

Latente Strukturanalyse

Joachim Krauth

1. Einführung

Ausgangspunkt sei ein Fragebogen mit n binären Ja-Nein-Items, der von einer Stichprobe von Probanden beantwortet wird. Jedem Probanden entspricht dann ein Antwortmuster aus Einsen und Nullen, das den Mustern der Ja- und Nein-Antworten entspricht. Für jedes Muster wird die relative Häufigkeit bestimmt, mit der es aufgetreten ist.

Das Ziel einer **latenten Klassenanalyse** besteht darin, die Verteilung der Antwortmuster durch das Vorhandensein von zwei oder mehr **latenten** Klassen zu erklären. Man geht also davon aus, daß die Probandenstichprobe nicht homogen ist, sondern bezüglich einer möglicherweise nicht direkt meßbaren **latenten Variablen** variiert. Unterschiedliche latente Klassen entsprechen dann unterschiedlichen **Werten** der latenten Variablen.

Die latenten Klassen werden beschrieben durch eine Menge von latenten **Parametern**. Derartige Parameter geben für jedes Item und jede Klasse die Wahrscheinlichkeit für die Antwort „Ja“ an. Man macht die Voraussetzung, daß man die Wahrscheinlichkeiten für die **manifesten Variablen** kennt, d.h. die Verteilung für die verschiedenen Antwortmuster. Mit Hilfe der sogenannten **Berechnungsgleichungen** (accounting equations) bestimmt man aus den manifesten Wahrscheinlichkeiten die latenten Wahrscheinlichkeiten. Um die Gleichungen aufstellen zu können, benötigt man das sogenannte **Prinzip der lokalen Unabhängigkeit**.

Um sich die Grundidee klarzumachen, ist es gut, an ein einfaches Beispiel aus der beschreibenden Statistik anzuknüpfen, durch welches man Studienanfängern die Möglichkeit von Fehlinterpretationen von empirischen Korrelations- und Regressionskoeffizienten erläutern kann. Ausgangspunkt sei eine Zufallsstichprobe aus den Bewohnern einer Stadt. Für jeden dieser Bewohner seien die zwei Merkmale Schuhgröße und Einkommen erhoben worden, also Merk-

male, zwischen denen absolut kein Kausalzusammenhang unterstellt werden kann. Tatsächlich ergibt sich jedoch eine positive Korrelation der beiden Merkmale. Die Erklärung wird sofort einsichtig, wenn man die Meßwertpaare in eine Punktwolke (Scatter-Diagramm) einträgt und die Datenpunkte für Männer und Frauen unterschiedlich kennzeichnet. Man sieht dann einerseits, daß innerhalb der Gruppe der Männer kein Zusammenhang zwischen Schuhgröße und Einkommen besteht. Das gleiche gilt für die Gruppe der Frauen. In der Graphik ergibt sich jeweils eine ungefähr kreisförmige Punktwolke für jedes Geschlecht. Andererseits haben in der Regel Männer ein höheres Einkommen und eine höhere Schuhgröße als Frauen. Dieses führt dazu, daß die kreisförmigen Punktwolken für die beiden Geschlechter bei Fallenlassen der Unterscheidung der Geschlechter zu einer ovalen Punktwolke verschmelzen und eine positive Korrelation zwischen Schuhgröße und Einkommen vortäuschen. Falls man also eine Korrelation vorliegen hat, die man auf das Wirken einer latenten Variablen zurückführen will, wie in diesem Fall auf das Geschlecht, so muß man die vorliegende Punktwolke in kreisförmige Teilbereiche zerlegen, d.h. Klassen suchen, für die die betrachteten Merkmale lokal unabhängig sind.

In der Praxis hat man dabei natürlich größere Schwierigkeiten als in unserem Beispiel, da man den Einfluß von Geschlecht, Alter, Intelligenz usw., d.h. der offensichtlichen latenten Variablen, meist leicht berücksichtigen kann. Die latente Strukturanalyse stellt sich die Aufgabe, gerade solche nicht offensichtlichen latenten Variablen aufzuspüren.

Wie Madansky (1968) ausführt, ist die Grundidee der latenten Strukturanalyse schon sehr alt. Schon bei Cournot (1838), Weinberg (1902), Benini (1928) und de Meo (1934) werden Modelle diskutiert, mit denen man auf nicht beobachtbare (latente) Variablen mit Hilfe von polytomen beobachtbaren (manifesten) Variablen rückschließen will. Bei diesen Modellen wird die multivariate Verteilung der manifesten Variablen dargestellt als eine Mischung von multivariaten Verteilungen. Dabei ist die zu bestimmende Mischungsverteilung gerade die Verteilung der latenten Variablen.

Eine Teilklasse dieser Modelle sind diejenigen, bei denen man annimmt, daß die Variablen innerhalb jeder der Komponenten der Mischverteilung als unabhängig anzusehen sind. In Lazarsfeld (1950a) wird für solche Modelle der Begriff **latente Strukturmodelle** eingeführt. Anschließend an die beiden grundlegenden Arbeiten von Lazarsfeld (1950a, 1950b) wurden von dem gleichen Autor später noch andere Einführungs- und Übersichtsartikel verfaßt (Lazarsfeld, 1954, 1955, 1959). Einen systematischen Überblick mit einer Einbettung in die Skalierungstheorie findet man bei Torgerson (1958, S. 360-395). Einen Kurzüberblick gibt Madansky (1968).

Einführungen in deutscher Sprache geben von der Lippe (1973), Fischer (1974, S. 160-179) und als Kurzüberblick Wottawa (1979, S. 51-58).

Eine umfassende Darstellung der Theorie mit einer ausführlichen Biographie bis 1967 enthält die Monographie von Lazarsfeld & Henry (1968). Der Übersichtsartikel von Fielding (1977) berücksichtigt Literatur bis 1973.

2. Grundbegriffe der latenten Strukturanalyse

Die hier vorkommenden Probleme treten immer dann auf, wenn die Variablen, die einen interessieren, nicht direkt meßbar sind und man andere Variablen messen muß, um aus den Ergebnissen Rückschlüsse auf die interessierenden Variablen ziehen zu können. Beispielsweise mag ein Psychiater bei einem Klienten das Vorhandensein oder Nichtvorhandensein gewisser Symptome beobachten und daraus Rückschlüsse darauf ziehen wollen, ob der Klient schizophran ist oder nicht. Schizophrene Klienten zeigen dabei nicht unbedingt immer alle Symptome, während sich manche Symptome auch bei nicht-schizophrenen Klienten zeigen können. Mit Hilfe der latenten Strukturanalyse versucht man nun die latenten Klassen der Schizophrenen und Nichtschizophrenen zu identifizieren.

Die einfachste Form der Analyse geht von einer Menge von dichotomen Ja-Nein-Items aus, wobei man diese Voraussetzung auch zu polytomen Items abschwächen kann. Man postuliert die Existenz eines **latenten Raumes**, in dem sich die Mitglieder einer Population befinden, wobei die Position eines Mitgliedes völlig die Wahrscheinlichkeit für eine Ja-Antwort für irgendein Item bestimmt. Der latente Raum ist somit gerade der Raum, auf dem die uns interessierende latente Variable definiert ist. Je nach der spezifischen Struktur des latenten Raumes und der Variabilität der **Itemwahrscheinlichkeiten** innerhalb dieses Raumes sprechen wir von einem speziellen **latenten Strukturmodell**. Der latente Raum kann definiert werden als diejenige Klassifikation, die für die statistischen Interaktionen zwischen den manifesten Variablen verantwortlich ist. Durch diese Klassifikation wird eine gegebene Population in homogene Teilpopulationen ‚entmischt‘ (Lazarsfeld, 1959).

Falls der latente Raum nur aus einer endlichen Anzahl von Punkten besteht, so spricht man von **diskreten Klassenmodellen**. Im obigen Beispiel würde bei n Symptomen, die jeweils nur vorhanden oder nicht vorhanden sein können, der latente Raum aus 2^n möglichen Punkten bestehen. Falls wir in dem Anfangsbeispiel nur zwischen hohem (1) und niedrigem (0) Einkommen bzw. Schuhgrößen unterscheiden würden, so würde der latente Raum aus den $2^2 = 4$ Punkten (0,0), (0,1), (1,0) und (1,1) bestehen. Ein mögliches Modell würde dann gegeben durch

$$p_{00} = p_{11} = 0.5, p_{01} = p_{10} = 0.$$

In diesem Modell würde demnach die Kombination hoher Schuhgröße mit hohem Einkommen und niedriger Schuhgröße mit niedrigem Einkommen mit

Wahrscheinlichkeit 0.5 auftreten, während die beiden Kombinationen, bei denen niedrige Werte der einen Variablen mit hohen Werten der anderen Variablen zusammentreffen, nur die Wahrscheinlichkeit 0 haben.

Falls man einen Test aus n dichotomen Items verwendet, um eine Fähigkeit zu messen, so kann man davon ausgehen, daß die latente Fähigkeit stetig verteilt ist, so daß der latente Raum z.B. durch die reelle Achse beschrieben werden kann. Die wahren Fähigkeitswerte der Probanden können dann z.B. als normalverteilt angesehen werden. Für jedes Item und jeden Punkt der Geraden, d.h. für jeden Fähigkeitswert, gibt es dann eine Wahrscheinlichkeit für eine Ja-Antwort. Diese Wahrscheinlichkeiten für jeden Punkt der Geraden kann man durch eine Funktion, die **Itemcharakteristikkfunktion** oder **Spurfunktion** (trace-line) beschreiben. In diesem Fall spricht man von einem **stetigen Modell**. Die Unterscheidung zwischen stetigen und diskreten Modellen bezieht sich also auf die Struktur des latenten Raumes und nicht auf die Struktur der Items, die in beiden Fällen diskrete Ausprägungen haben.

Um weitere Grundbegriffe einführen zu können, nehmen wir an, daß wir in unserem Schizophreniebeispiel insgesamt $n = 10$ Symptome beobachten, jeweils mit den Ausprägungen vorhanden (1) und nichtvorhanden (0). Weiterhin nehmen wir an, daß es zwei Klassen von Probanden gibt, nämlich Schizophrenie (1) und Nichtschizophrene (2). Das **Axiom der lokalen Unabhängigkeit** besagt:

Innerhalb jeder der beiden latenten Klassen (indiziert durch 1 und 2) sind die Reaktionen auf verschiedene Items unabhängig. Die klasseninterne Wahrscheinlichkeit jedes Reaktionsmusters für jede Teilmenge von Items ist das Produkt der entsprechenden eindimensionalen Randwahrscheinlichkeiten.

Dieses führt zu einer gewissen Anzahl von Bedingungsgleichungen, z.B.

$$p_{13}^{(1)} = p_{1\cdot}^{(1)} \cdot p_{\cdot 3}^{(1)}, p_{\bar{1}\bar{3}}^{(1)} = p_{\bar{1}\cdot}^{(1)} \cdot p_{\cdot \bar{3}}^{(1)} = (1 - p_{1\cdot}^{(1)}) p_{\cdot 3}^{(1)}, p_{125}^{(2)} = (1 - p_{1\cdot}^{(2)}) p_{\cdot 2}^{(2)} p_{\cdot 5}^{(2)} \text{ usw.}$$

Dabei bedeutet die erste Gleichung, daß sich innerhalb der Klasse 1 der Schizophrenen aufgrund der Unabhängigkeit des Auftretens von Symptom 1 und Symptom 3 die gemeinsame Wahrscheinlichkeit als Produkt der Randwahrscheinlichkeiten ergibt. Die zweite Gleichung bedeutet, daß die Wahrscheinlichkeit dafür, daß Symptom 1 nicht auftritt, während gleichzeitig Symptom 3 auftritt, sich innerhalb der Klasse 1 der Schizophrenen als Produkt der entsprechenden Randwahrscheinlichkeiten ergibt. Schließlich wird in der dritten Gleichung die Wahrscheinlichkeit für das gleichzeitige Nichtauftreten von Symptom 1 und das Auftreten von Symptom 2 und Symptom 5 innerhalb der Klasse 2 der Nichtschizophrenen als Produkt der entsprechenden Randwahrscheinlichkeiten dargestellt.

Über das Axiom der bedingten lokalen Unabhängigkeit und seine Bedeutung auch für andere Skalierungsansätze informieren Lord & Novick (1968, S. 538-540). Es zeigt sich, daß es schwer ist, auf diese Annahme zu verzichten. In der Faktoranalyse kann man dieses Postulat zwar auf die lineare lokale Unabhängigkeit abschwächen, darf dafür aber auch nur Interaktionen 1. Ordnung zulassen. Harper (1972) diskutiert, wie man die lokale Unabhängigkeit wenigstens soweit abschwächen kann, daß eine paarweise Abhängigkeit zwischen den Items zugelassen ist.

