Latent Class Growth Analysis

1

Latent Class Growth Analysis

- Setting
  - Longitudinal data
  - A single item measured repeatedly
  - Hypothesized trajectory classes (categorical latent variable)
- Aim
  - Estimate trajectory shapes
  - Estimate trajectory class probabilities
  - Relate class probabilities to covariates
  - Classify individuals into classes (posterior probabilities)
- Applications
  - Single process
  - Two processes

2

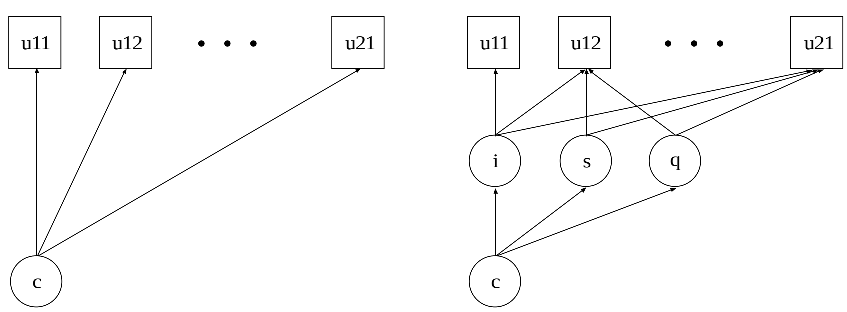
Single Process Latent Class Growth Analysis: Cambridge Delinquency Data

- 411 boys in a working class section of London
  (n = 403 due to 8 boys who died)
- Ages 10 to 32 (ages 11 - 21 used here)
- Outcome is number of convictions in the last 2 years, modeled as an ordered polytomous variable scored 0 for 0 convictions, 1 for one conviction, and 2 for more than one conviction

  Sources: Farrington & West (1990); Nagin & Land (1993); Roeder, Lynch & Nagin (1999); Muthen (2004)

3

LCA Vs LCGA



Number of parameters (11 *u*'s, 3 classes):

	LCA	LCGA
binary *u*	35	11
3-categ. *u*	68	12

4

Latent Class Analysis With 3 Classes On Cambridge Data

LogL = -1,032 (68 parameters), BIC = 2,472

5

Input LCGA On Cambridge Data

```
TITLE:      LCGA
            ordered polytomous variables for conviction at each
            age11-21
            dep. variable 0, 1 , 2 (0, 1, or more convictions)

DATA:       FILE = naginordered.dat;

VARIABLE:   NAMES = u11 u12 u13 u14 u15 u16 u17 u18 u19 u20
                    u21 c1 c2 c3 c4;
            USEVAR = u11-u21;
            CATEGORICAL = u11-u21;
            CLASSES = c(3);

ANALYSIS:   TYPE = MIXTURE;
```

6

Input LCGA On Cambridge Data (Continued)

```
MODEL:      %OVERALL%

            i s q | u11@-.6  u12@-.5   u13@-.4  u14@-.3  u15@-.2
            u16@-.1  u17@0  u18@.1  u19@.2  u20@.3  u21@.4;

OUTPUT:     TECH1 TECH8;

PLOT:       SERIES = u11-u21(s);
            TYPE = PLOT3;
```

7

Output Excerpts LCGA On Cambridge Data

Model Results

```
                      Estimates      S.E.     Est./S.E.
Latent Class 1

I   |
  u11                     1.000      0.000        0.000
  u12                     1.000      0.000        0.000
  u13                     1.000      0.000        0.000
  u14                     1.000      0.000        0.000
  u15                     1.000      0.000        0.000
  u16                     1.000      0.000        0.000
  u17                     1.000      0.000        0.000
  u18                     1.000      0.000        0.000
  u19                     1.000      0.000        0.000
  u20                     1.000      0.000        0.000
  u21                     1.000      0.000        0.000
```

