Mischverteilungsmodelle

Jürgen Rost und Edgar Erdfelder

Statistische Modelle basieren häufig auf der Annahme, daß die Elemente einer Stichprobe – im Regelfall Versuchspersonen (Vpn) in einer bestimmten Untersuchungssituation – Beobachtungsdaten liefern, die einer für alle Stichprobenelemente identischen Verteilung folgen. Die Parameter statistischer Modelle, die auf dieser Annahme basieren, charakterisieren somit ausnahmslos die gesamte zugrundeliegende Population. So beschreiben beispielsweise die Ladungszahlen bei einer Faktorenanalyse den linearen Zusammenhang einer manifesten und einer latenten Variablen in einer bestimmten Grundgesamtheit (vgl. Bentler, Wu & Houck, Rietz, Rudinger & Andres sowie Schönemann & Borg, in diesem Band). Analog sind Regressionsgewichte, Pfadkoeffizienten, Koeffizienten in log-linearen Modellen und Itemschwierigkeiten Parameter, die jeweils die gesamte zugrundeliegende Personenpopulation kennzeichnen (vgl. dazu die Beiträge von Andersen, Andres und Roskam, in diesem Band). Ein großer Teil der Statistik befaßt sich zwar auch mit dem Vergleich verschiedener Stichproben, die möglicherweise unterschiedlichen Populationen entstammen. Allerdings sind diese Populationen dann empirisch eindeutig definiert, d.h. durch beobachtbare Tatbestände wie z.B. die Versuchsbedingung oder die Ausprägung bestimmter Variablen wie Alter, Bildungsstand etc. festgelegt (vgl. die Kapitel von Bortz, Diepgen und Willmes, in diesem Band).

Mischverteilungsmodellen liegt demgegenüber die Idee zugrunde, daß die beobachteten Daten eine Mischung aus verschiedenen Teilpopulationen darstellen, die durch unterschiedliche Verteilungsparameter gekennzeichnet sind. Trifft diese Annahme zu, so ist es offenbar nicht sinnvoll, eines der o.g. statistischen Modelle für alle Elemente einer Stichprobe anzupassen. Eine naheliegende Alternative wäre, die Stichprobe zunächst einmal zu „entmischen“. Dies bedeutet, sie derart in homogene Substichproben zu zerlegen, daß ein bestimmtes statistisches Modell für jede der Substichproben gilt, aber mit unterschiedlichen Parameterwerten (Faktorladungen, Regressionskoeffizienten etc.) zwischen den Substichproben.

Ein Modell, das unterschiedliche Parameterwerte eines bestimmten „lokalen“ (d.h. subpopulationsspezifischen) Verteilungsmodells in verschiedenen Teilpopulationen zuläßt, ist entweder ein *Moderatorvariablenmodell* oder ein *Mischverteilungsmodell.* Von einem Moderatorvariablenmodell spricht man, wenn die Subpopulationen durch Ausprägungen einer beobachtbaren („manifesten“) Variablen definiert sind, eben der „Moderatorvariablen“. Moderatorvariablen sind somit „manifeste mischende Variablen“. Die Berücksichtigung von Moderatorvariablen stellt insofern keine hohen Ansprüche an die statistische Modellbildung, als ja die Parameter eines bestimmten statistischen Modells im Prinzip für jede (bekannte) Teilgruppe getrennt geschätzt werden können. Das Problem besteht hier also weniger in der Modellbil-

https://doi.org/10.25521/HQM26

dung und den damit assoziierten statistischen Problemen (s. Abschnitt 3), sondern in der *Identifizierung* relevanter Moderatorvariablen für eine konkrete substanzwissenschaftliche Fragestellung. Dies kann natürlich ebenfalls ein sehr komplexes und für die sozial- und verhaltenswissenschaftliche Forschung sehr wichtiges Problem darstellen (vgl. Schmitt, 1992).

Ein Mischverteilungsmodell (MVM) bezieht sich dagegen auf den Fall, daß die mischende Variable latent, d.h. nicht beobachtbar ist. Anders als Moderatorvariablen können latente mischende Variablen nicht unabhängig von einem konkreten MVM definiert werden. Latente mischende Variablen sind dadurch implizit definiert, daß für jede ihrer Ausprägungen bestimmte statistische Modelleigenschaften gelten, die in einer Mischung von Individuen mit verschiedenen Ausprägungen dieser latenten Variablen verlorengehen. Die einzige Möglichkeit, die unbeobachtbare mischende Variable näher zu bestimmen, ist daher die Suche nach latenten Subpopulationen, innerhalb derer ein vorgegebenes statistisches Modell paßt. Damit eröffnen MVM die Möglichkeit, interindividuelle Unterschiede bei der statistischen Modellbildung zu berücksichtigen.

Ganz sicher ist somit die Methodologie der MVM von besonderer Bedeutung für die Differentielle und die Diagnostische Psychologie. Allerdings wäre es falsch, die Bedeutung dieser Modellklasse hierauf einzuschränken. Die Stichproben, die auf der Grundlage von MVM analysiert werden, müssen nicht notwendigerweise Daten verschiedener Vpn sein, sondern können auch Daten einer einzelnen, über verschiedene Aufgaben oder Items hinweg untersuchten Vp sein. Man kann also MVM auch zur Analyse von „personenspezifischen Mischverteilungen“ einsetzen, die z.B. dadurch zustande kommen können, daß eine bestimmte Vp unterschiedliche kognitive Prozesse bei der Aufgabenbearbeitung einsetzt. Somit sind MVM auch für die allgemeinpsychologische, insbesondere die kognitionspsychologische Forschung sehr bedeutsam (vgl. Yantis, Meyer & Smith, 1991). Andere Teildisziplinen der Psychologie – beispielsweise die Entwicklungspsychologie – können ebenfalls von Anwendungen der MVM profitieren, was in Abschnitt 2.2 anhand eines Beispiels verdeutlicht wird (vgl. auch Everitt, 1986).

Nach diesen einleitenden Bemerkungen werden wir zunächst die mathematische Struktur von MVM auf allgemeiner Ebene beschreiben (Abschnitt 1), bevor wir einige für die Psychologie besonders nützliche spezielle MVM vorstellen und ihre Anwendungsmöglichkeiten anhand von Beispielen illustrieren (Abschnitt 2). Der dritte Abschnitt behandelt Probleme der statistischen Analyse von MVM (Identifizierbarkeitsprüfung, Parameterschätzung und Modellgeltungsprüfung). Abschließend wird weiterführende Literatur genannt, und es wird auf Computerprogramme verwiesen, die die aufwendigen numerischen Verfahren der Analyse von MVM realisieren (Abschnitt 4).