Bei n dichotomen Items wird ein Zweiklassenmodell durch $2n + 2$ Parameter charakterisiert. Für die erste Klasse sind dies die n latenten Wahrscheinlichkeiten bzw. latenten Randwahrscheinlichkeiten $p_1^{(1)}, \dots, p_n^{(1)}$, für die zweite Klasse die n latenten Wahrscheinlichkeiten $p_1^{(2)}, \dots, p_n^{(2)}$. Zusätzlich benötigt man noch die Wahrscheinlichkeit $v^{(1)}$ bzw. $v^{(2)}$, mit der ein Proband zu der Klasse 1 bzw. der Klasse 2 gehört. Die Summe der Klassenwahrscheinlichkeiten ergibt immer 1:

$$1 = v^{(1)} + v^{(2)}.$$

Dieses ist die erste Modellgleichung. Die weiteren Berechnungsgleichungen ergeben sich mit Hilfe der Formel von der totalen Wahrscheinlichkeit, denn es gilt z.B. für die Wahrscheinlichkeit p_1 für das Auftreten des Symptoms 1 in der Population

$$p_1 = p_1^{(1)}v^{(1)} + p_1^{(2)}v^{(2)}.$$

Dieses liegt daran, daß z.B. der unbekannte latente Parameter $p_2^{(1)}$ die bedingte Wahrscheinlichkeit dafür ist, daß ein Proband, der zu Klasse 1 gehört, das Symptom 2 aufweist. Entsprechend ergibt sich z.B. die Wahrscheinlichkeit dafür, daß ein Proband in der Population das Merkmal 1 aufweist und das Merkmal 3 nicht aufweist, zu

$$p_{1\bar{3}} = p_{1\bar{3}}^{(1)}v^{(1)} + p_{1\bar{3}}^{(2)}v^{(2)} = p_1^{(1)}(1 - p_3^{(1)})v^{(1)} + p_1^{(2)}(1 - p_3^{(2)})v^{(2)}.$$

Beobachtbar sind nur die relativen Häufigkeiten der verschiedenen Auftretenskombinationen der Symptome. Diese verwendet man als Schätzungen für die entsprechenden Wahrscheinlichkeiten. So ist die relative Häufigkeit der Personen, bei denen Symptom 1 und Symptom 5 nicht auftreten, jedoch das Symptom 3, eine Schätzung für die entsprechende nichtbedingte Wahrscheinlichkeit $p_{1\bar{3}\bar{5}}$. Indem man diese Schätzungen anstelle der unbekannteren Wahrscheinlichkeiten in die Berechnungsgleichungen einsetzt, bekommt man ein Gleichungssystem für die unbekannteren latenten Parameter $v^{(1)}, v^{(2)}, p_1^{(1)}, \dots, p_n^{(1)}, p_1^{(2)}, \dots, p_n^{(2)}$. Durch Auflösen dieses Gleichungssystems erhält man Schätzungen für diese latenten Parameter.

Wie man leicht-einsieht, ergeben sich bei n dichotomen Items und 2 latenten Klassen insgesamt 2^n Berechnungsgleichungen (Lazarsfeld & Henry, 1968, S.

27). Für $n=2$ Items gibt es $2^2=4$ Gleichungen für $2n+2=2\cdot 2+2=6$ Parameter. In diesem Fall kann man die Parameter und damit auch das Modell nicht **identifizieren**. Da in den Gleichungen die gesamte Information über das Modell enthalten ist, besteht keine Möglichkeit, eine eindeutige Lösung zu finden, es sei denn, wir erlegen den latenten Parametern noch weitere Restriktionen auf.

Für $n=3$ Items gibt es $2^3=8$ Gleichungen für $2\cdot 3+2=8$ Parameter. Hier besteht zumindest die Möglichkeit, eine eindeutige Lösung zu erhalten. Bei der Lösung dieses nichtlinearen Gleichungssystems geht man zunächst von zwei Annahmen aus. Man nimmt an, daß das spezielle latente Strukturmodell mit zwei Klassen tatsächlich vorliegt und daß man die **manifesten** nichtbedingten Wahrscheinlichkeiten auf den linken Seiten der Gleichungen tatsächlich kennt. Dann reduziert sich das Problem auf eine rein algebraische Fragestellung. Die Schätzung der latenten Parameter, falls man nur Schätzungen der manifesten Wahrscheinlichkeiten hat, und die Anpassungsgüte des latenten Strukturmodells an die beobachteten Daten sind eine danach zu behandelnde Frage.

Die algebraische Frage der Lösung der Gleichungssysteme findet man behandelt in Lazarsfeld (1961) und in Lazarsfeld & Henry (1968). Für die manifesten Wahrscheinlichkeiten werden dazu sogenannte **Kreuzprodukte** der Items i und j

$$[ij] = p_{ij} - p_i p_j$$

und sogenannte **geschichtete Kreuzprodukte**

$$[ij; k] = p_{ijk} p_k - p_{ik} p_{jk}$$

eingeführt.

Mit Hilfe dieser Ausdrücke werden die Berechnungsgleichungen umgeschrieben und dadurch formal leichter lösbar. Die Lösungen brauchen jedoch keineswegs Werte zwischen 0 und 1 anzunehmen. Mit Hilfe der Kreuzprodukte lassen sich eine Reihe von notwendigen Bedingungen formulieren, die für ein vorgegebenes latentes Strukturmodell erfüllt sein müssen. Ein weiterer Schritt ist die Einführung sogenannter **symmetrischer Parameter**. Ein symmetrischer Parameter 3. Ordnung ist definiert durch

$$[ijk] = p_{ijk} - p_i[jk] - p_j[ik] - p_k[ij] - p_i p_j p_k.$$

Falls man mit X_i die Zufallsvariable bezeichnet, die gleich 1 ist, falls eine Ja-Antwort bei Item i vorliegt und gleich 0 sonst, so gilt für die Kovarianz

$$[ij] = E[(X_i - p_i)(X_j - p_j)] = \text{Cov}[X_i, X_j]$$

und für das zentrale Moment 3. Ordnung

$$[ijk] = E[(X_i - p_i)(X_j - p_j)(X_k - p_k)].$$

Entsprechendes gilt für symmetrische Parameter höherer Ordnung. Derartige Beziehungen werden sowohl bei Strukturuntersuchungen als auch bei der Herleitung von Schätzverfahren verwendet.

Mit Hilfe der latenten Parameter kann man für jedes mögliche Reaktionsmuster die Wahrscheinlichkeit dafür berechnen, daß ein Proband, der dieses Muster gezeigt hat, zu einer bestimmten Klasse gehört. Mit Hilfe dieser **Rekrutierungswahrscheinlichkeiten** (recruitment probabilities) kann man Probanden in der Weise klassifizieren, daß man sie der für sie wahrscheinlichsten Klasse zuordnet. Für $n=3$ ergäbe sich z.B. für das Muster (Ja, Nein, Ja) die Wahrscheinlichkeit

$$p_{1\bar{2}3} = p_{1\bar{2}3}^{(1)} v^{(1)} + p_{1\bar{2}3}^{(2)} v^{(2)} = p_1^{(1)} (1-p_2^{(1)}) p_3^{(1)} v^{(1)} + p_1^{(2)} (1-p_2^{(2)}) p_3^{(2)} v^{(2)}.$$

Der erste Summand ist die Wahrscheinlichkeit dafür, daß ein Proband aus der Klasse 1 das Muster aufweist, der zweite Summand die entsprechende Wahrscheinlichkeit für einen Probanden der Klasse 2. Man ordnet den Probanden der Klasse mit der größeren Wahrscheinlichkeit zu. Addiert man diese größeren Wahrscheinlichkeiten über alle Muster, so erhält man die totale Wahrscheinlichkeit dafür, einen Probanden richtig zu klassifizieren. Zieht man diesen Wert von 1 ab, so ergibt sich die Wahrscheinlichkeit einer Fehlklassifikation.

Weist ein Proband das obige Muster auf, so ist die Wahrscheinlichkeit dafür, daß er dann zu der Klasse 1 gehört, gegeben durch

$$P(\text{Klasse 1} \mid \text{Muster (Ja, Nein, Ja)}) = v^{(1)} p_{1\bar{2}3}^{(1)} / p_{1\bar{2}3}.$$

Für mehr als $n=3$ Items erhalten wir mehr Berechnungsgleichungen als unbekannte latente Parameter. Z.B. ergeben sich für $n=4$ Items $2^4=16$ Gleichungen für $2 \cdot 4 + 2 = 10$ Parameter. Damit handelt es sich nicht mehr um ein **Interpolationsproblem**, bei dem ein Gleichungssystem aufzulösen ist, sondern um ein **Approximationsproblem**, bei dem $2n+2$ Parameterschätzwerte gesucht sind, die die 2^n Gleichungen mit den $2^n - 1$ beobachteten relativen Häufigkeiten möglichst gut erfüllen. Jedoch wurde in der ersten Zeit bei der Schätzung der latenten Parameter immer versucht, auch bei mehr als $n=3$ Items die Interpolationsmethode zu verwenden, indem man aus der Menge der Gleichungen immer nur so viel Gleichungen auswählte, wie unbekannte latente Parameter zu schätzen waren (algebraische Schätzverfahren).

3. Allgemeines Vorgehen bei der latenten Strukturanalyse

In Lazarsfeld (1959) wird ein Schema aus 9 Schritten diskutiert, die bei einer latenten Strukturanalyse durchzuführen sind.

Der erste Schritt betrifft die **Auswahl** und **Spezifizierung** des Modells. Dieses betrifft Annahmen über die Spurfunktionen und über die Verteilung der Population über dem latenten Raum.

Als zweiter Schritt werden die **Berechnungsgleichungen** für das spezielle Modell aufgestellt. Diese verknüpfen die Verteilung der manifesten Variablen mit der der latenten Variablen.

Der dritte Schritt betrifft die sogenannten **Reduzierbarkeitsbedingungen**. Für mehr als $n=3$ Items erhält man mehr Gleichungen als unbekannte latente Parameter. Dieses führt dazu, daß die manifesten Wahrscheinlichkeiten gewissen zusätzlichen Restriktionen unterliegen, wenn ein bestimmtes Strukturmodell vorliegt. Diese Restriktionen haben die Form von Gleichungen, die besonders einfach mit Hilfe der Kreuzprodukte und geschichteten Kreuzprodukte formuliert werden können. Man spricht von Reduzierbarkeitsbedingungen, weil das System der Berechnungsgleichungen nur bei Erfülltsein dieser Bedingungen gelöst werden kann. Sind die Bedingungen nicht erfüllt, so ist das Modell falsch.

Im vierten Schritt muß die **Identifizierbarkeit** überprüft werden, d.h. es muß geklärt werden, ob die vorliegenden manifesten Parameter ausreichen, um die Werte der latenten Parameter festzulegen. Wie sich im dritten Schritt gezeigt hat, reicht es dazu nicht aus, zu überprüfen, ob mehr Gleichungen als Unbekannte vorliegen. Aus den Reduzierbarkeitsbedingungen geht hervor, daß viele der manifesten Parameter aus anderen manifesten Parametern berechnet werden können. Die Berechnungsgleichungen sind also nur zum Teil unabhängig.

Im fünften Schritt geht es um die **Identifikation** des Modells. Hierbei ist nach der tatsächlichen Auflösung der Berechnungsgleichungen nach den latenten Parametern gefragt. Man stellt bei der Identifikation die latenten Parameter dar durch Funktionen der manifesten Parameter. Die Darstellung wird wieder vereinfacht durch die Verwendung der Kreuzprodukte.

Der sechste Schritt betrifft die Anpassungsprozedur oder das **Schätzverfahren**. Die bisherigen Schritte bezogen sich alle auf algebraische Fragestellungen innerhalb eines bestimmten Modells. Bei der Anpassung konkreter Daten hat man es zusätzlich mit Stichprobenschwankungen zu tun. Während im Modell ein bestimmtes Kreuzprodukt immer dasselbe ist, in welchen Berechnungsgleichungen es auch auftreten mag, können sich empirische Kreuzprodukte, in die anstelle der Randwahrscheinlichkeiten die entsprechenden relativen Häufigkeiten eingesetzt werden, durchaus voneinander unterscheiden. Dieses führte zu dem Vorschlag, gewisse Mittelwertbildungen über die empirischen Kreuzprodukte vorzunehmen. Andere Schätzverfahren werden noch besprochen.

Im siebten Schritt ist die **Anpassungsgüte** zu überprüfen, d.h. ein Maß dafür anzugeben, wie gut die Daten zu dem angepaßten Modell passen. Da die Daten kaum jeweils die Reduzierbarkeitsbedingungen exakt erfüllen, werden die aufgrund des Modells vorhergesagten manifesten Wahrscheinlichkeiten auch nicht mit den beobachteten relativen Häufigkeiten exakt übereinstimmen. Für ein dem Problem angemessenes Modell sollten aber die Abweichungen klein und zufällig verteilt sein.

Im achten Schritt wird das **Rekrutierungsmuster** bestimmt. Damit soll die Frage angesprochen werden, wo sich ein Individuum mit einem bestimmten Antwortmuster im latenten Raum befindet. Prinzipiell kann sich die Person in jedem Punkt des latenten Raumes befinden. Im allgemeinen wird sie aber mit großer Wahrscheinlichkeit aus einem bestimmten Bereich des latenten Raumes stammen und mit geringer Wahrscheinlichkeit aus anderen Bereichen. In diesem Sinne entspricht jedem Antwortmuster ein Vektor von Aufenthaltswahrscheinlichkeiten, das Rekrutierungsmuster.

Im neunten und letzten Schritt geht es um die Frage der **Klassifikation** der Probanden bzw. um die Zuordnung von **Skalenwerten**. Man sucht dann nach einem Kriterium, um einen Probanden aufgrund seines Antwortmusters einem bestimmten Punkt des latenten Raumes, der besonders typisch für ihn ist, zuzuordnen. Gelegentlich begnügt man sich auch damit, die Antwortmuster gemäß einem solchen Kriterium anzuordnen. Zur Schätzung solcher Skalenwerte reicht die Kenntnis der ersten Momente, z.B. des Erwartungswertes, der Rekrutierungsverteilung schon aus. Falls man einer Person den für sie wahrscheinlichsten Punkt im latenten Raum als Skalenwert zuordnen will, so benötigt man die Kenntnis der ganzen Rekrutierungsverteilung.

In diesem Zusammenhang interessiert man sich oft noch für ein anderes duales Problem, die **Skalierung der Items**. Man möchte den Items Skalenwerte zuordnen, die angeben, wie weit ein einzelnes Item zu den Skalenwerten der Individuen beiträgt bzw. wie weit jedes Item zwischen Individuen diskriminiert, die an verschiedenen Punkten des latenten Raumes lokalisiert sind. Diese Fragestellung wird in Lazarsfeld (1954) diskutiert.

