

# Alternative nicht parametrische Verfahren der multivariaten Analyse

Robert Holz (Wuppertal)

## Inhalt

1. Einleitung
2. Darstellung der Techniken
  - 2.1 Empirische Verteilungsfunktion, Quantile und Marktskalen
  - 2.2 Clusterverfahren
    - 2.2.1 Der allgemeine Ansatz
    - 2.2.2 Besonderheiten des Fuzzy C Means Clusterverfahrens
  - 2.3 Relationale Diskriminanzanalyse
    - 2.3.1 Segment-Schichtung
    - 2.3.2 Segment-Statistik
    - 2.3.3 Vage Klassifikation
    - 2.3.3 Mehrdimensionalität
    - 2.3.4 Verfahrensabgrenzung
      - 2.3.4.1 Copulas und t-Normen
      - 2.3.4.2 Bestimmtheit
      - 2.3.4.3 Korrelation
  - 2.4 Das Problem der impliziten Gewichte
    - 2.4.1 Vergleichbarkeit von Kennzahlen
    - 2.4.2 A-priori Bewertung von Kennzahlen
  - 2.5 Evolutorische Skalenbetrachtungen
  - 2.6 Zur Technik der multivariaten Analyse mittels Bayes-Regeln
    - 2.6.1 Bayes-Regeln
    - 2.6.2 Dempster-Shafer Updating

## Literatur

### *1. Einleitung*

Die folgenden Betrachtungen stellen alternative Verfahren der multivariaten Analyse bereit, die an erster Stelle eine transparentere Entscheidungsfindung unterstützen großteils aber auch angemessenere im Sinne von genaueren Verfahren darstellen.

Wesentlich betrachtet werden die Informationstechniken Fuzzy C Means Clusterung, relationale Diskriminanzanalyse, evolutorische Skalenbetrachtung und verschiedene nicht parametrische Verfahren der Ermittlung von A-posteriori- sowie gemeinsamen Verteilungen, hierbei wird unter anderem im Vergleich zu Rating-Techniken auch auf das Konzept der Copula eingegangen.

Die Vorteile relationaler Betrachtungen bei Datenbeständen, die häufig wesentlich auch fehlende Daten enthalten und die Möglichkeiten der a-priori Komplexitätsreduktion mittels OLAP-Techniken, die andererseits zur Einbindung verborgener Information relationale Betrachtungen notwendig machen, sind nur am Rande angesprochen.

Ziel der Darstellung ist besonders auch die Sensibilisierung für die mit der statistischen Erkenntnisgewinnung verbundenen Unschärfe-Probleme, die aus der Berücksichtigung sowohl der Anzahl der zur Verfügung stehenden Daten, wie auch der verzerrten metrischen Berücksichtigung der Daten resultieren.

## 2. Darstellung der Techniken

### 2.1. Empirische Verteilungsfunktion, Quantile und Marktskalen

Wir nennen

$$F_n: \mathbb{R} \rightarrow [0; 1] \quad \text{mit} \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{[x_i, \infty)}(x), \quad n \in \mathbb{N}$$

die **empirische Verteilungsfunktion** einer gegebenen Wertemenge  $X_n = \{x_1, \dots, x_n\}$ , wobei o.B.d.A.  $x_{i+1} \geq x_i$  für  $i=1, \dots, n-1$  sei.

Die empirische Verteilungsfunktion zählt also aus einer gegebenen Wertemenge die Elemente kleiner oder gleich einem gegebenem Wert und normiert die so erhaltene Anzahl bezüglich der Gesamtanzahl der nicht notwendig verschiedenen Werte der betrachteten Wertemenge.

Sei weiter  $X_k = \{x_1, \dots, x_k\}$ ,  $k \leq n$ , o.B.d.A.  $x_{l+1} \geq x_l$  für  $l=1, \dots, k-1$  die Teilmenge der verschiedenen  $x \in X_n$  und

$$p_l = \frac{|\{x \in X_n : x = x_l\}|}{n}, \quad l = 1, \dots, k$$

die relative Häufigkeit der verschiedenen  $x$  in  $X_n$ . Als  **$\alpha$ -Quantil**<sup>1</sup> wird dann allgemein der Wert

$$q(\alpha) = \inf\{x: F_n(x) \geq \alpha\}$$

bezeichnet. Sofern kein  $x_i$  der Wertemenge existiert für das  $F_n(x_i) = \alpha$  gilt, kann hiervon abweichend - so in der Quantilsfunktion unter Excel -

---

<sup>1</sup> Vgl. etwa Elpelt, Hartung: „Multivariate Statistik“, Oldenburg 1995, S. 33.

$$q(\alpha) = x_i + \frac{\alpha - \sum_{l=1}^i p_l}{p_{i+1}} (x_{i+1} - x_i), \quad \text{falls } \alpha \in \left( \sum_{l=1}^i p_l; \sum_{l=1}^{i+1} p_l \right)$$

festgelegt werden, womit die Quantile als Funktion von  $\alpha$  verstetigt werden.