1 Die allgemeine Struktur von Mischverteilungsmodellen

Die Wahrscheinlichkeit eines einzelnen beobachtbaren Stichprobendatums, d.h. die Wahrscheinlichkeit einer bestimmten Ausprägung $\mathbf{y}$ einer möglicherweise vektoriellen (mehrdimensionalen) ZV $\mathbf{Y}$, sei durch die Wahrscheinlichkeitsfunktion $p(\mathbf{Y} = \mathbf{y})$ oder im Fall einer stetigen („kontinuierlichen“) ZV $\mathbf{Y}$ durch die Dichtefunktion

$f(\mathbf{Y} = \mathbf{y})$ beschrieben. Diese „globalen" Wahrscheinlichkeiten oder Wahrscheinlichkeitsdichten werden bei MVM auf bedingte („lokale") Wahrscheinlichkeiten $p(\mathbf{Y} = \mathbf{y}|X = x)$ bzw. Dichten $f(\mathbf{Y} = \mathbf{y}|X = x)$ zurückgeführt, wobei X die latente mischende Variable[1] ist und $p(\mathbf{Y} = \mathbf{y}|X = x)$ sowie $f(\mathbf{Y} = \mathbf{y}|X = x)$ die Verteilung der beobachteten ZV $\mathbf{Y}$ bei gegebener Ausprägung x von X beschreiben. Die globale Wahrscheinlichkeitsfunktion $p(\mathbf{Y} = \mathbf{y})$ oder Dichtefunktion $f(\mathbf{Y} = \mathbf{y})$ hängt natürlich nicht nur von den lokalen Verteilungen $p(\mathbf{Y} = \mathbf{y}|X = x)$ bzw. $f(\mathbf{Y} = \mathbf{y}|X = x)$ innerhalb der latenten Subpopulationen ab, sondern auch von der sogenannten *mischenden Verteilung*, d.h. von der Verteilung der Variablen X, die für den Fall einer diskreten mischenden Variablen X durch $p(X = x)$ beschrieben sei. Die Darstellung der globalen Verteilung von $\mathbf{Y}$ als Funktion der lokalen Verteilungen und der mischenden Verteilung erhält man dadurch, daß man die globale Verteilung als eine *Randverteilung* der gemeinsamen Verteilung von X und $\mathbf{Y}$ auffaßt, die bei Aggregierung über die Ausprägungen von X resultiert. Weist die mischende Variable die Ausprägungen $x = 1, 2, \ldots, c$ auf, so ergibt sich die globale Wahrscheinlichkeitsfunktion von $\mathbf{Y}$ aus der folgenden Summe:

$$p(\mathbf{Y} = \mathbf{y}) = \sum_{x=1}^{c} p(\mathbf{Y} = \mathbf{y} \cap X = x) = \sum_{x=1}^{c} p(\mathbf{Y} = \mathbf{y}|X = x) \cdot p(X = x) \; . \tag{1}$$

Ist $\mathbf{Y}$ nicht diskret, sondern stetig, so erhält man analog zu Gleichung (1) die globale Dichtefunktion:

$$f(\mathbf{Y} = \mathbf{y}) = \sum_{x=1}^{c} f(\mathbf{Y} = \mathbf{y} \cap X = x) = \sum_{x=1}^{c} f(\mathbf{Y} = \mathbf{y}|X = x) \cdot p(X = x) \; . \tag{2}$$

Unabhängig davon, ob $\mathbf{Y}$ diskret oder stetig ist, spricht man von einem *diskreten* oder *finiten MVM*, wenn die mischende Variable X diskret ist. Dies trifft sowohl auf Gleichung (1) (finites MVM für eine diskrete ZV $\mathbf{Y}$) als auch auf Gleichung (2) (finites MVM für eine stetige ZV $\mathbf{Y}$) zu, so daß in beiden Gleichungen nur über eine endliche Anzahl von Ausprägungen der mischenden Variablen X summiert werden muß. Ist die mischende Variable X dagegen stetig (mit reellwertigen Ausprägungen x und Dichtefunktion $f(X = x)$), so muß bei der Bestimmung der globalen Wahrscheinlichkeitsfunktion einer diskreten ZV $\mathbf{Y}$ über die möglichen Realisationen von X integriert statt summiert werden, was der folgenden Modellgleichung entspricht:

$$p(\mathbf{Y} = \mathbf{y}) = \int_{x=-\infty}^{+\infty} p(\mathbf{Y} = \mathbf{y}|X = x) \cdot f(X = x) \, dx \; . \tag{3}$$

Ist nicht nur X, sondern auch $\mathbf{Y}$ stetig, so erhält man analog zu Gleichung (3) die globale Dichtefunktion:

$$f(\mathbf{Y} = \mathbf{y}) = \int_{x=-\infty}^{+\infty} f(\mathbf{Y} = \mathbf{y}|X = x) \cdot f(X = x) \, dx \; . \tag{4}$$

[1] Im Prinzip kann die mischende Variable ebenfalls vektoriell (mehrdimensional) sein. In den meisten MVM wird jedoch eine eindimensionale mischende Variable angenommen. Um die Darstellung möglichst einfach zu halten, wollen wir uns in diesem Kapitel auf diesen Fall beschränken.

In den beiden letztgenannten Fällen spricht man von *kontinuierlichen* oder *stetigen MVM*, weil die mischende Variable stetig ist.

Verteilungen, die durch eine der Gleichungen (1) bis (4) definiert sind, bezeichnet man als (diskrete bzw. stetige) *Mischverteilungen.* Zu beachten ist allerdings, daß *jede* Wahrscheinlichkeits(dichte)funktion in einer der o.g. Formen geschrieben werden kann, solange man keinerlei Vorannahmen über die lokalen Verteilungen und die mischende Verteilung macht. So kann man z.B. eine beliebige Normalverteilung in der Form (2) darstellen, indem man beispielsweise die reellen Zahlen in c disjunkte und exhaustive Intervalle zerlegt, die den Subpopulationen $x = 1, 2, \ldots, c$ entsprechen, und dann geeignete Wahrscheinlichkeiten $p(X = x)$ sowie lokale Verteilungen für diese Intervalle wählt. Dieses Beispiel macht deutlich, daß es ohne Bezugnahme auf bestimmte lokale Verteilungsmodelle gar nicht möglich ist, „Mischpopulationen" und „homogene Populationen" zu unterscheiden. Mischverteilungen werden also erst interessant und von anderen Verteilungen abgrenzbar, wenn man zu speziellen MVM übergeht, d.h. den in den Gleichungen (1) bis (4) ausgedrückten Strukturannahmen weitere Annahmen über die mischende und/oder die lokalen Verteilungen hinzufügt.

Stetige MVM spielen in der psychologischen Methodenlehre schon seit längerem eine wichtige Rolle, auch wenn dies durch die Bezeichnung der Modelle häufig nicht deutlich gemacht wird. Ein prominentes Beispiel sind etwa alle parametrischen *Item-Response-Modelle* (IRT-Modelle; vgl. das Kapitel von Scheiblechner und das Kapitel von Roskam über *Latent-Trait*-Modelle, in diesem Band). Hier stellt die (globale) Verteilung der möglichen Antwortmuster $\mathbf{y} = (y_1, y_2, \ldots, y_j, \ldots, y_k)$ eines Tests mit k dichotomen Items eine Mischung von lokalen Wahrscheinlichkeiten $p(\mathbf{Y} = \mathbf{y} | X = x)$ dar, die zusätzlich von der Verteilung der stetigen latenten Variablen X (des „*latent trait*") in der Population abhängt. Die globale Wahrscheinlichkeitsfunktion der Antwortmuster hat daher die typische Struktur des stetigen MVM gemäß Gleichung (3) (vgl. Rost & Strauß, 1992). Beim zweiparametrigen logistischen Modell (Birnbaum-Modell) für dichotome Items ($y_j \in \{0, 1\}$) gilt beispielsweise für die lokale Wahrscheinlichkeitsfunktion (vgl. Kubinger sowie Roskam, in diesem Band):

$$p(\mathbf{Y} = \mathbf{y} | X = x) = \prod_{j=1}^{k} \frac{exp(y_j \alpha_j (x - \beta_j))}{1 + exp(\alpha_j (x - \beta_j))} \,. \tag{5}$$

Setzt man Gleichung (5) in Gleichung (3) ein und fügt eine Annahme über die Verteilung $f(X = x)$ des mischenden „*latent trait*" hinzu, so erhält man die globale Wahrscheinlichkeitsfunktion der Antwortmuster (vgl. Andersen & Madsen, 1977; Holland, 1990).