8

Output Excerpts LCGA On Cambridge Data (Continued)

```
                    Estimates       S.E.    Est./S.E.
Latent Class 1

S   |
  u11                  -0.600      0.000        0.000
  u12                  -0.500      0.000        0.000
  u13                  -0.400      0.000        0.000
  u14                  -0.300      0.000        0.000
  u15                  -0.200      0.000        0.000
  u16                  -0.100      0.000        0.000
  u17                   0.000      0.000        0.000
  u18                   0.100      0.000        0.000
  u19                   0.200      0.000        0.000
  u20                   0.300      0.000        0.000
  u21                   0.400      0.000        0.000
```

9

Output Excerpts LCGA On Cambridge Data (Continued)

```
                    Estimates       S.E.    Est./S.E.
Latent Class 1

Q   |
  u11                   0.360      0.000        0.000
  u12                   0.250      0.000        0.000
  u13                   0.160      0.000        0.000
  u14                   0.090      0.000        0.000
  u15                   0.040      0.000        0.000
  u16                   0.010      0.000        0.000
  u17                   0.000      0.000        0.000
  u18                   0.010      0.000        0.000
  u19                   0.040      0.000        0.000
  u20                   0.090      0.000        0.000
  u21                   0.160      0.000        0.000

Means
I                      -1.633      0.329       -4.955
S                       0.264      0.431        0.613
Q                      -7.376      0.249       -5.906
```

10

Output Excerpts LCGA On Cambridge Data (Continued)

	Estimates	S.E.	Est./S.E.
Thresholds			
u11$1	-0.917	0.319	-2.876
u11$2	0.827	0.304	2.716
u12$1	-0.917	0.319	-2.876
u12$2	0.827	0.304	2.716
u13$1	-0.917	0.319	-2.876
u13$2	0.827	0.304	2.716
u14$1	-0.917	0.319	-2.876
u14$2	0.827	0.304	2.716
u15$1	-0.917	0.319	-2.876
u15$2	0.827	0.304	2.716
u16$1	-0.917	0.319	-2.876
u16$2	0.827	0.304	2.716
u17$1	-0.917	0.319	-2.876
u17$2	0.827	0.304	2.716
u18$1	-0.917	0.319	-2.876
u18$2	0.827	0.304	2.716
u19$1	-0.917	0.319	-2.876
u19$2	0.827	0.304	2.716
u20$1	-0.917	0.319	-2.876
u20$2	0.827	0.304	2.716
u21$1	-0.917	0.319	-2.876
u21$2	0.827	0.304	2.716

11

Output Excerpts LCGA On Cambridge Data (Continued)

	Estimates	S.E.	Est./S.E.
Latent Class 2			
Q |			
u11	0.360	0.000	0.000
u12	0.250	0.000	0.000
u13	0.160	0.000	0.000
u14	0.090	0.000	0.000
u15	0.040	0.000	0.000
u16	0.010	0.000	0.000
u17	0.000	0.000	0.000
u18	0.010	0.000	0.000
u19	0.040	0.000	0.000
u20	0.090	0.000	0.000
u21	0.160	0.000	0.000
Means			
I	-5.246	0.509	-10.297
S	0.802	0.836	0.959
Q	-2.796	2.515	-1.112

12

Output Excerpts LCGA On Cambridge Data (Continued)

	Estimates	S.E.	Est./S.E.
Thresholds			
u11$1	-0.917	0.319	-2.876
u11$2	0.827	0.304	2.716
u12$1	-0.917	0.319	-2.876
u12$2	0.827	0.304	2.716
u13$1	-0.917	0.319	-2.876
u13$2	0.827	0.304	2.716
u14$1	-0.917	0.319	-2.876
u14$2	0.827	0.304	2.716
u15$1	-0.917	0.319	-2.876
u15$2	0.827	0.304	2.716
u16$1	-0.917	0.319	-2.876
u16$2	0.827	0.304	2.716
u17$1	-0.917	0.319	-2.876
u17$2	0.827	0.304	2.716
u18$1	-0.917	0.319	-2.876
u18$2	0.827	0.304	2.716
u19$1	-0.917	0.319	-2.876
u19$2	0.827	0.304	2.716
u20$1	-0.917	0.319	-2.876
u20$2	0.827	0.304	2.716
u21$1	-0.917	0.319	-2.876
u21$2	0.827	0.304	2.716

