Building Latent Class Growth Trees

Mattis van den Bergh & Jeroen K. Vermunt

February 1, 2017

Abstract

Researchers use latent class growth (LCG) analysis to detect meaningful subpopulations that display different growth curves. However, especially when the number of classes required to obtain a good fit is large, interpretation of the encountered class-specific curves may not be straightforward. To overcome this problem, we propose an alternative way of performing LCG analysis, which we call LCG tree (LCGT) modeling. For this purpose, a recursive partitioning procedure similar to divisive hierarchical cluster analysis is used: classes are split until a certain criterion indicates that the fit does not improve. The advantage of the LCGT approach compared to the standard LCG approach is that it gives a clear insight into how the latent classes are formed and how solutions with different numbers of classes relate. The practical use of the approach is illustrated using applications on drugs use during adolescence and mood regulation during the day.

Longitudinal data are used by social scientist to study development of behaviors or other phenomena. The analysis will often be done with latent growth curve models (MacCallum & Austin, 2000), with the aim to assess inter-individual differences in intra-individual change over time (Nesselroade, 1991). The typical growth model can be described as a multilevel model (Raudenbush & Bryk, 2002), in which the intercept and slopes of the time variables are allowed to

vary across individuals. This heterogeneity is captured using random effects, which are basically continuous latent variables (Jung & Wickrama, 2008). This approach assumes that the growth trajectories of all individuals can be appropriately described by a single set of the growth parameters, and thus that all individuals come from a single population. Growth mixture modeling relaxes this assumption by allowing for differences in growth parameters across unobserved subpopulations; that is, each latent class has a separate growth model. However, fully unrestricted growth mixture models are seldom used in practice, in part due to frequent estimation problems, as well as the preference for simpler, restricted models. Probably the most widely used form of growth mixture modeling is Latent Class Growth (LCG) analysis, whereby the (co-)variances of the growth factors within classes are fixed to zero (Jones, Nagin, & Roeder, 2001; Nagin & Land, 1993). This assumes that all individuals within a class follow the same trajectory and thus that there is no residual heterogeneity within classes.

When a LCG model is applied, two key modeling decisions need to be made; that is, on the number of classes and on the shape of the class-specific trajectories. In general, the decision on the number of classes is of more importance than the decision on the shape of the trajectory of each class as long as the shape is flexible enough (Nagin, 2005). Typically, researchers estimate LCG models with different numbers of classes and select the best model using likelihood-based statistics, usually with information criteria like AIC or BIC, which weigh model fit and complexity. Although there is nothing wrong with such a procedure, in practice it is often perceived as being problematic, especially when the model is applied with a large data set; that is, when the number of time points and/or the number of subjects is large. One problem occurring in such situations is that the selected number of classes may be rather large (Francis, Elliott, & Weldon, 2016). This causes the class trajectories to pick up very specific aspects of the data, which might not be interesting for the research question at

hand. Moreover, these specific trajectories are hard to interpret substantively and compare to each other. A second problem results from the fact that usually one would select a different number of classes depending on the model selection criterion used. Because of this, one may wish to inspect multiple solutions, as each of them may reveal specific relevant features in the data. However, it is fully unclear how solutions with different numbers of classes are connected, making it impossible to see what a model with more classes adds to a model with less classes.

To circumvent the issues mentioned above, it is most convenient to have models with differing numbers of classes that are substantively related; in other words, a model with $K+1$ classes is a refined version of a model with K classes, where one of the classes is split in two parts. Such an approach would result in a hierarchical structure, comparable to hierarchical cluster analysis (Everitt, Landau, Leese, & Stahl, 2001) or regression trees (Friedman, Hastie, & Tibshirani, 2001). Van der Palm, van der Ark, and Vermunt (2015) developed an algorithm for hierarchical latent class analysis that can be used for this purpose. While they focused on density estimation, with some adaptations their algorithm has also been used to build so called latent class trees for substantive interpretation (Van den Bergh, Schmittmann, & Vermunt, 2016). In this paper, this procedure will be extended to the longitudinal framework to construct Latent Class Growth Trees (LCGT).

With LCGT analysis a hierarchical structure is imposed on the latent classes by estimating 1- and 2-class models on a 'parent' node, which initially comprised the full data. If the 2-class model is preferred according to a certain information criterion, the data is split into 'child' nodes and separate data sets are constructed for each of the child nodes. The split is based on the posterior class membership probabilities; hence, the data patterns in each new data set will be the same as the original data set, but with weights equal to the posterior class membership probabilities for the child class concerned. Subsequently, each new

3

child node is treated as a parent and it is checked again whether a 2-class model provides a better fit than a 1-class model on the corresponding weighted data set. This procedure continues until no node is split up anymore. Because of this sequential algorithm, the classes at different levels of the tree can be substantively related, since child classes are subclasses of a parent class. Therefore, LCGT modelling allows for direct interpretation of the relationship between solutions with different numbers of classes, while still retaining the same statistical basis.

The remainder of the paper is set up as follows. In the next section, we discuss the basic LCG model and show how it can be used to build a LCGT. Also split criteria and guidelines for deviating from a binary split at the root of the tree will be discussed, together with an entropy measure for the post-hoc evaluation of the quality of splits. Two empirical data sets are used to illustrate LCGT analysis. The paper concludes with final remarks by the authors.

Method

Latent Class Growth models

Let y_{it} denote the response of individual i at time point t, T_i the number of measurements of person i, and $\mathbf{y}_i$ the full response vector of person i. Moreover, let X be the discrete latent class variable, k a particular latent class, and K the number of latent classes. A LCG model is, in fact, a regression model for the responses y_{it}, where time variables are used as predictors and where intercept and slope parameters differ across latent classes. We will define the LCG model within the framework on the generalized linear model, which allows dealing with different scale types of the response variable (Muthén, 2004; Vermunt, 2007).

