Chapter 12
Mixture Models: Latent Profile and Latent Class Analysis

Daniel Oberski

Abstract Latent class analysis (LCA) and latent profile analysis (LPA) are techniques that aim to recover hidden groups from observed data. They are similar to clustering techniques but more flexible because they are based on an explicit model of the data, and allow you to account for the fact that the recovered groups are uncertain. LCA and LPA are useful when you want to reduce a large number of continuous (LPA) or categorical (LCA) variables to a few subgroups. They can also help experimenters in situations where the treatment effect is different for different people, but we do not know which people. This chapter explains how LPA and LCA work, what assumptions are behind the techniques, and how you can use `R` to apply them.

12.1 Introduction

Mixture modeling is the art of unscrambling eggs: it recovers hidden groups from observed data. By making assumptions about what the hidden groups look like, it is possible to get back at distributions within such groups, and to obtain the probability that each person belongs to one of the groups. This is useful when:

- You measured one thing but were really interested in another. For example, how students answer exam questions is indicative of whether or not they have mastered the material, and how somebody you chat with online reacts to your messages is indicative of them being human or a bot;
- You fit a model but suspect that it may work differently for different people, and you are interested in how. For example, when designing the information given to vistors of a birdwatching site, putting up signs with just the Latin names of birds is helpful to some people and likely to annoy others. When investigating the effect of putting up such signs it might be helpful to take this into account.

D. Oberski (✉)
Tilburg University, Tilburg, The Netherlands
e-mail: D.L.Oberski@uvt.nl

J. Robertson and M. Kaptein (eds.), *Modern Statistical Methods for HCI*,
Human–Computer Interaction Series, DOI 10.1007/978-3-319-26633-6_12

Table 12.1 Names of different kinds of latent variable models

Observed	Models for means		Regression models	
	Latent		Latent	
	Continuous	Discrete	Continuous	Discrete
Continuous	Factor analysis	**Latent profile analysis**	Random effects	Regression mixture
Discrete	Item response theory	**Latent class analysis**	Logistic ran. eff.	Logistic reg. mix.

- You have a lot of different variables—too many to handle and interpret—and would like to reduce these to a few easily interpretable groups. This is often done in marketing where such groups are called "segments".

There are many other uses of mixture modeling—too many to explain here. Suffice to say that by understanding mixture modeling, you will make a start at understanding a host of other statistical procedures that can be very useful, such as regression mixture modeling, noncompliance in experiments, capture-recapture models, randomized response, and many more. Moreover, mixture models are popular tools in computer vision, such as face detection and hand gesture recognition (e.g., Yang and Ahuja 2001). While these applications go beyond the scope of this chapter, it may be helpful to keep in mind that they are extensions of the models we discuss here.

Mixture modeling is a kind of latent variable modeling: it helps you to deal with situations where some of the variables are unobserved. The specific thing about mixture modeling is that is concerns latent variables that are discrete. You can think of these as groups, "classes", or "mixture components"—or as categories of an unobserved nominal or ordinal variable. Depending on whether the observed variables are continuous or categorical, mixture models have different names. These names, together with the names of the other kinds of latent variable models, are shown in Table 12.1, in which the rows correspond to continuous or discrete *observed* variables, and the columns to continuous or discrete *latent* variables. The left part of the table concerns models in which the groups are based on differences in means, and the right part concerns models in which the groups are based on differences in regression-type coefficients. The two types of models dealt with in this chapter are indicated in bold: "latent profile analysis", which tries to recover hidden groups based on the means of continuous observed variables, and "latent class analysis", which does the same for categorical variables.[1] Some of the other models in the table are explained in other chapters.

A different name for latent profile analysis is "gaussian (finite) mixture model" and a different name for latent class analysis is "binomial (finite) mixture model". Its Bayesian version is popular in the computer science literature as "latent Dirichlet allocation". Here we will stick to the terminology LCA/LPA, which is more common in the social sciences.

[1] Confusingly, sometimes latent class analysis is used as a broader term for mixture models.