Als **empirische Quantilsfunktion (oder Quantilsfunktion unter Excel) von  $F_n$**  bezeichnen wir dann die Abbildung

$$q : [0; 1] \rightarrow \mathbb{R} \quad \text{mit } q(\alpha) = \begin{cases} x_i \in X_k, & \text{falls } \alpha = \sum_{l=1}^i p_l \\ x_i + \frac{\alpha - \sum_{l=1}^i p_l}{p_{i+1}} (x_{i+1} - x_i), & \text{falls } \alpha \in \left( \sum_{l=1}^i p_l; \sum_{l=1}^{i+1} p_l \right) \\ x_1, & \text{falls } \alpha = 0 \end{cases}$$

wobei  $\sum_{l=1}^0 p_l := 0$  gelte.

Wir sprechen der Kürze halber dann auch einfach von der Quantilsfunktion sofern die zu Grunde liegende empirische Verteilungsfunktion ersichtlich ist.

Mit dem Begriff **Marktskala der Feinheit  $\delta$** ,  $\delta = 1/m$ ,  $m \leq n$ , wird weiter im Folgenden die an den äquidistanten Stützstellen  $0, \delta, 2\delta, 3\delta, \dots, m\delta$  betrachtete Quantilsfunktion verbunden.

Also beispielsweise für  $m=10$  die Anordnung

$$q(0) \leq q(0,1) \leq \dots \leq q(1).$$

Wegen des eher zufälligen Charakters sind die Randpunkte  $q(0)$  und  $q(1)$  der Quantilsfunktionen zumeist weniger von Interesse.

Unter der **Kupierung** einer (empirischen) Verteilungsfunktion  $F_n$  **zum Niveau  $\gamma$** , verstehen wir weiter die Transformation

$$\gamma q(\alpha) = \begin{cases} q(\gamma), & \text{falls } \alpha \leq \gamma \\ q(\alpha), & \text{falls } \gamma < \alpha < 1 - \gamma \\ q(1 - \gamma), & \text{falls } 1 - \gamma \leq \alpha \end{cases}$$

der Quantilsfunktion zu  $F_n$ . Wir sprechen von einer **links- bzw. rechtsseitigen Stützung zum Niveau  $\gamma$** , wenn die mit der Kupierung verbundene Transformation der Quantilsfunktion nur am unteren bzw. oberen Ende der Quantilsfunktion vorgenommen wird.

Die Quantilsfunktion stellt offensichtlich eine an Mengen von Objekten orientierte Einteilung im Gegensatz zu beispielsweise äquidistanten metrischen Einteilungen einer Wertemenge  $X_n = \{x_1, \dots, x_n\}$  her und eignet sich so besonders zur Betrachtung ordinal skalierten Wertemengen, wie sie auch durch Rating-Symbole gegeben sind.

Es werden im Folgenden sowohl mittels Zählmaßen Verfahren die sich ausschließlich auf die ordinale Struktur von Wertemengen beziehen betrachtet, so etwa die Rangkorrelation, wie auch Verfahren die sich auf die Metrik von Wertemengen stützen, wobei letzteres in empirischer Anwendung immer sowohl ein Berücksichtigen der Metrik wie implizit auch des Zählmaßes zur Folge hat, so bei den folgenden Clusterverfahren, die ausserdem eine metrische Skalierung der Merkmale voraussetzen.

