Proceedings of Statistics Canada Symposium 2009
Longitudinal Surveys: from Design to Analysis

Longitudinal Mixed-Membership Models for Survey Data on Disability

Daniel Manrique-Vallier, Stephen E. Fienberg[1]

Abstract

When analyzing longitudinal data we need to balance our understanding of individual variability with the production of meaningful and interpretable summaries of overall population tendencies. This is specially true when those in the target population are know to be heterogeneous in their ways of progressing over time due to unobserved individual traits. Additional complications arise when the data are discrete and multivariate so that the resulting contingency tables are very sparse. We propose a new family of models to analyze such data by combining features from a version of the cross-sectional Grade of Membership Model (Erosheva et al., 2007) and from the longitudinal Multivariate Latent Trajectory Model (Connor, 2006) and them to data the National Long Term Care Survey (NLTCS), a longitudinal survey with six completed waves aimed to assess the state and characteristics of disability among U.S. citizens age 65 and above. These models assume the existence of a small number of "typical" or "extreme" classes of individuals and model their evolution over time. We regard individuals as belonging to all of these classes in different degree, by considering them as convex weighted combinations of the extreme classes. In this way, we are able to describe distinct general tendencies (the extreme cases) while accounting for the individual variability. We propose a full Bayesian specification and estimation methods based on Markov chain Monte Carlo sampling. We illustrate the our methods using data from the NLTCS.

Key Words: Bayesian Hierarchical Model, Grade Of Membership Model, Latent Trajectories, Markov Chain Monte Carlo.

1. Introduction

In this paper we propose models and estimation procedures to deal with discrete multivariate longitudinal data obtained from a heterogeneous population. This work is motivated by the analysis of data arising from the National Long Term Care Survey (NLTCS), a longitudinal panel survey instrument aimed to assess chronic disability among the elderly population in the United States. Through the analysis of the NLTCS data, researchers seek to answer important questions related to the aging process and disability prevalence in the U.S.: How many elder Americans will live with disabilities? What is the of duration of disability episodes? What is the age of onset of disability? Is it changing for younger generations? (see e.g. Connor et al. (2006)). Answers to these questions are of great importance in public policy design due to, among other reasons, the increased public and private expenditure for disabled people in contrast with their able peers (Manton et al., 1997).

Many of the relevant public policy questions for which the NLTCS can potentially provide answers have to do with changes over time: changes during the life of an individual ("how is this individual likely to age?") or comparing people across different generations ("are people from later generations acquiring disabilities differently than people born 20 years before?"). To answer these questions we need to look at the data longitudinally. In addition, elderly American people are known to be a heterogeneous population, as not everyone could be expected to age the same way. Thus models for longitudinal disability data need to be capable of representing such heterogeneity.

In this paper we present models and methods for their analysis that seek to capture both the longitudinal nature of the NLTCS and the complexity in the heterogeneity of the human aging process, combining the ideas of mixed membership from the Grade of Membership model (Woodbury et al., 1978; Erosheva et al., 2007) and the longitudinal descriptions of the aging process from the Latent Trajectory family (Nagin 1999; Connor, 2006). We illustrate the methods using data from the six waves of the NLTCS.

[1] Daniel Manrique-Vallier, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. USA; Stephen E. Fienberg, Department of Statistics and Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213. USA.

2. Data - The National Long Term Care Survey

The National Long Term Care Survey (NLTCS) is a longitudinal panel survey aimed at assessing chronic disability among elderly population in the United States. Its target population are people aged 65 years and older that present functional limitations lasting or expected to last 90 or more days (White, 2008). So far the survey has gone through six waves conducted in 1982, 1984, 1989, 1994, 1999 and 2004.