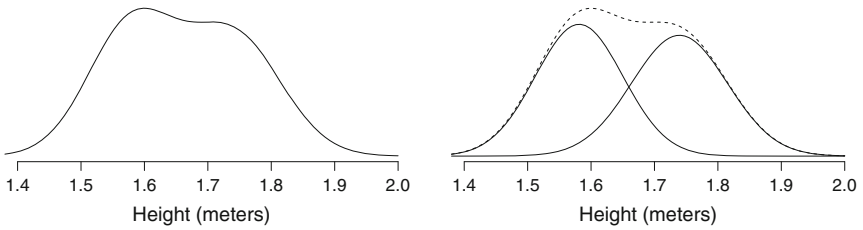

Fig. 12.1 Peoples' height. *Left* observed distribution. *Right* men and women separate, with the total shown as a *dotted line*

12.2 Mixtures of Continuous Variables: Latent Profile Analysis

For this chapter, I measured the height of every human being on the planet.[2] I then plotted the distribution of heights measured in Fig. 12.1 (right-hand side).

Interestingly, the height distribution on the left-hand side of Fig. 12.1 is clearly not normal. It has two peaks that look like they might come from two groups. Now, you may have happened to guess what these groups are: women and men. If I had recorded each person's sex, I could confirm this by creating my picture separately for men and women, as shown on the right-hand side of Fig. 12.1. Women differ in their height according to a normal distribution, as do men—it is just when you combine the two that the non-normal picture emerges.

Unfortunately, I forgot to record peoples' sex, and now it is too late. When you perform an experiment this might happen to you too—or it might have happened to the people who collected your data. For example, a usability test might omit peoples' handedness. Even more commonly, you may need information that is difficult or impossible to get at directly, such as an attitude, feeling, or socially desirable behavior. In all of these cases you will be in a similar spot as I am with the height data: I think there might be hidden groups, but I have not actually recorded them. Latent class analysis is concerned precisely with recovering such hidden ("latent") groups.

So how can we recover the picture on the right-hand side of Fig. 12.1? One idea is to split the sample on some guess about peoples' sex and make histograms within each guessed group. Unfortunately, it can be shown that this will never give the picture on the right (McLachlan and Peel 2004, pp. 26–28). Without extra information, the only way to get exactly that picture is to guess each person's sex correctly, but the only information I have to guess sex is height. And although a random man is likely to be taller than a random woman, inevitably people like Siddiqa Parveen, who is currently the world's tallest woman, and Chandra Bahadur Dangi, the shortest man, will cause me to count incorrectly, and create a picture that is at least slightly different from the "true" one on the right.[3]

[2] This is not true, but the rest of the chapter is.

[3] Apparently, Ms. Parveen is 213.4 cm and Mr. Dangi is 48.3 cm.

This is where mixture modeling comes to the rescue. Because it turns out that, while we will never know each person's sex for certain, we can still recover a picture very close to that on the right-hand side of Fig. 12.1. So we can discover the distributions of height for men and women without ever having observed sex! Even more astoundingly, as more and more data are gathered, we will more and more correctly surmise what the distributions for men and women look like *exactly*.

There is a well-known saying that "there is no such thing as a free lunch". I have personally falsified this statement on several—sometimes delicious—occasions. But while false in real life, in mathematics this statement is law. We will pay in unverifiable assumptions—on this occasion the assumption that height is normally distributed for men and women. This assumption is unverifiable from just the observed data because the problem is exactly that we do not know sex. So when faced with data that produce a picture like the one on the left-hand side of Fig. 12.1, we will need to simply assume that this picture was produced by mixing together two perfect normal distributions, without us being able to check this assumption.

