# Mixture and Group-Based Trajectory Models – part 1.

## Transcript

Video: https://youtu.be/8Xa4q4jNQ3Y

Hello, my name is Oliver Perra, and I’m going to give a presentation where I will talk about mixture and group-based trajectory models. This is the first of a series of three presentations.

So in this first presentation, I will explain the logic of trajectory-based growth models by comparing this type of analysis against growth models. I will then outline the key characteristics and parameters of trajectory-based growth models and finally, I will illustrate the criteria for determining the number of groups that we identify in this type of analysis. In the material provided, there will be also some exercises and some further material that you can read.

Many studies we are interested in developmental trajectories, that say that is how an outcome like a behaviour or a competence changes over time. And in many instances, we may consider an example like the one I represent here in this figure, where we may be satisfied that most participants follow a single trajectory on average, with some variability around this trajectory that represents sources of variation unaccounted for. A similar trajectory is defined by parameters that describe the change of the variable and we call these parameters growth parameters. These include an inter-set which represents the variable status at the point where we decide to start our observations, and slope parameters that represent the outcome rate of change per units of time. I will discuss these parameters later in the presentation.

The main point here is that we use a model that we can call ‘one size fits all’ because we assume that participants generally follow this trajectory. And variations across individuals’ growth parameters follow a predefined distribution, for example, a normal distribution around the average parameters.

In many fields, theories assume that the ‘one size fits all’ approach that I just described is inadequate. In clinical developmental psychology, for example, there is a long tradition of theories that assume there are typical developmental trajectories where people do not develop significant mental health issues. But other individuals follow atypical trajectories where they display significant mental health issues at various ages or at various times, for example, following a traumatic event. In this graph here, for example, I uploaded fictional data where the lines represent trajectories of different individuals, and the graph is trying to highlight how there may be different trajectory groups, see, for example, the difference between the red and the black trajectories, the red and the black lines.

Theories that assume different groups of people with different development trajectories, sometimes assume that there are different aetiologies that is different causal mechanisms that can explain these differences, or that different trajectories are the result of exposure to different events over time, and often they assume that these groups of individuals can respond differently to various interventions and treatments.

Here I wanted to give an example of a similar theory, and that’s Moffitt’s Taxonomy of Antisocial Behaviour. I also put the QR code that directs you to this paper. And Moffitt hypothesized that there may be two qualitatively different groups of adolescents that engage in antisocial behaviour – one small group includes adolescents that engage in antisocial behaviour from an earlier age, and continue to engage in antisocial behaviour after adolescence. Here represented this group with the red line in the graph. And Moffitt suggested that the causal mechanisms that lead to this life course, antisocial behaviour, involve early neuropsychological problems that lead to an accumulation of risk factors, for example, family conflict and so on.

A larger group of adolescents instead, the green line here in this graph, engage in antisocial behaviour only during adolescence, and they do so because they want to emulate antisocial peers that are perceived as being more independent at a time when affirming their own independence is very important. Once these adolescents transition into adult roles and acquire financial and effective autonomy, they are no longer motivated to engage in antisocial behaviour.

[00:05:25]

So this is an example, a famous example of a taxonomy of developmental trajectories, a theory that supposes that there are different groups of people that follow different trajectories in antisocial behaviour in this case.

So if we assume that there are different trajectories, trajectory groups in a sample or in a population, how can we identify these groups? Traditionally, researchers had used arbitrary categorisation rules, in the example in this graph, we can plot each individual’s starting point of the outcome, and estimate the variables, the outcome change over time, and come up with some sensible criteria for classification. The problems with arbitrary criteria or ad hoc rules are that we must assume before we start a study that these trajectory groups exist, and we do not have formal statistical tests to check if there is enough variation in the sample we are serving to support the existence of different trajectory groups. Furthermore, arbitrary criteria or ad hoc rules do not provide information about the adequacy and precision of these classifications. And finally, when we are advising ad hoc rules or arbitrary rules, we would rely on their served trajectories, with few means to disentangle the true variation we observe from … variations within the data.