Let $E(y_{it}|X = k)$ denote the expected value of the response at time point t for latent class k. After an appropriate transformation $g(\cdot)$, which mainly

depends on the measurement level of the response variable, $E(y_{it}|X = k)$ is modelled as a linear function of time variables. The most common approach is to use polynomial growth curves, which yields the following regression model for latent class k:

$$g[E(y_{it}|X = k)] = \beta_{0k} + \beta_{1k} \cdot t + \beta_{2k} \cdot t^2 + ... + \beta_{sk} \cdot t^s \qquad (1)$$

The choice of the degree of the polynomial (the value of s) is usually an empirical matter, though polynomials of degree larger than three are seldom used. Recently, Francis et al. (2016) proposed an alternative approach involving the use of baseline splines in LCG models.

To complete the model formulation for the response vector $\mathbf{y}_i$, we have to define the form of the class-specific densities $f(y_{it}|X = k)$, which could be univariate normal for a continuous response, binomial for a binary response, etc.. The response density for class k is a function of the expected value $E(y_{it}|X = k)$ and for continuous variables also of the residual variance. The LCG model for $\mathbf{y}_i$ can now be defined as follows:

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} P(X = k) \prod_{t=1}^{T_i} f(y_{it}|X = k), \qquad (2)$$

where the size of class k is represented by $P(X = k)$. A graphical representation of a LCG model with $K = 3$ can be seen in Figure 1.

The model estimates (the β parameters and class sizes) can be obtained by maximizing the following log-likelihood function:

$$\log L(\theta; \mathbf{y}) = \sum_{i=1}^{N} \log f(\mathbf{y}_i), \qquad (3)$$

where $f(\mathbf{y}_i)$ takes the form defined in Equation (2) and N denotes the total sample size. Maximization is usually achieved through an EM algorithm (Dempster, Laird, & Rubin, 1977), possibly combined with a Newton-type algorithm (Vermunt & Magidson, 2013).
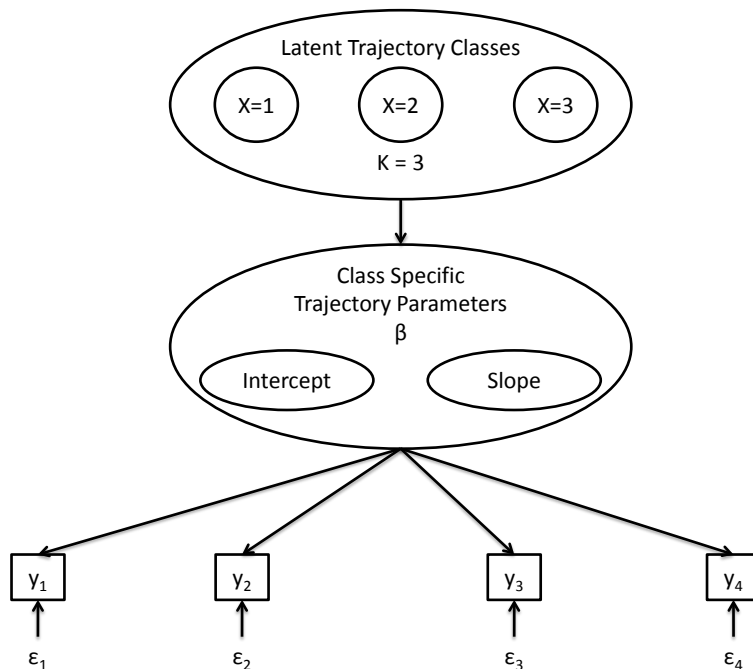
Figure 1: Graphical representation of a LCG model with three trajectory classes.

After selecting a particular model, individuals may be assigned to latent classes based on their the posterior class membership probabilities. Using the Bayes theorem, these probabilities are obtained as follows:

$$P(X = k|\mathbf{y}_i) = \frac{P(X = k) \prod_{t=1}^{T_i} f(y_{it}|X = k)}{f(\mathbf{y}_i)}. \qquad (4)$$

Latent Class Growth Tree models

Using an algorithm similar to the algorithm developed by Van der Palm et al. (2015) for divisive latent class analysis, a LCG model can also be constructed in a tree form. Such a LCGT has the advantages that increasing K classes to $K + 1$ classes results in directly related classes. This is because newly formed classes are obtained by splitting one of the K classes. Due to this direct relation, models with different numbers of classes can be substantively related, while still retaining the same statistical basis. Below we first describe the algorithm for

constructing a LCGT in more detail, and subsequently discuss various statistics that can be used during this process.

A LCGT consists of parent and child nodes. Every set of child nodes is based on one parent node and the first parent node consists of the root node containing the complete data set. At each parent node, standard LCG models are used and its child nodes are the classes assessed with the selected parent model. At the next level of the tree, these child nodes, in their turn, become parent nodes, and conditional on each new parent node a new set of LCG models is defined. This process continues until a stopping criterion is reached, for example, when the BIC does no longer decrease when splitting.

The basic equations of the growth curves of a LCGT model do not differ from those of a standard LCG model (e.g., Equation 1). The fact that the LCGT model is based on LCG models at parent nodes can be formulated as follows:

$$Pf(\mathbf{y}_i|X_{parent}) = \sum_{k=1}^{K} P(X_{child} = k|X_{parent}) \prod_{t=1}^{T} f(y_{it}|X_{child} = k, X_{parent}),$$
$$(5)$$

where X_{parent} represents the parent class at level l and X_{child} represents one of the K possible newly formed classes at level $l + 1$, with in general K being 2. Furthermore, $P(X_{child} = k|X_{parent})$ represents the size of a class, given the parent node, while $f(y_{it}|X_{child} = k, X_{parent})$ represents the class-specific response density at timepoint t, given the parent class. In other words, as in a standard LCG analysis, a model for $\mathbf{y}_i$ is defined, but now conditioned on belonging to the parent class concerned.