Während stetige MVM in der psychologischen Methodenlehre schon einen etablierten Platz einnehmen, wird die Bedeutung finiter MVM für die psychologische Forschung erst in jüngerer Zeit betont (z.B. Erdfelder, 1990; Rost & Langeheine, 1991; Sixtl, 1985, 1993; Yantis, Meyer & Smith, 1991). Dies ist insofern erstaunlich, als gerade diese Modellklasse sich in substanzwissenschaftlicher Hinsicht als sehr fruchtbar erweist. Wir werden aus diesem Grund in diesem Kapitel vorwiegend Beispiele für finite MVM ansprechen und stetige MVM nur am Rande erwähnen.

Obwohl finite MVM ähnlich wie Clusteranalysen (vgl. Meiser & Humburg, in diesem Band) von der Annahme einer Subgruppenstruktur ausgehen, werden Elemente der Stichprobe anders als bei Clusteranalysen einer bestimmten Subgruppe nicht einfach zugeordnet. Vielmehr kann für jedes Individuum (bzw. jede Beobachtung $\mathbf{Y} = \mathbf{y}$) eine Zugehörigkeitswahrscheinlichkeit zu jeder der c latenten Klassen angegeben werden. Diese Zugehörigkeitswahrscheinlichkeit einer Beobachtung $\mathbf{y}$ zu einer Klasse x wird als „*recruitment probability*" bezeichnet und läßt sich mit Hilfe des Bayesschen Theorems (vgl. auch Molenaar & Lewis, in diesem Band) im diskreten Fall aus der Wahrscheinlichkeit der Beobachtung $\mathbf{y}$ in Klasse x, $p(\mathbf{Y} = \mathbf{y}|X = x)$, wie folgt berechnen (vgl. Gleichung (1)):

$$p(X = x|\mathbf{Y} = \mathbf{y}) = \frac{p(\mathbf{Y} = \mathbf{y} \cap X = x)}{p(\mathbf{Y} = \mathbf{y})} = \frac{p(\mathbf{Y} = \mathbf{y}|X = x) \cdot p(X = x)}{\sum_{j=1}^{c} p(\mathbf{Y} = \mathbf{y}|X = j) \cdot p(X = j)} . \quad (6)$$

Die *recruitment probabilities* addieren sich zu 1, d.h. es gilt für jede Beobachtung $\mathbf{y}$

$$\sum_{x=1}^{c} p(X = x|\mathbf{Y} = \mathbf{y}) = 1 , \quad (7)$$

was letztlich Ausdruck der Modellannahme ist, daß es sich um disjunkte und exhaustive Klassen handelt. Analoge Resultate ergeben sich, falls $\mathbf{Y}$ nicht diskret, sondern stetig ist. In Gleichung (6) sind dann lediglich die entsprechenden Wahrscheinlichkeitsfunktionen durch Wahrscheinlichkeitsdichtefunktionen zu ersetzen.

Möchte man trotz der Tatsache, daß die *recruitment probabilities* in den seltensten Fällen 0 oder 1 sind, analog zur Clusteranalyse manifeste Klassen bilden, d.h. jedes Individuum genau einer Klasse zuordnen, so weist man die Individuen zweckmäßigerweise jeweils derjenigen Klasse x zu, für die die Schätzung von $p(X = x|\mathbf{Y} = \mathbf{y})$, welche man durch Einsetzen von Parameterschätzungen in Gleichung (6) erhält, bei dem betreffenden Individuum maximal wird. Dabei kann es natürlich passieren, daß die relativen Größen der so gebildeten manifesten Klassen nicht genau den geschätzten Klassengrößen der latenten Klassen, $\hat{p}(X = x)$, entsprechen.

Für jede der manifesten Klassen kann man die mittleren Zuordnungswahrscheinlichkeitsschätzungen aller zugeordneten Individuen berechnen, die sich sehr gut als ein anschauliches Maß dafür interpretieren lassen, wie gut die Stichprobe „entmischt" werden konnte, und somit auch dafür, wie gut von den beobachteten Daten auf die Klassenzugehörigkeit rückgeschlossen werden kann: Je mehr sich die mittleren Zuordnungswahrscheinlichkeitsschätzungen dem Wert 1 nähern, desto perfekter gelingt die Diagnose der Klassenzugehörigkeit aufgrund des Datenvektors $\mathbf{y}$.

2 Spezielle Mischverteilungsmodelle

In diesem Abschnitt sollen einige spezielle MVM vorgestellt werden. Die zu besprechenden MVM wurden nicht nach statistischen, sondern nach substanzwissenschaftlichen Kriterien ausgewählt. Deshalb fehlen hier viele Modelle, die in der einschlägigen statistischen Literatur (z.B. Everitt & Hand, 1981; McLachlan & Basford, 1988; Sixtl, 1993; Titterington, Smith & Makov, 1985) ausgiebig untersucht wurden, aber bislang keine oder nur vereinzelte Anwendungen in der Psychologie aufzuweisen haben. Umgekehrt wird man einige MVM erwähnt finden, die speziell auf psychologische Probleme und Fragestellungen zugeschnitten sind, jedoch in der allgemeinen statistischen Literatur bislang unberücksichtigt geblieben sind.

2.1 Mischungen uni- und multivariater Normalverteilungen

Finite Normalverteilungsmischungen entstehen aus Modellgleichung (2), indem für die lokalen Dichten $f(\mathbf{Y} = \mathbf{y}|X = x)$ uni- bzw. multivariat normale Dichten $f(\mathbf{Y} = \mathbf{y}; E_x(\mathbf{Y}), Var_x(\mathbf{Y}))$ eingesetzt werden, wobei der Mittelwert bzw. Centroid $E_x(\mathbf{Y})$ und eventuell auch die Varianz bzw. Varianz-Kovarianzmatrix $Var_x(\mathbf{Y})$ zwischen den Stufen $x = 1, 2, \ldots, c$ von X variieren können. Man erhält dann die Modellgleichung

$$f(\mathbf{Y} = \mathbf{y}) = \sum_{x=1}^{c} f(\mathbf{Y} = \mathbf{y}; E_x(\mathbf{Y}), Var_x(\mathbf{Y})) \cdot p(X = x) \; . \qquad (8)$$

Finite Normalverteilungsmischungen sind häufig, aber keineswegs immer auch multimodale Verteilungen. Eine hinreichende Bedingung für Unimodalität einer Normalverteilungsmischung mit $c = 2$ Subpopulationen haben Everitt und Hand (1981) angegeben.