4. Modelle der latenten Strukturanalyse

4.1 Allgemeines Modell

In Anlehnung an Anderson (1959) beschreibt Fielding (1977) das allgemeine latente Strukturmodell (siehe auch McDonald, 1962). Es seien X_1, \dots, X_n manifeste Variablen und entsprechend Y_1, \dots, Y_m latente Variablen. Die Variablen können sich prinzipiell auf beliebigem Skalenniveau befinden. Dabei ist ge-

wöhnlich m als viel kleiner als n angenommen, meist sogar $m=1$. Man nimmt eine bedingte Wahrscheinlichkeitsverteilung $f(x_1, \dots, x_n \mid y_1, \dots, y_m)$ der X 's bei gegebenen Y 's an, wobei f eine Dichte ist für stetige X 's und eine Menge von Wahrscheinlichkeiten für diskrete X 's. Entsprechend sei eine Randverteilung der Y 's gegeben durch $g(y_1, \dots, y_m)$. Damit folgt für die nichtbedingte Randverteilung von X_1, \dots, X_n

$$h(x_1, \dots, x_n) = \int f(x_1, \dots, x_n \mid y_1, \dots, y_m) g(y_1, \dots, y_m) dy_1 \dots dy_m.$$

Nach der Formel von Bayes folgt für ein Individuum mit den Werten x_{o1}, \dots, x_{on} der manifesten Variablen für die bedingte Dichte bzw. bedingten Wahrscheinlichkeiten bezüglich der Werte y_1, \dots, y_m des Individuums im latenten Raum

$$k(y_1, \dots, y_m \mid x_{o1}, \dots, x_{on}) = \frac{f(x_{o1}, \dots, x_{on} \mid y_1, \dots, y_m) g(y_1, \dots, y_m)}{h(x_{o1}, \dots, x_{on})}.$$

Aufgrund von x_{o1}, \dots, x_{on} kann man mit Hilfe dieser Gleichung ein Individuum einem Punkt im latenten Raum zuweisen. Naheliegend ist die Maximum-Likelihood-Methode, nach der ein Individuum mit (x_{o1}, \dots, x_{on}) den Punkt (y_{o1}, \dots, y_{om}) zugewiesen bekommt, für den die Funktion k maximal wird. Falls man Fehlklassifikationen für verschiedene Punkte für unterschiedlich schwerwiegend hält, so wird man die Punkte unterschiedlich gewichten, d.h. Entscheidungsregeln mit geeigneten Verlustfunktionen verwenden.

Falls man die obige Gleichung verwenden will, muß man davon ausgehen, daß man allenfalls die Funktion h für die manifesten Variablen kennt, während f und g unbekannt sind und auch nicht eindeutig aus h gefolgert werden können. Deshalb verwendet man das **Axiom der lokalen Unabhängigkeit**, welches besagt, daß für einen festen Punkt im latenten Raum die manifesten Variablen unabhängig sind. Für $m=1$ latente Variable bedeutet dieses

$$f(x_1, \dots, x_n \mid y) = f_1(x_1 \mid y) \dots f_n(x_n \mid y)$$

bzw.

$$h(x_1, \dots, x_n) = \int f_1(x_1 \mid y) \dots f_n(x_n \mid y) g(y) dy.$$

Inhaltlich bedeutet das Axiom, daß Individuen mit gleichen latenten Ausprägungen einander ähnlich sind und daß die gesamte relevante Information in den latenten Variablen liegt, so daß nach deren Festlegung jedes Verhalten nur noch zufällig ist. Man geht also davon aus, daß die sich zwischen den manifesten Variablen zeigenden Abhängigkeiten allein auf Abhängigkeiten in den latenten Variablen zurückzuführen sind. Das Axiom wird genauer diskutiert bei Lord & Novick (1968, S. 538-540). Eine Abschwächung des Axioms zumindest für das latente Klassenmodell in dem Sinne, daß die Items paarweise abhängig sein dürfen, sieht Harper (1972) vor.

Die unter den Werten der manifesten Variablen bedingten Dichten bzw. Wahrscheinlichkeiten $f_i(x_i | y)$ heißen **Spurfunktionen**. Diese versucht man zusammen mit der latenten Verteilung g zu ermitteln, um die latente Struktur zu erhalten. Dann weiß man, wie die manifesten Variablen von den latenten Variablen abhängen und kann Individuen skalieren und die Parameter der Funktionen schätzen.

Darstellungen des allgemeinen latenten Strukturmodells findet man außer in den oben erwähnten Arbeiten auch in Torgerson (1958, S. 361-367) und Green (1952).

4.2 Existenzproblem

Bei der Konstruktion eines Modells geht man von einer manifesten Wahrscheinlichkeitsfunktion $h(x_1, \dots, x_n)$ aus, die noch von gewissen unbekanntem aber schätzbaren Parametern abhängt. Beispielsweise kann h die Dichte einer n -dimensionalen Normalverteilung sein. Ebenso wählt man die Spurfunktionen f_i und die latente Verteilung g als Wahrscheinlichkeitsfunktionen vorgegebener Form mit noch freien Parametern. Auch die Anzahl m der latenten Variablen muß spezifiziert sein. Es ist dann durchaus nicht gesagt, daß es Parameter gibt, so daß die obige Beziehung zwischen h , f_1, \dots, f_n und g erfüllbar ist. z.B. existiert kein latentes Strukturmodell, falls h einer Normalverteilung und die f_i Cauchyverteilungen entsprechen (Anderson, 1959). Unter Umständen kann die Frage der Existenz eines latenten Strukturmodells durch die Frage der Gültigkeit gewisser Reduzierbarkeitsbedingungen beantwortet werden. Man vergleiche dazu Lazarsfeld (1959, S. 512-515).

4.3 Identifikationsproblem

Falls man h kennt, so reicht die durch das Axiom der lokalen Unabhängigkeit gegebene Beziehung zwischen h , f_1, \dots, f_n und g nicht aus, um f_1, \dots, f_n und g eindeutig zu identifizieren. Man benötigt Zusatzannahmen über die Form der Spurfunktionen bzw. der latenten Verteilung, um die Struktur zu identifizieren. Je nach den speziellen Annahmen ergeben sich dann die einzelnen Modelle.

Selbst innerhalb dieser Untermodelle ist es möglich, daß verschiedene Strukturen, die verschiedenen Parameterkombinationen entsprechen, mit einer Verteilung f verträglich sind. Viele Arbeiten haben sich damit beschäftigt, Bedingungen herauszufinden, unter denen sich bei einem speziellen Modell eine Struktur, gegeben durch f_1, \dots, f_n und g , eindeutig identifizieren läßt. Dazu gehören z.B. McHugh (1956, 1958), Madansky (1960) und Goodman (1974a).

Insbesondere das allgemeine Identifikationsproblem wurde untersucht von Koopmans & Reiersol (1950) und Koopmans (1949, 1951).

Häufig ist es nicht möglich, für ein gegebenes Untermodell die Struktur zu identifizieren. Für praktische Zwecke ausreichend ist dann oft die eindeutige Identifikation gewisser Strukturaspekte, z.B. allein die Identifikation gewisser Erwartungswerte anstelle einer ganzen Verteilung (Fielding, 1977). Auch können nichtidentifizierbare Modelle durch Einschränkungen der möglichen Parameter oft identifizierbar gemacht werden (Goodman, 1974a).

4.4 Strukturproblem

Falls man unter Kenntnis von h das Existenzproblem und das Identifikationsproblem gelöst hat, so sind als nächstes die zu f_1, \dots, f_n und g gehörigen Parameter zu bestimmen (Anderson, 1959). Dieses Problem ist für jedes Untermodell getrennt zu untersuchen. In der Sprechweise von Lazarsfeld (1959) ist dieses das Identifikationsproblem.

4.5 Latentes Klassenmodell

Eines der einfachsten Modelle ist das latente Klassenmodell von Lazarsfeld (1950a). Hier nimmt Y eine endliche Menge von Werten, sogenannte *latente Klassen*, an. Die latente Verteilung g ist eine Multinomialverteilung, während die Spurfunktion $f_i(x_i | y)$ die Wahrscheinlichkeit für ein Individuum aus der Klasse y ist, für das Item i die Reaktion x_i zu zeigen. Gesucht ist in diesem Falle also ein Modell mit q latenten Klassen, den Werten der latenten Variablen Y , die so geartet sind, daß innerhalb der Klassen die n manifesten Variablen (Items) unabhängig sind. Wenn solche Klassen konstruiert sind, z.B. die Klassen der Schizophrenen und Nichtschizophrenen, so sind alle Interaktionen zwischen den manifesten Variablen aufgeklärt. Auf dieses Modell (latent class model) war im Anfang schon eingegangen worden, und es ist auch in der Literatur am häufigsten betrachtet worden.

Besonderes Interesse hat das Identifikationsproblem, d.h. die Frage der eindeutigen Lösbarkeit der Berechnungsgleichungen gefunden. Bei n manifesten Variablen und q latenten Klassen ergeben sich 2^n Gleichungen für $q(n+1)$ latente Parameter. Für $2^n > q(n+1)$ sind verschiedene Lösungsvorschläge gemacht worden. Dabei versuchte man meist durch geeignete Auswahl der Gleichungen das Ziel einer eindeutigen Parameterbestimmung zu erreichen.

Ein erster Lösungsansatz für das Strukturproblem scheint der von Green (1951) zu sein. Nach Umschreibung der Berechnungsgleichungen in Matrizenform wird analog zur Faktoranalyse eine Lösung durch Faktorisierung von zwei Matrizen angestrebt.

Im einzelnen betrachtet Green (1951) die $(n+1) \times (n+1)$ -Matrix P_0 , deren erstes Element eine 1 ist und in deren 1. Zeile und 1. Spalte ansonsten die eindimensionalen Randwahrscheinlichkeiten stehen. Hierbei ist n die Anzahl der Items. In den übrigen Zeilen bzw. Spalten befinden sich die zweidimensionalen Randwahrscheinlichkeiten. Weiterhin wird die (qxq) -Diagonalmatrix V definiert, die in der Diagonale die Wahrscheinlichkeiten $v^{(1)}, \dots, v^{(q)}$ enthält, die angeben, mit welcher Wahrscheinlichkeit ein zufällig ausgewählter Proband zu einer der q Klassen gehört. Schließlich sei L eine $(n+1) \times q$ -Matrix mit den Elementen $p_i^{(j)}$ die die bedingten Wahrscheinlichkeiten dafür bezeichnen, daß eine Person aus der latenten Klasse j das Item i positiv beantwortet.

Die Matrix A , definiert durch

$$A = LV^{1/2},$$

enthält dann die unbekanntenen latenten Parameter bzw. man kann diese bei Kenntnis von A leicht bestimmen. Aufgrund der Berechnungsgleichungen gilt

$$P_0 = AA',$$

wobei A' die Transponierte von A ist. Mit Hilfe eines faktoranalytischen Ansatzes läßt sich P_0 darstellen als

$$P_0 = BB',$$

wobei

$$BE = A$$

gilt mit einer orthogonalen Matrix E . Um E zu bestimmen, faktorisiert man in ähnlicher Weise die Matrix P_1 , die die dreidimensionalen Randwahrscheinlichkeiten in symmetrischer Form enthält:

$$P_1 = CC'.$$

Aus der Matrix

$$TT' = (B'B)^{-1}B'C$$

ergibt sich dann bis auf die Anordnung der Items eindeutig die gesuchte Matrix E mit Hilfe einer Hauptkomponentenanalyse. Eine auf T. W. Anderson zurückgehende Vereinfachung (vgl. Green, 1951) bestimmt TT' aus

$$TT' = (B'B)^{-1}B'P_1B(B'B)^{-1}.$$

Damit erspart man sich die Faktorisierung von P_1 .

Eine Schwierigkeit bereiten die Randwahrscheinlichkeiten mit wiederholt auftretenden gleichen Indizes in den Matrizen P_0 und P_1 . Dieses sind artifizielle

Größen ohne realen Bezug zu den manifesten Wahrscheinlichkeiten, was zu gewissen Schätzproblemen führt.

Eine andere Matrixmethode für dieses Problem wurde zuerst von Lazarsfeld & Dudman (1951) sowie unabhängig davon durch Koopmans (1951) vorgeschlagen. Diese Methode, die auf die Bestimmung der Nullstellen einer Determinantengleichung hinausläuft, wurde von Anderson (1954) weiterentwickelt. Durch Gibson (1955) wurde das Verfahren so modifiziert, daß alle Items simultan berücksichtigt wurden. Darauf aufbauend konnte Madansky (1960) das Verfahren verallgemeinern, insbesondere in Hinblick auf notwendige und hinreichende Bedingungen für die Identifizierbarkeit. Gibson (1962a) diskutierte die Möglichkeit, zusätzliche Items nachträglich in eine latente Klassenanalyse mit einzubeziehen. Die hier beschriebene Matrixmethode wird auch als **Basismethode** bezeichnet. Außer in den angegebenen Originalarbeiten wird sie in Lazarsfeld & Henry (1968, S. 46-48) und in Fielding (1977) dargestellt.

Die wesentliche Bedingung für die Identifizierbarkeit des latenten Klassenmodells besteht darin, daß die Anzahl der Gleichungen mindestens so groß ist wie die Anzahl der Unbekannten, d.h. $2^n \geq q(n+1)$. Für eine hinreichend große Anzahl von manifesten Variablen sollte eine Lösung immer dadurch möglich sein, daß man nur Gleichungen auswählt, die sich auf Randverteilungen niedriger Dimensionen beziehen, z.B. höchstens der zweiten Dimension. In diesem Fall würden also nur die Berechnungsgleichungen für die p_i und die p_{ij} berücksichtigt. Alternativ könnte man weniger manifeste Variablen und dafür Randverteilungen höherer Dimensionen betrachten (Gibson, 1955).