Estimation of the LCG model at the parent node X_{parent} involves maximizing the following weighted log-likelihood function:

$$\log L(\theta; \mathbf{y}, X_{parent}) = \sum_{i=1}^{N} w_{i,X_{parent}} P(\mathbf{y}_i|X_{parent}),$$
$$(6)$$

where $w_{i,X_{parent}}$ is the weight for person i at the parent class, which equals this person's posterior probability of belonging to the parent class concerned. So, building a LCGT involves estimating a series of LCG model using weighted data sets.

To see how the weights $w_{i,X_{parent}}$ are constructed, let us first look at the posterior class membership probabilities for the child nodes, conditional on the corresponding parent node. Assuming a split is accepted, the posteriors are obtained as follows:

$$P(X_{child} = k|\mathbf{y}_i; X_{parent}) = \frac{P(X_{child} = k|X_{parent}) \prod_{t=1}^{T_i} f(y_{it}|X_{child} = k, X_{parent})}{P(\mathbf{y}_i|X_{parent})}.$$
(7)

As proposed by Van der Palm et al. (2015), we use a proportional split based on these posterior class membership probabilities for the K child nodes conditional on the parent node, denoted by $k = 1, 2, ..., K$. If a split in two classes is performed, the weights for the two newly formed classes at the next level are obtained as follows:

$$w_{i,X_{child}=1} = w_{i,X_{parent}} P(X_{child} = 1|\mathbf{y}_i; X_{parent}) \tag{8}$$

$$w_{i,X_{child}=2} = w_{i,X_{parent}} P(X_{child} = 2|\mathbf{y}_i; X_{parent}). \tag{9}$$

In other words, a weight for individual i at a particular node equals the weight at the parent node times the posterior probability of belonging to the child node concerned conditional on belonging to the parent node. As an example, the weights $w_{i,X_1=2}$ used for investigating a possible split of class $X_1 = 2$ are constructed as follows:

$$w_{i,X_{12}} = w_{i,X=1} P(X_1 = 2|\mathbf{y}_i, X = 1), \tag{10}$$

where in turn $w_{i,X=1} = P(X = 1|\mathbf{y}_i)$. This implies:

$$w_{i,X_{12}} = P(X = 1|\mathbf{y}_i)P(X_1 = 2|\mathbf{y}_i, X = 1), \tag{11}$$

which shows that a weight at level two is in fact a product of two posterior class membership probabilities.

Construction of a LCGT can be performed using standard software for LC analysis, namely by running a series of LC models with the appropriate weights. After each accepted split a new data set is constructed and the procedure repeats itself. We developed an R routine in which this process is fully automated. It calls the Latent GOLD program (Vermunt & Magidson, 2013) in batch mode to estimate 1- and 2-class models, evaluates whether a split should be made, and keeps track of the weights when a split is accepted. In addition, it creates various graphical displays which facilitates the interpretation of the LCGT (see among others Figure 2). A novel graphical display is a tree depicting the class-specific growth curves for the newly formed child classes (for an example, see Figure 5). In the trees, the name of a child class equals the name of the parent class plus an additional digit, a 1 or a 2. To prevent that the structure of the tree will be affected by label switching resulting from the fact that the order of the newly formed classes depends on the random starting values, when building the LCGT we locate the larger class at the left branch with number 1 and the smaller class at the right branch with number 2.

Statistics for building and evaluating the LCGT

Different types of statistics can be used to determine whether a split should be accepted or rejected. Here, we will use the BIC (Schwarz, 1978), which is defined as follows:

$$BIC = -2\log L(\theta; \mathbf{y}, X_{parent}) + \log(N)P, \tag{12}$$

where $\log L(.)$ represents the log-likelihood at the parent node concerned, N the

total sample size, and P the number of parameters of the model at hand. Thus, a split is performed if at a parent node concerned the BIC for the 2-class model is lower than the one of the 1-class model. Note that using a less strict criterion (e.g. AIC) will yield the same splits as the BIC, but possible also additional splits, and thus a larger tree.

Special attention needs to be dedicated to the first split at the root node of the tree, in which one picks up the most dominant features in the data. In many situations, a binary split at the root may be too much of a simplification, and one would prefer allowing for more than two classes in the first split. For this purpose, we cannot use the usual criteria like a AIC or BIC, as this would boil down to using again a standard LCG model. Instead, for the decision to use more than two classes at the root node, we propose looking at the relative improvement of fit compared to the improvement between the 1- and 2-class model. When using the log-likelihood value as the fit measure, this implies assessing the increase in log-likelihood between, say, the 2- and 3-class model and compare it to the increase between the 1- and 2-class model. More explicitly, the relative improvement between models with K and $K + 1$ classes $(RI_{K,K+1})$ can be computed as:

$$RI_{K,K+1} = \frac{\log L_{K+1} - \log L_K}{\log L_2 - \log L_1}, \tag{13}$$

which yields a number between 0 and 1, where a small value indicates that the K-class model can be used as the first split, while a larger value indicates that the tree might improve with an additional class at the root of the tree. Note that instead of an increase in log-likelihood, in Equation 13 one may use other measures of improvement of fit, such as the decrease of the BIC or the AIC. Screeplots depicting the difference in log-likelihood (or BIC or AIC) for models with one class difference can also be used to judge whether the relative improvement is large, as will be illustrated in the empirical examples presented

10

below.