The corresponding mathematical model is

$$p(\text{height}) = Pr(\text{man})\text{Normal}(\mu_{\text{man}}, \sigma_{\text{man}}) + Pr(\text{woman})\text{Normal}(\mu_{\text{woman}}, \sigma_{\text{woman}}), \tag{12.1}$$

which I will write as

$$p(\text{height}) = \pi_1^X \text{Normal}(\mu_1, \sigma_1) + (1 - \pi_1^X)\text{Normal}(\mu_2, \sigma_2). \tag{12.2}$$

So we see that the the probability curve for height is made up of the weighted sum of two normal curves,[4] which is exactly what the right-hand side of Fig. 12.1 shows. There are two reasons for writing π_1^X instead of $Pr(\text{man})$: first, when fitting a mixture model, I can never be sure which of the "components" (classes) corresponds to which sex. This is called the "label switching problem". Actually, it is not really a problem, but just means that X is a dummy variable that could be coded $1 = \text{man}, 2 = \text{woman}$ or vice versa. The second reason is that by using a symbol such as π_1^X, I am indicating that this probability is an unknown parameter that I would like to estimate from the data. Note that the superscript does not mean "to the power of" here, but is just means "π_1^X is the probability that variable X takes on value 1". This way of writing equations to do with latent class analysis is very common in the literature and especially useful with categorical data, as we will see in the next section.

Assuming that both men's and women's weights follow a normal distribution, the problem is now to find the means and standard deviations of these distributions: the within-class parameters (i.e. μ_1, μ_2, σ_1, and σ_2).[5] The trick to doing that is in starting with some initial starting guesses of the means and standard deviations. Based on these guesses we will assign a *posterior probability* of being a man or woman to each person. These posterior probabilities are then used to update our guess of the within-class parameters, which, in turn are used to update the posteriors, and so

[4] As can be gleaned from the figures, by "normal curve" I mean the probability density function.

[5] We also need to know the proportion of men/women π_1^X but I will ignore that for the moment.

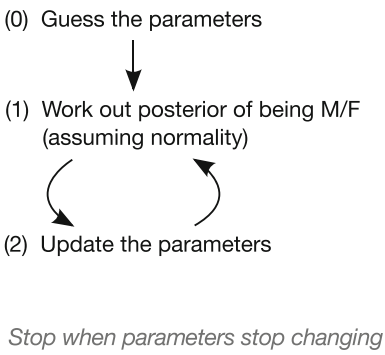

Fig. 12.2 EM algorithm estimating the distribution of height for men and women separately without knowing peoples' sex

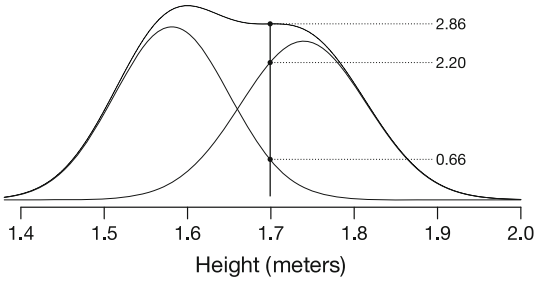

Fig. 12.3 Tall woman or short man? The probability density of observing a person 1.7 m tall is a weighted sum of that for men and women separately

on until nothing seems to change much anymore. This algorithm, the "expectation-maximization" (EM) algorithm, is depicted in Fig. 12.2. The online appendix contains `R` code (`simulation_height.R`) that allows you to execute and play with this algorithm.

How can the posterior probabilities be obtained (step 1), when the problem is exactly that we do not know sex? This is where our unverifiable assumption comes in. Suppose that my current guess for the means and standard deviations of men and women is given by the curves in Fig. 12.3, and I observe that you are 1.7 m tall. What then is the posterior probability that you are a man? Figure 12.3 illustrates that this posterior is easy to calculate. The overall probability density of observing a person of 1.7 m is 2.86, which we have assumed (in Eq. 12.2) is just the average of two numbers: the height of the normal curve for men plus that for women. The picture shows that a height of 1.7 m is much more likely if you are a man than if you are a woman. In fact, the posterior is simply the part of the vertical line that is made up by the normal curve for men which is $2.20/(2.20 + 0.66) = 2.20/2.86 \approx 0.77$. So, if all I know is that you are 1.7 m tall, then given the current guess of normal curves for men and women, the posterior probability that you are a man is about 0.77. Of course, the posterior probability that you are a woman is then just $1 - 0.77 = 0.23$, since both probabilities sum to one.