Bei der Basismethode muß man aus der Menge der n manifesten Variablen drei disjunkte Teilmengen auswählen, von denen zwei je $q-1$ und eine genau eine Variable enthält. Dabei sei q die Anzahl der latenten Klassen. Die eine ausgezeichnete Variable bezeichnet man auch als **Schichtungsvariable** (stratifier). Man stellt die ausgewählten Berechnungsgleichungen auf die folgende Weise in Matrixform dar. Auf der linken Seite steht eine Matrix B_{hv} , die sogenannte **Basismatrix**. Diese ist von der Form, daß in der ersten Spalte die eindimensionalen Randwahrscheinlichkeiten für die positive Beantwortung eines Items der ersten Teilmenge stehen und in der ersten Zeile die entsprechenden Größen für die zweite Teilmenge. Im Schnittpunkt, d.h. im ersten Element der Matrix steht eine Eins. Die restlichen Elemente der Matrix geben die zugehörigen zweidimensionalen Randwahrscheinlichkeiten wieder. Diese $(q \times q)$ -Matrix läßt sich aufgrund der zugehörigen Berechnungsgleichungen darstellen in der Form

$$B_{hv} = L_h V L'_v.$$

Hier ist L_h eine Matrix, deren Zeilen jeweils die latenten Parameter für die Items aus der ersten Teilmenge enthalten, und L_v die entsprechende Matrix für

die Items aus der zweiten Teilmenge. Zusätzlich enthalten L_h und L_v ein weiteres artifizielles Item an erster Stelle, für das alle latenten Wahrscheinlichkeiten Eins sind, d.h. die erste Zeile dieser Matrizen enthält nur Einsen. Dieses führt dazu, daß L_h und auch L'_v , die Transponierte von L_v , ($q \times q$)-Matrizen sind. Mit V wird wieder eine ($q \times q$)-Diagonalmatrix bezeichnet, die in der Diagonale die Wahrscheinlichkeiten $v^{(1)}, \dots, v^{(q)}$ enthält. Die Aufgabe besteht darin, die Matrizen L_h , V und L'_v zu bestimmen, da diese die unbekannteren latenten Parameter enthalten. Dazu stellt man eine weitere Matrixgleichung auf, gegeben durch

$$B_{hv;k} = L_h V D_k L'_v.$$

Hier ist $B_{hv;k}$ die **geschichtete Basismatrix**. Diese ergibt sich aus der Basismatrix B_{hv} durch **Schichtung** bezüglich des **Schichtungsitems** (stratifier item), das hier durch k indiziert wird. Die Schichtung erfolgt so, daß in B_{hv} jede eindimensionale Randwahrscheinlichkeit durch eine entsprechende zweidimensionale Randwahrscheinlichkeit ersetzt wird, indem man als zweite Variable die Schichtungsvariable hinzunimmt. Entsprechend werden die zweidimensionalen Randwahrscheinlichkeiten durch dreidimensionale ersetzt. Das erste Element von B_{hv} , eine Eins, wird durch p_k ersetzt, d.h. durch die Wahrscheinlichkeit für eine positive Beantwortung des Schichtungsitems. Die Matrix D_k ist eine ($q \times q$)-Diagonalmatrix, die in der Diagonale die latenten Wahrscheinlichkeiten des Schichtungsitems und sonst Nullen enthält. Es ergibt sich, daß die Lösungen t der Determinantengleichung

$$\left| L_h D_k L_h^{-1} - tI \right| = 0,$$

wobei I die Einheitsmatrix ist, gerade die latenten Wahrscheinlichkeiten für das Schichtungsitem sind. Da B_{hv} und $B_{hv;k}$ als bekannt vorausgesetzt werden und

$$L_h D_k L_h^{-1} = B_{hv;k} B_{hv}^{-1}$$

gilt, erhält man auf diese Weise die latenten Parameter für das Schichtungsitem. Aufgrund der Beziehungen

$$(B_{hv;k} B_{hv}^{-1} - I p_k^{(i)}) L_h^{(i)} = 0, \quad i=1, \dots, q,$$

wobei $L_h^{(i)}$ der i -te Spaltenvektor von L_h ist, erhält man durch Auflösung dieser homogenen linearen Gleichungssysteme die Matrix L_h . Die Lösungen dieser Gleichungssysteme sind eigentlich nur bis auf konstante Faktoren eindeutig bestimmt, jedoch wird Eindeutigkeit dadurch erzwungen, daß das erste Element von $L_h^{(i)}$ immer 1 sein muß. Ferner gilt

$$V L'_v = L_h^{-1} B_{hv}.$$

Da V eine Diagonalmatrix ist und die erste Spalte von L'_v aus Einsen besteht, enthält die erste Spalte von VL'_v die Klassenwahrscheinlichkeiten $v^{(1)}, \dots, v^{(q)}$. Falls man auf diese Weise V bestimmt hat, ergibt sich schließlich L'_v aus

$$L'_v = V^{-1}L_h^{-1}B_{hv}.$$

Liegen mehr als $(q-1) + (q-1) + 1 = 2q-1$ Items vor und bezeichnet r die Menge dieser zusätzlichen Items, so kann man die nicht notwendig quadratische Matrix P_{hr} aufstellen, die die manifesten zweidimensionalen Randwahrscheinlichkeiten p_{ij} enthält, wobei i zu h und j zu r gehört. Die fehlenden latenten Parameter ergeben sich dann aus

$$L'_r = V^{-1}L_h^{-1}P_{hr}.$$

Es ergeben sich vier Bedingungen, die erfüllt sein müssen, damit ein latentes Klassenmodell mit q Klassen bei n Items mit Hilfe der Basismethode eindeutig gelöst werden kann (Lazarsfeld & Henry, 1968, S. 56):

- (1) Es muß $n \geq 2q-1$ gelten.
- (2) Alle Klassen müssen nichtleer sein.
- (3) Es gibt mindestens ein Item, das unterschiedliche latente Wahrscheinlichkeiten hat. Dieses wird als Schichtungsitem verwendet.
- (4) Es gibt zwei disjunkte Teilmengen der Items aus jeweils $q-1$ Elementen, die das Schichtungsitem nicht enthalten. Diese Mengen sind so geartet, daß die Matrizen L_h und L_v Inverse haben.

Praktische Kritikpunkte an der Basismethode bestehen darin, daß sie wesentlich von der Auswahl der $2q-1$ Items und deren Zerlegung in drei disjunkte Teilmengen abhängt. Dieses führt zu einer Abhängigkeit von der Anordnung der Items. Weiterhin ist nicht garantiert, daß die Lösungen reelle Zahlen zwischen 0 und 1 sind.

Eine Erweiterung des obigen Ansatzes beruht auf der Verwendung der sogenannten **ansteigenden Matrizen** (ascending matrices) (Lazarsfeld & Henry, 1968, S. 60-62), die nicht von den Items sondern von den Antwortmustern her definiert werden. Dabei werden Randverteilungen höherer Ordnung mit einbezogen und weniger manifeste Variablen benötigt (Madansky, 1960).

Von Gibson (1955) wurde eine zusammengesetzte Schichtungsvariable vorgeschlagen. Dieses bedeutet, daß in der Diagonalen der Matrix D_k nicht mehr die latenten Wahrscheinlichkeiten eines einzelnen Items sondern gewichtete oder ungewichtete Summen der latenten Wahrscheinlichkeiten von mehreren Schichtungsvariablen stehen. Damit sollen die Lösungen der Determinantengleichung besser getrennt werden können.

Einen alternativen Ansatz bei Vorliegen von zwei manifesten Variablen, der auf der sogenannten **singular value decomposition** einer Matrix beruht, geben

Good (1969) und Gilula (1979) an. Speziell leitet Gilula (1979) notwendige und hinreichende Bedingungen für die Existenz von dichotomen latenten Variablen, eine Technik zur Identifikation der Verteilungen dieser Variablen sowie die bedingten Verteilungen der manifesten Variablen bei gegebenen latenten Variablen her.

Die Probleme, die auftreten, falls man die ursprünglichen Items informationsmäßig äquivalenten Transformationen unterwirft und dann latente Klassenmodelle anpaßt, diskutieren Batchelder & Narens (1977). Eine Spezifikation des latenten Klassenmodells zum Zwecke der Itemanalyse bei Leistungstests betrachtet Wilcox (1979a, 1979b).

4.6 Latentes Polynommodell

Man geht von einer stetigen latenten Variablen Y aus, die über einem gewissen Intervall definiert ist, und nimmt an, daß ein Individuum mit dem Wert y der latenten Variablen bei Item i die Reaktion $x_i = 1$ mit der Wahrscheinlichkeit

$$f_i(1 | y) = a_{0i} + a_{1i}y + a_{2i}y^2 + \dots + a_{ri}y^r$$

zeigt. Der Grad r des Polynoms, durch das die Spurfunktion beschrieben wird, soll bekannt sein in der gleichen Weise wie die Anzahl der Klassen im latenten Klassenmodell. Die Wahl dieses Grades wird zweckmäßigerweise so getroffen, daß die Anpassungsgüte des Modells an die Daten hinreichend gut ist. Die latente Verteilung g wird oft als Beta-Dichte angenommen. Dieses Polynommodell (latent polynomial model) wird ausführlich in Lazarsfeld & Henry (1968, S. 197-217) behandelt.

Auf den ersten Blick erscheint es sinnlos, Spurfunktionen als Polynome zu wählen, weil die bedingten Wahrscheinlichkeiten $f_i(1 | y)$ Zahlen zwischen 0 und 1 sein müssen. Dieses ist sicher nicht erfüllt, falls man Polynome betrachtet, die über der ganzen reellen Achse definiert sind. Deshalb nimmt man an, daß die latente Variable y nur über einem gewissen Intervall $\alpha < y < \beta$ absolut stetig verteilt ist, d.h. dort eine Wahrscheinlichkeitsdichte besitzt, und daß die $f_i(1 | y)$ Zahlen zwischen 0 und 1 sind für y aus diesem Intervall. Dieses führt zu gewissen Restriktionen für die Polynome, d.h. zu Restriktionen für die Polynomkoeffizienten. Die letzteren werden gelegentlich, um die Darstellungen zu erleichtern, als von Null verschieden angenommen. Dieses ist aber keine notwendige Voraussetzung.

Die Vorteile des Polynommodells bestehen darin, daß man durch Polynome viele stetige Spurfunktionen approximieren kann und daß die manifesten Wahrscheinlichkeiten nur von den Momenten der latenten Verteilung abhängen.

Eine Auflösung der Berechnungsgleichungen ergibt sich wieder über eine Matrixdarstellung. Die Basismatrix B_{hv} der manifesten Wahrscheinlichkeiten wird wieder als Produkt dreier Matrizen dargestellt:

$$B_{hv} = A'_h M A_v.$$

Hier sind A_h und A_v Matrizen der jeweiligen Polynomkoeffizienten und M ist eine Momentenmatrix. Die Auflösung wird dadurch erschwert, daß die Matrix D_k für die Schichtungsvariable keine Diagonalmatrix mehr ist sondern eine **Semi-Diagonalmatrix**, in der neben der Hauptdiagonalen auch Nachbardiagonalen besetzt sind.

Speziell sind lineare und quadratische Spurfunktionen untersucht worden, weil sowohl die Lösungsmethoden als auch die Interpretationsmöglichkeiten hier am einfachsten sind. Darstellungen hierfür findet man bei Lazarsfeld & Henry (1968, S. 206-211) und speziell für lineare Spurfunktionen bei Torger-son (1958, S. 367-374).

4.7 Lokalisiertes Klassenmodell

Man geht wieder von polynomialen Spurfunktionen aus, nimmt man aber an, daß die latente Verteilung g eine diskrete Verteilung mit einer endlichen Anzahl von Stützstellen ist. Es liegt also ein latentes Kontinuum vor, auf dem die Klassen an gewissen Punkten lokalisiert sind (located class model). Falls die Anzahl der Stützstellen höchstens gleich $r-1$ ist ($r = \text{Grad der Polynome}$), so ergibt sich wieder das latente Klassenmodell. Im anderen Fall bestehen Einschränkungen für die Werte der Spurfunktionen. Dieses Modell wird in Lazarsfeld & Henry (1968, S. 148-156) behandelt.

Die Annahme diskreter Klassen und einer stetigen Skala mag widersprüchlich erscheinen. Jedoch kann es einerseits sein, daß diskrete angeordnete Klassen existieren, über deren Abstand man Genaueres wissen möchte. Die andere Interpretation wäre, daß die Klassenstruktur eine Approximation an eine stetige Skala ist.

Ebenso wie im Polynommodell sind auch im lokalisierten Klassenmodell insbesondere die linearen (Lazarsfeld & Henry 1968, S. 149-153) und quadratischen (Lazarsfeld & Henry 1968, S. 153-156) Spurfunktionen untersucht worden.

4.8 Latentes Inhaltsmodell

In diesem Modell verwendet man anstelle der Polynome die Funktionen

$$f_i(1 | y) = a_i + b_i y d_i$$

und nimmt für die latente Dichte g eine Gleichverteilung über dem Intervall $(0,1)$ an. Eine Darstellung des latenten Inhaltsmodells (latent content model) wird gegeben in Lazarsfeld & Henry (1968, S. 160-181). Ausführlicher ist Somers (1961).

Im latenten Inhaltsmodell wird angenommen, daß die latente Variable Y nur Werte zwischen 0 und 1 annehmen kann. Damit auch die bedingten Wahrscheinlichkeiten $f_i(1 | y)$ zwischen 0 und 1 liegen, müssen die Koeffizienten a_i , b_i und d_i folgenden Restriktionen unterliegen:

$$d_i \geq 0, 0 < a_i < 1, 0 < a_i + b_i < 1.$$

Außerdem ist es zweckmäßig

$$b_i > 0$$

zu fordern, damit mit Wachsändern y die Wahrscheinlichkeit für die positive Beantwortung eines Items wächst. Diese Wachstumsrate kann von Item zu Item in Abhängigkeit von d_i unterschiedlich sein.

Das latente Inhaltsmodell geht von einer stetigen latenten Variablen aus, und man kann deshalb annehmen, daß sich Datensätze, für die dieses Modell geeignet ist, wesentlich von Datensätzen unterscheiden, für die das latente Klassenmodell anzuwenden ist. Die Lösung der Berechnungsgleichungen erfordert für dieses Modell allerdings einige neue Überlegungen. Hierbei gibt es Probleme mit nicht eindeutigen Lösungen.