The BIC and $RI_{K,K+1}$ statistics are used to determine whether and how splits should be performed. However, often we are also interested in evaluating the quality of splits in terms of the amount of separation between the newly formed classes; that is, to determine how different the classes are. In other words, is a split substantively important yes or no. This is also relevant if one would like to assign individuals to the classes resulting from a LCGT. Note that the assignment of individuals to the two child classes is more certain when the larger of the posterior probabilities $P(X_{child} = k|\mathbf{y}_i; X_{parent})$ is closer to 1. A measure to express this is the entropy; that is,

$$Entropy(X_{child}|\mathbf{y}) = \sum_{i=1}^{N} w_{i|X_{parent}} \sum_{k=1}^{2} -P(X_{child} = k|\mathbf{y}_i; X_{parent}) \log P(X_{child} = k|\mathbf{y}_i; X_{parent}).$$

(14)

Typically $Entropy(X_{child}|\mathbf{y})$ is rescaled to lie between 0 and 1 by expressing it in terms of the reduction compared to $Entropy(X_{child})$, which is the entropy computed using the unconditional class membership probabilities $P(X_{child} = k|X_{parent})$. This so-called $R^2_{Entropy}$ is obtained as follows:

$$R^2_{Entropy} = \frac{Entropy(X_{child}) - Entropy(X_{child}|\mathbf{y})}{Entropy(X_{child})}$$

(15)

The closer $R^2_{Entropy}$ is to one, the better the separation between the child classes in the split concerned.

Empirical examples

The proposed LCGT methodology will be illustrated by the analyses of two longitudinal data sets. The data set in the first example contains a yearly dichotomous response on drugs use collected using a panel design. The second data set contains an ordinal mood measure, recorded using an experience sam-
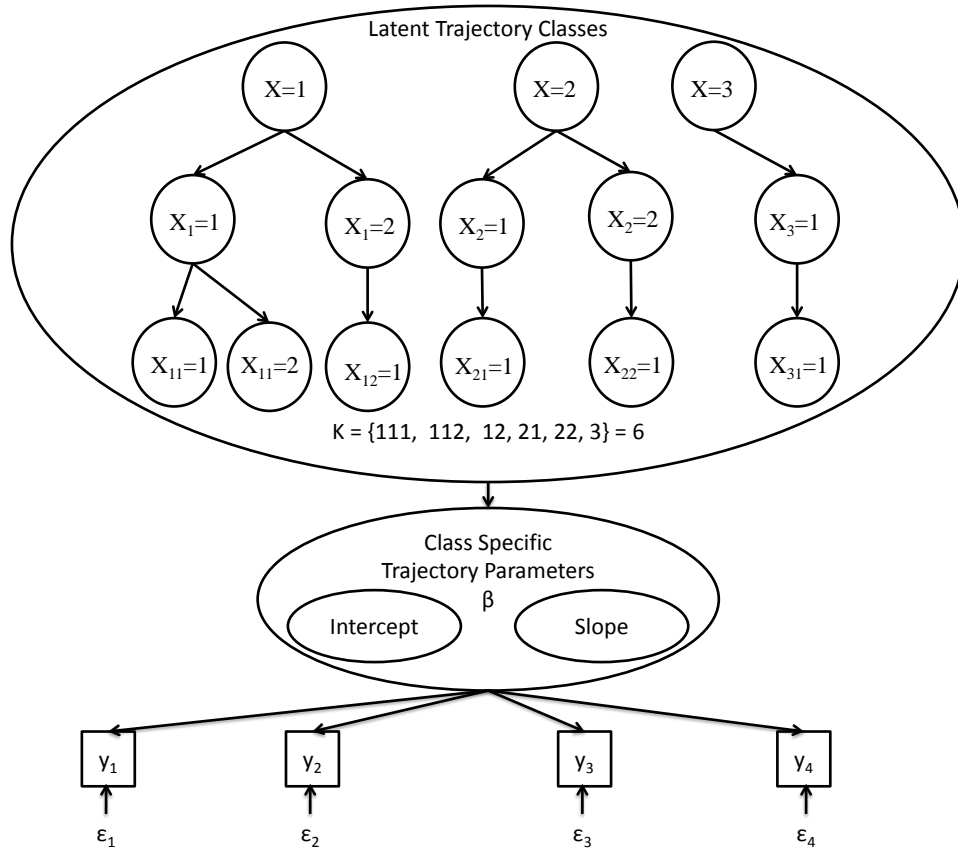
Figure 2: Graphical example of a LCGT model with a root of three classes.

pling design with eight measures per day during one week. The two data sets illustrate LCGT analyses, differing in the number of classes at their root node. For both examples, the quality of the splits will also be evaluated using the entropy-based R-squared.

Example 1: Drugs use

The first data set stems from the National Youth Survey (Elliot, Huizinga & Menard, 1989). It contains nine waves, from 1976 to 1980 yearly and from 1980 to 1992 with three year intervals. The age at the first wave of the 1725

respondents (53% man and 47% women) varied between 11 and 17 years. We use age at the panel wave concerned as the time variable, which takes on values ranging from age 11 to 23. Each respondent has been observed at most nine times (on average 7.93 times). The dichotomous dependent variable of interest in our example will be whether the respondent used drugs or not during the past year.

Because the trees based on second and third degree polynomial growth curves were almost identical, the simpler one using a second degree polynomial was retained. The tree structure and the class sizes at the splits [1] are presented in Figure 3. As can be seen, there are four binary splits, which result in a total of five latent classes at the end nodes.

To determine whether it would be better to increase the number of classes at the root of the tree, we can look at the relative improvement in fit of models with more than 2 classes according to the likelihood, BIC, and AIC as reported in Table 1. As can be seen, the relative improvement with a third class is around 10%. As this is quite low, we retain the tree with a binary split at the root. This conclusion is supported by the screeplots in Figure 4.