Univariate finite Normalverteilungsmischungen dürften die ersten Mischungen gewesen sein, die untersucht wurden (Pearson, 1894). In der Entwicklungspsychologie gibt es einige Anwendungen derartiger Mischungen, welche die Überprüfung von Stufentheorien der Entwicklung zum Ziel haben. Erdfelder (1990, S. 492f.) berichtet hierüber zusammenfassend. *Multivariate finite Normalverteilungsmischungen* wurden von Wolfe (1970) als clusteranalytisches Verfahren vorgeschlagen (vgl. Meiser & Humburg, in diesem Band). Die gleiche Art von Modellbildung läßt sich, wie Thomas (1983, 1985) gezeigt hat, im Rahmen der Verhaltensgenetik zur Prüfung der Hypothese geschlechtsspezifischer Vererbungsprozesse einsetzen. Weitere Anwendungsbeispiele aus dem klinisch- und forensisch-psychologischen Bereich berichtet Everitt (1986).

2.2 Mischungen von Binomialverteilungen

Mischungen von Binomialverteilungen erhält man, indem in die Modellgleichung (1) (finite Binomialmischungen) oder (3) (stetige Binomialmischungen) Binomialwahrscheinlichkeiten für die lokalen Wahrscheinlichkeiten $p(Y = y|X = x)$ einer univariaten, kategorialen ZV Y eingesetzt werden. Hat Y $m + 1$ verschiedene Kategorien $y = 0, 1, \ldots, m$, so folgt Y für die Ausprägung x von X der Verteilung

$$p(Y = y|X = x) = \binom{m}{y} \pi_x^y (1 - \pi_x)^{m-y} \; , \qquad (9)$$

wobei der Binomialparameter π_x (die „Erfolgswahrscheinlichkeit“) zwischen den Ausprägungen von X variiert.

Binomialmischungen wurden ausgiebig von Sixtl (1985, 1993) untersucht. Eine theorietestende Anwendung finiter und stetiger Binomialmischungen im Bereich der Entwicklungspsychologie stellt Erdfelder (1990, S. 498–503) vor: Um herauszufinden, ob die Wahrscheinlichkeiten der Lösung von Klasseninklusionsproblemen nach Piaget und der Beherrschung des Umkehrlernens nach Kendler im Verlaufe der Ontogenese kontinuierlich wachsen oder aber – wie das nach Piagets Stufentheorie der kognitiven Entwicklung zu erwarten ist – sprunghaft von einem Stadium der Nichtbeherrschung zu einem Stadium der Beherrschung ansteigen, wurden Daten einer

Querschnittsuntersuchung von Tabor und Kendler (1981) herangezogen. Stimmt die Stufentheorie, so sollten die Kinder der verschiedenen Altersgruppen jeweils einer von zwei latenten Subpopulationen angehören, nämlich – je nach Entwicklungsstadium – entweder der Gruppe der „wissenden“ oder der Gruppe der „nichtwissenden“ Kinder. Wenn die Entwicklung der zugrundeliegenden Fähigkeiten dagegen kontinuierlich verläuft, sollte das Modell einer stetigen Binomialmischung die Daten besser beschreiben. Als Dichteverteilung $f(X = x)$ im Modell der stetigen Binomialmischung wählte Erdfelder (1990) die Beta-Verteilung, so daß als Mischverteilung die Beta-Binomialverteilung resultiert (vgl. Sixtl, 1993, S. 434ff). Eine beta-verteilte ZV X kann beliebige reelle Werte zwischen 0 und 1 annehmen, und ihre Verteilung ist durch zwei positive Parameter a und b definiert, die den Erwartungswert

$$E(X) = a/(a+b) \tag{10}$$

und die Varianz

$$Var(X) = ab(a+b)^{-2}(a+b+1)^{-1} \tag{11}$$

festlegen. Die Beta-Binomialmischung hat die Wahrscheinlichkeitsfunktion

$$p(Y = y) = \binom{m}{y} \cdot \frac{B(y+a, m-y+b)}{B(a,b)}\ , \tag{12}$$

wobei $B(v,w)$ die sogenannte Beta-Funktion

$$B(v,w) := \int_{z=0}^{1} z^{(v-1)}(1-z)^{(w-1)} \quad dz \tag{13}$$

für positiv-reellwertiges v und w ist.

Die Anwendung dieser Modelle auf die Daten von Tabor und Kendler (1981) zeigte, daß kein Modell generell überlegen ist. Umkehrlernprobleme lassen sich besser durch Beta-Binomialmischungen beschreiben, was für die Hypothese der kontinuierlichen Entwicklung spricht. Dagegen paßt das Modell einer finiten Binomialmischung mit $c = 2$ Subpopulationen besser auf die Klasseninklusionsdaten, was diesbezüglich für Piagets Stufenkonzept spricht (Erdfelder, 1990, S. 498f).

2.3 Mischungen von Produkten von Binomialverteilungen

Eine multivariate Verallgemeinerung gemischter Binomialverteilungen und zugleich einen Spezialfall der im folgenden Abschnitt zu schildernden LCA stellen die „*binomial response models*“ dar. Hierbei wird angenommen, daß die Antwortvariablen $Y_1, Y_2, \ldots, Y_k$ einer Menge von k Items (oder beliebiger anderer Variablen) binomialverteilt sind *und* daß Unabhängigkeit der Variablen innerhalb sogenannter latenter Klassen (Subpopulationen) gilt (Rost, 1985). Damit entspricht die lokale Verteilung der vektoriellen ZVn $\mathbf{Y} := (Y_1, Y_2, \ldots, Y_k)$ für $X = x$ dem Produkt von Binomialverteilungen, wobei die Einzelvariablen Y_j natürlich unterschiedliche Binomialparameter m_j und π_{jx} haben dürfen:

$$p(\mathbf{Y} = (y_1, y_2, \ldots, y_k)|X = x) = \prod_{j=1}^{k} \binom{m_j}{y_j} \pi_{jx}^{y_j}(1-\pi_{jx})^{(m_j-y_j)}\ . \tag{14}$$

Binomialverteilte Antwortvariablen kann man in vielen Fällen sowohl bei Leistungstests als auch bei Fragebögen mit Ratingskalen sinnvollerweise annehmen: Erfordert ein Item z.B. multiple Reaktionen, die alle unabhängig voneinander und mit gleicher Wahrscheinlichkeit (korrekt) auftreten oder nicht, so ist die Summe der gezeigten (korrekten) Reaktionen binomialverteilt. Für Ratingdaten beschreibt Rost (1985, 1988a, S. 190) einen hypothetischen Antwortprozeß, der zur Binomialannahme führt.

Ein entsprechendes *binomial response model* gibt es auch als stetiges MVM in Form einer Verallgemeinerung des dichotomen Rasch-Modells (Andrich, 1978a; Douglas, 1978). In der Praxis hat sich jedoch die Binomialannahme oft als zu restriktiv erwiesen, was seine Ursachen z.B. darin haben kann, daß die Einzelreaktionen, die zu einem Summenwert aufaddiert werden, eben doch nicht unabhängig voneinander sind. Da sich solche *binomial response models* auch als Spezialfälle von Schwellenwertmodellen darstellen lassen (vgl. dazu Langeheine & Rost, in diesem Band, Gleichungen (5) und (7); Wright & Masters, 1982; Rost, 1988a, S. 189ff) ist es in der Regel sinnvoller, mit letzteren zu arbeiten (vgl. Rost, 1988b).