Wegen der auftretenden Schwierigkeiten wurden einige Spezialfälle genauer untersucht: 1. Der Fall gleicher Krümmungen, d.h. gleicher Werte d_i ; 2. der Fall, daß die Spurfunktionen aller Items in Null beginnen, d.h. alle Werte a_i sind gleich Null; 3. der Fall von symmetrischen Spurfunktionen (vgl. Lazarsfeld & Henry 1968, S. 191-194).

Als Alternative zu der stetigen Gleichverteilung der latenten Variablen wird in Lazarsfeld & Henry (1968, S. 195-196) auch die Betaverteilung betrachtet. Damit soll eine stärkere Gewichtung für einen Punkt der Skala (Modalwert der Betaverteilung) erreicht werden.

4.9 Latentes Distanzmodell

Im latenten Distanzmodell (latent distance model) wird ähnlich wie im lokalisierten Klassenmodell versucht, eine gewisse Anordnung der Klassen zu berücksichtigen. Es handelt sich um eine probabilistische Version der sogenannten **Guttman-Skalierung**. In Torgerson (1958, S. 374) wird darauf hingewiesen, daß auch jedes andere Modell mit monotonen Spurfunktionen und hinreichend gut diskriminierenden Items als Analogon zur Guttman-Skalierung dienen kann. In dem deterministischen Modell von Guttman wird angenommen, daß falls ein Individuum ein Item i richtig beantwortet und das nächstschwerere Item $i+1$ falsch beantwortet, es alle Items, die leichter als i sind, richtig beantwortet und alle Items, die schwerer als $i+1$ sind, falsch beantwortet.

Die latente Variable Y sei stetig verteilt und

$$f_i(1 | y) = a_i \text{ für } y \leq c_i, \quad f_i(1 | y) = b_i \text{ für } y > c_i,$$

wobei gilt

$$c_1 < c_2 < \dots < c_n, \quad a_i < b_i.$$

Das so definierte latente Distanzmodell ist ein eingeschränktes latentes Klassenmodell. Ein Proband wird für einen sehr niedrigen Wert von y alle Items mit hoher Wahrscheinlichkeit falsch beantworten. Für Werte y mit $c_1 < y \leq c_2$ wird er Item 1 mit hoher Wahrscheinlichkeit richtig und die anderen Items falsch beantworten, usw. Im deterministischen Grenzfall mit $a_i = 0$, $b_i = 1$ ergibt sich das Guttman-Modell.

In Lazarsfeld & Henry (1968, S. 140-142) und in Hays & Borgatta (1954) wird auch der Spezialfall mit der Restriktion $a_i = 1 - b_i$ näher untersucht. Das latente Distanzmodell wird allgemein in Lazarsfeld & Henry (1968, S. 123-148), in Hays & Borgatta (1954), in Torgerson (1958, S. 374-385) und Proctor (1970) diskutiert.

4.10 Testtheoretisches Modell

Es sei Y eine stetige latente Variable und $f_i(1 | y)$ die Dichte einer $N(a_i, b_i^2)$ -Verteilung. Man nimmt an, daß Y gemäß $N(0,1)$ verteilt ist. Eine andere Formulierung des Modells geht von einer Darstellung (Anderson, 1959)

$$Z = cY + a + U$$

aus, wobei Z , c , a und U Vektoren der Dimension n sind, U_i normalverteilt mit der Varianz b_i^2 und Y gemäß $N(0,1)$ verteilt ist. Dieses ist ein Faktoranalysemodell mit einem gemeinsamen Faktor. Dieses Modell wird in der psycholo-

gischen Testtheorie verwendet, wobei ein Alternativmodell anstelle der Normalverteilung die logistische Verteilung verwendet.

Diskutiert werden die testtheoretischen Modelle, darunter auch das **Rasch-Modell**, von Lazarsfeld & Henry (1968, S. 217-225). Eine Gegenüberstellung von latenter Strukturanalyse und Testtheorie findet man bei Lazarsfeld (1959, 1960).

Während man das testtheoretische Modell in der obigen Form in die latente Strukturanalyse einbetten kann, werden in der Testtheorie gewisse Modifikationen des Modells vorgenommen, um es den dort gestellten Anforderungen besser anzupassen. So bemerken Lord & Novick (1968, S. 358-394) in ihrem Vergleich mit den latenten Strukturmodellen, daß anders als in der latenten Strukturanalyse in der Testtheorie nicht alle Antwortmuster sondern nur gewisse Statistiken dieser Muster, die die für testtheoretische Fragestellungen wichtigen Informationen enthalten, betrachtet werden. So wird z.B. anstelle des Vektors der Einzelantworten nur die Statistik ‚Anzahl der richtigen Antworten‘ verwendet, die unter gewissen Voraussetzungen suffizient ist, d.h. die gesamte relevante Information enthält.

Eine Anwendung der latenten Strukturanalyse in der Testtheorie gibt Meredith (1965) aufbauend auf McDonald (1962). In der Arbeit wird unter anderem die empirische Bestimmung von Spurfunktionen und die Schätzung der Itemreliabilitäten diskutiert.

In Torgerson (1958, S. 385-395) wird das testtheoretische Modell auf der Basis von Lord (1953) und Tucker (1952) diskutiert. Dabei ergeben sich das Normal-Ogiven-Modell von Tucker (1952) und das Intelligenzskalenmodell von Thurstone (1925) als Spezialfälle eines allgemeinen Normal-Ogiven-Modells.

4.11 Latentes Profilmodell

Im latenten Profilmodell (latent profile model) geht man von q latenten Klassen, d.h. q diskreten Werten der latenten Variablen Y aus, nimmt aber an, daß die manifesten Variablen $X_1 \dots X_n$ stetig verteilt sind. Für die bedingten Verteilungen der X_i werden keine speziellen Annahmen gemacht, aber gefordert, daß einige der bedingten Momente, z.B. $E[X_i | y_j]$ für $i=1, \dots, n$, $j=1, \dots, q$, eindeutig spezifiziert sind. Für jede manifeste Variable erhält man auf diese Weise einen Vektor der bedingten Erwartungswerte bei festen Klassen

$$(E[X_i | y_1], \dots, E[X_i | y_q]),$$

der als **latentes Profil** bezeichnet wird. Damit eine Struktur identifizierbar ist, muß man aus der als bekannt vorausgesetzten gemeinsamen Verteilung der

X_1, \dots, X_n das Profil, den Vektor der bedingten Varianzen und die durch $v^{(1)}, \dots, v^{(4)}$ gegebene latente Verteilung bestimmen.

Die latente Profilanalyse ergibt für jedes Item ein Profil, d.h. gibt für jedes Item an, welchen Erwartungswert es in jeder der latenten Klassen hat. Man wendet zur Lösung der Berechnungsgleichungen die Basismethode auf die standardisierten manifesten Variablen an und erhält deshalb als Lösung nicht das latente Profil sondern das standardisierte latente Profil. Durch Multiplikation mit der Standardabweichung und Addition des Erwartungswertes für das jeweilige Item ergibt sich aber sofort das nichtstandardisierte latente Profil.

Das latente Profilmodell wird besprochen bei Lazarsfeld & Henry (1968, S. 228-239) und Fielding (1977). In Gibson (1959) erfolgt die Verallgemeinerung der latenten Klassenanalyse zur latenten Profilanalyse, um Beziehungen zwischen quantitativen Variablen zu untersuchen. McDonald (1962) untersucht den Zusammenhang mit dem latenten Klassenmodell. In Takane (1976) werden für die bedingten Verteilungen der manifesten Variablen bei festen latenten Klassen speziell Normalverteilungen angenommen. Eine Verallgemeinerung des latenten Profilmodells derart, daß nicht nur die manifesten sondern auch die latenten Variablen stetig verteilt sind, wird in Lazarsfeld & Henry (1968, S. 239-244) versucht.

4.12 Andere Modelle

Eine naheliegende Verallgemeinerung des latenten Klassenmodells mit dichotomen Items ist die Erweiterung auf polytome Items. Man vergleiche hierzu Lazarsfeld & Henry (1968, S. 226-228) und Formann (1978b). In diese Richtung gehören auch die Ansätze, die die latente Klassenanalyse in die Theorie der multivariaten Kontingenztafeln einbetten. Beispiele dafür sind die Arbeiten von Goodman (1974a, 1974b, 1978) und Haberman (1974, 1977). In Goodman (1974b) ist vor allem der Ansatz der **Pfadigrammanalyse** bemerkenswert, während die anderen Artikel die Theorie der **loglinearen Modelle** verwenden.

Eine andere Erweiterung der Theorie betrifft den Einsatz von latenten Klassenmodellen bei der Analyse von Verläufen, d.h. in Meßwiederholungsplänen. Hier wurde ein latentes **Markov-Ketten-Modell** in Lazarsfeld & Henry (1968, S. 252-263) betrachtet. Man nimmt dabei an, daß die Population zu jedem Zeitpunkt in latente Klassen zerfällt und daß die Antwortwahrscheinlichkeiten nur durch diese Klassen bestimmt werden. Zusätzlich sollen sich die Individuen entsprechend einer homogenen Markovkette von Klasse zu Klasse bewegen können. Bei q latenten Klassen führt dieses zu einer (qxq) -Matrix der Übergangswahrscheinlichkeiten. Ein ähnliches, wenn auch nicht identisches Modell schlug Coleman (1964) vor. Einen weiteren Ansatz zu einer

latentem Markov-Ketten-Analyse mit einer Anwendung auf das Gefangenen-Dilemma entwickelt von der Sanden (1978).

5. Statistische Fragestellungen

5.1 Parameterschätzung

5.1.1 Einführung

Der ursprüngliche Ansatz der latenten Strukturanalyse war algebraischer und nicht statistischer Natur. Man ging davon aus, daß die gemeinsame Verteilung $h(x_1, \dots, x_n)$ der manifesten Variablen bekannt ist und daß das speziell betrachtete Modell die tatsächliche latente Struktur richtig wiedergibt. Das Problem der Bestimmung der latenten Parameter bestand somit vor allem darin, Bedingungen anzugeben und Gleichungen auszuwählen, die eine Identifikation der Parameter ermöglichten. Bei der Anwendung der latenten Strukturanalyse kann man jedoch weder davon ausgehen, daß man die Verteilung der manifesten Variablen kennt, noch daß das angenommene Modell exakt zutrifft. Die verfügbaren Daten weichen aufgrund von Meß- und Stichprobenfehlern von den zu schätzenden Parametern ab und sind als Realisationen von manifesten Zufallsvariablen anzusehen. Es besteht also das Bedürfnis nach Schätzverfahren für die latenten Parameter, die gewissen Gütekriterien genügen.

5.1.2 Algebraische Verfahren

Bei den sogenannten **algebraischen Verfahren** ersetzt man die unbekanntem ein- und mehrdimensionalen manifesten Randwahrscheinlichkeiten p_i , p_{ik} , p_{ijk} usw. durch die entsprechenden relativen Häufigkeiten und verwendet dann die für die Parameter entwickelten algebraischen Methoden, um die Berechnungsgleichungen nach den unbekanntem latenten Parametern aufzulösen. Weil auf den linken Seiten der Gleichungen jetzt nicht mehr die manifesten Randwahrscheinlichkeiten stehen, sondern nur deren Schätzungen, erhält man als Lösungen ebenfalls nur Schätzungen der latenten Parameter.

Schwierigkeiten entstehen dadurch, daß verschiedene Bedingungen an die manifesten Parameter gestellt werden müssen, damit die algebraischen Techniken verwendet werden können. Es handelt sich hierbei um Bedingungen, die die Identifizierbarkeit der Parameter und die Invertierbarkeit gewisser Matrizen sicherstellen. Eine erste Voraussetzung ist also die Wahl von Items, deren zugehörige Lösungswahrscheinlichkeiten die angesprochenen Bedingungen erfüllen. Eine zweite Voraussetzung betrifft den vorgegebenen Datensatz. Auch die vorgefundenen relativen Häufigkeiten müssen die Bedingungen er-

füllen, d.h. dürfen nicht allzuweit von den zugehörigen Wahrscheinlichkeiten abweichen. Dieses setzt im Verhältnis zu der Anzahl der betrachteten Items große Stichprobenumfänge voraus. Jedoch muß man in jedem Falle mit einer gewissen Wahrscheinlichkeit damit rechnen, daß die Verfahren versagen, denn auch bei noch so großen Stichprobenumfängen besteht eine gewisse positive Wahrscheinlichkeit dafür, daß z.B. alle Versuchspersonen die Items in der gleichen Weise beantworten oder sich eine sonstige ungünstige Konstellation ergibt, die keinen Rückschluß auf die latenten Parameter zuläßt.

Eine weitere Schwierigkeit ist die Möglichkeit unzulässiger Lösungen. Die Verfahren lassen nämlich durchaus zu, daß man negative oder Eins übersteigende Schätzungen für die latenten Wahrscheinlichkeiten erhält. Selbst nicht-reelle Zahlen sind u.U. als Lösungen möglich.

Schließlich kann man nicht unbedingt damit rechnen, eindeutige Lösungen zu erhalten. Je nachdem, welche Gleichungen oder welche Schichtungsvariablen man auswählt, kann man unterschiedliche Lösungen erhalten. Man benötigt deshalb zusätzliche Selektionskriterien. Solche Kriterien werden diskutiert in Anderson (1959) und darauf basierend in Fielding (1977). So erfordert die Basislösung die Inversion von Matrizen, deren Determinanten demnach von Null verschieden sein müssen. Es scheint deshalb naheliegend zu sein, bei der Basismethode die Itemmenge so in die drei Teilmengen von manifesten Variablen zu unterteilen, daß die Determinanten groß werden. Dieses Vorgehen scheinen Simulationsstudien zu bestätigen. Ein Vorschlag ist, diejenigen M derartigen Zerlegungen auszuwählen, die die größten Werte der Determinanten liefern. Für jede der M Zerlegungen ist dann das Gleichungssystem zu lösen, wobei nichtzulässige Lösungen ausgeschieden werden. Schließlich mittelt man über alle zulässigen Lösungen.