To interpret the encountered classes, the growth curves can be plotted for the two newly formed classes at each node of the tree. This is displayed in Figure 5. As can be seen, the first split results in a class with a low probability to use drugs (class 1) and a class with a high probability to use drugs (class 2). Subsequently both of these classes are split further. Class 1 is split into class 11 with a very low probability of using drugs (on average 0.01%) and class 12 with a low probability during the first few years, but with a slight increase from age 20 to 33. Class 2 is split into class 21 and 22, which mainly differ in the moment

---

[1]Every split should sum up to the class size of its parent node. However, because the allocation is carried out on the basis of the posterior probabilities, the class sizes are not integers. For convenience, these numbers have been rounded, which causes slight deviations where the sum of two child nodes does not exactly add up to the parent node.

at which the probability of drugs use is the highest: Respondents of class 21 start using drugs a few years earlier than respondents of class 22. Finally, class 21 is split further, where class 211 has a moderate probability (around 0.6) to use drugs at an early age, but this probability also quickly declines. Class 212 has a very high probability (around 0.95) to start using drugs at an early age and this probability stays quite constant up to age 25.
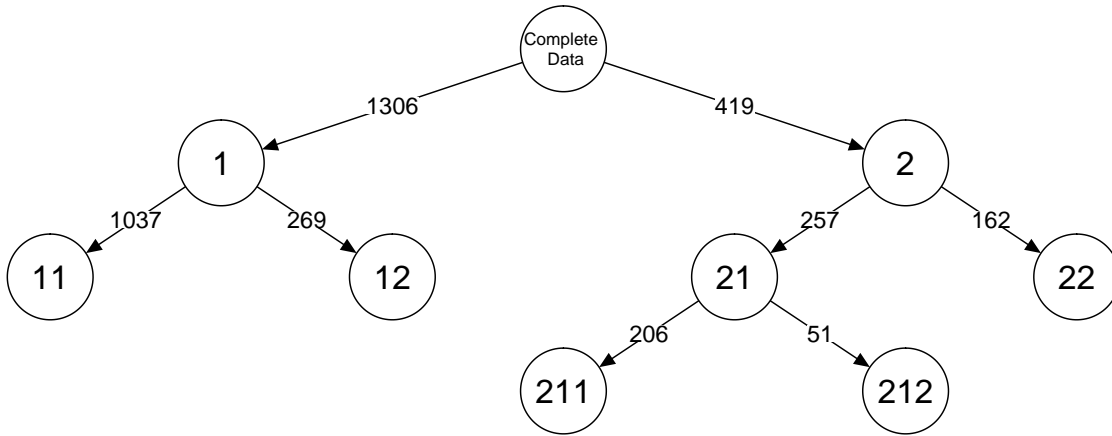

Figure 3: Layout of a LCGT with a root of two classes on Drugs use.

Table 1: Fit statistics of a traditional LC growth model with 1 to 6 classes.

	$\log L$	P	BIC	AIC	$RI_{\log L}$	RI_{BIC}	RI_{AIC}
1	-5089	3	10200	10183			
2	-4246	7	8543	8505			
3	-4156	11	8394	8334	0.106	0.090	0.102
4	-4086	15	8284	8202	0.083	0.067	0.079
5	-4046	19	8233	8129	0.048	0.031	0.043
6	-4028	23	8228	8102	0.021	0.003	0.016

Table 2 shows the entropy R-squared values for all splits. The encountered values confirm what could also be seen from the depicted growth curves: The
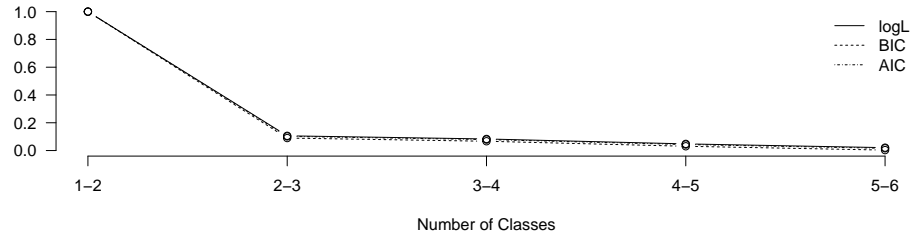
Figure 4: Screeplots of the difference in likelihood and BIC of succesive LCG models for the data on drugs use.
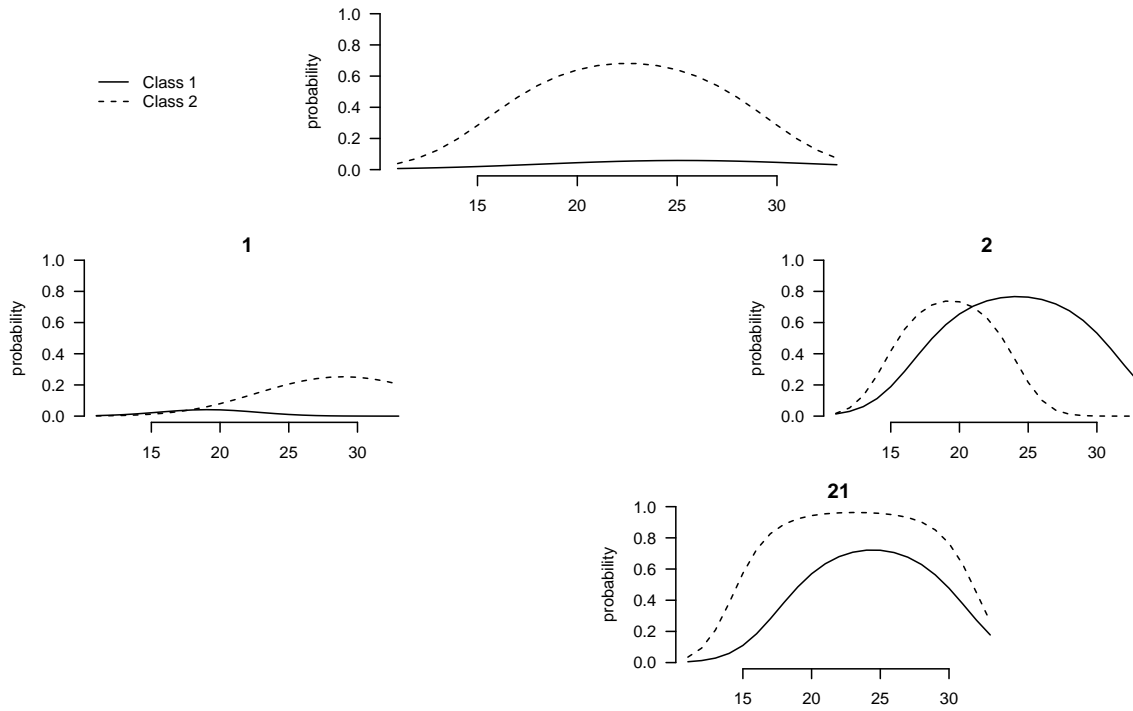