13

Output Excerpts LCGA On Cambridge Data (Continued)

	Estimates	S.E.	Est./S.E.
Latent Class 3			
Q |			
u11	0.360	0.000	0.000
u12	0.250	0.000	0.000
u13	0.160	0.000	0.000
u14	0.090	0.000	0.000
u15	0.040	0.000	0.000
u16	0.010	0.000	0.000
u17	0.000	0.000	0.000
u18	0.010	0.000	0.000
u19	0.040	0.000	0.000
u20	0.090	0.000	0.000
u21	0.160	0.000	0.000
Means			
I	0.000	0.000	0.000
S	0.311	1.012	0.308
Q	-3.853	0.943	-1.983

14

Output Excerpts LCGA On Cambridge Data (Continued)

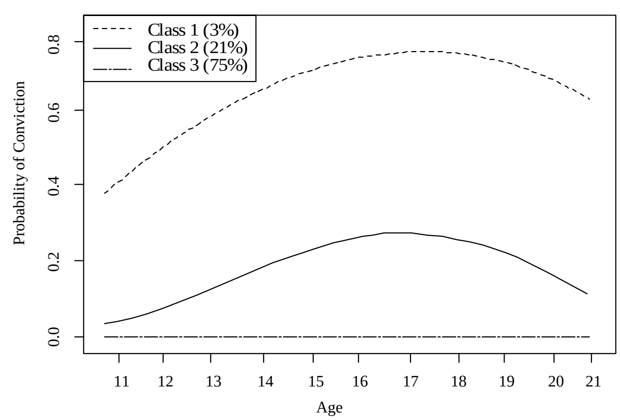
	Estimates	S.E.	Est./S.E.
Thresholds			
u11$1	-0.917	0.319	-2.876
u11$2	0.827	0.304	2.716
u12$1	-0.917	0.319	-2.876
u12$2	0.827	0.304	2.716
u13$1	-0.917	0.319	-2.876
u13$2	0.827	0.304	2.716
u14$1	-0.917	0.319	-2.876
u14$2	0.827	0.304	2.716
u15$1	-0.917	0.319	-2.876
u15$2	0.827	0.304	2.716
u16$1	-0.917	0.319	-2.876
u16$2	0.827	0.304	2.716
u17$1	-0.917	0.319	-2.876
u17$2	0.827	0.304	2.716
u18$1	-0.917	0.319	-2.876
u18$2	0.827	0.304	2.716
u19$1	-0.917	0.319	-2.876
u19$2	0.827	0.304	2.716
u20$1	-0.917	0.319	-2.876
u20$2	0.827	0.304	2.716
u21$1	-0.917	0.319	-2.876
u21$2	0.827	0.304	2.716

15

LCGA On Cambridge Data (Continued)

3-class LCGA
LogL = -1,072
(12 parameters)
BIC = 2,215

3-class LCA
LogL = -1,032
(68 parameters)
BIC = 2,472



16

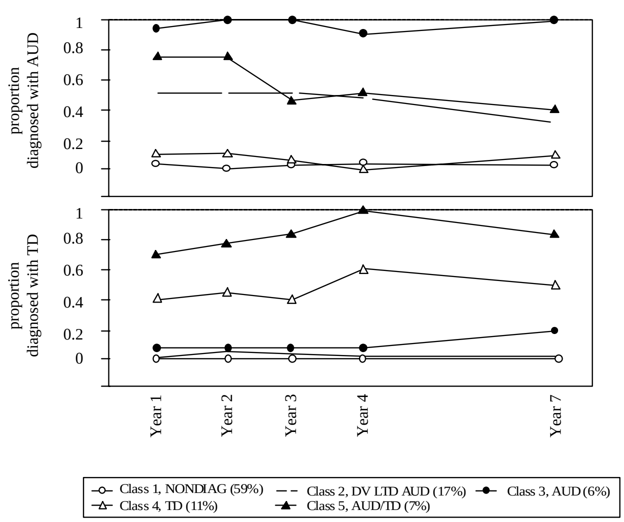
Multiple Process LCGA: Relating Trajectory Class Variables For Different Outcomes