Now for step (2) in Fig. 12.2. I apply my earlier idea: guess people's gender, and then count their height towards men's or women's means and standard deviations, depending on the guess. Recognizing that I am not 100 % certain about any one

person's sex, however, instead of guessing "man" or "woman", I will use the posterior probabilities. For example, if a 1.7 m-tall person has a 0.77 posterior chance of being a man and (therefore) a 0.23 chance of being a woman, I will add (0.77)(1.7) to my estimate of the mean of men but also (0.23)(1.7) to that for women. So each person's observed value contributes mostly to the mean of the sex I guessed for them but also some to the other sex, depending on how strongly I believe they are a man or woman. If I have no idea whether a person is man or woman, that is, when the probability is 0.50:0.50, the height contributes equally to the guessed means of both. In Fig. 12.3 this occurs where the two curves cross, at about 1.65. By using the posterior probabilities as weights, I obtain new guesses for the means and standard deviations, which allow me to go back to step (1) and update the posteriors until convergence. Note that whereas earlier in the section this idea of guessing then calculating would not work, the assumptions allow me to get the posterior probabilities necessary to perform this step.

This section has shown how you might recover hidden groups from observed continuous data. By making assumptions about what the hidden groups look like, it is possible to get back at distributions within such groups, and to obtain a posterior probability that each person belongs to one of the groups. There are several software packages that can do latent profile analysis, including Mplus (Muthén and Muthén 2007) and Latent GOLD (Vermunt and Magidson 2013a). In `R`, this model could be fitted using the `mclust` library (Fraley and Raftery 1999):

```
library(mclust)
height_fit_mclust <- Mclust(height)
summary(height_fit_mclust, parameters = TRUE)
```

As mentioned above, you can play with this example using the online appendix (`simulation_height.R`).

With just one variable, we needed to pay a hefty price for this wonderful result: an unverifiable assumption. It is possible to do better by incorporating more than one variable at the same time; for example, not just height but also estrogen level. Both are imperfect indicators of sex but using them together allows me to guess the hidden group better. The next section gives an example of modeling with several categorical variables.

12.3 Mixtures of Categorical Variables: Latent Class Analysis

Latent class analysis (LCA) is similar to latent profile analysis: it also tries to recover hidden groups. The difference, as you can see in Table 12.1, is that LCA deals with categorical observed variables. Another difference between LCA and LPA is that no specific distribution is assumed for the variables: each of the observed variables' categories has an unknown probability of being selected without this probability

following any particular functional form. But by retracting our assumption, we also retract the necessary payment to buy the possibility of estimating latent classes. In LCA, this payment is made, not by assuming that the variables are distributed in any particular way, but by assuming that *within each class, the observed variables are unrelated to each other*. This assumption is called "local independence", and tends to create classes in which the observations are similar to each other but different from those in other classes.

Typically, the researcher (1) determines the "best" number of latent classes K that sufficiently explain the observed data; and (2) summarizes the results. At times, researchers are also interested in (3) relating the most likely class membership to some external variables not used in the model. When performing this last task, however, the researcher should be very careful to account for the uncertainty in each observation's class membership: failure to do so will result in biased estimates of the class membership prediction. Modeling class membership while accounting for the uncertainty about it is called "three-step modeling" in the literature, and can be done in certain software packages such as Mplus and Latent GOLD. For more information, see Bakk et al. (2013).

Suppose there are three observed variables A, B, and C, all three categorical with six categories. The model for traditional latent class analysis is then typically written as

$$\pi_{abc}^{ABC} = \sum_{x} \pi_{x}^{X} \pi_{a}^{A|X} \pi_{b}^{B|X} \pi_{c}^{C|X}, \tag{12.3}$$

where X is the latent class variable, π_x^X the size of class x and, for example, $\pi_a^{A|X}$ is the probability that variable A takes on the value a in the latent class x. Equation 12.3 describes the probability of seeing any combination of values a, b, and c as depending solely on the differences in latent class sizes (π_x^X) combined with how different these classes are in terms of the observed variables. Within the classes, the variables are unrelated ("conditionally independent"), which is reflected in the product $\pi_a^{A|X} \pi_b^{B|X} \pi_c^{C|X}$.