Figure 5: LCGT with a root of 2 classes on drug use over age.

first split on the complete data set shows a large difference between the two classes. Furthermore, classes 11 and 12 are similar and thus badly separated,

15

whereas the differences between classes 21 and 22 and between classes 211 and 212 are substantial. Hence, after the first split, the branch of class 2 contains more important additional differences than the one of class 1.

Example 2: Mood regulation

The second data set stems from a momentary assessment study by Crayen, Eid, Lischetzke, Courvoisier, and Vermunt (2012). It contains 8 mood assessments per day during a period of one weeks among 164 respondents (88 women and 76 males, with a mean age of 23.7, SD = 3.31). Respondents answered a small number of questions on a handheld device at pseudo-random signals during their waking hours. The delay between adjacent signals could vary between 60 and 180 minutes (M [SD] = 100.24[20.36] minutes, min = 62 minutes, max = 173 minutes). Responses had to be made within a 30-minute time window after the signal, and were otherwise counted as missing. On average, the 164 participants responded to 51 (of 56) signals (M [SD] = 51.07 [6.05] signals, min = 19 signals, max = 56 signals). In total, there were 8374 non-missing measurements.

At each measurement occasion, participants rated their momentary mood on an adapted short version of the Multidimensional Mood Questionnaire (MMQ). Instead of the original monopolar mood items, a shorter bipolar version was used to fit the need for brief scales. Four items assessed pleasant-unpleasant mood (happy-unhappy, content-discontent, good-bad, and well-unwell). Participants rated how they momentarily feel on a 4-point bipolar intensity scales (e.g., very unhappy, rather unhappy, rather happy, very happy). For the current analysis, we focus on the item well-unwell. Preliminary analysis of the response category

Table 2: Relative entropy per split of the LCGT on drugs use.

	Complete Data	1	2	21
$R^2_{Entropy}$	0.746	0.268	0.545	0.619

frequencies showed that the lowest category (i.e., very unwell) was only chosen in approximately 1% of all occasions. Therefore the two lower categories were collapsed together into one unwell category. The following analysis is based on the recoded item with three categories (conform Crayen et al. (2012)).
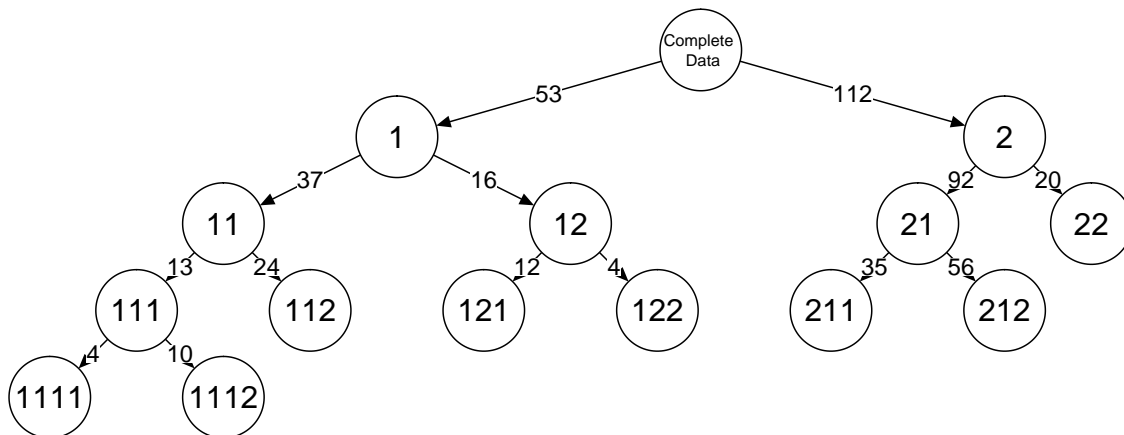
Figure 6: Layout of a LCGT with a root of two classes on mood regulation.

For the analysis, we used a LCG model based on an ordinal logit model. The time variable was the time during the day, meaning that we model the mood change during the day. There was a substantial difference between a tree based

Table 3: Likelihood, number of parameters, BIC and relative improvement of the likelihood and BIC of a traditional LC growth model with 1 to 6 classes.