2.4 Mischungen von Produkten von Multinomialverteilungen

Im Rahmen der sozialwissenschaftlichen Methodologie dürfte die von Paul Lazarsfeld (1950) entwickelte „*latent class analysis*“ (LCA, vgl. Langeheine & Rost, in diesem Band) das wohl älteste und inzwischen bekannteste MVM darstellen. Formal handelt es sich um eine finite Mischung von Produkten von Multinomialverteilungen, d.h. die beobachtete ZV ist ein Vektor von diskreten Variablen, deren Ausprägungen z.B. die (kategorialen) Antworten auf k Fragen eines Fragebogens sein können: $\mathbf{y} = (y_1, y_2, \ldots, y_k)$ mit $y_j \in \{0, 1, \ldots, m_j\}$. Aus der Annahme, daß sich die gemeinsame (multivariate) Verteilung dieser kategorialen Variablen nur in homogenen Subpopulationen – den latenten Klassen – als Produkt der einzelnen Multinomialverteilungen darstellen läßt, daß also Unabhängigkeit der Itemantworten nur innerhalb der latenten Klassen gilt (*Annahme der lokalen stochastischen Unabhängigkeit*), folgt das Modell der LCA, das offensichtlich einen Spezialfall der Modellgleichung (1) darstellt:

$$p(\mathbf{Y} = \mathbf{y}) = \sum_{x=1}^{c} \prod_{j=1}^{k} p(Y_j = y_j | X = x) \cdot p(X = x) \ . \qquad (15)$$

Aufgrund seiner Allgemeinheit bildet dieses Modell den Rahmen für eine Vielzahl von anderen Modellen (vgl. Formann, 1984, Lazarsfeld & Henry, 1968; Rost, 1988a; Langeheine & Rost, in diesem Band).

2.5 Mischungen von Markoff-Ketten

Der typische Anwendungsfall von Markoff-Ketten sind kategoriale Längsschnittdaten, bei denen die Übergangswahrscheinlichkeiten der Kategorien einer Variable Y von Zeitpunkt t zum Zeitpunkt $t + 1$ parametrisiert werden sollen. Da sich solche Markoff-Modelle oft als nicht auf die Daten passend erwiesen haben, liegt es nahe anzunehmen, daß die durch ein Markoff-Modell parametrisierten stochastischen Prozesse nur in Teilpopulationen gelten bzw. in verschiedenen Teilpopulationen in unterschiedlicher Weise (mit anderen Parametern). Gemischte Markoff-Modelle (Langeheine & van de Pol, 1992; vgl. auch Langeheine & Rost, in diesem Band) stellen daher diskrete MVM dar, in denen die lokalen Wahrscheinlichkeiten einer vektoriellen Variablen $\mathbf{Y}$ mit Ausprägungen $\mathbf{y} = (i, j, k, l)$ im Falle von 4 Zeitpunkten auf „Ausgangswahrscheinlichkeiten“ π_{ix} und „Übergangswahrscheinlichkeiten“ $\tau_{j|ix}, \tau_{k|jx}$ und $\tau_{l|kx}$ zurückgeführt werden:

$$p(\mathbf{Y} = \mathbf{y}|X = x) = \pi_{ix}\tau_{j|ix}\tau_{k|jx}\tau_{l|kx} \ . \tag{16}$$

Diese Modelle stellen insofern eine Erweiterung der normalen LCA dar (vgl. Abschnitt 2.4), als die Annahme lokaler stochastischer Unabhängigkeit der Variablen abgeschwächt ist: Durch die Übergangswahrscheinlichkeiten τ sind stochastische Abhängigkeiten zwischen Beobachtungen zu verschiedenen Zeitpunkten zugelassen. Anwendungsbeispiele und verschiedene Modellvarianten werden bei Langeheine und Rost (in diesem Band) vorgestellt.

2.6 Mischungen von Rasch-Modellen

In Abschnitt 1 wurde erwähnt, daß sich quantitative IRT-Modelle generell als stetige MVM darstellen lassen. Die zu messende latente Dimension (der *latent trait*) ist dabei die mischende Variable. Nun stellt das Rasch-Modell (und seine Verallgemeinerungen) unter den IRT-Modellen insofern einen Sonderfall dar, als man die Verteilung der latenten Variablen nicht kennen oder parametrisieren muß, um die globale Verteilung der Antwortmuster, $p(\mathbf{Y} = \mathbf{y})$ mit $\mathbf{y} = (y_1, y_2, \ldots, y_k)$ und $y_j \in \{0, 1\}$, unter Annahme der Modellgültigkeit zu spezifizieren. Konditioniert man die globale Antwortmuster-Wahrscheinlichkeit nämlich auf den mit diesem Antwortmuster verbundenen Summenwert $r = \sum_j y_j$, so erhält man eine Modellgleichung, in die nur die Itemparameter des Rasch-Modells σ_i und deren symmetrische Grundfunktionen γ_r eingehen (vgl. Fischer, 1974; Rost, 1996a):

$$p(\mathbf{Y} = \mathbf{y}) = p(\mathbf{Y} = \mathbf{y}|R = r) \cdot p(R = r) = p(R = r) \cdot exp(\sum_j y_j\sigma_j)/\gamma_r \ . \tag{17}$$

Die mischende Variable ist hier nicht mehr enthalten (weswegen man Rasch-Modelle auch „verteilungsfrei“ nennt), dafür aber die Wahrscheinlichkeitsfunktion der Summenwerte $p(R = r)$, die von der Verteilung der latenten Variablen abhängt. Diese Wahrscheinlichkeiten können aber direkt durch die beobachteten relativen Häufigkeiten der Summenwerte geschätzt werden, so daß die mischende Variable – wenn man so will – nicht latent, sondern „manifest“ ist (= die Summenwertvariable).

Diese Eigenschaften des Rasch-Modells sind zwar nicht unabdingbare Voraussetzung, aber doch sehr nützlich, wenn man das Rasch-Modell selbst zu einem diskreten MVM verallgemeinern will. Rost (1990) hat dieses Modell für dichotome Daten entwickelt, wobei nicht nur die Itemparameter σ_{ix}, sondern auch die Summenwertwahrscheinlichkeiten $\pi_{rx} = p(R = r|X = x)$ als klassenspezifische Parameter eingeführt werden (*mixed Rasch model*):

$$p(\mathbf{Y} = \mathbf{y}|X = x) = \pi_{rx} \cdot exp(\sum_j y_j\sigma_{jx})/\gamma_{rx} \ . \tag{18}$$

Mislevy und Verhelst (1990) diskutieren ein ähnliches Modell, jedoch ohne die besagte Konditionierung auf den Summenwert vorzunehmen, und Kelderman und Macready (1990) diskutieren diesen Ansatz im Rahmen log-linearer Modelle.