Let's apply LCA as an exploratory technique to the SuSScale example data with 3 time points. As a reminder, some example answers to the 10 questions in the questionnaire are shown in Table 12.2. These 10 variables are moderately related

Table 12.2 Example data from the SUS questionnaire

	Time	V01	V02	V03	V04	V05	V06	V07	V08	V09	V10
1	1	4	3	5	4	3	4	3	2	2	3
87	2	4	5	3	5	5	5	5	4	5	5
173	3	3	3	2	4	3	2	3	2	2	3
2	1	3	2	3	3	3	3	4	3	3	2
…	…	…	…	…	…	…	…	…	…	…	…

Table 12.3 Fit of the latent class models with increasing numbers of classes

# Classes	Log-likelihood	AIC	BIC
1	–3753.32	7614.64	7806.50
2	–3441.70	7103.39	7494.22
3	–3327.72	6987.45	7577.24
4	–3245.73	**6935.46**	**7724.22**

The lowest AIC and BIC are shown in **boldface**

with an average correlation of 0.4. Our goal is to find K classes such that these relationships are small within the classes, where we still need to find out the best number of classes K.

Because the data are categorical, we apply a latent class model for polytomous variables to the 258 observations. This is done using the `poLCA` package in `R` (Linzer and Lewis 2011; Core Team 2012):

```
library(poLCA)
f <- as.matrix(dplyr::select(scale, starts_with("V")))~Time
M4 <- poLCA(f, scale, nclass=4, nrep=50)
```

Here, the first command loads the library; the second command constructs an `R` formula with all of the observed indicators (variables that measure the class membership) V01–V10 on the left-hand side, and the covariate Time on the right-hand side to control for any time effects. The last command fits the model with four classes. The `nrep=50` argument tells `poLCA` to try out 50 different starting values so that we are certain that the best solution has been found. This is always recommended when performing latent class analysis.

We fit the model for a successively increasing number of classes, up to four: $K = \{1, 2, 3, 4\}$. With one class, the model in Eq. 12.3 simply says the variables are independent. Obviously if this one-class model fits the data well we are done already since there are no relationships between the variables in that case. The first order of business is therefore to evaluate how well the latent class models with different numbers of classes fit the data and to select one of them. There are many measures of model fit, the most common ones being the AIC ("Akaike information criterion") and BIC ("Bayesian information criterion"). These are shown for our four models in Table 12.3.

In Table 12.3, lower values of the fit measures are better. The best model appears to be the four-class model. We therefore pick that model as our preferred solution. This procedure of choosing the lowest BIC is the most used in LCA. However, this is an exploratory method: just as in exploratory factor analysis, it is therefore often also possible to pick a lower or a higher number of classes based on substantive concerns or ease of interpretation. We might only be interested in finding a small number of broad classes, for instance, and ignore any remaining distinctions within these classes (Fig. 12.4).

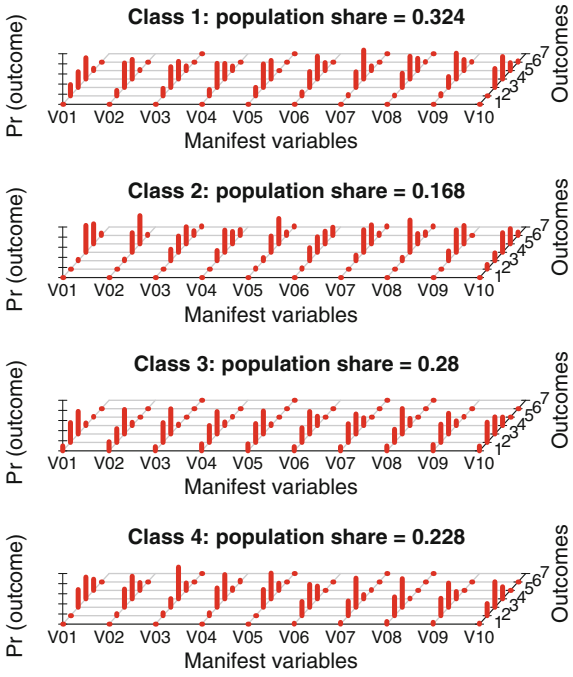

Fig. 12.4 Output of `poLCA` for the four-class model showing the estimated distribution of the ten observed variables V01–V10 within each of the four classes. The profile plot is easier to read

With the argument `graphs=TRUE`, `poLCA` produces graphs that show the estimated distributions of the observed variables in each latent class. These graphs can be very useful with fewer variables of nominal measurement level, but with ten variables having six categories each, this output is rather hard to read. Instead, I created a so-called "profile plot" for these variables in Fig. 12.5, which displays the estimated class means instead of their entire distribution. This plot is usually reserved for continuous variables, but with our six-point ordinal variables it still makes sense to display their means.