The co-occurrence of alcohol and tobacco use disorders (Jackson, Sher, Wood, 1999)

- Parallel processes
- College sample, n = 450

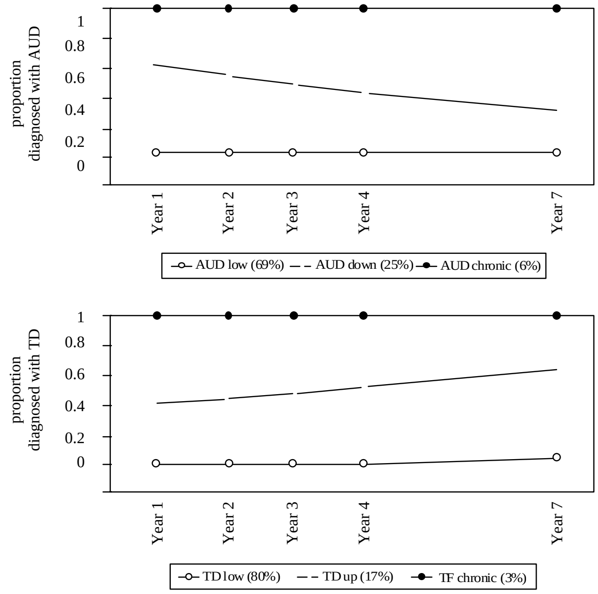
17

Co-Occurrence Of Alcohol And Tobacco Use Disorder

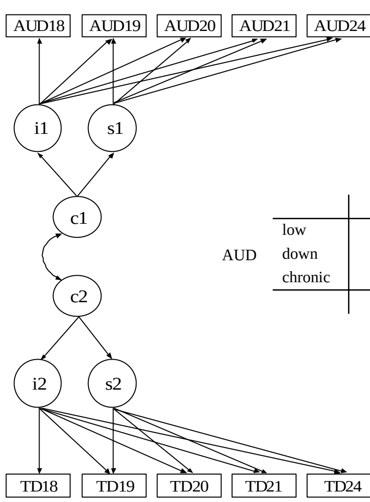


18

Co-Occurrence Of Alcohol And Tobacco Use Disorder



Co-Occurrence Of Alcohol And Tobacco Use Disorder



Class Probability Estimates

		TD low	TD up	TD chronic	
AUD	low	.61	.08	.001	.69
	down	.15	.07	.03	.25
	chronic	.04	.02	.003	.06
		.80	.17	.03	1.00

Further Readings On
Latent Class Growth Analysis

Jones, B.L., Nagin, D.S. & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. Sociological Methods & Research, 29, 374-393.

Land, K.C. (2001). Introduction to the special issue on finite mixture models. Sociological Methods & Research, 29, 275-281.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), New developments and techniques in structural equation modeling (pp. 1-33). Lawrence Erlbaum Associates. (#86)

Nagin, D.S. (1999). Analyzing developmental trajectories: a semi-parametric, group-based approach. Psychological Methods, 4, 139-157.

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

Further Readings On
Latent Class Growth Analysis (Continued)

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

Nagin, D.S. & Tremblay, R.E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and non violent juvenile delinquency. Child Development, 70, 1181-1196.

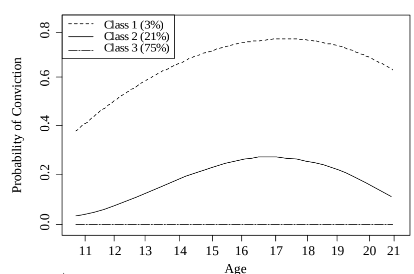
Nagin, D.S. & Tremblay, R.E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. Psychological Methods, 6, 18-34.