Rost (1991, 1996a,b) und von Davier und Rost (1995) beschreiben die Verallgemeinerung des Mischverteilungs-Rasch-Modells für mehrkategorielle, ordinale Daten. Dieses Modell ist zur Beantwortung sehr unterschiedlicher Fragestellungen eingesetzt worden. Köller, Rost und Köller (1994) sowie Rost und von Davier (1993) analysierten mit dem dichotomen *Mixed*-Rasch-Modell Tests zum räumlichen Vorstellungsvermögen und konnten die Hypothese bestätigen, daß derartige Aufgaben mit einer analytischen oder einer holistischen Strategie gelöst werden können. Es ließen sich in beiden Untersuchungen zwei latente Klassen identifizieren, an deren Parameterprofilen ablesbar ist, daß die zugehörigen Personen jeweils eine der beiden Strategien präferierten. Rost und Georg (1991) analysierten Daten aus einer nationalen Jugendstudie und konnten für die Skala „Jugendzentrismus" eine große (80 %) Klasse von „skalierbaren" und eine kleinere (20 %) Klasse von „unskalierbaren" Jugendlichen mittels des ordinalen *Mixed*-Rasch-Modells unterscheiden. Rost, Carstensen und von Davier (in press) analysierten Daten eines Persönlichkeitsfragebogens und fanden Klassen von Personen, die sich hinsichtlich ihres *response sets* bei der Beantwortung der Fragen anhand einer fünfstufigen Ratingskala unterscheiden. Von Davier und Rost (in press) haben das *Mixed*-Rasch-Modell und die normale *Latent-Class*-Analyse zu einem Hybrid-Modell verbunden und zeigen für das Konstrukt des *self monitoring*, daß sich eine Drei-Klassen-Struktur ergibt mit einer LCA-Klasse (*low self monitorer*) und zwei Rasch-Klassen unterschiedlicher Arten der Selbstpräsentation.

2.7 Weitere Ansätze

Die bisherigen Ausführungen haben deutlich gemacht, daß die Idee finiter MVM im Prinzip auf jedes der aus der psychologischen Methodenlehre bekannten statistischen Modelle übertragen werden kann. Das Grundprinzip ist immer gleich: Man nimmt an, daß ein bestimmtes statistisches Modell innerhalb bestimmter Subpopulationen unbekannter Größe gilt, aber mit variierenden Parameterwerten zwischen den Subpopulationen. So könnte man beispielsweise auch ein Mischverteilungs-Strukturgleichungsmodell konstruieren. Liegen „hybride" Datensätze basierend auf teilweise stetigen, teilweise diskreten ZVn Y_j vor, so lassen sich entsprechende hybride MVM konstruieren. Erdfelder (1988) hat beispielsweise ein hybrides Modell vorgeschlagen, das für die stetigen ZVn lokale multivariate Normalverteilungen und für die diskreten ZVn lokale unabhängige Multinomialverteilungen vorsieht. Wenn die Kovarianzen der stetigen Variablen Y_j zwischen den Realisationen der mischenden Variablen X variieren können, spielt X die Rolle einer „latenten Moderatorvariablen", die die Regressions- und Korrelationsbeziehungen zwischen den beobachtbaren stetigen Variablen verändert.

Heyer und Niederée (1992) haben gezeigt, daß die Idee finiter MVM sich auch zur „Probabilisierung" deterministischer Meßmodelle gut eignet. Bekanntlich sind deterministische Meßstrukturen (Niederée & Narens, in diesem Band) zumindest in der Psychologie empirisch selten haltbar. Nimmt man jedoch an, daß deterministische

Meßstrukturen „lokal" gelten und die beobachteten Daten Mischungen verschiedener deterministischer Strukturen darstellen, so erhält man ein probabilistisches Modell, das wesentlich eher geeignet ist, psychologische Daten adäquat zu beschreiben.

3 Statistische Probleme

Die statistische Anwendung von MVM beinhaltet grundsätzlich dreierlei: (1) die Identifizierbarkeitsprüfung des zu analysierenden Modells, (2) die Parameterschätzung und (3) die Prüfung der Anpassungsgüte.

3.1 Identifizierbarkeitsprüfung

Ein MVM, das durch Festlegung einer mischenden Verteilung und einer Familie von lokalen Verteilungen entsteht, ist genau dann global identifizierbar, wenn unterschiedlichen Konstellationen von Modellparametern immer auch unterschiedliche Verteilungen von $\mathbf{Y}$ zugeordnet werden. Ist ein MVM nicht global identifizierbar, so existieren modellkonforme Mischverteilungen, die durch *unterschiedliche* Modellparameterkonstellationen beschrieben werden können. In diesem Fall macht die Parameterschätzung im allgemeinen keinen Sinn, weil die Möglichkeit besteht, daß die Lösung nicht eindeutig ist. Ausnahmen können sich ergeben, wenn nur einige Modellparameter nicht identifizierbar sind und sich das substanzwissenschaftliche Interesse auf die identifizierbaren Parameter beschränkt. Für sehr viele Subklassen von MVM können die Bedingungen, unter denen *globale Identifizierbarkeit* gegeben ist, in der einschlägigen statistischen Literatur, insbesondere in der Monographie von Titterington et al. (1985, Kap. 3.1), nachgelesen werden.

Von der globalen Identifizierbarkeit ist die *lokale Identifizierbarkeit* zu unterscheiden. Globale Identifizierbarkeit impliziert immer auch lokale Identifizierbarkeit, während die Umkehrung dieser Aussage nicht gilt. Lokale Identifizierbarkeit ist gegeben, wenn in der „näheren Umgebung" einer gefundenen Lösung von *Maximum-Likelihood* (ML) Schätzgleichungen (vgl. Abschnitt 3.2) keine weiteren Lösungen existieren. Die Prüfung der lokalen Identifizierbarkeit setzt also nicht nur einen konkreten Datensatz (Realisation von $\mathbf{Y}$) voraus, sondern auch einen Vektor $\hat{\pi}$ von ML-Parameterschätzungen für alle (voneinander unabhängigen) Modellparameter. Hat die Matrix der ersten partiellen Ableitungen der Antwortmuster-Wahrscheinlichkeiten $p(\mathbf{Y} = \mathbf{y})$ nach den Modellparametern π vollen Spaltenrang, so ist das Modell lokal identifizierbar (vgl. Formann, 1984, S. 23ff, für Details dieses Verfahrens). Damit wird – anschaulich ausgedrückt – geprüft, ob jeder berechnete Schätzwert der Modellparameter wirklich auf den betreffenden Punkt fixiert ist, oder ob nicht ein ganzes Intervall den Bereich zulässiger, „gleich guter" Parameterwerte definiert.

3.2 Parameterschätzung

Bei finiten MVM sind grundsätzlich zwei Arten von Parametern zu schätzen, nämlich die „*mixing proportions*", d.h. die Wahrscheinlichkeitsfunktion $p(X = x)$ der mischenden Variablen, und die Parameter der lokalen Wahrscheinlichkeits(dichte)funktionen. Die Anzahl der Teilpopulationen c stellt dagegen keinen freien Parameter dar, sondern muß a priori festgelegt werden. Die empirische Bestimmung von c kann nur durch Vergleich der Modellanpassung unter verschiedenen Werten von c erfolgen. Formelle und informelle Techniken diskutieren Titterington et al. (1985, Kap. 5).

Karl Pearson (1894), der sich als erster mit Wahrscheinlichkeitsmischungen beschäftigt hat, schlug die sogenannte Momentmethode der Parameterschätzung vor. Sind k Modellparameter zu schätzen, so sind nach der Momentmethode die ersten k Momente der Mischverteilung als Funktion der Modellparameter auszudrücken und anschließend mit Schätzern dieser Momente (Stichprobenmittelwert, -varianz, etc.) gleichzusetzen. Die Lösung des resultierenden Gleichungssystems liefert Momentschätzer der Modellparameter. In statistischer Hinsicht haben diese Schätzer jedoch keine optimalen Eigenschaften (Titterington et al., 1985, Kap. 4.2).