The solution to the problem of ad hoc rules or arbitrary criteria for categorisation is provided by development of mixture and group-based models, which I am presenting here and articulating in the next presentations. I will particularly focus on growth mixture models and latent class growth analysis. And these are person-centred approaches that is they assume that a sample is made up of a mixture of individuals with different propensities for some behavioural of different propensities for change over time in the behaviour we observe in this case. The goal of these approaches is to identify a limited number of groups that adequately explain the variation in individuals’ trajectories we observe. The methods I will present are based on latent class analysis, and if you want to know more about latent class analysis, I have prepared other material for NCRM.

These methods, latent class analysis and growth mixture and latent class growth analysis are based on probability rules and therefore have many advantages compared to arbitrary rules for categorisation. One obvious advantage is that they allow to test a person-centred approach provides a better, more adequate representation of the data compared to a ‘one size fits all’ approach. They also allow us to assess the precision of the categorisation by returning probabilities of individuals being one category, one group or another. So in this way, we can check that the level of uncertainty in assigning participants to trajectory groups is adequate. They also allow to disentangle random variation in the variables from real variation.

And finally, by assuming different trajectory groups, these methods also help relax some restrictive assumptions in ‘one size fits all’ growth models, as I will explain later in this presentation. Before describing the trajectory group models, I wanted to provide an example of latent growth curve model, which is the basis for understanding the person-centred models. We start from an interesting longitudinal data, that is variables collected at least on three different occasions or more. In the figure h ere, the variable is collected over five occasions, for example. The variable can be continuous, like (inaudible 00:09:58), for example, depression versus non-depression, or can be an order categorical variable, for example, drug use from none to occasional to frequent drug use. The change across time can be represented by different models, for example, multilevel regression models, but here I focus on a latent variable model. The idea is that we can represent the intercept, the initial stages of the variable at time zero, the time we take as the start of our observations, as a latent variable. And in the figure here, you see that the arrow, the intercept through the variables are all fixed at one, which means that the relationship between these observed variables and the latent intercept is constant, that is the intercept represents a value of the outcome that does not change, since it is the level at which all individuals start, on average.

The slope is the second latent variable, and explains the rate of change in the outcome. The blue number in the arrows from the slope to the observed variables represents the time factors that multiple the effect of the slope. For example, the association between the latent slope and the variable at the start of the study is multiplied by zero, meaning that the level of the variable at the start of the study is only affected by the intercept. And the effect of the slope on the observed variables is then multiplied by numbers representing an increase in unit of times. So if the slope is estimated as being 1.5, for example, the level of the outcome at time 1 will increase by 1.5, while it will increase by 3 by time 2. So 1.5x2. The slope in this case represents a linear trajectory where for every year that passes after the start of the study, the level of the variable observed increases by a factor of 1.5 on average.

[00:12:19]

Note that these latent variables have variance, which means that each individual displayed some variation that is different values above or below the average initial status and above and below the rate of change. And this variation can also be correlated providing equal variance. And one assumption is usually that variation around the intercept and the slope are normally distributed. So that’s the basic growth curve model.

A growth mixture model takes the growth model that I’ve just described, but assumes that the sample is made up of a mixture of individuals that have different average growth parameters, different intercepts and slopes. These groups cannot be directly observed, but we can infer them by looking at the growth parameters of the different individuals. We can estimate different latent classes where individuals in each class share the same propensity for displaying a certain developmental trajectory. One important issue to notice here is that there is variance in the intercept and slopes, which means that individuals within one latent class will still display variation in the average trajectory. Within classes, the variation around the growth parameters is also supposed to be normally distributed. In other words, individuals within one class are supposed to share the same propensity for distinctive developmental trajectory and each trajectory class has its own distinctive average trajectory, but individuals within each class will also vary around this average trajectory.

The latent class growth analysis will provide a similar model to the one just described, one where we assume that there is a mixture of individuals in the sample, and each group or class of individuals had different propensities for distinct developmental trajectories. The key difference in latent class growth analysis that we assume that there is no variability within classes. All individuals within a class have the same growth parameters, or in other words, they follow the same developmental trajectory with variation being only due to measurement error.