The sampling frame of the NLTCS is the Medicare record system, which provides a good representation of the elderly population of the U.S. since near 97% of Americans aged 65 or older are included in it (Corder and Manton, 1991). After an initial selection, according to a complex sample design, every individual in the sample is screened to detect if he or she presents a functional limitation. Those who are screened-in are then given a detailed questionnaire, and re-interviewed at each survey wave until they die. Those who were screened-out are re-screened on subsequent waves to check if their functional status has changed. At each wave, a new cohort (approx. 5,000 individuals) is sampled to replace those who died, so that the sample size for each wave is kept at around 20,000 individuals (Clark, 1998). So far 45,009 unique individuals have been interviewed, considering all waves.

The NLTCS approaches disability through the measurement of each individual's capacity to perform a set of six “Activities of Daily Living" (ADL) such as eating, bathing or dressing and ten “Instrumental Activities of Daily Living" (IADL) such as preparing meals or maintaining finances. Broadly stated, ADLs seek to measure a person's ability to take care of him or herself at a fundamental level, while IADLs measure the ability of living independently within a community (Connor, 2006). The survey instrument registers these measurements as a series of answers to triggering questions that are then summarized into a set of binary responses. These binary responses indicate the presence or absence of impairments to perform such activities.

3. Methods

The goal of our analysis is to characterize typical progressions of acquisition of disabilities over time, while taking into consideration and characterizing the heterogeneity of the population. We have proceeded by combining two previously employed methods.

The first method is the latent trajectory model (Nagin, 1999), which is specially well suited in applications where the researcher wants to understand typical evolutions over time and suspects that the population is heterogeneous but a small number of homogeneous classes might exist. Connor (2006) adapted this technique for the analysis of multivariate discrete data and applied it to the NLTCS analysis, identifying latent trajectory curves of probability of acquiring a disability over time. This tool provides a flexible and easy to interpret representation of the data that allows for latent heterogeneity in the population, handling it by clustering the population into *exclusive* classes. In Connor's formulation, this assumption essentially says that, within a class, every single individual responds to the exact same underlying aging process. All the response variability within class is thus attributed to random fluctuations within that class, disregarding the fact that these classes are ideals to which quite possibly no real individuals actually belong (Kreuter and Muthén, 2008).

The Grade of Membership (GoM) family of models (Woodbury et al., 1978; Erosheva et al., 2007) provides a conceptually attractive way relaxing this assumption. Instead of forcing every single individual into one and only one class, the GoM model seeks to identify *pure types* or extreme profiles and then assumes that every individual belongs to more than one of them in different degree. In this way it retains the interpretative power of specifying a reduced number of “typical" or “extreme" profiles but adds extra flexibility by not assuming exclusive membership.

3.1 Notation and setup

We will use the following notation and structure:

1. There are N subjects in the sample, indexed by i, and N_i measurements for each subject $i \in \{1,...,N\}$;

2. For each individual, we measure *J* binary variables simultaneously in each measurement event. The manifest response vector for individual *i* and question *j* is $y_{ij*} = (y_{ij1},...,y_{ijN_i})$ [2];
3. Each individual has an associated covariate vector $X_{i.}$. In this application we will only consider a vector of time dependent covariates $X_{i*} = (X_{i1},...,X_{iN_i})$, although time invariant and other more complicated structures of covariates can also be considered.

3.2 A Grade of Membership multivariate trajectory model

We start modeling the marginal distribution of the response to question $j \in \{1,...,J\}$ at measurement time $t \in \{1,...,N_t\}$, y_{ijt}, for a full member of extreme profile *k* (i.e. an individual *i* such that g_{ik}=1 and $g_{ik'}$=0 for $k' \neq k$) as a function of some covariates registered at time *t*, X_{it},

$$Pr\left(Y_{ijt} = y_{ijt} \middle| \; g_{ik} = 1, X_i\right) \quad = \quad f_{\theta_{j|k}}\left(y_{ijt} \middle| \; X_{it}\right)$$

and model the same marginal distribution of response for a generic individual with membership vector $G_{i*} = (g_{i1},...,g_{iK})$ as the convex combination,

$$Pr\left(Y_{ijt} = y_{ijt} \middle| G_{i*} = (g_{i1},...,g_{iK}), X_{i*}\right) = \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijt} \mid X_{it})$$