Ein Alternativvorschlag besagt, daß man für jede Lösung ein Maß für die Anpassungsgüte des Modells bestimmt und dann die Lösung mit der besten Anpassung aussucht. Dazu vergleicht man etwa die beobachteten Häufigkeiten mit den aufgrund der Parameterschätzungen durch das Modell vorhergesagten Wahrscheinlichkeiten mit Hilfe eines Chiquadratkriteriums. Die Lösung mit dem kleinsten Chiquadratwert wird dann ausgewählt. Einen solchen Ansatz findet man bei Lazarsfeld & Henry (1968, S. 98).

Die sich bei den algebraischen Verfahren ergebenden Punktschätzungen sind **konsistent**. Dieses bedeutet, daß man durch genügend große Wahl des Stichprobenumfanges immer erreichen kann, daß die Wahrscheinlichkeit dafür, daß sich Schätzung und zu schätzender Parameter um mehr als eine vorgegebene Differenz (absolut genommen) unterscheiden, kleiner als eine vorgegebene Wahrscheinlichkeit ist. Auch bei einer konsistenten Schätzung kann es also mit einer gewissen Wahrscheinlichkeit vorkommen, daß sich Parameter und Schätzung auch bei großem Stichprobenumfang stark unterscheiden. Jedoch ist

diese Wahrscheinlichkeit bei konsistenten Punktschätzungen klein. In diesem Sinne sind relative Häufigkeiten konsistente Punktschätzungen für Wahrscheinlichkeiten. Sofern die Schätzungen für die latenten Parameter stetige Funktionen der relativen Häufigkeiten sind, erhält man auch konsistente Schätzungen für die latenten Parameter. Außerdem sind diese Schätzungen asymptotisch normalverteilte Zufallsvariablen. Für die latente Klassenanalyse findet man eine Herleitung von T. W. Anderson in Lazarsfeld & Henry (1968, S. 273-287).

Das bisher angesprochene algebraische Verfahren verwendet nur die Schätzungen der manifesten Randwahrscheinlichkeiten p_i , p_{ij} und p_{ijk} , d.h. nur die ersten drei Momente, und wurde ursprünglich von Lazarsfeld & Dudman (1951) vorgeschlagen und von Anderson (1954, 1959) erweitert. Es beruht im Wesentlichen auf der Bestimmung der Nullstellen einer Determinantengleichung. Mehr Information nutzt ein Ansatz von Gibson (1955) aus. Eine Verallgemeinerung durch Mitverwendung von höheren Momenten stammt von Madansky (1958, 1960). In Gibson (1962a) wird angegeben, wie die latenten Parameter von Items zu schätzen sind, die nachträglich hinzugenommen wurden und bei denen man dieselbe latente Struktur unterstellen kann.

5.1.3 Faktorisierungsmethoden

Von Gibson (1951, 1956) wurde ein Ansatz vorgeschlagen, der nur die Berechnungsgleichungen für die Randwahrscheinlichkeiten 1. und 2. Ordnung berücksichtigt und das Gleichungssystem durch Faktorisierung einer Matrix analog zu den Verfahren der Faktoranalyse zu lösen versucht. Eine kurze Darstellung gibt Anderson (1959). Ein Vorteil dieses Verfahrens ist darin zu sehen, daß es keine Auswahl von Berechnungsgleichungen voraussetzt wie im Falle der algebraischen Methoden. Dafür kommt eine Nichteindeutigkeit der Lösungen dadurch zustande, daß das Ergebnis einer solchen Faktorisierung nur bis auf orthogonale Rotationen eindeutig bestimmt ist. Hier liegt es nahe, die Berechnungsgleichungen höherer als 2. Ordnung mit heranzuziehen, um eine eindeutige Lösung zu erhalten. Ein Algorithmus hierfür wurde von Green (1970) für die latente Klassenanalyse vorgeschlagen, der dann von Mardberg (1971) und Isaacson (1972) für die latente Profilanalyse verwendet wurde. Der ursprüngliche Ansatz von Green (1951), der auf der Berücksichtigung der 3. Momente beruht, wird auch von Anderson (1959) diskutiert.

5.1.4 Maximum-Likelihood-Schätzungen

Sowohl die algebraischen Methoden als auch die Faktorisierungsmethoden bei der Schätzung der latenten Parameter verwenden ein in der Statistik unübliches

zweistufiges Vorgehen. Zunächst werden die manifesten Verteilungen als bekannt angesehen und eine algebraische Lösung des Gleichungssystems für die latenten Parameter angestrebt. Im zweiten Schritt werden dann in die Lösungsgleichungen für die latenten Parameter die beobachteten relativen Häufigkeiten anstelle der manifesten Wahrscheinlichkeiten eingesetzt. Dieses Verfahren führt zwar zu konsistenten und asymptotisch normalverteilten Schätzungen der latenten Parameter, wird aber in der Statistik deshalb kaum verwendet, weil es andere Verfahren gibt, die ebenfalls konsistente und asymptotisch normalverteilte Schätzungen liefern, die aber eine größere Effizienz haben. Damit ist gemeint, daß diese Verfahren für große Stichprobenumfänge Schätzungen liefern, die eine kleinere Varianz als die Schätzungen des obigen Verfahrens haben, d.h. Schätzungen die weniger um den zu schätzenden Parameter streuen. Falls man Schätzungen hat, die konsistent und asymptotisch normalverteilt sind, sowie asymptotisch die kleinstmögliche Varianz haben, so spricht man von BAN-Schätzern (best asymptotic normal). Derartige Schätzer liefert unter bestimmten Voraussetzungen die **Maximum-Likelihood-Methode**. Diese geht von einem bekannten Verteilungstyp für die manifesten Variablen aus, z.B. einer n-dimensionalen Multinomialverteilung (McHugh, 1956, 1958), und versucht Schätzungen der latenten Parameter so zu bestimmen, daß die Wahrscheinlichkeit dafür, daß gerade die vorgefundenen relativen Häufigkeiten auftreten, maximal wird. Man kann auch sagen, daß zwischen dem Vektor der relativen Häufigkeiten und dem Vektor der zugehörigen Wahrscheinlichkeiten ein gewisser Abstand besteht und man die Schätzungen so wählt, daß falls man sie anstelle der Wahrscheinlichkeiten in das Abstandsmaß einsetzt, dieser Abstand minimal wird. Bei den Maximum-Likelihood-Schätzungen ist der Abstand gegeben durch die „Unwahrscheinlichkeit“ mit der ein Vektor von relativen Häufigkeiten bei einem vorgegebenen Vektor von Wahrscheinlichkeiten auftritt. Andere Schätzverfahren, die ebenfalls zu BAN-Schätzern führen, verwenden als Abstandsmaß eine Chiquadratstatistik (Minimum-Chiquadrat-Methode, modifizierte Minimum-Chiquadrat-Methode) oder eine Informationsstatistik.

Bei der Maximum-Likelihood-Methode verfährt man so, daß man die **Likelihoodfunktion**, d.h. die Wahrscheinlichkeitsfunktion, in die man die vorgefundenen Meßwerte eingesetzt hat, nach den unbekanntem Parametern partiell differenziert und die Ableitungen gleich Null setzt. Diese sogenannten **Maximum-Likelihood-Gleichungen** sind dann nach den Schätzungen der gesuchten Parameter aufzulösen. Da eine explizite Lösung meist nicht möglich ist, verwendet man Iterationsverfahren. Die Ausgangswerte liefert dabei meist eines der oben angesprochenen Verfahren (Henry, 1966).

Ein Vorteil der Maximum-Likelihood-Methode ist sicher, daß die Willkürlichkeit der Lösung in Abhängigkeit von den ausgewählten Gleichungen nicht mehr auftritt. Jedoch setzt das Verfahren die Kenntnis der latenten Verteilung

und der Spurfunktionen voraus, was außer im Falle der latenten Klassenanalyse zu Schwierigkeiten führen kann.

Für die latente Klassenanalyse findet man die Herleitung der Schätzungen bei McHugh (1956) mit einer Korrektur in McHugh (1958). In Formann (1978a, 1978b) wird durch eine vorausgehende Transformation der latenten Parameter erreicht, daß diese zwischen Null und Eins liegen müssen. Die transformierten Parameter werden dann mit der Maximum-Likelihood-Methode geschätzt. Man erhält mit diesem Verfahren zwar immer zulässige Lösungen, jedoch im allgemeinen keine Maximum-Likelihood-Schätzungen für die ursprünglichen Parameter (vgl. Brunk, 1958). In Goodmann (1979b) wird im Gegensatz zu einer Behauptung von Formann (1978a) bewiesen, daß das Iterationsverfahren von Goodman (1974a, 1974b) zu Maximum-Likelihood-Schätzungen führt, die immer innerhalb des zulässigen Intervalls liegen.

Erweiterungen auf latente Strukturmodelle mit polytomen Items findet man bei Goodman (1974a, 1974b) und Haberman (1974, 1977). Für die latente Profilanalyse unter der Annahme, daß die bedingten Verteilungen der manifesten Variablen bei gegebenen latenten Klassen Normalverteilungen sind, findet man bei Takane (1976) Maximum-Likelihood-Schätzer hergeleitet.

5.1.5 Aufteilungsmethoden

Von Madansky (1958, 1959) wurde ein Verfahren vorgeschlagen, das direkt die Hauptvoraussetzung der latenten Strukturanalyse, nämlich die lokale Unabhängigkeit, ausnutzt und Schätzungen der latenten Parameter danach bewertet, wie gut dieses Axiom erfüllt ist. Im latenten Klassenmodell ordnet man jedes der N Individuen genau einer der q latenten Klassen zu, d.h. teilt die Stichprobe in q Teilstichproben auf. Für jede Teilstichprobe muß das Axiom der lokalen Unabhängigkeit erfüllt sein. Man kann dann für jede Teilstichprobe die Teststatistik eines Chiquadrattests auf Unabhängigkeit berechnen und diese Größen über alle q Klassen aufsummieren (Anderson, 1959). Diese Summe bildet ein Maß der bedingten Unabhängigkeit über alle Klassen. Eine Schätzprozedur besteht dann darin, unter allen möglichen Aufteilungen diejenige auszuwählen, die den kleinsten Wert der Summenchiquadratstatistik liefert. Dieses Minimum ist nicht notwendig asymptotisch chiquadratverteilt. Die latenten Parameter werden dann durch die üblichen in der Chiquadratstatistik verwendeten Schätzwerte geschätzt.

Ein numerisch einfacheres Maß wurde von Madansky (1958, 1959) anstelle der Chiquadratsumme vorgeschlagen (vgl. Anderson, 1959; Fielding, 1977). Die Minimierung dieses Maßes führt zu konsistenten Schätzungen für die latenten Parameter, und es lassen sich auch asymptotische Varianzen dieser Schätzer angeben.

Da die Anzahl der möglichen Aufteilungen mit wachsenden N und q sehr schnell anwächst, scheint das Verfahren nur für relativ kleine Stichprobenumfänge und wenige latente Klassen praktikabel zu sein. In Cassady, Miller & Dingman (1968) wird eine Version des Verfahrens betrachtet, die weniger Rechenzeit benötigt. Ein anderer Ansatz könnte darin bestehen, eine rechen-technisch noch zu bewältigende Menge von Zufallsaufteilungen zu verwenden und eines der Kriterien in einer solchen Menge zu minimieren.

5.1.6 Andere Schätzmethoden

Andere naheliegende Schätzverfahren wurden schon im Zusammenhang mit der Maximum-Likelihood-Methode diskutiert. Dieses waren die Chiquadrat-Minimum-Methode, die als Abstandsmaß die Chiquadratanpassungsstatistik verwendet (Lazarsfeld & Henry, 1968, S. 99-105; van der Waerden, 1971), ferner die modifizierte Chiquadrat-Minimum-Methode, bei der man den Nenner der Anpassungsstatistik vereinfacht (van der Waerden, 1971) und die von Madansky (1958) verwendet wurde sowie die auf der Informationsstatistik beruhenden Schätzer (Kuliback, 1959). Diese Verfahren führen ebenso wie die Maximum-Likelihood-Methode zu BAN-Schätzern. Eine Entscheidung für eines der Verfahren wäre demnach vor allem unter dem Gesichtspunkt des Rechenaufwandes zu treffen. Leider führen alle diese Verfahren zu komplizierten Schätzern.

Ein iteratives Verfahren von J. E. Moyal beschreibt Anderson (1959). Man wählt eine beliebige Zuordnung der Individuen zu den latenten Klassen. Diese führt direkt zu Schätzungen für die latenten Parameter durch relative Häufigkeiten. Aufgrund dieser Schätzungen ordnet man jedes Individuum der Klasse zu, für die die geschätzte bedingte Wahrscheinlichkeit am größten ist. Aufgrund dieser Zuordnung erhält man neue Schätzungen, usw. Dieses Verfahren ist zwar rechentechnisch einfach, führt aber gemäß Anderson (1959) zu nicht-konsistenten Schätzungen.

In Takane (1972) wird das latente Klassenmodell als Spezialfall der Faktoranalyse interpretiert und die Rekrutierungswahrscheinlichkeiten analog zu der Schätzung der Faktorenwerte geschätzt. Diese Schätzungen werden mit den direkten Schätzungen der Rekrutierungswahrscheinlichkeiten verglichen.

5.1.7 Programme und Algorithmen

Angesichts der Probleme, die das Schätzen der latenten Parameter mit sich bringt, ist es nicht verwunderlich, daß seit der Einführung der latenten Strukturanalyse zahlreiche Autoren bemüht waren, dem Anwender Algorithmen

bzw. fertige Programme zur Verfügung zu stellen. Eine Auswahl sei hier zitiert.

