Categories or Continua? The Correspondence Between Factor and Mixture Models

Eric Loken
HDFS Pennsylvania State University

Peter Molenaar
HDFS Pennsylvania State University

May 18/19 2006

Outline

- The case for mixture models
- Common factor model
- Latent profile model
- Formal correspondence between the two
- Empirical example
- Summary

Why Categories?

- Discrete types exist in nature
- Distinct causal structures
- Explain associations among observed variables
- Distinct predictions for different categories

Why Continua?

- Individual differences exist in nature
- Reflect underlying causal structures
- Explain associations with simpler structure
- Make predictions

Infant Temperament

- Kagan -- inhibited versus uninhibited
- Distinct patterns of reactivity in terms of motor activity and crying in infancy related to behaviors in later childhood
- But is it categorical or continuous?

Metabolic Syndrome

- Collection of risk factors associated with obesity, diabetes, cardiovascular problems
- Triglycerides, glucose, HDL cholesterol, systolic blood pressure, diastolic blood pressure, waist circumference
- Above threshold on 3 or more indicators sometimes considered diagnostic of having “metabolic syndrome”

Mixture Models in Measurement

- Modeling birth weight and gestational age of newborns in Finland
- Adding "guessing" parameter to IRT model
- Identifying over- and under-reporters in self-reports of dietary intake

Choosing Between Mixture and Factor Models

- Should be made on theoretical and substantive grounds
- Deep divisions between "camps" often not justified
- Some models with latent "categories" or "continua" are surprisingly similar in formal structure and provide similar fit to the data

Common Factor Model

$$
\begin{aligned}
Y_{1i} &= \lambda_{11}F_{1i} + \lambda_{12}F_{2i} + \lambda_{13}F_{3i} + \varepsilon_{1i} \\
Y_{2i} &= \lambda_{21}F_{1i} + \lambda_{22}F_{2i} + \lambda_{23}F_{3i} + \varepsilon_{2i} \\
Y_{3i} &= \lambda_{31}F_{1i} + \lambda_{32}F_{2i} + \lambda_{33}F_{3i} + \varepsilon_{3i} \\
Y_{4i} &= \lambda_{41}F_{1i} + \lambda_{42}F_{2i} + \lambda_{43}F_{3i} + \varepsilon_{4i} \\
Y_{5i} &= \lambda_{51}F_{1i} + \lambda_{52}F_{2i} + \lambda_{53}F_{3i} + \varepsilon_{5i} \\
\cdots & \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
Y_{pi} &= \lambda_{p1}F_{1i} + \lambda_{p2}F_{2i} + \lambda_{p3}F_{3i} + \varepsilon_{pi}
\end{aligned}
$$

Mixture model also implies a covariance structure

$$\Sigma = \Lambda \Phi \Lambda^T + \Theta$$

Σ = implied variance/covariance matrix

Λ = matrix of factor loadings

Φ = matrix factor variances and covariances

Θ = diagonal matrix of unique variances

Maximum Likelihood Estimation

$$LL(Y \mid \mathbf{\Lambda}, \Phi, \Theta) = -\frac{1}{2} n[\log|\mathbf{\Sigma}| + tr(\mathbf{S}\mathbf{\Sigma}^{-1})]$$

Common Factor Model

$$
\begin{aligned}
Y_{1i} &= \lambda_{11}F_{1i} + \lambda_{12}F_{2i} + \lambda_{13}F_{3i} + \varepsilon_{1i} \\
Y_{2i} &= \lambda_{21}F_{1i} + \lambda_{22}F_{2i} + \lambda_{23}F_{3i} + \varepsilon_{2i} \\
Y_{3i} &= \lambda_{31}F_{1i} + \lambda_{32}F_{2i} + \lambda_{33}F_{3i} + \varepsilon_{3i} \\
Y_{4i} &= \lambda_{41}F_{1i} + \lambda_{42}F_{2i} + \lambda_{43}F_{3i} + \varepsilon_{4i} \\
Y_{5i} &= \lambda_{51}F_{1i} + \lambda_{52}F_{2i} + \lambda_{53}F_{3i} + \varepsilon_{5i} \\
\cdots & \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
Y_{pi} &= \lambda_{p1}F_{1i} + \lambda_{p2}F_{2i} + \lambda_{p3}F_{3i} + \varepsilon_{pi}
\end{aligned}
$$

Multivariate Normal Mixture

$$Y \sim \pi_1 N(\mu_1, \Sigma_1) + \pi_2 N(\mu_2, \Sigma_2) + \ldots + \pi_k N(\mu_k, \Sigma_k)$$

Observed Y distribution is composed of K subgroups.