Now, assuming that conditional on the membership vector, g_i, and the covariates, X_{i*}, the responses are independent between items and measurements,

$$Pr\left(Y_{i**} = y_{i**} \middle| G_{i*} = (g_{i1},...,g_{iK}), X_{it}\right) \quad = \prod_{j=1}^{J} \prod_{t=1}^{N_i} \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijt} \mid X_{it})$$

which combined with the assumption of random sampling gives us the joint model

$$Pr\left(Y_{***} = y_{***} \middle| g_{**}, X_{**}\right) \quad = \quad \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{t=1}^{N_i} \sum_{k=1}^{K} g_{ik} f_{\theta_{j|k}}(y_{ijl} \mid X_{it})$$

This model is similar to the joint latent class trajectory model proposed by Connor (2006) where we are generalizing the clustering from full membership (i.e. $g_{ik} = 1$ for some *k*) to mixed membership.

Following Connor (2006), in this implementation we choose the distribution function $f_{\theta_{j|k}}(y \mid X_i)$ for the single response of pure-type individual of extreme profile *k* as $f_{\theta_{j|k}}(y_{ijt} \mid X_{i*}) = \lambda_{jt|k}(X_{it})^{y_{ijt}}\left(1 - \lambda_{jt|k}(X_{it})\right)^{1-y_{ijt}}$ with $\lambda_{jt|k}(X_{it}) = logit^{-1}\left(\beta_{0j|k} + \beta_{1j|k} Age_{it}\right)$, where Age_{it} is the age of the *i*th individual at measurement time *t*. Note that under this specification, $\lambda_{jt|k}$ is actually a time dependent function. This specification has the advantage of being relatively simple, with just $2 \times J$ parameters per extreme profile and of representing the intuitively sound notion that the underlying probability of disability is a monotonic (increasing) function of age. Other specifications are certainly possible.

We regard the *N* membership vectors, g_{i*}, to be iid realizations from a common distribution, G_α, with support on

---

[2] Throughout this document we will use this notation when we want to refer to the vector that results from fixing a subset of the sub indexes of an indexed variable while letting the rest to vary. For instance if we have the collection of scalar variables λ_{jk} with $j \in \{1,..,J\}$ and $k \in \{1,...,K\}$ we can write the vectors $\lambda_{*k} = (\lambda_{1k}, \lambda_{2k},...,\lambda_{Jk})$, and $\lambda_{j*} = (\lambda_{j1}, \lambda_{j2},...,\lambda_{jK})$.

the K–1 dimensional unit simplex, Δ_{K-1}. Similar to Erosheva et al. (2007), we model that distribution as $g_{i*} \mid \alpha \sim^{iid} Dirichlet(\alpha)$.

The Dirichlet distribution in this setting has some important properties. In the first place, it is conjugate to the multinomial distribution, facilitating a great deal the computations using Gibbs samplers; second, adopting the re-parametrization $\alpha = (\alpha_0 \cdot \xi_1, ..., \alpha_0 \cdot \xi_K)$ with $\alpha_0 > 0$, $\xi_k > 0$ and $\sum_k \xi_k = 1$ we can interpret the vector ξ_* as the average proportion of the population in the k-th extreme profile and α_0 as a parameter governing the spread of the distribution: as α_0 approaches 0, the samples from G_α are more and more concentrated towards the vertices of Δ_{K-1} and; as α_0 increases they are more concentrated near the distribution's average.