LCGA Vs GMM
Modeling Without Vs With Random Effects

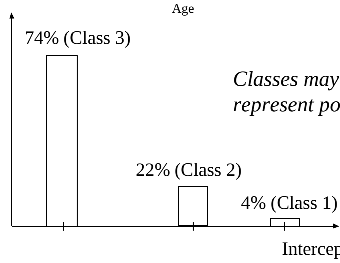
23

Non-Parametric View Of LCGA
Applied To The Cambridge Data

3-class LCGA
    LogL = -1,072
    (12 parameters)
    BIC = 2,215



3-class LCGA
only intercept class means
varying across classes
    LogL = -1,073
    (8 parameters)
    BIC = 2,194

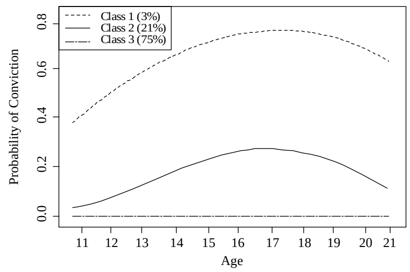


Classes may be "real" or only represent points on a continuum

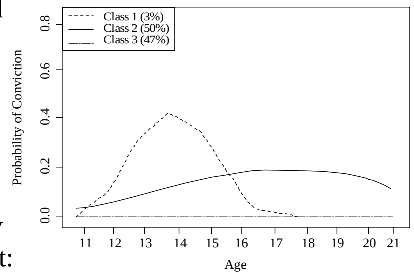
24

LCGA Vs GMM Applied To The Cambridge Data

3-class LCGA
LogL = -1,072
(12 parameters)
BIC = 2,215



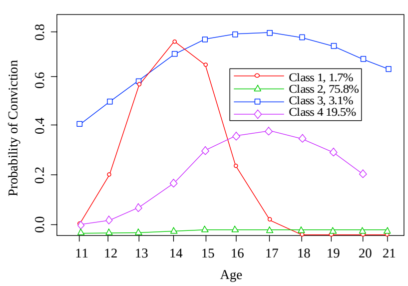
3-class GMM with a normal random intercept and a zero class
LogL = -1,067
(11 parameters)
BIC = 2,200

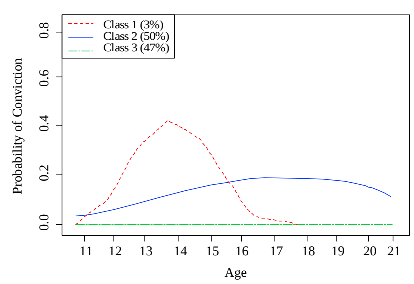


1-class GM with a normally distributed random intercept:
LogL = -1,075 (5 par's), BIC = 2,180

25

LCGA Vs GMM Applied To The Cambridge Data (Continued)





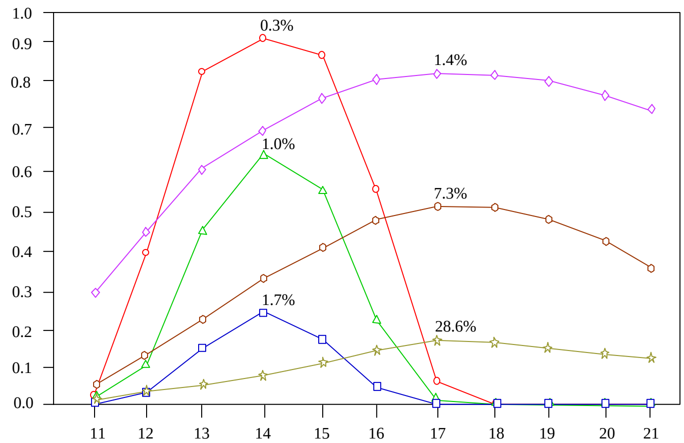
4-class LCGA
LogL = -1,066 (16 par's)
BIC = 2,228
Bootstrap LRT 3 vs. 4 classes: $p = 0.08$
→ can't reject 3 classes

3-class GMM and a normally distributed random intercept and a zero class
LogL = -1,067 (11 par's)
BIC = 2,200

– Is the late-peaking 3% class in the 4-class LCGA "real"?