Figure 12.5 shows the profile plots for all four fitted models. In a profile plot, there is one line for each class. The lines represent the estimated mean of the observed variable on the x-axis within that class. So the profile plot for the two-class solution shows that the two classes separate people who give high scores on all of the questions (class 2) from people who give low scores on all of them (class 1). Note that the class numbers or "labels" are arbitrary. In the three-class solution, there is also a class with low scores on all of the variables (class 1). The other two classes both have high scores on the first five variables but are different from each other regarding the last five variables: class 2 also has high scores on these whereas class 3 has low scores. The four-class solution, finally, has "overall high" (class 2) and "overall low" (class 3) classes, as well as "V01–V05 but not V06–V10" (class 4) and its opposite (class 1).

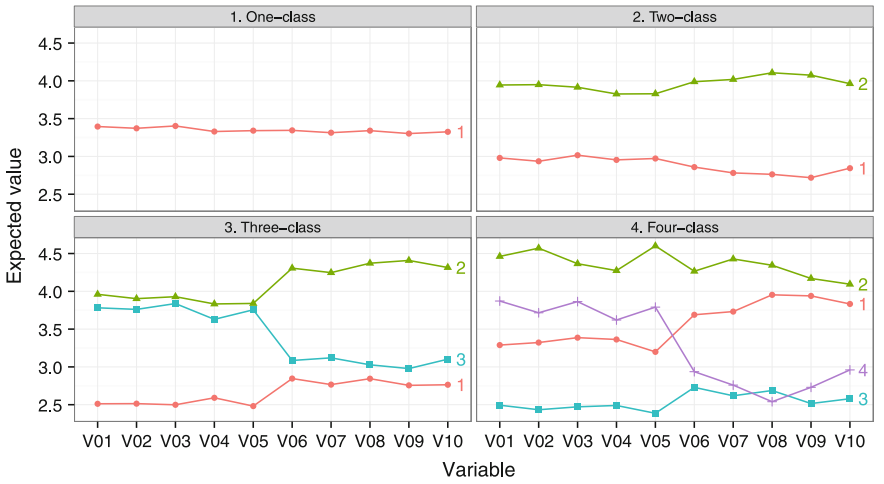

Fig. 12.5 "Profile plot": estimated means of the 10 observed variables in the SuSScale data within each of the K classes for $K = \{1, 2, 3, 4\}$

Although the four-class solution is the "best" solution, this does not necessarily mean that it is a good solution. In particular, if our local independence model is to provide an adequate explanation of the observed data, the relationships between the indicators after controlling for class membership should be small. An intuitive way of thinking about this is that the scatterplot between two indicators should show a relationship *before* conditioning on the class membership, but none *after* this conditioning.

Figure 12.6 demonstrates this with two example indicators, V01 and V10. To make the points easier to see in the figure they have been "jittered" (moved a bit). The top part of Fig. 12.6 is simply a scatterplot of V01 and V10 over all 258 observations. The shape of the points here corresponds to the most likely class membership of that point. For example, the upper-rightmost point is a triangle to indicate that its most likely class is the "overall high" class (class 2). It can be seen that there is a moderate relationship between these two indicators before accounting for the classes, with a linear correlation of 0.30. The bottom part of Fig. 12.6 splits up the same scatterplot by class, so that each graph contains only points with a particular shape. It can be seen that within each of the classes the relationship is almost non-existent. This is exactly what conditional independence means. Therefore the model fits well to the data for these two indicators. Graphs like Fig. 12.6 only make sense when the observed score (1–6 in our case) is approximately of interval level. For nominal variables, a different kind of fit measure is therefore necessary. One such measure is the "bivariate residual" (BVR), the chi-square in the residual cross-table between two indicators. At the time of writing, the BVR is not available in `R` but can be obtained from commercial software such as Latent GOLD (Vermunt and Magidson 2013b).