Yields profiles of means, component probabilities and conditional variances.

Conditional independence assumes within group covariance matrices are diagonal.

ML Estimation

$$L(Y \mid \pi, \mu, \Sigma) = \Pi(\pi_1 N(\mu_1, \Sigma_1) + \ldots + \pi_k N(\mu_k, \Sigma_k))$$

Similar Models?

- A formal correspondence between mixture and factor models has been noted by:
  - McDonald 1967
  - Bartholomew 1988
  - Waller & Meehl 1998
  - Molenaar & von Eye 1994

Similar Models

- Although often described completely differently, models actually highly similar
- Both imply a variance/covariance structure
- “factor loading” = “conditional mean”
- “factor variance” = “group probability”
- “uniqueness” = “pooled within variance”

First and Second Moments Implied by

Latent Profile Model

$$E(y_i) = \Sigma \pi_k \mu_{ik}$$

$$E(y_i^{\,2}) = \Sigma \pi_k \mu_{ik}^2 + \pi_k \sigma_{ik}^2$$

$$E(y_i y_j) = \Sigma \pi_k \mu_{ik} \mu_{jk}$$

Latent profile model also implies a covariance structure

$$\Sigma = \Lambda \Phi \Lambda^T + \Theta$$

Σ = implied variance/covariance matrix

Λ = matrix of group means (μ)

Φ = diagonal matrix with group probabilities (π)

Θ = diagonal matrix of pooled within group variances

Rewriting Latent Profile Model

$$Y_{1i} = \mu_{11}F_{1i} + \mu_{12}F_{2i} + \mu_{13}F_{3i} + \varepsilon_{1i}$$
$$Y_{2i} = \mu_{21}F_{1i} + \mu_{22}F_{2i} + \mu_{23}F_{3i} + \varepsilon_{2i}$$
$$Y_{3i} = \mu_{31}F_{1i} + \mu_{32}F_{2i} + \mu_{33}F_{3i} + \varepsilon_{3i}$$
$$Y_{4i} = \mu_{41}F_{1i} + \mu_{42}F_{2i} + \mu_{43}F_{3i} + \varepsilon_{4i}$$
$$Y_{5i} = \mu_{51}F_{1i} + \mu_{52}F_{2i} + \mu_{53}F_{3i} + \varepsilon_{5i}$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$
$$Y_{pi} = \mu_{p1}F_{1i} + \mu_{p2}F_{2i} + \mu_{p3}F_{3i} + \varepsilon_{pi}$$

Framed as Missing Data Problem

Y_1	Y_2	Y_3	Y_4	….	Y_p	F_1	F_2	….	F_k
5	1	2	8	….	5	?	?	….	?
4	3	2	7	….	6	?	?	….	?
5	6	4	7	….	4	?	?	….	?
6	2	4	9	….	5	?	?	….	?
3	1	5	6	….	4	?	?	….	?

Empirical Example

- Perceptions of Adolescents
- Respondents asked to rate on a scale of 1-10 how likely it was that an adolescent would display given attribute
- 44 attributes in all, subset of 9 selected for this particular analysis

Observed Covariance Matrix of Attributes

	1	2	3	4	5	6	7	8	9
Confused	2.98								
Considerate	-0.02	2.82							
Depressed	1.89	-0.08	4.56						
Emotional	1.36	-0.03	1.63	2.39					
Generous	0	1.93	0.17	0.09	2.66				
Hardworking	0.06	1.6	0.04	0.1	1.79	2.86			
Helpful	-0.04	1.68	-0.03	-0.02	1.82	1.89	2.6		
Intelligent	0.18	0.43	0.37	0.32	0.51	0.53	0.61	1.91	
Tests Limits	0.86	-0.28	1.12	0.73	0.14	-0.07	0.05	0.46	2.31

Factor Analysis

- Ratings covary due to underlying shared factor structure
- Determine K as dimensionality of factor space
- Factors represent salient dimensions along which college students perceive adolescents
- Decompose into shared variance (communality) and non-shared variance (uniqueness)

Two Factor ML Solution with Varimax Rotation
(solution presented in correlation metric)

	F1	F2
Confused		.71
Considerate	.77	
Depressed		.71
Emotional		.70
Generous	.85	
Hardworking	.78	
Helpful	.83	
Intelligent	.29	.18
Tests Limits		.48

Latent Profile Model

- Population of raters is composed of heterogeneous mix of qualitatively different perspectives
- Successively test models with increasing K number of subgroups
- Subgroups represent qualitatively different perspectives (or types of raters)