As Erosheva et al. (2007) and Airoldi et al. (2007) discuss, a priori setting the parameters α for the Dirichlet distribution might be too strong an assumption to do realistic modeling. We will estimate these parameters from the data specifying hyper-priors and computing posterior distributions. For this purpose we use hyper priors for α_0 and ξ_* similar to the ones in Erosheva (2002) and Erosheva et al. (2007): $\alpha \sim Gamma(\tau,\eta)$, $\xi \sim Dirichlet((1,...,1)_k)$ (Uniform over Δ_{K-1}) and complete the specification with the priors $\beta_{0j|k} \sim^{iid} N(\mu_0, \sigma_0^2)$ and $\beta_{1j|k} \sim^{iid} N(\mu_1, \sigma_1^2)$, with β_0 independent from β_1.

4. Results

To test our methods, we have applied them to data from six waves from the NLTCS (N=45,009), analyzing the responses for the six ADLs[3] (J=6) using a total of K=3 extreme profiles. The posterior estimation has been performed using a custom Markov chain Monte Carlo (MCMC) algorithm based on a latent class representation first proposed by Erosheva (2002) and Erosheva et al. (2008). For this application we have chosen the priors $\beta_{0j|k}, \beta_{1j|k} \sim^{iid} N(0,100)$ and $\alpha_0 \sim Gamma(1,5)$.

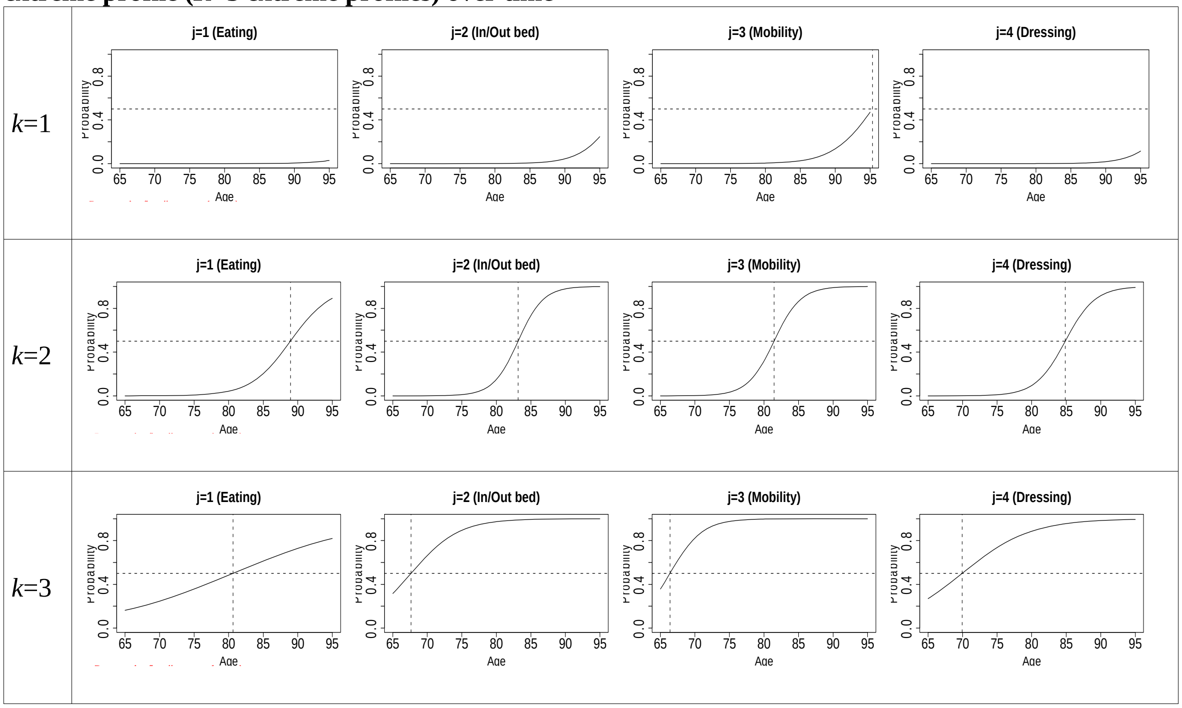
Figure 4-1 shows the curves of the estimated posterior probability of acquiring a disability in each ADL (only four are shown) as a function of age, for the *ideal* members of the three extreme profiles, similar to the ones in Connor (2006). They appear reflect some desirable features: the slopes of all extreme profiles are positive, showing the expected increasing tendency of the probability of suffering a disability as time passes and the ages where the idealized individuals of the extreme profiles reach a probability of 0.5 of acquiring a disability are within reasonable ranges. Also, the posterior distribution of the parameters are quite concentrated around a central value (not shown), opposed to their prior specification. From the picture, we can see that the method has identified three well separated profiles that reflect quite different aging processes: a class of people that live relatively healthy until very late (k=1); a class of people that remain healthy until around the age of 85, when they experiment a sudden increase in their probability of acquiring disabilities (k=2) and; a class of people that have an early increase of the probability of getting disabled (k=3).