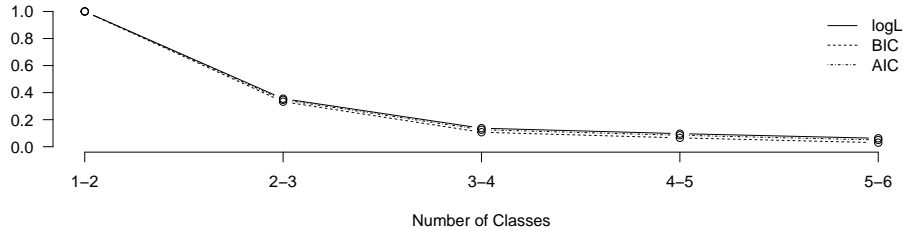|	$\log L$	P	BIC	AIC	$RI_{\log L}$	RI_{BIC}	RI_{AIC}
1	-7199	4	14424	14408			
2	-6741	9	13538	13504			
3	-6578	14	13244	13191	0.355	0.333	0.347
4	-6516	19	13149	13077	0.137	0.107	0.126
5	-6471	24	13091	13001	0.097	0.065	0.085
6	-6443	29	13064	12956	0.062	0.030	0.050

Figure 7: Screeplots of the difference in likelihood and BIC of succesive LCG models for the data on mood regulation.

on a second or a third degree polynomial, which indicates that developments are better described by cubic growth curves (see also the trajectory plots in Figure 9). Because there was no substantial difference between a tree based on a third or a fourth degree polynomial, a third degree polynomial was used. The LCGT model obtained with a root of two classes is quite large, with in total seven binary splits, resulting in a total of eight latent classes (Figure 6). A large tree already indicates that a larger number of classes at the root of the tree might be appropriate. Moreover, based on the relative improvement of the log-likelihood, BIC, and AIC (Table 3), it seems sensible to increase the number of classes at the root of the three. A screeplot of the relative change in log-likelihood, BIC, and AIC also show that that after three classes the relative gain is quite small for both measures.

The layout and size of the LCGT with 3 root classes can be seen in Figure 8 and its growth curve plots in Figure 9. The growth plots show that at the root of the tree, the three different classes all improve their mood during the day. They differ in their overall mood level, with class 3 having the lowest and class 2 the highest overall score. Moreover, class 1 seems to be more consistently increasing than the other two classes.

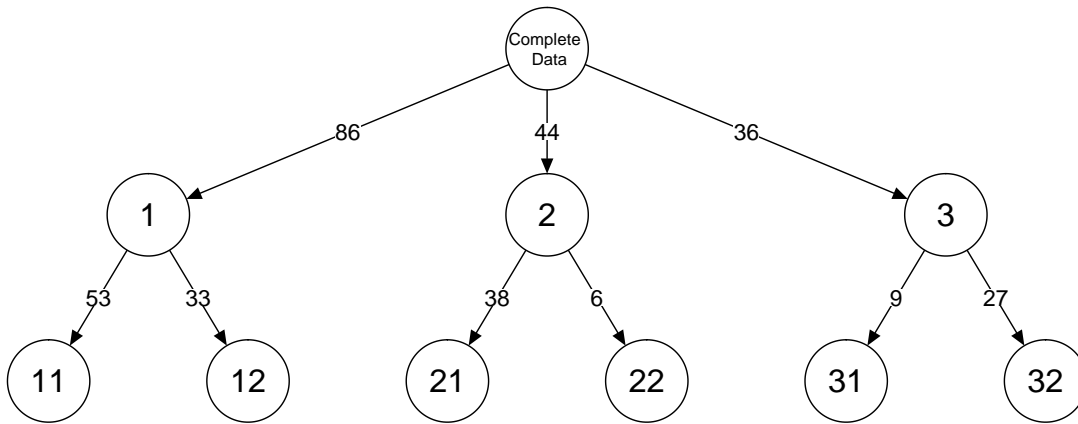These three classes can be split further. Class 1 splits into two classes with

18

Figure 8: Layout of a LCGT with a root of three classes on mood regulation during the day.
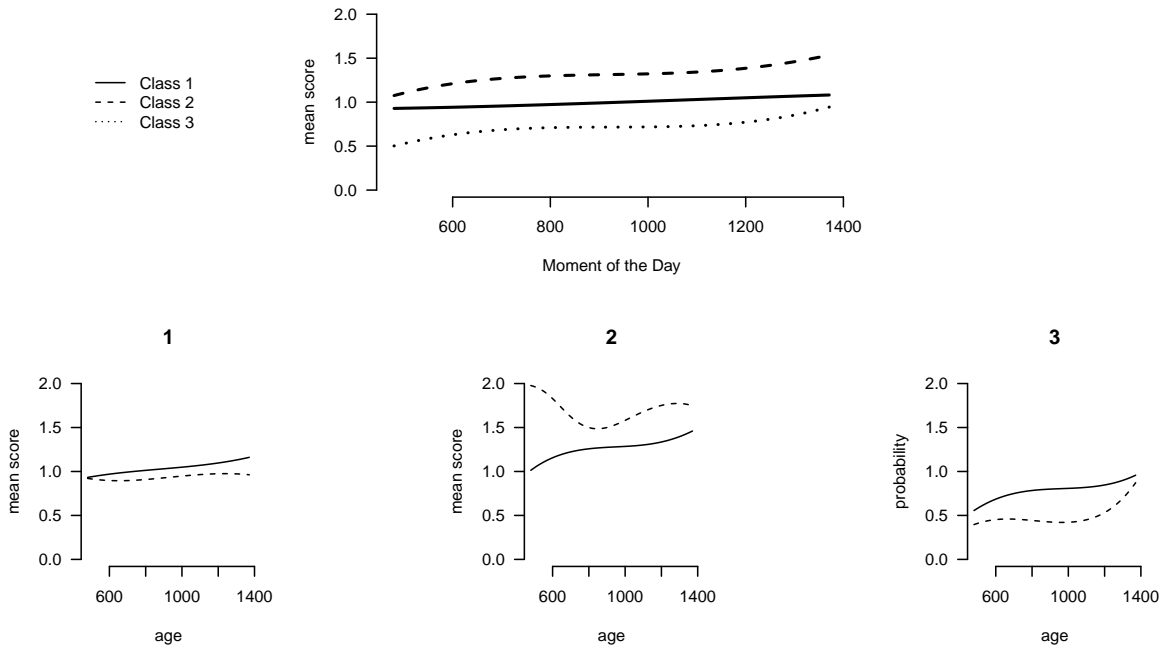

Figure 9: LCGT with a root of 3 classes on mood regulation of mood regulation during the day.