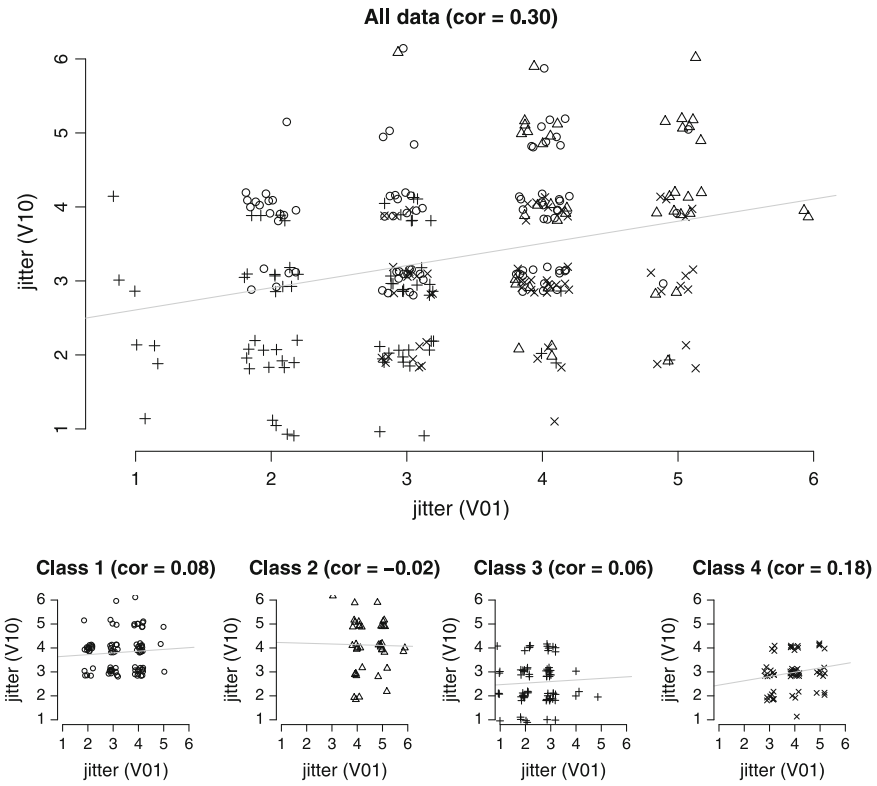

Fig. 12.6 *Top graph* the data with observed correlation 0.30 are modeled as a "mixture" of four classes (*circles*, *triangles*, *pluses*, and *crosses*). *Bottom graph* these classes are chosen such that the correlation between the variables is minimal within them

In this example we applied LCA as an exploratory method to artificial data that were generated according to a linear factor model. The profiles then recuperate the generated structure. In real data, there will often be additional classes. For example, if some respondents tend to use the extreme points of a scale ("extreme response style") while others use the whole scale, this will lead to an additional class in which the extremes are more likely to be chosen. Because this is a nonlinear effect, such a finding is not possible with linear models such as factor analysis.

12.4 Other Uses of Latent Class/Profile Analysis

While the traditional use of LCA/LPA is as an exploratory technique, latent class models can also be seen as a very general kind of latent variable modeling. Latent class models are then a type of structural equation modeling, factor analysis, or

random effects (hierarchical/multilevel) modeling in which the latent variable is discrete rather than continuous (Skrondal and Rabe-Hesketh 2004). The advantage of a discrete latent variable is that the distribution of the latent variable is estimated from the data rather than assumed to have some parametric form, such as a normal distribution. Hence, several special cases of latent class models are sometimes called "nonparametric". The term "nonparametric" here does not mean that there are no parameters or assumptions; rather, it means that the distribution of the latent variables is estimated. In fact, the relaxation of assumptions about the distribution of the latent variables usually means that there are *more* parameters to estimate, and that some of the assumptions, such as local independence rather than just uncorrelatedness, become necessary to identify parts of the model.

