

Aus dem Institut für Psychologie der Universität Kiel

## Multinomiale und Bayes-statistische konfigurale Klassifikation

VON WILFRIED HOMMERS, Kiel

### Zusammenfassung, Summary, Résumé

Ein Verfahren zur Klassifikation von Pbn aufgrund individueller Abweichung von der Annahme wiederholter multinomialer Zufallereignisse und aufgrund individueller maximaler Likelihood der Zugehörigkeit zu einer bestimmten Gruppe wird nebst Modelltest dargelegt. Als Anwendungsbeispiel wird über die Ergebnisse bei zwei denkbaren Klassifikationen in verschiedenen Therapiemotivationsgruppen berichtet.

### Multi-nomial and Bayes-statistical configural classification

The author suggests a method and a "model test" for the classification of subjects on grounds on individual deviation from the assumption of repeated multi-nomial chance encounters and on grounds of individual maximal likelihood of membership in a certain group. The author reports on the results reached with two conceivable classifications into different therapy-motivation groups. (L. Canders)

### Classification configurale multinomique et conforme aux statistiques de Bayes

L'auteur présente – avec test-modèle – un procédé permettant de classer les sujets selon leur tendance individuelle à s'écarter de la croyance aux événements fortuits multinomiques répétés et selon la likelihood maximale individuelle d'appartenance à un groupe précis. En guise d'application, on relate les résultats dans le cas de deux classifications possibles dans deux groupes différents de motivation thérapeutique. (J. Chanel)

Im folgenden wird ein Verfahren zur Klassifikation von Pbn aufgrund von individuell wiederholt vorliegender mehrdimensionaler Information dargestellt. Es verbindet die Zielsetzung der Konfigurationsfrequenzanalyse (KRAUTH und LIENERT 1973) und die Methoden der Bayes-Statistik<sup>1</sup> mit dem Konzept der multinomialen Zufallsverteilung von mehrkategorialen Ereignissen.

1 Reulicke und Rollett (1976) haben über eine eindimensionale Anwendung der Bayes-Statistik in der Diagnostik berichtet.

## Methode

## Definitionen und Zielsetzung

Benutzt werde ein Variablensatz  $x_1, \dots, x_m$ , von dem über eine Vp  $v$   $n_v$  Realisationen in Meßwertvektoren  $X_j = (x_{1j}, \dots, x_{mj})$  mit  $j = 1, \dots, n_v$  vorliegen. Eine Konfiguration  $K_i$  sei dann gegeben, wenn in einer bestimmten Teilmenge des Variablensatzes bei einer Vp  $v$  in einer Realisation  $j$  bestimmte Werte erscheinen:  $K_i = (x_{i1} = z_1, \dots, x_{i1} = z_1)$  mit  $1 \leq m$ .

Die Häufigkeit  $f_v(K_i) \leq n_v$  einer bestimmten Konfiguration  $K_i$  von einer Vp  $v$  soll nach Berücksichtigung statistischer Gesetzmäßigkeiten für die konfigurale Klassifikation der Vp ausschlaggebend sein. Dabei werden die Wahrscheinlichkeiten  $p(K_i)$  einer Konfiguration und die Häufigkeit  $n_v$  in Rechnung gestellt.

## Multinomiale konfigurale Klassifikation (MKK)

Von Interesse seien  $g-1$  Konfigurationen  $K_i$ , die sich gegenseitig ausschließen. Eine  $g$ -te Konfiguration sei als Restklasse definiert. Dann gilt

$$(1) \quad \sum f_v(K_i) = n_v \quad \text{und} \quad \sum p(K_i) = 1 \quad \text{mit } i = 1, 2, \dots, g.$$

Damit läßt sich der Vektor  $D_v = (f_v(K_1), \dots, f_v(K_g))$  der  $g$  Konfigurationenhäufigkeiten einer Vp als Realisation eines multinomial gesteuerten Zufallsprozesses auffassen. Für jede Vp ergibt sich dann die exakte Wahrscheinlichkeit ihres Konfigurationenhäufigkeitsvektors  $p(D)$  als

$$(2) \quad p(D) = \frac{n_v!}{\prod f_v(K_i)!} \prod p(K_i)^{f_v(K_i)}$$

mit  $i = 1, 2, \dots, g$ .

Ebenso lassen sich die exakten Wahrscheinlichkeiten von Vektoren  $D$  bestimmen, deren Häufigkeiten an einer bestimmten Stelle  $k$  mindestens genauso groß sind wie die eines von einer Vp vorliegenden Vektors  $D_v$ .

Eine multinomiale konfigurale Klassifikation (MKK) aufgrund der  $g$  Konfigurationenhäufigkeiten wird vorgenommen, wenn die Summe der exakten Wahrscheinlichkeiten von Datenvektoren  $D$ , die an einer bestimmten Stelle  $k < g$  mindestens so groß wie das vorliegende  $f_v(K_k)$  einer betrachteten Vp sind, nicht größer als ein Wert  $\alpha$  (z. B. 5%) wird. Dann wird diese Vp der Gruppe  $k$  zugeordnet. Dieses Vorgehen entspricht einer zufällig zu erwartenden Klassifikationshäufigkeit  $N_{MKK}$  in die  $(g-1)$  ausgezeichneten Klassen von

$$(3) \quad N_{MKK} = (g-1) \cdot 0,05 \cdot N_{\text{ges}}$$

Schätzung der Wahrscheinlichkeiten der Konfigurationen.

Die Schätzung von  $p(K_i)$  für die MKK erfolgt durch Mittelung der Konfigurationenhäufigkeitsvektoren aller  $V_{pn}$  der Gesamtstichprobe.

### Modelltest

#### – zentraler Modelltest

Unter einem zentralen Modelltest wird die Überprüfung der Zufallsabweichung des Klassifikationsergebnisses von der Annahme eines einzigen multinomialen Prozesses mit den durch Mittelung gewonnenen Parameterschätzungen  $p(K_i)$  verstanden. Wenn die Anzahl der multinomial klassifizierten  $V_{pn}$   $N_{MKK}$  nicht signifikant von der Erwartung für  $N_{MKK}$  (siehe (3)) abweicht, muß angenommen werden, daß die vorkommenden, also auch die zur Klassifikation führenden Konfigurationenhäufigkeiten zufällige Ergebnisse dieses zentralen multinomialen Zufallsprozesses darstellen. Dies bedeutet, daß von der Klassifikation gemäß der zugrundeliegenden spezifizierten Teilmenge Abstand genommen werden muß.

Ist dieser Teil des Modelltests dagegen signifikant ausgefallen, kann angenommen werden, daß die vorhandene Stichprobe von  $N_{ges}$   $V_{pn}$  nicht aus einer Population mit den zentralen multinomialen Parametern  $p(K_i)$  stammt. Es ist dann die implizite Hypothese des vorläufigen Klassifikationsergebnisses zu prüfen, daß die vorliegende Stichprobe aus mehreren Populationen  $j$  mit verschiedenen multinomialen Parametern  $p^{(j)}(K_i)$  stammt.

#### – reproduzierender Modelltest

Unter einem reproduzierenden Modelltest wird die Überprüfung der möglichen Zusammensetzung der Gesamtstichprobe aus  $g$  Populationen mit verschiedenen Parametervektoren verstanden.

Damit diese Zerlegung in  $g$  Teilstichproben als valide gelten kann, muß die beobachtete Häufigkeitsverteilung  $F_o$  aller möglichen Konfigurationshäufigkeitsvektoren  $D$  aus  $g$  Multinomialprozessen mit jeweils  $N_{K_j}$  ( $\sum N_{K_j} = N_{ges}$ ) Individuen und zugehörigen bedingten Wahrscheinlichkeitsvektoren  $p(K_i/H_j) = p^{(j)}(K_i)$  herstellbar sein.

Sei  $N_j$  der zugehörige Anteil von  $V_{pn}$  der Teilpopulation  $j$ , die  $\sum f_v(K_i) = n_v$  haben, dann läßt sich dieser bestimmen als

$$(4) \quad N_j = \frac{N_{K_j}}{N_{ges}} \cdot N_{n_v} \quad \text{mit} \quad N_{n_v} = \sum N_j .$$

Die erwartete Häufigkeit  $F_e(D)$  eines Vektors  $D_v$  ergibt sich nun bei jeder der  $g$  Teilpopulationen zu

$$(5) \quad F_e(D) = N_j \cdot p(D/H_j),$$

wobei sich  $p(D/H_j)$  aus Gleichung (2) durch Einsatz von  $p(K_i/H_j)$  für  $p(K_i)$  berechnen läßt. Die Abweichung von der über  $j$  summierten Erwartung läßt sich mit  $\chi^2$  auf Signifikanz prüfen. Dies erfordert allerdings, wenn man keine Klassen, d. h. Konfigurationshäufigkeitsvektoren, zusammenfassen will, mit zunehmendem  $g$  und zunehmenden möglichen  $n_v$  eine hinreichende Gesamtzahl von  $V_{pn}$ .

Beim reproduzierenden Modelltest wird die Schätzung der  $g$  multinomialen bedingten Parametervektoren notwendig. Diese ließen sich mit Hilfe der durch MKK klassifizierten  $V_{pn}$  für die  $g-1$  hervorgehobenen Klassen bestimmen, indem die jeweiligen mittleren Konfigurationenhäufigkeiten der jeweils klassifizierten  $V_{pn}$  berechnet werden

$$(6) \quad p(K_i/H_j) = \frac{\sum_k f_v(K_i)_k}{N_{K_j}} \quad \text{mit } k = 1, 2, \dots, N_{K_j}.$$

Für die Restklasse lassen sich als bedingte Parametervektoren die ursprünglichen  $p(K_i)$  verwenden.

– *reduzierter reproduzierender Modelltest*

Da die  $F_e(D)$  nur bei hinreichend hohem  $N_{\text{ges}}$  die für einen anschließenden  $\chi^2$ -Test gewünschte Ausprägung  $F_e(D) \geq 5$  erlangen, wird folgender Ersatz als Annäherung vorgeschlagen.

Man reduziere die  $(g-1)$  Klassen auf eine Klasse  $k$ , die die größte Varianz der Wahrscheinlichkeiten über  $p(K_i/H_j)$  hat und bestimme dann  $F_e(D)$ , wie schon angegeben, mit Hilfe eines Binomial-Prozesses. Es liegen dann zur Erklärung der empirischen Verteilungshäufigkeiten in einer Matrix mit den Zeilen  $n_v$  und den Spalten  $f_v(K_k)$  (Anzahl von  $K_k$  bei einer  $V_p$  mit  $n_v$ )  $g$  Stichproben vor mit  $P_j$  als jeweiligen Wahrscheinlichkeitsvektor.

$$(7) \quad P_j = (p(K_k/H_j), p(K_{i \neq k}/H_j)).$$

Die erwartete Häufigkeit  $F_r(D)^2$  dieses reduzierten Modelltests an einer Stelle  $n_v$  und  $f(K_k)$  ergibt sich dann mit

$$(8) \quad F_r(D) = \sum_j N_{n_v} \frac{N_{K_j}}{N_{\text{ges}}} \binom{n_v}{f(K_k)} p(K_k/H_j)^{f(K_k)} p(K_{i \neq k}/H_j)^{n_v - f(K_k)}$$

Es kann insbesondere geprüft werden, ob die  $g$  angenommenen Populationen in den durch MKK geschätzten Relationen diese reduzierte Ver-

2 Leider ergeben sich u. U. auch hierbei erwartete Häufigkeiten unter 5. Durch Zusammenfassung von Zellen in der Verteilungsmatrix innerhalb einer Zeile nach einem einheitlichen Prinzip, so daß auch noch über die Zeilen eine Zusammenfassung möglich ist, wäre dies dann weiterhin umgehbar.

teilungsmatrix besser erklären als die Annahme einer einzigen zentralen Binomialverteilung (reduzierter zentraler Modelltest).

Ist die empirische Verteilung nicht aus der Annahme von  $g$  Populationen mit den durch  $N_{kj}$  geschätzten Größenrelationen reproduzierbar, lassen sich zwei Möglichkeiten des Weitergehens unterscheiden.

– Geht man begründet davon aus, daß nur die betrachteten Konfigurationen relevant sein können, bietet sich in der im folgenden beschriebenen Bayes-statistischen konfiguralen Klassifikation ein weiterführender Weg.

– Im Falle, daß die bis hierhin betrachteten Konfigurationen keinen Ausschließlichkeitscharakter haben, sozusagen nur heuristisch angenommen wurden, kann eine neue Zusammenstellung von Konfigurationen eventuell auch aus einer neuen Teilmenge zu einem neuen Vektor vorgeommen werden. Für diese beginnt die Klassifikationsprozedur dann von Neuem.

#### Bayes-statistische konfigurale Klassifikation (BKK)

Wenn die Überprüfung der Abweichung gefundener Häufigkeiten von der Erwartung im zentralen Modelltest zu einem signifikanten Ergebnis führte, läßt sich die zugrundeliegende Stichprobe als aus mehreren Teilstichproben zusammengesetzt auffassen.

Die  $g-1$  hervorgehobenen Gruppen von interessierenden Merkmalsträgern sind durch die MKK u.U. in unzureichendem Umfang erfaßt. Die dort angewendete klassisch-statistische Verfahrensweise kann nur dann eine klassifikatorische Zuordnung begründen, wenn die Irrtumswahrscheinlichkeit gering ist. Das Verfahren der MKK hat aber auch den Nachteil, daß erst von einer hinreichend großen Anzahl  $n_v$  an überhaupt klassifikatorische Zuordnungen möglich sind. Dies ist z. B. der Fall, wenn schon der extremste Konfigurationenhäufigkeitsvektor – z. B.  $(n_v, 0, \dots, 0)$  – eine höhere Wahrscheinlichkeit als  $p = .05$  hat. Erhöht man deswegen die Irrtumswahrscheinlichkeit, ergibt sich eine weitere unverwünschte Eigenschaft, das Vorkommen von Doppelklassifikationen (z. B. wenn Vektoren  $(\frac{n_v}{2}, 0, \frac{n_v}{2}, 0, \dots, 0)$  vorkommen). Alles dies läßt sich vermeiden, wenn statt eines klassisch-statistischen Ansatzes ein Bayes-statistischer Ansatz gewählt wird. Dieser Ansatz kann auch im Falle des Scheiterns des reproduzierenden Modelltests für MKK verwendet werden. Dabei geht man davon aus, daß der Grund für das Scheitern dieses Modelltests in falscher Schätzung der Populationenrelationen beruhte.

Die Zugehörigkeit einer  $V_p$  zu einer bestimmten Gruppe  $k$  wird im Bayes-Ansatz als Gültigkeit einer bestimmten Hypothese  $H_k$  beschrieben. Die Wahrscheinlichkeit eines Konfigurationenhäufigkeitsvektors  $D$ , wenn

eine Hypothese gilt, sei wie bisher definiert durch einen multinomialen Zufallsprozeß

$$(9) \quad L_k = p(D_v/H_k) = \frac{n_v!}{\prod f_v(K_i)!} \prod p(K_i/H_k)^{f_v(K_i)}$$

mit  $i, k = 1, 2, \dots, g$ .

Die Wahrscheinlichkeit einer Konfiguration ist nun aber für jede der  $g$  Gruppen spezifisch als bedingte Wahrscheinlichkeit angegeben. Diese bedingten Wahrscheinlichkeiten werden wie beim reproduzierenden Modelltest aus den durch die MKK definierten Gruppen durch Mittelung der Vektoren  $D$  geschätzt. Die bedingte Datenwahrscheinlichkeit unter einer Hypothese wird als Likelihood  $L_k$  bezeichnet. Die Klassifikation in eine Gruppe erfolgt nun in die Gruppe, deren Likelihood maximal unter den  $g$  Hypothesen ist. Bei diesem Vorgehen wird ein spezifischer Fall des allgemeinen Bayes-Theorems für gleichwahrscheinliche Hypothesen angewendet:

$$(10) \quad P(H_k/D) = \frac{P(D/H_k) P(H_k)}{\sum P(D/H_j) P(H_j)}$$

Im Falle, daß etwas über die Grundraten der einzelnen Gruppen bekannt ist, ließe sich aufgrund dieses allgemeinen Ansatzes auch mit variierenden  $P(H_j)$  die maximale Hypothesenwahrscheinlichkeit bestimmen. Darauf wird hier aber verzichtet.

Als Modelltest ist an diese Klassifikation der reproduzierende Modelltest anzuschließen, wobei die sich mit der BKK ergebenden  $N_{K_j}$  die Populationsrelationen definieren.

### Anwendungsbeispiel

Das hier zur Demonstration verwendete Datenmaterial zielte darauf ab, die Therapiemotivation von potentiellen Klienten zu erfassen. Eine eingehende Darstellung des Problembereichs gibt STELLER 1974. Das Datenerhebungsverfahren ist bei STELLER und HOMMERS 1977 ausführlich beschrieben. Da im Rahmen dieser Arbeit nur die statistische Methode Gegenstand der Darstellung ist, wird auf eine ausführliche inhaltliche Erörterung verzichtet.

Von jeder  $V_p$  sind  $n_v$  Probleme als auf sie zutreffend bezeichnet worden. Diese sind durch Antworten auf jeweils vier Zusatzfragen näher gekennzeichnet ( $m = 4$ ):

$z_1$ : „Wie stark beschäftigt Sie das Problem?“

Antworten: „Gar nicht“ bis „Stark“ (INT = 0,1,2,3).

$z_2$ : „Wer hat am meisten Schuld daran?“

Antworten: LD = „Ich selbst“ und UZ = „Andere Personen oder Umstände“

$z_3$ : „Wer muß sich zuerst ändern?“

Antworten: ÄW = „Ich selbst“ und HW = „Andere Personen oder Umstände“

$z_4$ : „Glauben Sie daran, daß es eine Möglichkeit gibt, an dieser Sache etwas zu ändern?“

Antworten: „Gar nicht“ bis „Stark“ (EE = 0,1,2,3).

Diese Zusatzfragen bilden die Gesamtmenge von Variablen. Betrachtet werden im folgenden zwei Teilmengen des Variablensatzes aus ( $z_2$ ,  $z_3$ ) und ( $z_2$ ,  $z_3$ ,  $z_4$ ), die aufgrund von Literaturberichten über günstige und ungünstige Therapie-Voraussetzungen ausgewählt wurden.

Teilmenge aus  $z_2$  und  $z_3$

Die  $g-1 = 2$  im folgenden benannten Konfigurationen sollen zur Klassifikation herangezogen werden:

$K_1$ :  $z_2 = LD$  und  $z_3 = AW$  sowie

$K_2$ :  $z_2 = UZ$  und  $z_3 = HW$ .

Außerdem existiert die Restklasse  $K_3 = No$ .

Als Wahrscheinlichkeiten dieser Konfigurationen ergaben sich  $p(K_1) = .51$ ,  $p(K_2) = .30$  und  $p(K_3) = .19$ . Diese wurden aus der Gesamtstichprobe von 211 jugendlichen Inhaftierten und 207 Berufsschülern geschätzt ( $N_{ges} = 418$ ). In den beiden Teilstichproben unterschieden sich diese Werte nur geringfügig.

Ein Datenvektor  $D_v$  bei  $n_v = 6$  Problemen wäre z. B.  $D_v = (5,0,1)$ . Er besagt, daß die  $V_p$  5 ihrer Probleme mit „Ich bin schuld“ und „Ich muß mich ändern“ kennzeichnete. Die  $V_p$  beantwortete dagegen keine Zusatzfrage mit „Die anderen sind schuld“ und „Die anderen müssen sich ändern“.

Die Ergebnisse der MKK und BKK zeigt Tabelle 1. Als Schätzungen der bedingten Wahrscheinlichkeiten für die BKK und den reproduzierenden Modelltest wurden die im unteren Teil der Tabelle 1 aufgeführten Werte bestimmt.

Der reduzierte zentrale Modelltest für die MKK fiel mit  $\chi^2 = 86.35$  bei  $df = 2$  unter Zusammenfassung von Delinquenten und Berufsschülern höchst signifikant aus. Der reproduzierende Modelltest<sup>3</sup> bei Reduzierung

3 Die Modelltests wurden jeweils für den Bereich  $5 \leq n_v \leq 16$  durchgeführt. Die Freiheitsgrade ergeben sich aufgrund der angewendeten Klassenzusammenfassung innerhalb und über die 12 Zeilen der Verteilungsmatrix zu den 3 Klassen, die durch die Centile  $C_{5\%}$  und  $C_{95\%}$  begrenzt werden.

Tabelle 1

Ergebnisse der MKK und BKK unter Verwendung von spezifizierten Konfigurationen aus der Teilmenge  $z_2$  und  $z_3$ , sowie deren bedingte Konfigurationswahrscheinlichkeiten

		$N_{K_1}$	$N_{K_2}$	$N_{K_3}$
MKK	Berufsschüler	14	14	179
	Inhaftierte	40	30	141
BKK	Berufsschüler	80	39	88
	Inhaftierte	85	41	85
		$p(K_1/H_i)$	$p(K_2/H_i)$	$p(K_3/H_i)$
		$H_{K_1}$	.82	.09
		$H_{K_2}$	.20	.68
		$H_{K_3}$	.51	.30

auf die Konfiguration  $K_1$  fiel für die Populationen-Anteile der MKK mit  $\chi^2 = 12.8$  bei  $df = 2$  ebenso wie für die Anteile der BKK mit  $\chi^2 = 12.25$  bei  $df = 2$  signifikant aus. Deswegen erscheint die Berechtigung einer Erweiterung der Teilmenge auch nachträglich bestätigt.

Teilmenge aus  $z_2$ ,  $z_3$  und  $z_4$

Die  $g-1 = 4$  Konfigurationen sollen in ihren Häufigkeiten bei einer  $V_p$  bestimmt werden:

$K_1$ :  $z_2 = LD$  und  $z_3 = \ddot{A}W$  und  $z_4 = EE = 0,1$

$K_2$ :  $z_2 = LD$  und  $z_3 = \ddot{A}W$  und  $z_4 = EE = 2,3$

$K_3$ :  $z_2 = UZ$  und  $z_3 = HW$  und  $z_4 = EE = 0,1$

$K_4$ :  $z_2 = UZ$  und  $z_3 = HW$  und  $z_4 = EE = 2,3$

Außerdem existiert die Restklasse  $K_5 = No$ .

Als zentraler Wahrscheinlichkeitsvektor dieser Konfigurationen ergab sich  $P = (.14, .38, .14, .15, .19)$ . Ein Datenvektor bei  $n_v = 6$  Problemen wäre z.B.  $D_v = (1,4,0,0,1)$ . Er besagt, daß die meisten Probleme außer mit „Ich bin schuld“ und „Ich muß mich ändern“ mit starkem oder mittlerem Glauben an eine Änderungsmöglichkeit bezeichnet wurden.

Die Ergebnisse der MKK und BKK sowie die Schätzungen der bedingten Wahrscheinlichkeiten gibt Tabelle 2 an.