19

both an average score around one, class 11 just above and class 12 just below. Moreover, the increase in class 11 is larger than in class 12. The split of class 2 results in class 21 consisting of respondents with a very good mood in the morning, a quick decrease until mid-day, and a subsequent increase. In general the mean score of class 21 is high relative to the other classes. Class 22 starts with an average mean score and subsequently only increases. The splitting of class 3 results in two classes with a below average mood. Both classes increase, class 31 mainly in the beginning and class 32 mainly at the end of the day.

Table 4 shows the relative entropy for each split. Besides the root split, the relative entropy is largest for the split of class 2. This indicates that the differences between the subclasses 21 and 22 are larger than those between subclasses 31 and 32, while classes 11 and 12 differ the least.

Table 4: Relative entropy per split of each of the subsequent classes.

	Complete Data	1	2	3
$R^2_{Entropy}$	0.889	0.734	0.932	0.897

Discussion

LCG models are used by researchers who wish to identify (unobserved) subpopulations with different growth trajectories using longitudinal data. However, often the number of latent classes encountered is rather large, making interpretation of the results difficult. Moreover, because solutions with different number of classes are unrelated, a substantive comparison of models with different numbers of classes is not possible, which is especially problematic when different model selection criteria point at a different optimal number of classes. To resolve these issues, we proposed using LCGT models in which the identification of the latent classes is done in a sequential manner. The constructed hierarchical tree will show the most important distinctions in growth trajectories in the first

splits, and more detailed distinctions in latter splits. While we primarily used binary splits, we also showed how to decide about larger splits using relative improvement of fit measures. The latter is mainly of interest at the root of the tree. The proposed LCGT algorithm and graphical displays which are available as R code were illustrated with two empirical examples. The two illustrative examples showed that easily interpretable solutions are obtained using our new procedure.

Various extensions and variants of the proposed procedure are possible and worth to study in more detail. Whereas in the current paper we restricted ourselves to LCGTs with only binary splits after the split at the root of the tree, also at the second and next levels it may be of interest to use larger split sizes, which may result in a tree with different split sizes within branches. Because the size of the splits may strongly affect the structure of the constructed LCGT, we recommend deciding this separately per split rather than using a fully automated procedure. Note that at this stage more substantive information about the branch is available to guide a decision.

The BIC was used to decide whether or not to split a class, as it has been shown to perform well for standard LC and LCG analysis (Nylund, Asparouhov, & Muthén, 2007). However, other measures could be used as well, where their strictness will influence the likelihood to start a new branch within a tree. Therefore, the decision criterion used can affect the bottom part of the tree significantly. Note that the lower parts are also affected by the decision to increase the number of classes at the root of the tree. Moreover, the exact choice of a criterion depends on the required specificity of the encountered growth trajectories, where a less strict criterion may be used if one wishes to see more specific classes at the bottom of the tree.

While LCG models are becoming very popular among applied researchers, the use of these models is not easy at all (Van De Schoot, Sijbrandij, Winter, Depaoli, & Vermunt, 2016). We hope that the proposed LCGT methodology

will simplify the detection and interpretation of underlying growth trajectories. This does, of course, not mean that the standard LCG model is not useful anymore. In practice, a researcher may start with a standard LCG analysis, and switch to our LCGT approach when encountering difficulties in deciding about the number of classes or interpreting the differences between a possibly large number of classes.

References

Crayen, C., Eid, M., Lischetzke, T., Courvoisier, D. S., & Vermunt, J. K. (2012). Exploring dynamics in mood regulationmixture latent markov modeling of ambulatory assessment data. *Psychosomatic medicine*, *74*(4), 366–376.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2001). Hierarchical clustering. *Cluster Analysis, 5th Edition*, 71–110.

Francis, B., Elliott, A., & Weldon, M. (2016). Smoothing group-based trajectory models through b-splines. *Journal of Developmental and Life-Course Criminology*, *2*(1), 113–133.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer series in statistics Springer, Berlin.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A sas procedure based on mixture models for estimating developmental trajectories. *Sociological methods & research*, *29*(3), 374–393.

Jung, T., & Wickrama, K. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual review of psychology*, *51*(1), 201–226.

Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences. Thousand Oaks, CA: Sage Publications*, 345–68.

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

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model. *Criminology*, *31*(3), 327–362.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling*, *14*(4), 535–569.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Van den Bergh, M., Schmittmann, V. D., & Vermunt, J. K. (2016). Building latent class trees, with an application to a study of social capital. *Methodology, conditinal accepted*.

Van der Palm, D. W., van der Ark, L. A., & Vermunt, J. K. (2015). Divisive latent class modeling as a density estimation method for categorical data. *Journal of Classification*.

Van De Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2016). The grolts-checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–17.

Vermunt, J. K. (2007). Growth models for categorial response variables: Standard, latent-class, and hybrid approaches. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (p. 139-1580). Mahwah, NJ: Erlbaum.

Vermunt, J. K., & Magidson, J. (2013). Technical guide for latent gold 5.0: Basic, advanced, and syntax.