Für die Parameterschätzung bei der latenten Klassenanalyse entwickelten Lazarsfeld & Henry (1968, S. 74) ein Programm LASY, das auf einem algebraischen Verfahren beruhte. Von Henry (1966) stammt das Programm BAN, das unter den Namen NEWBAN geführt wird und Maximum-Likelihood-Schätzungen liefert. Von Cassady, Miller & Dingman (1968) stammt ein Algorithmus, der eine Verbesserung des von Madansky (1959) vorgeschlagenen Algorithmus zur Aufteilungsmethode darstellt.

Für die latente Klassenanalyse mit polytomen Items gibt Goodman (1974a) einen Algorithmus für Maximum-Likelihood-Schätzungen an. Eine Verallgemeinerung findet man bei Haberman (1977).

Für die latente Profilanalyse schlug Green (1970) einen Algorithmus basierend auf der Faktorisierungsmethode vor. Dieser wurde in dem Programm LPA von Mardberg (1971, 1975) und Isaacson (1972) verwendet. Von Takane (1976) wurde das Maximum-Likelihood-Schätzungen liefernde Programm MAXLPM entwickelt.

5.1.8 Probleme beim Schätzen

In Fielding (1977) werden eine Reihe von Problemen diskutiert, die beim Schätzen von latenten Parametern auftreten.

Zwei Hauptfragen sind das Vorhandensein einer nicht allzu zeitaufwendigen Berechnungstechnik und die Konvergenz von Iterationsmethoden, falls solche verwendet werden. Z.B. ist die Konvergenz des von Takane (1976) vorgeschlagenen Algorithmus theoretisch noch nicht untersucht und offensichtlich nur für geeignete Ausgangswerte gegeben. Konvergenzuntersuchungen derartiger Algorithmen findet man in Haberman (1977). Nach Anderson (1959) haben Simulationsstudien und theoretische Überlegungen gezeigt, daß zumindest bis zu Stichprobenumfängen von 1000 sich sehr ungenaue Schätzungen ergeben können. Auch die Faktorisierungsmethode, obwohl mehr Information ausnutzend und in der Version von Green (1970) eine Rotation vermeidend, konvergiert meist sehr langsam. Außerdem liefert sie Schätzungen mit großen Varianzen, d.h. ungenaue Schätzungen. Dieses kann dazu führen, daß Modelle mit ganz unterschiedlichen Parameterstrukturen zu dem gleichen Datensatz passen.

Auch für Maximum-Likelihood-Schätzungen benötigt man hinreichend große Stichprobenumfänge bzw. hinreichend große Besetzungszahlen für die Kontingenztafeln der Antwortmuster. Lazarsfeld & Henry (1968, S. 82) berichten,

daß bei Simulationsstudien mit dem Programm BAN die Folge der Schätzer nicht immer konvergierte. Insbesondere war dieses nicht der Fall, falls die latenten Itemparameter nahe bei 0 oder 1 lagen. Auch hier zeigten sich große Varianzen der Schätzungen.

5.2 Signifikanztests

Die meisten Signifikanztests, die in der latenten Strukturanalyse vorgeschlagen worden sind, beziehen sich auf das Anpassungsproblem, d.h. auf die Frage, ob das angenommene Modell mit den Daten verträglich ist. Dieses wird gelegentlich reduziert auf das Problem, ob die angenommene Anzahl der latenten Klassen ausreichend ist. Man muß sich darüber im klaren sein, daß die angegebenen Tests immer nur Tests bezüglich der Nullhypothese sind, daß sich die manifesten Parameter aufgrund des Axioms der lokalen Unabhängigkeit als Summen von Produkten der latenten Parameter darstellen lassen. Selbst wenn diese Nullhypothese nicht verworfen werden kann, ist es möglich, daß andere Modelle die Daten noch besser erklären könnten. Ein zusätzliches Kriterium zum Verwerfen eines Modells sind sicher latente Parameter, die nicht zwischen 0 und 1 liegen.

Für die latente Klassenanalyse schlägt McHugh (1956) einen Chiquadratanpassungstest vor (vgl. auch Fielding, 1977). Für die latente Klassenanalyse mit polytomen Items werden in Goodman (1974b) Chiquadratanpassungstests und in Goodman (1974a) Likelihoodquotiententests angegeben. Beide Möglichkeiten werden auch in Haberman (1977) und Formann (1978b) diskutiert. Für das latente Distanzmodell geben Lazarsfeld & Henry (1968, S. 148) einen Test an. In Gibson (1962) findet man einen Chiquadratanpassungstest für das Problem, ob zusätzliche Items zu einem aufgrund von anderen Items aufgestellten Modell passen.

Für die latente Profilanalyse gibt Takane (1976) einen Likelihoodquotiententest für die Anpassungshypothese an. Darauf aufbauend wird ein Test diskutiert, der die Nullhypothese, daß q Klassen vorliegen, gegen die Alternativhypothese testet, daß mindestens $q + 1$ Klassen vorliegen. Schließlich werden Tests, die spezielle Hypothesen bezüglich der Parameter betreffen, diskutiert.

6. Schätzung der latenten Variablen

Als Schätzung der latenten Variablen wird von Anderson (1959) der Rückschluß von dem vorgefundenen Antwortmuster eines Individuums auf seine Position im latenten Raum, speziell etwa seine Zuordnung zu einer bestimmten latenten Klasse, bezeichnet. In Torgerson (1958, S. 373) wird dieses als

Zuordnung von Skalenwerten an die Individuen, hier speziell für das lineare Polynommodell, diskutiert. Für die latente Klassenanalyse wird das Problem in Lazarsfeld & Henry (1968, S. 36-39, S. 68-70) behandelt. Spezialfälle, die die Skalierung von Antwortmustern betreffen, diskutiert Goodman (1975, 1978, 1979a). Für das latente Distanzmodell vergleiche man Lazarsfeld & Henry (1968, S. 138-139) und Torgerson (1958, S. 385), für das latente Profilmodell Lazarsfeld & Henry (1968, S. 238-239) und Gibson (1962b), für das latente Klassenmodell mit polytomen Items Formann (1978b).

Die Zuweisung von Individuen zu Positionen des latenten Raumes entspricht der Durchführung einer Cluster-Analyse (Fielding, 1977). In Skene (1978) wird versucht, diesen Aspekt der latenten Strukturanalyse auch auf Probleme der Diskriminanzanalyse anzuwenden für den Fall von fehlenden Daten. Hingegen verwendet Ikuzawa (1968) eine Diskriminanzanalyse, um Individuen im latenten Klassenmodell zu identifizieren.

7. Vergleich mit anderen Verfahren

7.1 Vergleich mit der Faktoranalyse

Seitdem die latente Strukturanalyse eingeführt wurde, ist sie immer wieder mit der schon vorher existierenden in Konkurrenz stehenden Faktoranalyse verglichen worden, und es ist versucht worden, die Zusammenhänge zwischen diesen beiden Verfahren herauszuarbeiten. In Green (1952) wird darauf hingewiesen, daß der Hauptunterschied der beiden Verfahren zu sein scheint, daß in der Faktoranalyse nur die Interaktionen 1. Ordnung berücksichtigt werden, in der latenten Strukturanalyse aber auch Interaktionen höherer Ordnung. Jedoch glaubt Gebert (1978) festgestellt zu haben, daß in den Beispielen, die in der Literatur zur latenten Klassenanalyse publiziert wurden, die Interaktionen höherer Ordnung alle exakt oder nahezu Null sind und daß die latente Klassenanalyse anscheinend nur in solchen Fällen sinnvolle Interpretationen liefert.

Wie schon aus Green (1952) geschlossen werden kann, besteht die Möglichkeit, die Faktoranalyse als Spezialfall in die latente Strukturanalyse einzubetten. Ähnlich finden Lazarsfeld & Henry (1968, S. 240-244) bei der Betrachtung stetiger latenter Räume eine Äquivalenz mit einem Einfaktormodell. In Gibson (1962a) wird auf die enge Beziehung der latenten Klassenanalyse zu dem faktoranalytischen Modell von Dwyer (1937) hingewiesen, während Gibson (1956) die Identität des latenten Strukturmodells von Lazarsfeld (1950a) und des Konzepts der Proportionalprofile in der Faktoranalyse bei Cattell (1944) und Cattell & Cattell (1955) nachweist. Hierhin gehört auch die Arbeit von Gibson (1959), in der als Vorteil der latenten Klassenanalyse gegenüber der Faktoranalyse angeführt wird, daß man auf diese Weise die Probleme der

Schätzung der Kommunalitäten, der Rotation und der Nichtlinearität vermeiden könne. Aus dieser Zeit ist auch der Vergleich der beiden Verfahren durch Lazarsfeld (1959). Einen empirischen Vergleich führten Miller, Eyman & Dingman (1961) durch. Eine Anwendung beider Verfahren bei der Klassifikation geistig zurückgebliebener Patienten ergab eine bessere Übereinstimmung mit der üblichen Typologie für die latente Klassenanalyse.

In Fielding (1977) wird ebenfalls die Einbettung der Faktoranalyse in das Modell der latenten Strukturanalyse diskutiert. Die Auffassung, daß es sich um zwei verschiedene Ansätze handelt, wird mit der praktischen Anwendung der latenten Strukturanalyse in Verbindung gebracht, bei der meist nur von einer latenten Variablen, die zudem noch klassiert ist, ausgegangen wird. Dieses führt nach Mardberg (1973) dazu, daß man von Faktoranalyse spricht, wenn es um die Art des Zusammenhangs der manifesten Variablen geht, während latente Klassen eine Möglichkeit bieten, Individuen zu klassifizieren.

In Anderson (1959) wird ein allgemeines Skalierungsmodell aufgestellt, als dessen Spezialfälle sich Faktoranalyse und latente Strukturanalyse ergeben. Daß in dieser Arbeit die Faktoranalyse nicht als Spezialfall der latenten Strukturanalyse angesehen wird, liegt an einer eingeschränkten Definition der latenten Strukturanalyse, bei der die manifesten Variablen speziell dichotom sein müssen. Dieses ist wohl auch ein Grund dafür, daß in Takane (1972) das latente Klassenmodell als Spezialfall der Faktoranalyse angesehen wird.

Wie in Lord & Novick (1968, S. 545) angemerkt wird, kann man bei der Faktoranalyse die Annahme der lokalen Unabhängigkeit abschwächen zu der Annahme der linearen lokalen Unabhängigkeit. Dieses bedeutet aber im Gegensatz zu der Meinung von Takane (1972) nicht eine größere Allgemeinheit der faktoranalytischen Modelle, sondern ist eine Folge der Nichtberücksichtigung höherer Interaktionen in der Faktoranalyse. Wie Harper (1972) gezeigt hat, läßt sich die lokale Unabhängigkeit im latenten Klassenmodell dadurch abschwächen, daß man paarweise Abhängigkeit von Items zuläßt.

7.2 Vergleich mit der Guttman-Skalierung

Die ersten Arbeiten von Lazarsfeld (1950a, 1950b) erschienen in enger Nachbarschaft mit Arbeiten zu Skalierungsverfahren, insbesondere der Guttman-Skalierung. Schon McCarthy (1951) erkannte, daß man die Guttman-Skalierung als deterministischen Grenzfall der latenten Klassenanalyse ansehen kann. In Hays & Borgatta (1954) wird dann die latente Distanzanalyse direkt als ein probabilistisches Modell für die nichtperfekte Guttman-Skala verwendet. Die Notwendigkeit für ein solches Modell wird in den Schwierigkeiten gesehen, die nichtperfekte Skalen bei Anwendung der Guttmantheorie machen.

8. Anwendungen

Die latente Strukturanalyse ist auf die unterschiedlichsten Fragestellungen angewandt worden. Eine Auswahl sei hier aufgeführt.

Die latente Klassenanalyse wurde von Gill (1976) zur Klassifikation von Sehbehinderten, von Wiggins (1973) zur Analyse von Wahleinstellungen und Verhaltensänderungen, von Kadushin (1966) zur Bildung von unterschiedlich motivierten Gruppen, die eine Psychotherapie wünschten (vgl. auch Lazarsfeld & Henry, 1968, S. 88-92), sowie von Lazarsfeld (1950b) zur Untersuchung der Einstellung eingezogener Rekruten (vgl. auch Lazarsfeld & Henry, 1968, S. 106-111) verwendet. In McHugh (1956) wird eine auf Schumacher, Maxson & Martinek zurückgehende Untersuchung über kreative Fähigkeit bei Maschinenkonstruktionen und in Gibson (1951) das Problem von Hörfunk-Programm-Präferenzen (vgl. auch Lazarsfeld & Henry, 1968, S. 113-120) mit Hilfe des latenten Klassenmodells analysiert. In Miller, Sabagh & Dingman (1962) wird dieses Modell im Rahmen der medizinischen Diagnostik zur Klassifikation auf der Basis unterschiedlicher Mortalität verwendet, während Miller, Eyman & Dingman (1961) es zur Klassifikation von geistig zurückgebliebenen Patienten benutzen. Schließlich versucht Baker (1962) mit Hilfe des Modells das Bibliotheksproblem zu lösen, insbesondere die Klassifikation von Dokumenten, und Green, Carmone & Wachspress (1976) verwenden es, um Kundenverhalten bezüglich eines neu eingerichteten Fernsprechdienstes zu untersuchen.

Das erweiterte latente Klassenmodell mit polytomen Items benutzt Hutchinson (1977) für Rückschlüsse auf die latente Unfallgeschwindigkeit bei Verkehrsunfällen und Formann (1978b) zur Bildung von Klassen von Suizidenten unter Verwendung einer Untersuchung von Bayreuther.

Die latente Profilanalyse wird von Magnusson & Ekehammer (1975) zur Analyse von Angstprofilen, von Goddard (1971) zur Untersuchung des Kontaktverhaltens von Firmenpersonal und von Mardberg (1972, 1973) zur Selektion und Klassifikation von Personal in der Industriepsychologie verwandt. Von Jonassen erhobenes Material über Charakteristiken von Gemeinden wird in Lazarsfeld & Henry (1968, S. 233-235) analysiert.