To illustrate the difference between these models, I have drawn these graphs. So in the first one on the left, I represented a ‘one size fits all’ model with similar distribution of the growth parameters, which by the way, is not normal. The growth mixture model can represent this distribution we observed by indicating different classes that have different average of the parameter. But they are all distributed normally around their respective average, as you can see all the distributions of the three classes here are all symmetrical and so on. So these number distribution represent the intraclass variation whereby individuals within each class can vary in their growth parameters.

Latent class growth analysis will instead represent the distribution by indicating different classes that have different parameters, but with no variation around these parameters. Since there is no variation within classes, more classes will be estimated to adequately represent the full distribution, as we can see here.

I will emphasise, once again, that these group based models as probabilistic, so the relationship between the observed and the latent variables are estimated with error. This also means that the classification of individuals in different trajectory groups is not certain, but there is some level of uncertainty, and it is important to take this into account, in particular when we want to test the association between trajectory groups and predictors.

This uncertainty in group classification is also reflected in some of the key parameters of the models, as well as the class specific raw parameters represented in these figures as I, S and Q for intercept, linear slope and quadratic slope respectively. A second key parameter is the latent class prevalence, that is the probability that a random individual will be in latent class one, two or three, and this probabilities sum up to one, which means that every individual is supposed to belong only to one class. In the example here for 45% of participant are estimated to belong to latent class one.

[00:17:21]

And the other important parameter is individuals’ probability of being in each of the latent classes estimated. We can see this individual highlighted here, for example, has 56% probability of being in latent class one, and 33% of being class two, and 21% of being latent class three. So the individual has a higher probability of being in class one, but you can see this affiliation is not certain and there is some margin of error. The posterior probabilities sum up to one, which means that despite this uncertainty, each individual is supposed to belong to one class or another. The statistic entropy represents the precision of this classification. Entropy varies from zero to one, where zero represents a case where all the participants have the same probability of being in class one, two or three and one represents a case where all the participants are certain to belong to one class or another. So the closer the entropy is to one, the more precise the classification.

And finally, I wanted to emphasize that in determining the number of trajectory classes that we need to adequately represent the interpersonal variability in the trajectories we observe, whether we apply growth mixture models or latent class growth analysis, we need to consider a series of statistics, which include model feed information criteria, and even entropy. And I refer to my presentation on latent class analysis if you want to know more. However, it’s important to exert some judgement and not blindly follow the statistics. And the judgement we need to exert is whether the models we are estimating represent the main features of the data in a parsimonious way, that is in a way that reduces complexity but also successfully represents key characteristics of the data and allows to evidence important features of the data. And some of the key criteria for successful model can be, for example, whether the groups we identify are characterised by different previous experiences, different variables that can explain the affiliation to different trajectory groups. And we may also identify groups that differ in these data outcomes, and also groups that show different trajectories in other processes. So the success in identifying an adequate number of groups should also determine by how useful and parsimonious are the groups and the models we identify.

Just before I finish, I wanted to provide some warnings, and one is that these models require samples with heterogeneity, which means that ideally the samples should be large or some people say, as a rule of thumb, at least 300 participants in the sample. The number of participants in the sample depends on other issues and the complexity of the models and so on, so it’s possible to apply this method to smaller samples than 300 participants, but the models need to be more constrained.

Estimating these models can also require a lot of computations, and especially in the case of growth mixture model, the estimation of the models can take a long time. And it’s also important to consider the issues with model estimation convergence and the fact that some of the solutions we find may not be that reliable, and I refer again to the presentation I made on latent class analysis.

So to summarise what I’ve said in this first presentation, growth models assuming that all participants follow the same trajectory may be inadequate or not appropriate for some data we’re looking at. And growth mixture models and latent class growth analysis provides person-centred approaches to identify groups with distinctive developmental trajectories in their outcomes and in the variables we observe. These models are based on probability and therefore provide robust and transparent methods for classification of individuals into different trajectory groups. And the difference between those two models is the growth mixture models assume variability within classes, whereas latent class growth analysis assumes that there is no variability within the classes, so the classes represent typical but different developmental trajectories.

And finally, I emphasized the importance of accepting some judgement in the selection of different models in identifying different number of groups. So thank you very much for your attention. Bye now.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK