+\delta_k W_{it}}}{1+e^{\beta_k A_{it}+\delta_k W_{it}}} \text{ and } \phi_{ikt}=e^{\zeta_k A_{\zeta,it}}, \tag{6}$$

where $A_{it}=(1,a_{it},a_{it}^2,\cdots,a_{it}^{n_\beta-1})^t$, $A_{\zeta,it}=(1,a_{it},a_{it}^2,\cdots,a_{it}^{n_\zeta-1})^t$, $W_{it}=(w_{i1},\cdots,w_{in_\delta})^t$, $\beta_k=(\beta_{k1},\cdots,\beta_{kn_\beta})$ and $\zeta_k=(\zeta_{k1},\cdots,\zeta_{kn_\zeta})$. The rho parameters introduced above are modeled as :

$$\log\left(\frac{\rho_{ikt}^{01}}{1-\rho_{ikt}^{01}}\right)=\nu_k^{01}A_{it} \tag{7}$$

and

$$\log\left(\frac{\rho_{ikt}^{1}}{1-\rho_{ikt}^{1}}\right)=\nu_k^{1}A_{it}. \tag{8}$$

Thus the log-likelihood of the model becomes

$$l(\psi;y)=\sum_{i=1}^{n}\log\left(\sum_{k=1}^{K}\pi_k g_k(y_i;\beta_k,\delta_k,\zeta_k,\nu_k^{01},\nu_k^{1}),\right) \tag{9}$$

where

$$g_k(y_i;\beta_k,\delta_k,\zeta_k,\nu_k^{01},\nu_k^{1})=\prod_{y_{it}=0}\rho_{ikt}^{01}(1-\rho_{ikt}^{1})\prod_{y_{it}=1}\rho_{ikt}^{01}\rho_{ikt}^{1} \tag{10}$$

$$\times\prod_{0<y_{it}<1}\left(1-\rho_{ikt}^{01}\right)\frac{\Gamma(\phi_{ikt})}{\Gamma(\mu_{ikt}\phi_{ikt})\Gamma((1-\mu_{ikt})\phi_{ikt})}y_{it}^{\mu_{ikt}\phi_{ikt}-1}(1-y_{it})^{(1-\mu_{ikt})\phi_{ikt}-1}. \tag{11}$$

5 The R package trajeR

The package **trajeR** is built around the core function `trajeR` which fits the model and computes its parameters for given degrees of the polynomial trajectories in the different groups. The function signature for `trajeR` is

```
R> trajeR(Y, A, Risk = NULL, TCOV = NULL, degre, degre.phi = 0,
+         Model, Method = "L",
+         ssigma = FALSE, ymax = max(Y) + 1, ymin = min(Y) - 1,
+         hessian = TRUE, itermax = 100, paraminit = NULL,
+         ProbIRLS = TRUE, refgr = 1, + fct = NULL, diffct = NULL,
+         nbvar = NULL, nls.lmiter = 50)
```

Some of these arguments are mandatory others optional.

The mandatory arguments are the main data matrices `Y`, `A`, as well as `degre`, `degre.phi`, `Model` and `Method`.

Here `Y` is the matrix containing the values of the variable of interest and `A` is the matrix containing the age or time variable. In most applications, this matrix just contains times of measurement that are the same for each individual in the sample, implying that all lines of the matrix `A` are equal, but this is not necessarily the case. `A` can for instance contain the age of the different individuals at the times of measurement, which is generally different for each individual in the sample.
`degre` is a vector indicating the degree of the polynomials describing the typical trajectories in the different groups. Implicitly, the dimension of this vector also indicates the number of groups into which we want to divide the population, so there is no no need for a seperate argument for the number of groups.
`degre.phi` is a vector indicating the degree of the polynomials describing the precision parameters in the different groups.
`Model` is a string defining the underlying distribution used in the model. The possible choices are LOGIT for the Logistic Regression Mixture Model, CNORM for the Censored Normal Mixture Model, ZIP for the Zero Inflated Poisson Mixture model, BETA for the BETA model and ZOIB for the zero-one inflated Beta Model.
`Method`, finally, is a string to decide which algorithm is used for estimating the model parameters. In case of the BETA and ZOIB models, only `L` for direct optimization is possible.

The optional arguments are `Risk`, `TCOV`, `degre.nu`, `ssigma`, `ymax`, `ymin`, `hessian`, `itermax`, `paraminit`, `ProbIRLS`, `refgr`, `fct`, `diffct`, `nls.lmiter`, `ng.nl` and `nbvar`.

`Risk` is a data matrix that contains the values of the covariate X modifying the group membership probability. By default, there is no such variable and `Risk` is a one-column matrix with value 1.
`ProbIRLS` allows to decide which method is used to compute the predictor probabilities. If its value is TRUE (default setting) we use the IRLS method and if it is FALSE we use the optimization method.
`TCOV` is an optional data matrix containing a time-dependent covariate W that influences the trajectories themselves. By default its value is NULL.

To ensure the identifiability of the parameters of the predictor, we have to fix a reference group. This can be done by the `refgr` command. It's default value is 1.
`hessian` indicates if we want to calculate the Hessian matrix, the default value being FALSE. If the method used is direct optimization, the Hessian matrix is computed by inverting the Fisher Information Matrix.
`itermax` gives the maximal number of iterations for the `optim` function or for the EM algorithm.
The choice of the initial parameters is very important in optimization problems. We can specify these initial parameters by `paraminit`. By default `trajeR` calculates the initial value based on the range or the standard deviation of the data (for the details, see Noel (2024)).

The output of `trajeR` for the ZOIB model is an object of class `Trajectory.ZOIB`.

5.1 Numerical application with simulated data

We generate 50 samples of 200 individuals, each of which follows a beta distribution containg zeros and ones. These individuals are divided into two groups with the following characteristics :

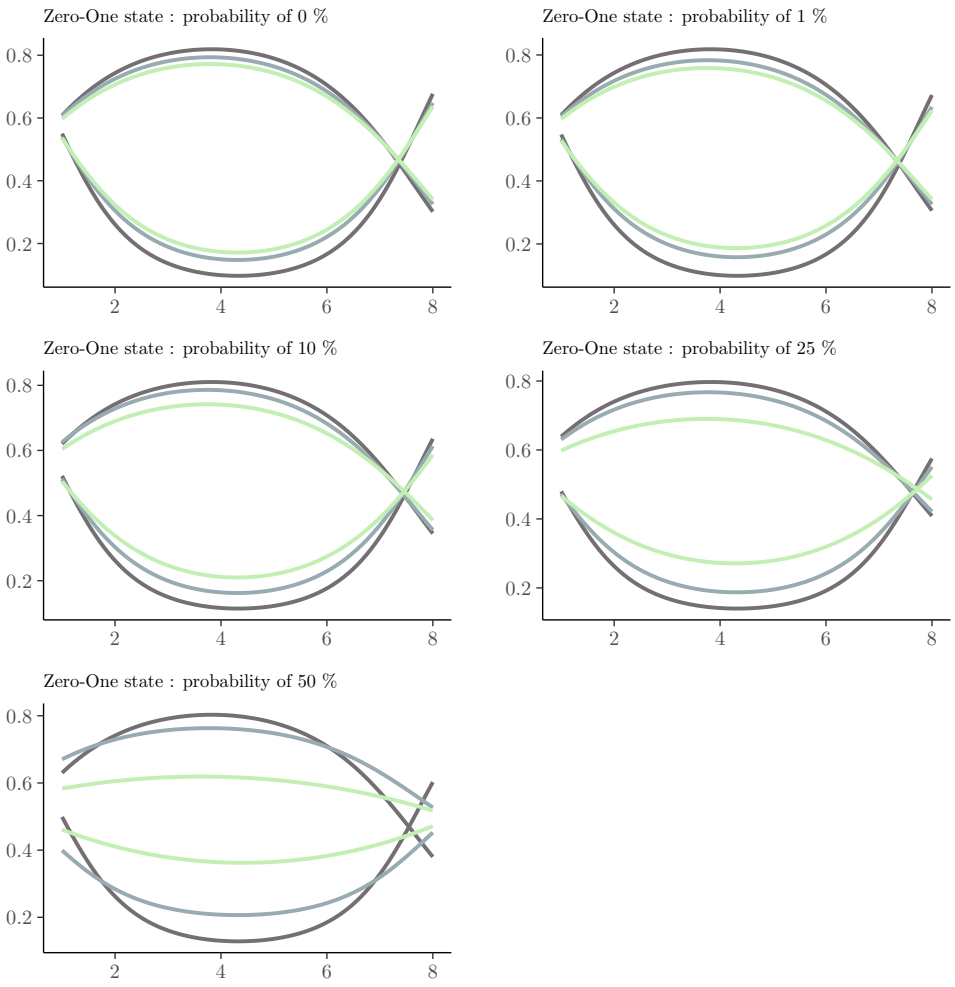
- Group 1 : $\pi_1 = 0.4$, $\beta_1 = (1.8833, -1.90201, 0.219884)$
- Group 2 : $\pi_2 = 0.6$, $\beta_2 = (-0.456576, 1.02901, -0.134615)$
- For group 1, $\nu_1^1 = -1$, and for group 2, $\nu_2^1 = 1$, so that the probabilities for a given value to be either 0 or 1 are 0.73 and 0.27 respectively.

We test several hypotheses about ν^{01} to measure the accuracy of the ZOIB model compared to the Beta model. We set ν_1^{01} and ν_2^{01} to 0, -1.1, -2.19, -4.6, and -50, making the probability of belonging to the zero-one state approximately 50%, 25%, 10%, 1%, and 0%, respectively. This allows us to assess the model's performance as the number of zeros or ones increases.

We fit the classical model with underlying beta distribution, as well as our model with underlying zero-one inflated beta distribution and we compare the values of the parameters.

We note that when the number of zeros or ones is very small, the two models perform similarly. However, as the number of zeros or ones increases, the results of the two models begin to diverge significantly. The Beta model tends to be less accurate than the ZOIB model in these cases. To capture the excess of zeros or ones, the Beta model must increase its variability to be more flexible. Conversely, the ZOIB model captures this excess state through its logistic component, allowing the beta component to more accurately capture the variability of the data.

Finite Mixture Models for an underlying zero-one inflated Beta distribution



Références

Bauer, D. J. et P. J. Curran (2003). Distributional assumptions of growth mixture models : implications for overextraction of latent trajectory classes. *Psychological methods 8*(3), 338.

Elmer, J., B. L. Jones, et D. S. Nagin (2018). Using the beta distribution in group-based trajectory models. *BMC medical research methodology 18*, 1–5.

Ferrari, S. et F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics 31*(7), 799–815.

Jones, B., D. Nagin, et K. Roeder (2001). A sas procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research Method Research 29*, 374–393.

Jones, B. L. et D. Nagin (2005). Advances in group-based trajectory modeling and a sas procedure for estimating them. *Ann Am Acad Pol Soc Sci 602*, 82–117.

L Jones, B. et D. Nagin. A stata plugin for estimating group-based trajectory models.

Nagin, D. (2005). *Group-based modeling of development*. Harvard University Press.

Noel, C. (2024). *On a generalisation of Nagin's finite mixture model*. Phd thesis, University of Luxembourg.

Noel, C. et J. Schiltz (2022). trajer, an r package for cluster analysis of time series. *Working Paper*.

Noel, C. et J. Schiltz (2024). Finite mixture models for an underlying beta distribution with an application to covid-19 data. *M. Stemmler, A. Von Eye & W. Wiedermann (eds.) Dependent Data in Social Sciences Research. Second Edition.*.

Schiltz, J. (2015). A generalization of nagin's finite mixture model. *M. Stemmler, A. Von Eye & W. Wiedermann (eds.) Dependent Data in Social Sciences Research.*, 107–126.

Smithson, M. et J. Verkuilen (2006). A better lemon squeezer ? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods 11*(1), 54.

Summary

Current versions of finite mixture model with underlying Beta distributions have the problem that the data y have to verify $0 < y < 1$. We resolve that problem by using the zero-one inflated Beta distribution instead and present our **R** package **trajeR** which allows to calibrate the model. We illustrate some of the possibilities of **trajeR** by means of an example with simulated data.