26

3-Class GMM With Non-Parametric Random Intercept Distribution



3-class GMM with non-parametric random intercept distribution
logL = -1,066 (15 par's), BIC = 2,222

27

Input Three-Class GMM On The Cambridge Data

```
TITLE:      GMM, 3 classes, 1 zero class
            ordered polytomous variables for conviction at each
            age 11-21
            dep. variable 0, 1 , 2 (0, 1, or more convictions)

DATA:       FILE IS naginordered.dat;

VARIABLE:   NAMES ARE u11 u12 u13 u14 u15 u16 u17 u18 u19 u20 u21
            c1 c2 c3 c4 c;
            USEVAR = u11-u21;
            CATEGORICAL = u11-u21;
            CLASSES = c(3);

ANALYSIS:   TYPE = MIXTURE;
            ALGORITHM = INTEGRATION;
            STARTS = 500 10; STITER = 20;
```

28

Input Three-Class GMM On The Cambridge Data (Continued)

```
MODEL:      %OVERALL%
            i s q |  u11@-.6   u12@-.5   u13@-.4 u14@-.3 u15@-.2
            u16@-.1  u17@0   u18@.1 u19@.2 u20@.3 u21@.4;
            s-q@0;
            %c#1%
            [i*2 s q];
            i;
            %c#2%
            [i@0 s q];
            i;
            %c#3%
            [u11$1-u21$1@15];
            [u11$2-u21$2@16];
            [i-q@0];
            i@0;

OUTPUT:     RESIDUAL TECH1 TECH7 TECH8;

PLOT:       TYPE = PLOT3;
            SERIES = u11-u21(s);
```

29

Input Three-Class GMM On The Cambridge Data Using A Nonparametric Approach

```
TITLE:      GMM, 3 classes, 1 zero class nonparametric
            ordered polytomous variables for conviction at each
            age 11-21
            dep. variable 0, 1 , 2 (0, 1, or more convictions)

DATA:       FILE IS naginordered.dat;

VARIABLE:   NAMES ARE u11 u12 u13 u14 u15
            u16 u17 u18 u19 u20
            u21 c1 c2 c3 c4 c;
            USEV = u11-u21;
            CATEGORICAL = u11-u21;
            CLASSES = c(7);

ANALYSIS:   TYPE = MIXTURE;
            STARTS = 500 10; STITER = 20;
```

30

Input Three-Class GMM On The Cambridge Data Using A Nonparametric Approach (Continued)

```
MODEL:     %OVERALL%
           i s q |  u11@-.6   u12@-.5   u13@-.4 u14@-.3 u15@-.2
                    u16@-.1  u17@0   u18@.1 u19@.2 u20@.3
                    u21@.4;

           %c#1% !c1ci1
           [i] (a);
           [s] (1);
           [q] (2);

            %c#2% !c1ci2
           [i] (ac1);
           [s] (1);
           [q] (2);

            %c#3% !c1ci3
           [i] (ac2);
           [s] (1);
           [q] (2);
```


Input Three-Class GMM On The Cambridge Data Using A Nonparametric Approach (Continued)

```
           %c#4% !c2ci1
           [i@0];
           [s] (3);
           [q] (4);

           %c#5% !c2ci2
           [i] (bc1);
           [s] (3);
           [q] (4);

           %c#6% !c2ci3
           [i] (bc2);
           [s] (3);
           [q] (4);

           %c#7%
           [u11$1-u21$1@15];
           [u11$2-u21$2@16];

           [i-q@0];
```

Input Three-Class GMM On The Cambridge Data Using A Nonparametric Approach (Continued)

```
MODEL CONSTRAINT:
           NEW(c1*0 c2*0);
           ac1 = a + c1;
           ac2 = a + c2;
           bc1 = c1;
           bc2 = c3;
```

33