Mixture models are also useful for analyzing experimental data. For example, when the effect of the treatment is thought to be different in different groups, but these groups are not directly observed: in this case "regression mixture modeling" can be used. In `R`, the `flexmix` package (Gruen et al. 2013) is especially useful for regression mixture modeling.

Another situation that often occurs in randomized experiments with people is that the people do not do what they are supposed to do. For example, a patient assigned to take a pill might neglect taking it, or a person receiving different versions of a website based on cookies might have blocking software installed that prevents the assigned version from coming up. When participants do not follow the randomly assigned treatment regime, this is called "noncompliance". For people in the treatment group, we can often see whether they did or did not actually receive the treatment. But simply deleting the other subjects would break the randomization, causing a selection effect. Therefore these people should be compared, not with the entire group of controls, but with a hidden subgroup of controls who *would have taken the treatment if they had been assigned to it*. The fact that we cannot observe this hypothetical hidden group leads to a latent class (mixture) model, and methods to deal with noncompliance in randomized experiments are special cases of the models discussed here. The Latent GOLD software contains several examples demonstrating how to deal with noncompliance. In `R`, a package containing some basic noncompliance functionality is `experiment` (Imai 2013).

12.5 Further References

An accessible and short introduction to latent class analysis and its variations is given by Vermunt and Magidson (2004). More background information and details on the various types of models that can be fitted is found in Hagenaars and McCutcheon (2002). A comprehensive introduction-level textbook is Collins and Lanza (2013). The manuals and examples of software that can fit these models, especially Mplus and Latent GOLD, are another great source of information. Some examples of applications of LCA and LPA to human-computer interaction are Nagygyörgy et al. (2013)

and Hussain et al. (2015). For an application to computer detection of human skin color, see Yang and Ahuja (2001, Chap. 4).

References

Bakk Z, Tekle FB, Vermunt JK (2013) Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. Sociol Methodol 43(1):272–311

Collins LM, Lanza ST (2013) Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences, vol 718. Wiley, New York

Fraley C, Raftery AE (1999) Mclust: software for model-based cluster analysis. J Classif 16(2): 297–306

Gruen B, Leisch F, Sarkar D (2013) flexmix: Flexible mixture modeling. http://CRAN.R-project.org/package=flexmix. R package version, pp 2–3

Hagenaars JA, McCutcheon AL (2002) Applied latent class analysis. Cambridge University Press, Cambridge

Hussain Z, Williams GA, Griffiths MD (2015) An exploratory study of the association between online gaming addiction and enjoyment motivations for playing massively multiplayer online role-playing games. Comput Hum Behav 50:221–230

Imai K (2013) Experiment: R package for designing and analyzing randomized experiments. R package version 1.1-1

Linzer DA, Lewis JB (2011) poLCA: an R package for polytomous variable latent class analysis. J Stat Softw 42(10):1–29

McLachlan, G. and Peel, D. (2004). Finite mixture models. John Wiley & Sons, New York

Muthén LK, Muthén B (2007) Mplus user's guide. Muthén & Muthén, Los Angeles

Nagygyörgy K, Urbán R, Farkas J, Griffiths MD, Zilahy D, Kökönyei G, Mervó B, Reindl A, Ágoston C, Kertész A et al (2013) Typology and sociodemographic characteristics of massively multiplayer online game players. Int J Hum-Comput Interaction 29(3):192–200

R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

Skrondal A, Rabe-Hesketh S (2004) Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Interdisciplinary statistics series. Chapman & Hall/CRC, Boca Raton

Vermunt JK, Magidson J (2004) Latent class analysis. The sage encyclopedia of social sciences research methods, pp 549–553

Vermunt J, Magidson J (2013a) LG-Syntax user's guide: manual for Latent GOLD 5.0 Syntax Module. Statistical Innovations Inc., Belmont

Vermunt JK, Magidson J (2013b) Technical guide for Latent GOLD 5.0: basic, advanced, and syntax. Statistical Innovations Inc., Belmont

Yang M-H, Ahuja N (2001) Face detection and gesture recognition for human-computer interaction. Springer Science & Business Media