Table 4-1 shows point estimates (posterior means) for parameters α_0 and ξ, reflecting how particular individuals age, opposed to the ideal ones. Parameter estimate ξ_* = (0.65, 0.25, 0.1) indicates the relative order of importance of each of the three extreme profiles, showing that more people are closer to profile k=1, followed by k=2 and k=3. Parameter estimate α_0= 0.264 indicates that the distribution over the population is quite concentrated towards the vertices of the simplex Δ_{K-1}, although not as much to make the model behave like a regular mixture model. Figure 4-2 shows an example of how the individual trajectories are formed from the extreme profiles. As can be seen, the model is quite flexible, allowing quite varied individual trajectories, but extracting just a few simple extreme curves that are easy to characterize and interpret.

---

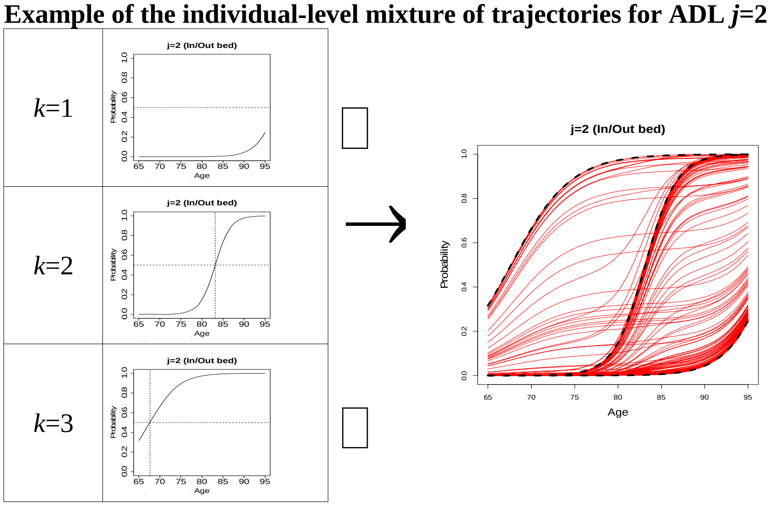
[3] The six ADLs are: Eating (j=1), getting in or out of bed (j=2), inside mobility (j=3), dressing (j=4), bathing (j=5), toileting (j=6).

Figure 4-1
Trajectories of the probability of acquiring a disability in the first four ADLs, for ideal members of each extreme profile (*K*=3 extreme profiles) over time



The intersection of the straight lines indicates the point where the probability of acquiring a disability for ADL *j* reaches 50%.

Figure 4-2
Example of the individual-level mixture of trajectories for ADL *j*=2 (getting in or out of bed)



The plot on the right superimposes the three extreme trajectories and presents a sample of 100 individual trajectories.

Table 4-1
Posterior means for population-level parameters for model with *K*=3 extreme profiles

Parameter	Estimate [sd.]
α_0	0.264 [0.00489]
(ξ_1, ξ_2, ξ_3)	(0.65 [0.004], 0.25 [0.003], 0.104 [0.002])

Numbers between brackets are posterior standard deviations.