Wann immer möglich, sollte deshalb auf die ML-Schätzmethode zurückgegriffen werden, die im Falle der Modellgültigkeit Schätzer mit günstigen statistischen Eigenschaften garantiert (siehe dazu Klauer, in diesem Band). Daß dies erst in den letzten beiden Jahrzehnten zunehmend realisiert wird, liegt daran, daß im Falle von MVM häufig keine expliziten Lösungen der ML-Schätzgleichungen existieren und numerische Verfahren der ML-Schätzung lange Zeit mit den unterschiedlichsten Problemen zu kämpfen hatten. Dies betrifft z.B. die Schätzung der *mixing proportions*, die jeweils im reellen Intervall $(0, 1)$ liegen und sich zu 1 aufaddieren müssen. Die Implementation einer derart komplexen Parameterrestriktion erwies sich lange Zeit als sehr schwierig, so daß beispielsweise negative Schätzer für einige *mixing proportions* resultieren konnten. Die Situation änderte sich mit der Verbreitung des von Dempster, Laird und Rubin (1977) als *expectation maximization* (EM) Algorithmus bezeichneten iterativen Verfahrens, das leicht zu implementieren und vor allen Dingen auch bei komplizierten Parameterrestriktionen sehr robust ist. Goodman (1974) hat diesen Algorithmus schon einige Jahre zuvor für das LCA-Modell entwickelt. Eine knappe Darstellung der Anwendung des EM-Algorithmus auf Parameterschätzprobleme im Rahmen finiter MVM findet man z.B. bei Rost (1988b, 1996a) oder Erdfelder (1990).

Dempster et al. (1977) haben gezeigt, daß der EM-Algorithmus unter sehr allgemeinen Randbedingungen auf ein Maximum der *Likelihood*-Funktion hin monoton konvergiert. Lediglich zwei Nachteile des Verfahrens sind zu konstatieren: (1) Im Falle multipler (lokaler) Maxima kann der EM-Algorithmus – wie andere numerische Verfahren auch – auf ein lokales statt auf das globale Maximum der *Likelihood*-Funktion hin konvergieren; (2) die Konvergenzgeschwindigkeit ist im allgemeinen recht langsam. Beide Nachteile sind bei der Verwendung schneller Computer nicht sehr gewichtig. Das Problem lokaler Maxima läßt sich praktisch nämlich dadurch entschärfen, daß man den Algorithmus mehrmals von unterschiedlichen Zufallsstartwerten aus startet. Ein Vergleich der erhaltenen Lösungen gibt Aufschluß darüber, ob mehrere lokale Maxima existieren und, wenn ja, welches mit dem höchsten *Likelihood*-Wert verbunden ist (und daher das „globale“ sein dürfte).

3.3 Prüfung der Anpassungsgüte

Ist das zu prüfende MVM an der Stelle der ML-Lösung lokal identifizierbar, so liegt es nahe, den *Likelihoodquotiententest* zur Prüfung der Modellgültigkeit heranzuziehen. Nahezu alle Computerprogramme zu MVM legen den Benutzern diesen Test oder einen anderen Test aus der Familie der *power divergence* χ^2-Statistiken – z.B. den Pearson-χ^2-Test – nahe (vgl. Read & Cressie, 1988). Unglücklicherweise ist dieser Test aber häufig nicht gerechtfertigt, weil die Regularitätsbedingungen, die in die Ableitung der asymptotischen χ^2-Verteilung der betreffenden Teststatistik eingehen, nicht erfüllt sind. Dies ist etwa dann der Fall, wenn zwei finite MVM mit $c = m, m \in \mathbb{N}$, und $c = m + 1$ gegeneinander getestet werden sollen (McLachlan & Basford, 1988, Kap. 1.10, Titterington et al., 1985, Kap. 5.4). Einige Autoren (z.B. Everitt & Hand, 1981) empfehlen trotz dieser Probleme die routinemäßige Anwendung der χ^2-Tests. Im allgemeinen dürfte es jedoch sicherer sein, eine transformierte χ^2-Statistik zu benutzen, deren Stichprobenverteilung sich in Simulationsstudien als gut mittels χ^2 approximierbar erwiesen hat (vgl. z.B. Wolfe, 1978), oder von vornherein die Referenzverteilung über eine Monte-Carlo-Studie zu generieren (vgl. Aitkin, Anderson & Hinde, 1981). Beide Wege sind extrem aufwendig, da sie im Grunde für jede neue Anwendung von MVM eine eigene Monte-Carlo-Studie erforderlich machen.

Ist die inferenzstatistische Absicherung der Modellgültigkeit nicht von zentraler Bedeutung oder praktisch nicht möglich, so gibt es neben graphischen Evaluationstechniken (vgl. Titterington et al., 1985) auch die Möglichkeit, konkurrierende Modelle mit informationstheoretischen Maßen hinsichtlich ihrer Anpassungsgüte zu vergleichen. Solche Maße, wie z.B. *Akaike's Information Criterion* (AIC) oder das *Bayesian Information Criterion* (BIC, vgl. Bozdogan, 1987), setzen den Wert der *Likelihood*-Funktion unter einem Modell mit der Anzahl der „investierten" Modellparameter in Form einer Nutzenfunktion in Beziehung und erlauben (im Gegensatz zu χ^2-Tests) auch den Vergleich von Modellen, die nicht in einer hierarchischen Beziehung zueinander stehen.

4 Weiterführende Literatur

Literatur, die das Gesamtgebiet der MVM abdeckt, existiert praktisch nicht. Für das Gebiet der finiten MVM können drei Monographien empfohlen werden: Everitt und Hand (1981), McLachlan und Basford (1988) sowie Titterington et al. (1985). Im letztgenannten Buch werden die statistischen Grundlagen und Probleme von MVM am ausführlichsten diskutiert. Everitt und Hand (1981) diskutieren die statistische Analyse spezieller finiter MVM für stetige (u.a. normale, exponentiale, χ^2-, F-, und Beta-Mischungen) und diskrete ZVn (binomiale und Poisson-Mischungen), die für Psychologen von besonderem Interesse sind. McLachlan und Basford (1988) schließlich beleuchten finite MVM vor allem unter clusteranalytischen Gesichtspunkten. Eine gute deutschsprachige Einführung in ausgewählte Probleme der Analyse von unterschiedlichen MVM bietet auch der Teil D des Buches von Sixtl (1993).

Eine schwierige Situation kann entstehen, wenn man keine theoretisch oder empirisch begründeten Annahmen über die lokalen Verteilungen innerhalb der Stufen der mischenden Variablen machen kann: Man behauptet lediglich die Existenz einer c-komponentigen Mischung, sagt aber nichts über die funktionale Form der lokalen Verteilungsfunktionen aus. Dieser interessante Fall, den wir im vorliegenden Beitrag ausgeklammert haben, ist dann nicht hoffnungslos, wenn man neben einer Stichprobe aus der Mischpopulation auch Stichproben aus den Subpopulationen gewinnen kann. Methodische Details und Anwendungsmöglichkeiten dieses Ansatzes werden ausgiebig von Yantis et al. (1991) diskutiert.