Der reduzierte zentrale Modelltest für die MKK fiel mit  $\chi^2 = 38,41$  bei  $df = 2$  unter Zusammenfassung der Inhaftierten und Berufsschüler höchst signifikant aus. Der reproduzierende Modelltest bei Reduzierung auf die varianzreichste Konfiguration  $K_2$  war für die MKK mit  $\chi^2 = 4,44$



Tabelle 2

Ergebnisse der MKK und BKK unter Verwendung von spezifizierten Konfigurationen aus der Teilmenge  $z_2$ ,  $z_3$  und  $z_4$ , sowie deren bedingte Konfigurationswahrscheinlichkeiten

MKK	$N_{K_1}$	$N_{K_2}$	$N_{K_3}$	$N_{K_4}$	$N_{K_5}$
Berufsschüler	7	11	7	14	168
Inhaftierte	26	22	28	21	114
BKK					
Berufsschüler	16	47	25	28	91
Inhaftierte	33	39	32	25	82
	$p(K_1/H_i)$	$p(K_2/H_i)$	$p(K_3/H_i)$	$p(K_4/H_i)$	$p(K_5/H_i)$
$H_{K_1}$	.48	.21	.12	.05	.14
$H_{K_2}$	.05	.80	.02	.06	.08
$H_{K_3}$	.16	.10	.47	.14	.13
$H_{K_4}$	.03	.22	.13	.51	.11
$H_{K_5}$	.14	.38	.14	.15	.19

bei  $df = 2$  nicht signifikant auf dem 5%-Niveau, während sich beim reproduzierenden Modelltest für BKK mit  $\chi^2 = 10.05$  bei  $df = 2$  ein signifikanter Wert ergab<sup>4</sup>.

Dies kann dahingehend interpretiert werden, daß das Ergebnis der MKK als Annahme von vier im Umfang durch die MKK definierten Thememotivationsgruppen ausreicht, die gefundene Abweichung von einer zentralen multinomialen Zufallsverteilung zu erklären. Wenngleich damit nicht ausgeschlossen ist, daß auch andere Konfigurationen sinnvoll sind, kann hier eine empirische Kreuzvalidierung dieser Gruppenbildung durch tatsächlich nachweisbare unterschiedliche Wirkung von Therapie angeschlossen.

Konsequenz dieses Ergebnisses ist weiterhin, daß für eine Zuweisung in eine Gruppe nicht nur ein bestimmtes Vorherrschen von Beantwortungen in den Zusatzfragen, sondern auch mindestens vier Probleme nötig waren. Bei weniger als vier Problemauswahlen konnte keine  $V_p$  in der MKK klassifiziert werden, weil keine  $V_p$  alle ihre Probleme einheitlich kennzeichnete. Von den 30  $V_{pn}$  mit vier Problemangaben konnte genau eine klassifiziert werden.

4 Die Zusammenfassung von Spalten- und Zeilen-Erwartungen erfolgte hier für  $4 < n_v < 16$ .

Die sich aufdrängende Hypothese, daß Vpn mit drei oder vier Problemen in der auf  $K_2$  reduzierten Verteilungsmatrix wie aufgrund der zentralen multinomialen Verteilung zu erwarten verteilt waren, wurde geprüft und konnte wegen nicht signifikanter Abweichungen der empirischen Verteilung beibehalten werden. Es ergeben sich für den zentralen Modelltest  $\chi^2 = 14.23$  und für den des BKK-Ergebnisses  $\chi^2 = 17.12$  bei jeweils  $df = 5$ .

#### Literatur

- KRAUTH, J., und G. A. LIENERT: Die Konfigurationsfrequenzanalyse. Freiburg 1973.
- REULICKE, W., und B. ROLLETT: Pädagogische Diagnostik und lernzielorientierte Tests. In: K. Pawlik (Hrsg.), *Diagnose der Diagnostik*. Stuttgart 1976.
- STELLER, M.: „Leidensdruck“ als Indikation für Sozialtherapie? Unveröff. Diss. Kiel 1974.
- STELLER, M., und W. HOMMERS: Zur Diagnose der Therapiemotivation durch konfigurale Klassifikation. *Diagnostica* 23, 266–280, 1977.

Dr. phil. Wilfried Hommers,  
Institut für Psychologie  
Neue Universität  
2400 Kiel