5. Discussion

Our preliminary results are interesting because they show the potential of our methods. In our application, using longitudinal data from the NLTCS, we have been able to characterize well separated and intuitively sound extreme profiles that can be understood as typical ways of aging, while at the same time characterizing the heterogeneity of the population using a very simple device that allows to construct individualized curves from the extreme profiles. This way of handling heterogeneity, although slightly more complicated that the one proposed in Connor (2006), allows us to be able to keep the number of extreme profiles low and interpretable while avoiding the introduction of too strong in-class homogeneity considerations.

The model presented in this paper is a basic implementation of the general idea of combining complete time dependent trajectories using a mixed membership device. Depending on the problem at hand, there are a number of obvious extensions that can be worked out, some of which we are developing at the moment. In our application, the the NTLCS, some of these natural extensions are the inclusion of other covariates at the group membership level and at the extreme profile level and the joint formulation with survival models to study the relationship of disability and mortality. Many of these extension will be included in Manrique-Vallier (2010).

For purposes of illustration, we have chosen to illustrate the methodology with *K*=3 extreme profiles. More generally, we need to incorporate methodology for deciding on an optimal value of *K*. We have carried out full computation for a series of values of *K*, running from 2 through 5. While the fit of the model, as measured in terms of the posterior predictive responses, increases with *K*, we observed less separation of profiles for *K*=4 and *K*=5, and a less satisfactory interpretation of the shape and structure of the profiles. Choosing an appropriate value of *K* remains an open problem in our work that will be addressed in Manrique-Vallier (2010).

References

Airoldi, E., Fienberg, S., Joutard, C. and Love, T. (2007). Discovering Latent Patterns with Hierarchical Bayesian Mixed-Membership Models, *Data Mining Patterns: New Methods and Applications*, 240–275.

Clark, R. (1998). An Introduction to the National Long-Term Care Survey, Office of Disability, Agigng, and Long-Term Care Policy within the U.S. Department of Health and Human Services.

Connor, J., Fienberg, S., Erosheva, A. and White, T. (2006). Towards a restructuring of the national long term care survey: A longitudinal perspective, Tech. rep.

Connor, J. T. (2006). Multivariate Mixture Models to Describe Longitudinal Patterns of Frailty in American Seniors, Ph.D. thesis, Department of Statistics & H. John Heinz III School of Public Policy & Management. Carnegie Mellon University.

Corder, L. and Manton, K. (1991). National surveys and the health and functioning of the elderly: the effects of design and content, *Journal of the American Statistical Association*, 86, 513–525.

Erosheva, E. (2002). Grade of membership and latent structures with application to disability survey data, Ph.D. thesis, Department of Statistics. Carnegie Mellon University.

Erosheva, E., Fienberg, S. and Joutard, C. (2007). Describing disability through individual-level mixture models for multivariate binary data, *Annals of Applied Statistics*, 1, 502–537.

Kreuter, F. and Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling, *J Quant Criminol*, 24, 1–31.

Manrique-Vallier, D. (2010). Longitudinal Mixed Membership Models With Applications, Ph.D. Dissertation, Department of Statistics, Carnegie Mellon University, expected May 2010.

Manton, K., Corder, L. and Stallard, E. (1997). Chronic disability trends in elderly United States populations: 1982-1994, *Proceedings of the National Academy of Sciences*, 94, 2593–2598.

Nagin, D. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach, *Psychological Methods*, 4, 139–157.

White, T. (2008). Extensions of Latent Class Transition Models with Application to Chronic Disability Survey Data, Ph.D. thesis, University of Washington.

Woodbury, M., Clive, J. and Garson Jr, A. (1978). Mathematical typology: A grade of membership technique for obtaining disease definition. *Computers in Biomedical Research*, 11, 277–98.