Estimating Latent Profiles

- Successive models with K = 1,2,3 yield increasingly better fitting models
- For K = 3 multiple local modes found
- Solution found with the highest likelihood represented below in terms of conditional means, group probabilities and conditional variances
- However, this could also be represented as the factor loadings and uniquenesses

	Group Means			Pooled Within Variance
	1	2	3	
Confused	-0.37	0.02	0.23	2.91
Considerate	-1.54	1.53	-0.04	1.46
Depressed	-0.34	-0.02	0.23	4.48
Emotional	-0.53	-0.06	0.39	2.23
Generous	-1.66	1.66	-0.05	1.07
Hardworking	-1.53	1.56	-0.06	1.48
Helpful	-1.61	1.76	-0.16	0.96
Intelligent	-0.43	0.65	-0.16	1.71
Tests Limits	0.05	0.25	-0.21	2.26
Probabilities	0.28	0.29	0.43	

Implications

- Similar, but definitely not equal in fit
- Rotational indeterminacy?
- Clues to fitting mixture models better (start values, over fitting)

Similar… but not equal in fit

- Although the ML 3-class solution can be written as a 3 – factor model, the actual fit to the observed covariance matrix is poor
- The problem is immediately obvious if we remain flexible in viewing the mixture solution both as a factor model and as a mixture model
- The communality of the two factor solution is very different than the "between group" variance in the three group mixture

	Group Means			Pooled Within Variance
	1	2	3	
Confused	-0.37	0.02	0.23	2.91
Considerate	-1.54	1.53	-0.04	1.46
Depressed	-0.34	-0.02	0.23	4.48
Emotional	-0.53	-0.06	0.39	2.23
Generous	-1.66	1.66	-0.05	1.07
Hardworking	-1.53	1.56	-0.06	1.48
Helpful	-1.61	1.76	-0.16	0.96
Intelligent	-0.43	0.65	-0.16	1.71
Tests Limits	0.05	0.25	-0.21	2.26
Probabilities	0.28	0.29	0.43	

Two Factor ML Solution with Varimax Rotation

	F1	F2
Confused		.71
Considerate	.77	
Depressed		.71
Emotional		.70
Generous	.85	
Hardworking	.78	
Helpful	.83	
Intelligent	.29	.18
Tests Limits		.48

Rotational Indeterminacy?

- If the conditional means and group probabilities can be represented as factor loadings and factor variances, then are there an infinite number of mixture solutions? Yes and no.

Rotations

$$\Sigma = \Lambda \Phi \Lambda^T + \Theta$$

$$\Sigma = \Lambda T T^T \Phi T T^T \Lambda^T + \Theta$$

	Group Means 1	2	3	Pooled Within Variance
Confused	-0.37	0.02	0.23	2.91
Considerate	-1.54	1.53	-0.04	1.46
Depressed	-0.34	-0.02	0.23	4.48
Emotional	-0.53	-0.06	0.39	2.23
Generous	-1.66	1.66	-0.05	1.07
Hardworking	-1.53	1.56	-0.06	1.48
Helpful	-1.61	1.76	-0.16	0.96
Intelligent	-0.43	0.65	-0.16	1.71
Tests Limits	0.05	0.25	-0.21	2.26
Probabilities	0.28	0.29	0.43	

“Equivalent” Mixture Solution

	Group Means			Pooled Within Variance
	1	2	3	
Confused	-1.47	-0.14	0.23	2.91
Considerate	-10.39	0.25	-0.04	1.46
Depressed	-1.25	-0.15	0.23	4.48
Emotional	-1.87	-0.25	0.39	2.23
Generous	-11.22	0.28	-0.05	1.07
Hardworking	-10.44	0.27	-0.06	1.48
Helpful	-11.28	0.37	-0.16	0.96
Intelligent	-3.53	0.20	-0.16	1.71
Tests Limits	-0.51	0.17	-0.20	2.26
Probabilities	0.01	0.56	0.43	

Insights into Fitting Models

- Suggests starting values
- Fitting too many classes to mixture model overdetermines the covariance structure
- Connection with clustering solutions as discussed in other fields (computer science and math)

Choosing Start Values

- Mixture models often multimodal
- Suggestion is to start from variety of initial values
- Factor structure provides excellent choice of starting values
- In this example, four class mixture starting at hi/lo on two factor model fits very well

Summary

- Up to first and second moments, mixture model solutions can be rewritten as factor models
- Mixture solution is not necessarily a good fit of covariance structure of the data
- Rotational indeterminacy of factor solution will not fit mixture model likelihood equivalently

Summary

- Complementary view of mixture and factor models yields insight for estimation and interpretation
- Many areas for future research on this topic (estimation, model diagnostics, extensions to other models etc.)