Fast alle Computerprogramme zu MVM verwenden eine Variante des oben erwähnten EM-Algorithmus für die Parameterschätzung. Leider gibt es dennoch kein allgemein anwendbares Programm, das auf alle MVM – oder wenigstens auf alle finiten MVM – anwendbar wäre. Man muß daher nach speziellen Programmen für die jeweilige Fragestellung suchen. Eine nicht mehr ganz aktuelle Programmübersicht findet man bei Erdfelder (1990, S. 504-506). Neuere Programme nennt und evaluiert Haughton (in press), darunter auch Programme zu finiten und stetigen Binomialmischungen (Erdfelder, 1993). Anwendungen der LCA auf kategoriale und ordinale Daten sowie Parameterschätzungen für Mischverteilungs-Rasch-Modelle können mit dem Programm WINMIRA (von Davier, 1995) durchgeführt werden.

Literaturverzeichnis

Aitkin, M., Anderson, D. & Hinde, J. (1981). Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society, A144*, 419–461.

Andersen, E. B. & Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 42*, 357–363.

Andrich, D. (1978). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology, 31*, 84–98.

Bozdogan, H. (1987). Model selection for Akaike's information criterion. *Psychometrika, 53*, 345–370.

von Davier, M. (1995). *WINMIRA: A program system for analyses with the latent class analysis and with the mixed Rasch model* (computer program). Groningen: iec PROGAMMA.

von Davier, M. & Rost, J. (1995). Polytomous mixed Rasch models. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 371–379). Berlin: Springer.

von Davier, M. & Rost, J. (in press). Self monitoring – A class variable? In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences.* Münster: Waxmann.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B39*, 1–38.

Douglas, G. A. (1978). Conditional maximum-likelihood estimation for a multiplicative binomial response model. *British Journal of Mathematical and Statistical Psychology, 31*, 73–83.

Erdfelder, E. (1988). Die Analyse latenter Moderatorvariablen. In W. Schönpflug (Hrsg.), *Bericht über den 36. Kongreß der Deutschen Gesellschaft für Psychologie in Berlin 1988*, Band 1 (S. 37–38). Göttingen: Hogrefe.

Erdfelder, E. (1990). Deterministic developmental hypotheses, probabilistic rules of manifestation, and the analysis of finite mixture distributions. In A. von Eye (Ed.), *Statistical methods in longitudinal research. Vol. II: Time series and categorical longitudinal data* (pp. 471–509). Boston: Academic Press.

Erdfelder, E. (1993). BINOMIX: A BASIC program for maximum likelihood analyses of finite and beta-binomial mixture distributions. *Behavior Research Methods, Instruments, & Computers, 25*, 416–418.

Everitt, B. S. (1986). Finite mixture distributions as models for group structure. In A. D. Lovie (Ed.), *New developments in statistics for psychology and the social sciences* (pp. 113–128). London: The British Psychological Society and Methuen.

Everitt, B. S. & Hand, D. J. (1981). *Finite mixture distributions.* London: Chapman and Hall.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests.* Bern: Huber.

Formann, A. K. (1984). *Die Latent-Class-Analyse.* Weinheim: Beltz.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Haughton, D. (in press). Packages for estimating finite mixtures: A review. *The American Statistician.*

Heyer, D. & Niederée, R. (1992). Generalizing the concept of binary choice systems induced by rankings: One way of probabilizing deterministic measurement structures. *Mathematical Social Sciences, 23*, 31–44.

Holland, P. W. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika, 55*, 5–18.

Kelderman, H. & Macready, G. B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement, 27*, 307–327.

Köller, O., Rost, J. & Köller, M. (1994). Individuelle Unterschiede beim Lösen von Raumvorstellungsaufgaben aus dem IST-70 bzw. IST-70 Untertest „Würfelaufgaben". *Zeitschrift für Psychologie, 202*, 65–85.

Langeheine, R. & van de Pol, F. (1992). Recent developments in discrete space and discrete time Markov modeling. In F. Faulbaum (Hrsg.), *SoftStat '91. Advances in statistical software* (S. 119–125). Stuttgart: Fischer.

Lazarsfeld, P. F. (1950). Logical and mathematical foundations of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, J. A. Clausen, *Measurement and Prediction* (= Studies in social psychology in World War II, Vol. IV, pp. 362–412). Princeton: Princeton University Press.

Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

McLachlan, G. J. & Basford, K. E. (1988). *Mixture models. Inference and applications to clustering.* New York: Dekker.

Mislevy, R. J. & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195–215.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London, A185*, 71–110.

Read, T. R. C. & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer.

Rost, J. (1985). A latent-class model for rating data. *Psychometrika, 50*, 37–49.

Rost, J. (1988a). *Quantitative und qualitative probabilistische Testtheorie.* Bern: Huber.

Rost, J. (1988b). Rating scale analysis with latent class models. *Psychometrika, 53*, 327–348.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271–282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology, 44*, 75–92.

Rost, J. (1996a). *Lehrbuch Testtheorie Testkonstruktion.* Bern: Huber.

Rost, J. (1996b). Logistic mixture models. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 441–455). Berlin: Springer.

Rost, J., Carstensen, C. & von Davier, M. (in press). Applying the mixed Rasch model to personality questionnaires. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences.* Münster: Waxmann.

Rost, J. & Georg, W. (1991). Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala „Jugendzentrismus". *Zentral-Archiv-Information, 28,* 52–74.

Rost, J. & Langeheine, R. (1991). Mischverteilungsmodelle: Die Methodologie der kommenden Jahre. In D. Frey (Hrsg.), *Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990,* Band 2 (S. 622–626). Göttingen: Hogrefe.

Rost, J. & Strauß, B. (1992). Review: Recent developments in psychometrics and test theory. *The German Journal of Psychology, 16,* 91–119.

Rost, J. & von Davier, M. (1993). Measuring different traits in different populations with the same items. In R. Steyer, K. F. Wender & K. F. Widaman, *Psychometric methodology. Proceedings of the 7th European Meeting of the Psychometric Society in Trier* (pp. 446–450). Stuttgart: Fischer.

Schmitt, M. (1992). Interindividuelle Konsistenzunterschiede als Herausforderung für die differentielle Psychologie. *Psychologische Rundschau, 43,* 30–45.

Sixtl, F. (1985). Notwendigkeit und Möglichkeit einer neuen Methodenlehre der Psychologie. *Zeitschrift für experimentelle und angewandte Psychologie, 32,* 320–339.

Sixtl, F. (1993). *Der Mythos des Mittelwerts. Neue Methodenlehre der Statistik.* München: Oldenbourg.

Tabor, L. E. & Kendler, T. S. (1981). Testing for developmental continuity or discontinuity: Class inclusion and reversal shifts. *Developmental Review, 1,* 330–343.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* Chichester: Wiley.

Thomas, H. (1983). Familial correlational analyses, sex-differences, and the x-linked gene hypothesis. *Psychological Bulletin, 93,* 427–440.

Thomas, H. (1985). A theory of high mathematical aptitude. *Journal of Mathematical Psychology, 29,* 231–242.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research, 5,* 329–350.

Wolfe, J. H. (1978). Comparative cluster analysis of patterns of vocational interest. *Multivariate Behavioral Research, 13,* 33–44.

Wright, B. & Masters, G. N. (1992). *Rating scale analysis: Rasch measurement.* Chicago: Mesa Press.

Yantis, S., Meyer, D. E. & Smith, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin, 110,* 350–374.