Das latente Distanzmodell verwenden Stouffer & Toby (1951), um Rollenkonflikte zu analysieren (vgl. Lazarsfeld & Henry, 1968, S. 142-148), während das latente Inhaltsmodell von Somers (1961) bei der Messung sozialer Rigidität verwendet wird (vgl. Lazarsfeld & Henry, 1968, S. 185-191). In Katz & Procter (1959) werden Soziogramme von Schulkindern mit Hilfe eines Markovkettenansatzes analysiert. Eine Reanalyse dieser Daten mit Hilfe des latenten Markovkettenmodells führen Lazarsfeld & Henry (1968, S. 258-263) durch. Ein solches Modell wendet van der Sanden (1978) auf Spiele des Gefangen-Dilemma-Typs an.

Literatur

- Anderson, T. W. 1954. On estimation of parameters in latent structure analysis. *Psychometrika*, 19, 1-10.
- Anderson, T. W. 1959. Some scaling methods and estimation procedures in the latent class models. In U. Grenander (Hrsg.). *Probability and statistics*. New York: Wiley. 9-38.
- Baker, F. 1962. Information retrieval based on latent class analysis. *Journal of the Association for Computing Machinery*, 9, 512-521.
- Batchelder, W. H. & Narens, L. 1977. A critical examination of the analysis of dichotomous data. *Philosophy of Science*, 44, 113-135.
- Benini, R. 1928. Gruppi chiusi e gruppi aperti in alcuni fatti collettivi di combinazioni. *International Statistical Institute. Bulletin*, 23, 362-383.
- Brunk, H. D. 1958. On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, 29, 437-454.
- Cassady, J. M., Miller, C. R. & Dingman, H. F. 1968. Latent class analysis: A direct approach. *Proceedings, 76th Annual APA Convention*, 209-210.
- Cattell, R. B. 1944. „Parallel proportional profiles“ and other principles for determining the choice of factors by rotation. *Psychometrika*, 9, 267-283.
- Cattell, R. B. & Cattell, A. K. S. 1955. Factor rotation for proportional profiles: analytical Solution and an example. *British Journal of Statistical Psychology*, 8, 83-92.
- Coleman, J. S. 1964. *Models of change and response uncertainty*. Englewood Cliffs, N. J.: Prentice-Hall.
- Cournot, A. A. 1838. Mémoire sur les applications du calcul des chances à la statistique judiciaire. *Journal de mathématiques pures et appliquées*, 3, 257-334.
- Dwyer, P. 1937. The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 2, 173-178.
- Fielding, A. 1977. Latent structure models. In: C. A. O’Muircheartaigh & C. Payne (Hrsg.). *Exploring data structures*. New York: Wiley. 125-157.
- Fischer, G. H. 1974. *Einführung in die Theorie psychologischer Tests*. Stuttgart/Wien: Huber.
- Formann, A. K. 1978a. A note on parameter estimation for Lazarsfeld’s latent class analysis. *Psychometrika*, 43, 123-126.
- Formann, A. K. 1978b. The latent class analysis of polychotomous data. *Biometrical Journal*, 20, 755-771.
- Gebert, A. 1978. Assoziationsstrukturanalyse und latent-class-model. *Zeitschrift für experimentelle und angewandte Psychologie*, 25, 46-54.
- Gibson, W. A. 1951. Applications of the mathematics of multiple factor analysis to the problems of latent structure analysis. Unveröffentlichte Dissertation. University of Chicago.

- Gibson, W. A. 1955. An extension of Anderson's solution for the latent structure equations. *Psychometrika*, 20, 69-73.
- Gibson, W. A. 1956. Proportional profiles and latent structure. *Psychometrika*, 21, 135-144.
- Gibson, W. A. 1959. Three multivariate models: factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, 24, 229-252.
- Gibson, W. A. 1962a. Extending latent class solutions to other variables. *Psychometrika*, 27, 73-81.
- Gibson, W. A. 1962b. Class assignment in the latent profile model. *Journal of Applied Psychology*, 46, 399-400.
- Gill, R. D. 1976. The model of latent structure analysis. *Statistica Neerlandica*, 30, **143-149**.
- Gilula, Z. 1979. Singular value decomposition of probability matrices: Probabilistic aspects of latent dichotomous variables. *Biometrika*, 66, 339-344.
- Goddard, J. B. 1971. Office communications patterns in central London. In: South east economic planning council: Office linkages in central London. Vol. II.
- Good, I. J. 1969. Some applications of the singular decomposition of a matrix. *Technometrics*, 11, 823-833.
- Goodman, L. A. 1974a. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Goodman, L. A. 1974b. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - a modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.
- Goodman, L. A. 1975. A new model for scaling response patterns: An application of the quasi-independence concept. *Journal of the American Statistical Association*, 70, 755-768.
- Goodman, L. A. 1978. Analyzing qualitative/categorical data: Log-linear models and latent-structure analysis. London: Addison-Wesley.
- Goodman, L. A. 1979a. The analysis of qualitative variables using more parsimonious quasi-independence models, scaling models, and latent structures that fit the observed data. In: R. M. Merton, J. S. Coleman & P. H. Rossi (Hrsg.). *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld*. New York: Free Press.
- Goodman, L. A. 1979b. On the estimation of Parameters in latent structure analysis. *Psychometrika*, 44, 123-128.
- Green, B. F. Jr. 1951. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 16, 151-166.
- Green, B. F. Jr. 1952. Latent structure analysis and its relation to factor analysis. *Journal of the American Statistical Association*, 47, 71-76.
- Green, B. F. Jr. 1970. Parameter estimation in latent class analysis. Department of Psychology. Manuscript. Baltimore: John Hopkins University.

- Green, P. E., Carmone, F. J. & Wachspress, D. P. 1976. Consumer Segmentation via latent class analysis. *Journal of Consumer Research*, 3, 170-174.
- Haberman, S. J. 1974. Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. *The Annals of Statistics*, 2, **911-924**.
- Haberman, S. J. 1977. Product models for frequency tables involving indirect observation. *The Annals of Statistics*, 5, 1124-1147.
- Harper, D. 1972. Local dependence latent structure models. *Psychometrika*, 37, **53-59**.
- Hays, D. G. & Borgatta, E. F. 1954. An empirical comparison of restricted and general latent distance analysis. *Psychometrika*, 19, 271-279.
- Henry, N. W. 1966. BAN, a Fortran program. Cambridge: Computing Laboratory.
- Hutchinson, T. P. 1977. Latent structure models applied to the joint distribution of drivers' injuries in road accidents. *Statistica Neerlandica*, 31, 105-111.
- Kuzawa, M. 1968. Individual identification in latent class analysis: Application of canonical analysis and discriminant analysis. *Shinrigakuhyoron*, 11, 171-198.
- Isaacson, A. 1972. Latent profile analysis, a brief description. Technical Report No. 54. Stockholm: Royal Institute of Technology.
- Kadushin, C. 1966. The friends and supporters of psychotherapy. *American Sociological Review*, 31, 786-802.
- Katz, L. & Procter, C. 1959. The concept of configuration of interpersonal relations in a group as a time dependent stochastic process. *Psychometrika*, 24, 317-327.
- Koopmans, T. C. 1949. Identification problems in economic model construction. *Econometrica*, 17, 125-138.
- Koopmans, T. C. 1951. Identification problems in latent structure analysis. Manuscript. Cowles Commission Discussion Paper: Statistics, No. 360.
- Koopmans, T. C. & Reiersol, O. 1950. The identification of structural characteristics. *Annals of Mathematical Statistics*, 21, 165-181.
- Kulback, S. 1959. *Information theory and statistics*. New York: Wiley.
- Lazarsfeld, P. F. 1950a. The logical and mathematical foundations of latent structure analysis. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. H. Star & J. A. Clausen (Hrsg.). *Measurement and prediction*. Princeton: Princeton University Press. Chapter 10.
- Lazarsfeld, P. F. 1950b. Some latent structures. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. H. Star & J. A. Clausen (Hrsg.). *Measurement and prediction*. Princeton: Princeton University Press. Chapter 11.
- Lazarsfeld, P. F. 1954. A conceptual introduction to latent structure analysis. In: P. F. Lazarsfeld (Hrsg.). *Mathematical thinking in the social sciences*. New York: The Free Press. Chapter 7.
- Lazarsfeld, P. F. 1955. Recent developments in latent structure analysis. *Sociometry*, 18, 391-403.

- Lazarsfeld, P. F. 1959. Latent structure analysis. In: S. Koch (Hrsg.). *Psychology: A study of a science*. New York: McGraw-Hill. Vol. 3, 476-535.
- Lazarsfeld, P. F. 1960. Latent structure analysis and test theory. In: H. Gulliksen & S. Messick. *Psychological scaling theory and applications*. New York: Wiley. Chapter 8.
- Lazarsfeld, P. F. 1961. The algebra of dichotomous systems. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. Stanford: Stanford University Press. **111-157.**
- Lazarsfeld, P. F. & Dudman, J. 1951. The general solution of the latent class case. In: P. F. Lazarsfeld (Hrsg.). *The use of mathematical models in the measurement of attitudes*. Paper 5. RAND Research Memorandum No. 455.
- Lazarsfeld, P. F. & Henry, N. W. 1968. *Latent structure analysis*. Boston: Houghton Mifflin.
- Lippe, P. von der. 1973. Zur Skalierung komplexer Variablen mit der „Latent Structure Analysis“. *Allgemeines Statistisches Archiv*, 57, 333-356.
- Lord, F. M. 1953. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 18, 57-77.
- Lord, F. M. & Novick, M. R. 1968. *Statistical theories of mental scores*. Reading, Massachusetts: Addison-Wesley.
- Madansky, A. 1958. Identification and estimation in latent class analysis. Unveröffentlichte Dissertation. University of Chicago.
- Madansky, A. 1959. Partitioning methods of latent class analysis. Report P 1644. Rand Corporation.
- Madansky, A. 1960. Determinantal methods in latent class analysis. *Psychometrika*, **25**, **183-198.**
- Madansky, A. 1968. Latent structure. In: *International Encyclopedia of the Social Sciences*. New York: The Macmillan Company. Vol. 9, S. 33-38.
- Magnusson, D. & Ekehammer, B. 1975. Anxiety profiles based on both situational and response factors. *Multivariate Behavioral Research*, 10, 27-43.
- Mardberg, B. 1971. L. P. A., a preliminary program for the solution of latent profile analysis, according to Green's method. Manuskript. Stockholm: University of Stockholm.
- Mardberg, B. 1972. Clustering jobs to general requirement profiles. Stockholm: Swedish Council for Personnel Administration.
- Mardberg, B. 1973. A model for selection and classification in industrial psychology. Supplement 19. *Psychological Laboratories*. University of Stockholm.
- Mardberg, B. 1975. LPA 2: A FORTRAN V Computer program for Green's Solution of latent class analysis applied to latent profile analysis. *Educational and Psychological Measurement*, 35, 163-166.
- McCarthy, P. J. 1951. A special review of „The American Soldier, Vol. IV“. *Psychometrika*, 16, 247-269.

- McDonald, R. P. 1962. A note on the derivation of the general latent class model. *Psychometrika*, 27, 203-206.
- McHugh, R. B. 1956. Efficient estimation and local identification in latent class analysis. *Psychometrika*, 21, 331-347.
- McHugh, R. B. 1958. Note on „Efficient estimation and local identification in latent class analysis“. *Psychometrika*, 23, 273-274.
- Meo, G. de. 1934. Su di alcuni indici atti a misurare l'attrazione matrimoniale in classificazioni dicotome. *Accademia delle Scienze Fisiche e Matematiche. Naples. Rendiconto*, 73, 62-77.
- Meredith, W. 1965. Some results based on a general stochastic model for mental tests. *Psychometrika*, 30, 419-440.
- Miller, C. R., Eyman, R. K. & Dingman, H. F. 1961. Factor analysis, latent structure analysis and mental typology. *The British Journal of Statistical Psychology*, 14, 29-34.
- Miller, C. R., Sabagh, H. & Dingman, H. F. 1962. Latent class analysis and differential mortality. *Journal of the American Statistical Association*, 57, 430-438.
- Procter, C. H. 1970. A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, 35, 73-78.
- Sanden, A. L. van der. 1978. Latent Markov chain analysis of a value conflict in prisoner's dilemma games. *The British Journal of Mathematical & Statistical Psychology*, 31, 126-143.
- Skene, A. M. 1978. Discrimination using latent structure models. *Compstat 1978. Proceedings in computational statistics. Wien: Physica-Verlag.* 199-204.
- Somers, R. H. 1961. Latent content model of latent structure. Unveröffentlichte Dissertation. Columbia University.
- Stouffer, S. A. & Toby, J. 1951. Role conflict and personality. *American Journal of Sociology*, 56, 395-406.
- Takane, Y. 1972. Estimation of the recruitment latent class by least squares methods. *Japanese Psychological Research*, 14, 87-102.
- Takane, Y. 1976. A statistical procedure for the latent profile model. *Japanese Psychological Research*, 18, 82-90.
- Thurstone, L. L. 1925. A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451.
- Torgerson, W. S. 1958. *Theory and methods of scaling.* New York: Wiley.
- Tucker, L. R. 1952. A level of proficiency scale for a unidimensional skill. *American Psychologist*, 7, 408.
- Waerden, B. L. van der. 1971. *Mathematische Statistik.* Berlin: Springer.
- Weinberg, W. 1902. Beiträge zur Psychologie und Pathologie der Mehrlingsgeburten beim Menschen. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 88, 346-430.

- Wiggins, L. 1973. Latent probabilities for attitude and behaviour processes. Amsterdam: Elsevier.
- Wilcox, R. R. 1979a. An alternative interpretation of three stability models. *Educational and Psychological Measurement*, **39**, 311-315.
- Wilcox, R. R. 1979b. Achievement tests and latent structure models. *British Journal of Mathematical and Statistical Psychology*, **32**, 61-71.
- Wottawa, H. 1979. Grundlagen und Probleme von Dimensionen in der Psychologie. Meisenheim: Hain.