Beliebige o.B.d.A aufsteigend geordnete sogenannte Qualitäten  $l_1 < l_2 < \dots < l_n$  einer Menge  $L$  mit  $n$  verschiedenen Qualitäten können aber ohne weiteres Wissen vermöge der Transformation

$$l_i \rightarrow \frac{2i-1}{2n}, \quad i = 1, \dots, n$$

äquidistant in das Intervall  $[0;1]$  abgebildet werden. Die getroffene Festlegung minimiert dann den maximalen Abstand eines beliebigen Punktes im Einheitsintervall mit  $1/2n$  zur naheliegenden Qualität. Sind die Qualitäten nicht als äquidistant anzunehmen, so können die Qualitäten bezüglich des dann notwendig vorhandenen Wissens im Einheitsintervall genauer abgebildet werden, womit auch ordinale Merkmale in den folgenden Metrik erwartenden Verfahren, verwendet werden können, was bezüglich der Überinterpretation der durch die Ordnung gegebenen Information zumindest zur Überprüfung der Ordnung sinnvoll Verwendung finden kann, sofern die Ordnung auf Teilinformationen begründet ist.

## 2.2. Clusterverfahren

Clusterverfahren bilden anhand gegebener Merkmale bzw. Kennzahlen  $K_i: U \rightarrow \mathbb{R}$  betrachteter Objekte  $u \in U$  Gruppen ähnlicher Objekte. Die Techniken hier setzen durch metrische Merkmale charakterisierte Objekte voraus, womit eine Ähnlichkeit der Objekte direkt mittels üblicher Abstandsmaße beurteilt werden kann. Es ist im folgenden weniger Ziel mögliche Optimierungsansätze zu diskutieren als die Vorteile besonders des FCM-Clusterverfahrens zur Visualisierung in multivariaten Problemen zu nutzen.

### 2.2.1 Ein allgemeiner Ansatz für Clusterungen mittels Distanzmaßen

Die Ähnlichkeit zweier Objekte  $u_1$  und  $u_2$  sei mittels einer Abstandsfunktion  $d(u_1, u_2)$  beurteilt. Geben wir weiter eine Gruppen- bzw. Clusteranzahl  $c$  vor und charakterisieren wir die Cluster durch Clustermittelpunkte  $v_1, \dots, v_c$  und legen wir außerdem die Zugehörigkeit eines Objektes  $u$  zu genau einem Cluster  $v$  mittels der Zugehörigkeitsfunktion

$$z(u, v) = \begin{cases} 1, & \text{falls } v \text{ das benachbarteste Cluster von } u \text{ ist} \\ 0, & \text{sonst} \end{cases}$$

fest<sup>2</sup>, so besteht der Clusteransatz in der Aufgabe, die Zugehörigkeiten  $z(.,.)$  und die  $c$  Clustermittelpunkte  $v_j$  einer Anzahl  $n$  betrachteter Objekte  $u_r$  so festzulegen, dass die Summe der Abstände der Objekte von ihren zugeordneten Clustermittelpunkten minimiert wird. D.h. die Zielfunktion

$$\sum_{r=1}^n \sum_{j=1}^c (z(u_r, v_j) d(u_r, v_j))^2,$$

ist bezüglich  $z(.,.)$  und  $v_1, \dots, v_c$  zu minimieren.

Sowohl die Objekte  $u$  wie auch die Clustermittelpunkte  $v$  sind als Vektoren gegeben und mehrdimensionale Abstandsmaße beruhen in der Zusammenführung der Abstände der Ausprägungen der einzelnen Merkmale im wesentlichen auf einer (ungewichteten) Addition der eindimensionalen Merkmals-Abstände, womit es großteils zur Vermeidung unbewußter Gewichtungen (vgl. auch Abschnitt 2.4 „implizite Gewichte“) notwendig wird sich auf transformierte normierte Merkmals- bzw. Kennzahlenausprägungen zu beziehen, um eine Vergleichbarkeit der Merkmale in der Abstandsfunktion zu erhalten, wie im Folgenden noch deutlicher wird.

---

<sup>2</sup> Sollte es mehrere benachbartste Clustermittelpunkte  $v$  von  $u$  geben, so sei zunächst eine beliebige Zuordnung zu einem der benachbartsten Mittelpunkte zugelassen.

Betrachten wir hier die Euklid-Norm und damit die Abstandsfunktion

$$d_2(u_r, v_j) := \sqrt{\sum_{i=1}^p (u_{r(i)} - v_{j(i)})^2},$$

wobei  $p$  die Dimension der Objekte also etwa die Anzahl berücksichtigter Kennzahlen oder Merkmale angibt. Der Mittelpunkt zu bildender Cluster kann dann als Vektor der Mittelwerte der Ausprägungen zugeordneter Objekte festgelegt werden:

$$v_j = \begin{pmatrix} \frac{\sum_{r=1}^n z(u_r, v_j) u_{r(1)}}{\sum_{r=1}^n z(u_r, v_j)} \\ \vdots \\ \frac{\sum_{r=1}^n z(u_r, v_j) u_{r(p)}}{\sum_{r=1}^n z(u_r, v_j)} \end{pmatrix}, \quad j=1, \dots, c, \quad c \in \{1, 2, 3, \dots\}$$

In der Literatur finden sich vielfältig Vorschläge sowohl zur Lösung des Optimierungsproblems als auch zur Festlegung einer geeigneten Clusteranzahl sowie auch zur Wahl und Bedeutung anderer Festlegungen der Mittelpunkte<sup>3</sup>. Die jeweils erzielten Ergebnisse sind nicht unabhängig von den eingeschlagenen Vorgehensweisen.

### 2.2.2 Besonderheiten des Fuzzy-C-Means Clusterverfahrens

Das sogenannte Fuzzy-C-Means Clusterproblem (FCM-Clusterproblem)<sup>4</sup> modifiziert das beschriebene Clusterproblem. So ist wesentlich die gleichzeitige Zuordnung der betrachteten Objekte zu mehreren Clustern zugelassen, wobei aber für die modifizierte Zugehörigkeitsfunktion  $\tilde{Z}(u, v)$  gefordert wird

$$\tilde{Z}(u, v) \in [0; 1] \quad \text{und} \quad \sum_{j=1}^c \tilde{Z}(u, v_j) = 1$$

Das dem allgemeinen Ansatz oben entsprechende FCM-Clusterproblem entspricht dann mit Verwendung der Euklid-Norm der Minimierung der Zielfunktion

<sup>3</sup> Vgl. etwa B. Elpelt, J. Hartung: „Multivariate Statistik“, Oldenburg 1995.

<sup>4</sup> Vgl zur FCM-Clusterung auch R.Holz: „Fuzzy Sets in der Tarifierung“, Shaker 1996 oder besser J.C. Bezdek: „Pattern recognition with fuzzy objective function algorithms“, Plenum Press, New York 1981.

$$\sum_{r=1}^n \sum_{j=1}^c \tilde{z}^2(u_r, \tilde{v}_j) d_2^2(u_r, \tilde{v}_j)$$

bezüglich  $\tilde{Z}(u, v)$  und modifizierten Clusterzentren  $\tilde{v}_1, \dots, \tilde{v}_c$ , unter der Nebenbedingung

$$\sum_{j=1}^c \tilde{z}(u, v_j) = 1,$$

bei beliebigem aber festem  $c > 1$ . Mit den modifizierten Clustermittelpunkten

$$\tilde{v}_j = \begin{pmatrix} \frac{\sum_{r=1}^n \tilde{z}^2(u_r, \tilde{v}_j) u_{r(1)}}{\sum_{r=1}^n \tilde{z}^2(u_r, \tilde{v}_j)} \\ \vdots \\ \frac{\sum_{r=1}^n \tilde{z}^2(u_r, \tilde{v}_j) u_{r(p)}}{\sum_{r=1}^n \tilde{z}^2(u_r, \tilde{v}_j)} \end{pmatrix}, \quad j=1, \dots, c \quad c \in \{1, 2, 3, \dots\}$$

stellen sich dann die optimalen Zugehörigkeitswerte in Abhängigkeit von  $\tilde{v}_1, \dots, \tilde{v}_c$  wie folgt dar:

$$\tilde{z}(u, \tilde{v}_i) = \frac{1/d_2^2(u, \tilde{v}_i)}{\sum_{j=1}^c 1/d_2^2(u, \tilde{v}_j)}, \quad i = 1, \dots, c, \quad u \in U.$$

Als Vorteile der FCM-Clusterproblematik gegenüber der originären Clusterproblematik zeigen sich, dass es beispielsweise mit dem sogenannten ISODATA-Algorithmus möglich ist, unabhängig von einer Strategie und selbst weitgehend unabhängig von der Festlegung der Ausgangsclustermittelpunkte ein lokales Extremum der Zielfunktion zu erhalten, wobei im ISODATA-Algorithmus im wesentlichen ausgehend von einer Anfangsbelegung der Clustermittelpunkte, sukzessive bis nur noch minimale Veränderungen zu beobachten sind, die Zugehörigkeiten und dann die Clustermittelpunkte neu berechnet werden.

Die genannten Vorteile beruhen auf der mit der FCM-Clusterung erzeugten größeren Information. Die verallgemeinerten Zugehörigkeitswerte  $\tilde{Z}(u, v)$  bilden eine Ordnung der Zugehörigkeit eines Objektes zu den verschiedenen Clustermittelpunkten und vernetzen so die Objekte mit allen Clustern, womit in jedem Iterationsschritt die gesamte Information erhalten bleibt und so die bei den üblichen agglomerativen oder divisiven Clusterstrategien bestehende Gefahr des „unbewußten Verlaufs“ zu Nebenoptima zumindest abgeschwächt ist.

Die verallgemeinerten Zugehörigkeiten verschiedener Objekte zu einem Cluster sind jedoch nicht unmittelbar vergleichbar<sup>5</sup>, weshalb wir im Folgenden bei der Betrachtung von Clustergewichten den genannten Vorteil zwar beim Auffinden der optimalen Clustermittelpunkte nutzen, die Objekte für die Gewichtsbildung aber jeweils einem Cluster mit höchster Zugehörigkeit streng zuordnen.

Die Zuordnung der Objekte zu allen Clustern bei der Festlegung der Clustermittelpunkte mildert darüber hinaus den Einfluß von untypischen Objekten an den Rändern ab. Die mit der Euklid-Norm neben der „Belohnung“ der Nähe zu Clustermittelpunkten verbundene „Bestrafung“ der bezüglich einzelner Komponenten entfernt liegenden Objekte - aufgrund der Quadrierung der Abstände - führt beim originären Clusterproblem zu starken Verfälschungen einzelner Clustermittelpunkte. Bei der FCM-Clusterung werden vermeintliche Ausreißer hingegen bei der Festlegung von Clustermittelpunkten  $\tilde{v}$  auf alle Cluster verteilt und dennoch dem nächstliegenden Cluster zugeordnet, was die FCM-Clusterung robust gegenüber Ausreißern macht.

Wir verwenden die FCM-Clusterung auch dazu - ähnlich dem Hilfsmittel der Korrelation - Zusammenhänge betrachteter Kennzahlen zu prüfen, indem wir - von den einzelnen Objekten abstrahierend - funktionale Zusammenhänge der Clustermittelpunkte betrachten. Hierzu berücksichtigen wir in der Abstandsfunktion bewußte Gewichtungen, um die Betrachtung aussagekräftiger in Bezug auf wesentliche Blickrichtungen gestalten zu können, d.h. wir verwenden die gegenüber oben modifizierte Abstandsfunktion

$$d_2(u_r, v_j) := \sqrt{\sum_{i=1}^p g_i (u_{r(i)} - v_{j(i)})^2}, \quad g_i \in \mathbb{R},$$

was allein schon die angesprochene und im Abschnitt 2.4 weiter begründete Vergleichbarkeit der Merkmale der Objekte erfordert. Die Realisierung des Verfahrens in der Datei Analyse.xls ermöglicht auch das Festhalten von Merkmalsausprägungen im Iterations-Prozess, womit zusätzlich spezielle Blickrichtungen erzwungen werden können.

---

<sup>5</sup> Vgl. ebenda R.Holz: „Fuzzy Sets ...“, Shaker 1996, S. 72ff

Zur Prüfung von Abhängigkeiten wird das Verfahren außerdem zur Clusterung von Randverteilungen und deren Copula verwendet.

### 2.3. Relationale Diskriminanzanalyse

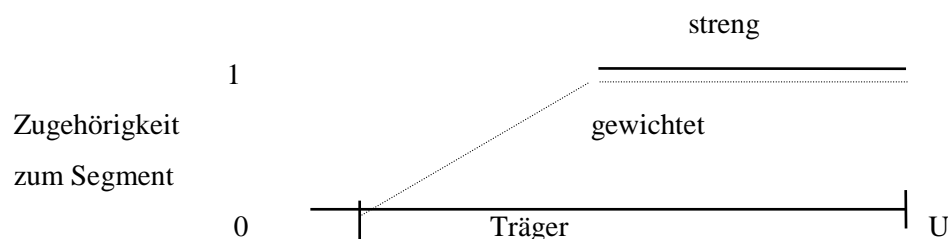
Die Besonderheit des Verfahrens ist einerseits die der Clusterung ähnliche Abstrahierung von funktionalen Zusammenhängen und besonders auch die effiziente Nutzung des vorhandenen Datenmaterials. Es werden gegensätzliche Objektesegmente betrachtet und innerhalb dieser Segmente die im Verhältnis zu allen Objekten unterschiedlichen Vorkommen der Merkmalsausprägungen in Bezug auf die die Segmente definierenden Merkmale betrachtet, um mittels Streuungen und quasi-funktionalen Zusammenhängen Einblicke in Abhängigkeiten zwischen den Merkmalen von Objekten zu gewinnen.

#### 2.3.1 Segment-Schichtung

Die vorhandenen Informationen zu gebildeten Segmenten werden geschichtet betrachtet, um auch innere Aussagen zu den betrachteten Segmenten erhalten zu können. Hierzu wird das jeweils betrachtete Segment in Objekte aufgeteilt, die ihm vollständig angehören und in solche die die Segmenteigenschaften mehr oder weniger erfüllen, dem sogenannten unschärfe oder fuzzy Bereich<sup>6</sup>. So kann mit einfachen Mitteln eine Segment-Statistik gebildet werden, die die an die Korrelation angelehnte **Blickrichtung**

„je mehr die Eigenschaften des Segmentes gegeben sind, desto ...“

für beliebige ein- oder mehrdimensionale Objekteausschnitte ermöglicht.



Die betrachteten Objekte  $u$  erhalten also insgesamt bezüglich der Segment definierenden (abhängigen) Merkmale einen Zugehörigkeitswert  $\mu(u) \in [0; 1]$  zu dem jeweils betrachteten Segment.

<sup>6</sup> Für eine Einordnung der Fuzzy Set Theorie beachte etwa R. Holz: „Rating, Ranking, Scoring und Fuzzy Sets“, Blätter der DGVM Heft 3/98, S. 363-384 oder auch R. Seising (Hrsg.): „Fuzzy Theorie und Stochastik“, Gabler 1999 und für die angesprochene Technik R. Holz: „Großstädte Ranking 2000“, Shaker-Verlag Aachen 2000.

### 2.3.2 Segment-Statistik

Für die **Segment-Statistik** werden dann sowohl in den Segment definierenden Merkmalen wie auch in den übrigen (erklärenden bzw. unabhängigen) Merkmalen

- mit dem **Träger** alle Objekte betrachtet, die dem Segment mit einer Zugehörigkeit  $\mu(u) > 0$  zugeordnet sind und deren Mittelwert  $m_1$  sowie deren Streuung  $s_1$  ermittelt.
- Weiter wird der bezüglich der Zugehörigkeiten **gewichtete** Mittelwert  $m_2$  betrachtet
- und schließlich der Mittelwert  $m_3$  und die Streuung  $s_3$  der Objekte, die dem Segment mit Zugehörigkeit  $\mu(u) = 1$  **streng** zugeordnet sind.

So läßt sich mittels verringerter Streuungen gegenüber der Gesamtstreuung  $s_0$  der jeweiligen Kennzahlwerte beurteilen ob Beobachtungen **charakteristisch** für das betrachtete Segment sind, während die Mittelwerte  $m_1$  bis  $m_3$  für sich und im Vergleich zum Mittelwert  $m_0$  aller Objekte eine eventuell vorhandene **Systematik** bezüglich der vorgegebenen Blickrichtung erkennbar machen. Wir nennen Mittelwerte  $m_1$  bis  $m_3$  **systematisch im Sinne der Blickrichtung**, wenn  $m_0 \leq m_1 \leq m_2 \leq m_3$  oder  $m_0 \geq m_1 \geq m_2 \geq m_3$  gilt.

Ein wesentlicher Vorteil der Segmentstatistik gegenüber beispielsweise Korrelationen ist – wie schon angedeutet – der wesentlich geringere Informationsverlust der Betrachtung. So wird es beispielsweise auch möglich über die Betrachtung der Veränderung eines abhängigen Merkmals als erklärendes Merkmal zu beurteilen in wie weit aktuell beobachtete Phänomene aus den Veränderungen resultieren und in soweit zufälligen Charakters sind.

### 2.3.3 Vage Klassifikation

Die Objekte  $u$  einer Gesamtmenge betrachteter Objekte  $U$  sind über Kennzahlen  $K_i$  bzw. über deren Ausprägungen  $K_i(u)$  mit

$$K_i : U \rightarrow \mathbb{R} \cup \{-\}, i = 1, \dots, p$$

charakterisiert, über die mittels sogenannter vager Konzepte  $k_i$  ( $k_i$  (Konzept):  $U \rightarrow [0; 1]$ ) im Sinne eines verallgemeinerten Wahrheitswertes Objekte mit speziellen Kennzahleigenschaften gefiltert werden

können. Mit ‘-‘ lassen wir wie üblich auch nicht vorhandene Werte zu. Etwa mit den abstrakten Eigenschaften hohe, normale und geringe Kennzahlwerte - genauer mit den vagen Konzepten  $k_{i(\text{hoch})}(u)$ ,  $k_{i(\text{normal})}(u)$  und  $k_{i(\text{gering})}(u)$  – können dann eindimensionale Objektsegmente gebildet werden.

In der aufgezeigten Weise festgelegte vage Konzepte bezeichnen wir als linguistische oder **vage Klassifikation der Kennzahl  $K_i$** .

Eine vage Klassifikation ist so weder mit einem Vollständigkeits-Kriterium<sup>7</sup> noch mit anderen objektivierenden Kriterien verbunden. So verwenden wir vage Klassifikationen im wesentlichen zur Erzielung ausreichender und ausgeglichener statistischer Basen für eine Diskriminanzanalyse.

In einer Diskriminanzanalyse im Sinne hier – deren Zielrichtung<sup>8</sup> vielfältig sein kann - werden also über filternde vage Konzepte gegensätzliche Segmente so gebildet, dass ihnen ausreichend und vergleichbar viele Objekte zugeordnet sind. Im Konzept hohe Kennzahlwerte erhalten beispielsweise die 20% höchsten Werte den Wahrheitswert 1 und die 20% nächst niedrigeren Werte linear zugeordnete Wahrheitswerte aus (0; 1). Analog wird mit dem Konzept gering verfahren. Wie in der Fuzzy Set Theorie allgemein üblich ist die Negation einer Eigenschaft  $k_{i(\text{Konzept})}(u)$  formell mit  $1 - k_{i(\text{Konzept})}(u)$  gegeben. Zumeist wird es dann nicht sinnvoll sein ein Konzept zusammen mit dessen formeller Negation in einer Diskriminanzanalyse zu betrachten.

Mit den mittels der Quantilsfunktion in Excel berechneten  $\alpha$ -Quantilen  $q_i(\alpha)$  der Kennzahlwerte  $\{K_i(u):u \in U\}$  ist also beispielsweise

$$k_{i(\text{hoch})}(u) = \begin{cases} 1, & \text{falls } K_i(u) \geq q_i(0,8) \\ \frac{K_i(u) - q_i(0,6)}{q_i(0,8) - q_i(0,6)}, & \text{falls } q_{0,6} < K_i(u) < q_i(0,8) \\ 0, & \text{sonst} \end{cases}$$

und

<sup>7</sup> Vollständigkeits-Begriffe können beispielsweise top down im Sinne einer vollständigen Berücksichtigung der betrachteten Objekte durch die Klassifikation oder bottom up durch eine vollständige Berücksichtigung relevanter Perspektiven zu Klassifikationen gebildet werden, woraus sich dann auch Qualitäten beispielsweise für Rating-Verfahren ableiten lassen.

<sup>8</sup> Zumeist ist die Analyse unterschiedlicher oder gemeinsamer Einflüsse auf die betrachteten Segmente das Ziel einer Diskriminanzanalyse. Ziel kann aber auch die auf höherem Abstraktions-Niveau durchgeführte Nachahmung bekannter sonstiger multivariater Verfahren sein, wobei mit dem höheren Abstraktions-Niveau besonders eine stabilere Modellbildung erreichbar sein kann und die Überprüfung komplexer mathematisch-statistischer Kalküle insbesondere bezüglich der Wirkung von Skalenverzerrungen stellt ein mögliches Ziel dar.

$$k_{i(\text{gering})}(u) = \begin{cases} 1, & \text{falls } K_i(u) \leq q_i(0,2) \\ \frac{q_i(0,4) - K_i(u)}{q_i(0,4) - q_i(0,2)}, & \text{falls } q_i(0,2) < K_i(u) < q_i(0,4) \\ 0, & \text{sonst} \end{cases}$$

Das Konzept normal ist für die hier betrachtete Technik zumeist weniger von Bedeutung. Mit einer in Bezug auf die Segment-Stärke den gewählten Konzepten hoch und gering vergleichbaren Festlegung der Form

$$k_{i(\text{normal})}(u) = \begin{cases} 1, & \text{falls } q_i(0,4) \leq K_i(u) \leq q_i(0,6) \\ \frac{K_i(u) - q_i(0,3)}{q_i(0,4) - q_i(0,3)}, & \text{falls } q_i(0,3) < K_i(u) < q_i(0,4) \\ \frac{q_i(0,7) - K_i(u)}{q_i(0,7) - q_i(0,6)}, & \text{falls } q_i(0,6) < K_i(u) < q_i(0,7) \\ 0, & \text{sonst} \end{cases}$$

ist die Segment-Statistik bezüglich der oben skizzierten Segment-Schichtung aber in gleicher Weise anwendbar. Auch muß das Konzept normal nicht zwingend symmetrisch sein<sup>9</sup>.

### 2.3.4 Mehrdimensionalität

Sofern die Segmente mittels mehrerer Kennzahlen gebildet werden, sind die eindimensionalen Konzepte also die Zugehörigkeitswerte der Objekte zu den betrachteten Kriterien geeignet zu Verknüpfen. Zu Bedeutung gelangen hier insbesondere die in [Holz 98]<sup>10</sup> betrachteten und motivierten sogenannten Durchschnittsoperatoren. Der Vollständigkeit halber sind die vorgestellten Operatoren hier nochmals aufgeführt.

#### **Definition:**

Eine Abbildung  $D: [0;1]^P \rightarrow [0;1]$  nennen wir **(p-stelligen) Durchschnittsoperator**,  $p \in \{1,2,\dots\}$ , falls gilt:

**Idempotenz:**  $D(a,\dots,a) = a$

<sup>9</sup> In den Versionen der die Techniken hier bereitstellenden Software des Autors ist das Konzept normal wegen des geringeren Parametrisierungsaufwandes symmetrisch konzipiert, über alternative Verknüpfungen sind aber auch nicht symmetrische zentrale Konzepte abbildbar.

<sup>10</sup> R. Holz: „Rating, Ranking, Scoring und Fuzzy Sets“, Blätter der DGVM Heft 3/98, 363-384.

**Monotonie:**  $D(a_1, \dots, a_p) \leq D(b_1, \dots, b_p)$  sofern  $a_i \leq b_i, i = 1, \dots, p$

Man weist leicht nach, dass  $\text{Min}(a_1, \dots, a_p)$  und  $\text{Max}(a_1, \dots, a_p)$  Durchschnittsoperatoren sind. Darüber hinaus gilt<sup>11</sup>

$$\text{Min}(a_1, \dots, a_p) \leq D(a_1, \dots, a_p) \leq \text{Max}(a_1, \dots, a_p)$$

für alle Durchschnittsoperatoren, was allein aus der Idempotenz und der Monotonie folgt. Andererseits folgt die Idempotenz aus der Eigenschaft der Durchschnittsoperatoren bezüglich  $\text{Min}(\cdot)$  und  $\text{Max}(\cdot)$  innere Operatoren zu sein.

Markante Familien von Durchschnittsoperatoren sind:

a) Der sogenannte **verallgemeinerte Mittelwert**,

$$D_\alpha(a_1, \dots, a_p) = \left( \frac{a_1^\alpha + \dots + a_p^\alpha}{p} \right)^{\frac{1}{\alpha}}, \alpha > 0.$$

Für  $\alpha=1$  erhalten wir so offensichtlich den arithmetischen Mittelwert der Zugehörigkeitswerte. Für  $\alpha=2$  erinnert dieser Durchschnittsoperator an die Euklid-Norm, womit deutlich wird, dass  $\alpha$  zur Betonung der Skalenpole verwendet werden kann.

b) Der sogenannte **Min-Max-Operator**,

$$D_{\text{min/max}}(a_1, \dots, a_p) \cong \gamma \text{Min}(a_1, \dots, a_p) + (1 - \gamma) \text{Max}(a_1, \dots, a_p), \gamma \in [0, 1].$$

Mit dem Min-Max-Operator können die zu aggregierenden Kriterien graduell zwischen den Extremen  $\text{Min}(\cdot)$  und  $\text{Max}(\cdot)$  abgestuft werden.

c) Der **gewichtete Durchschnitt**,

$$D_{\text{WA}}(a_1, \dots, a_p) = \sum_{i=1}^p w_i a_i, \sum_{i=1}^p w_i = 1, w_i \in (0, 1).$$

d) **Bereichsdurchschnitte**,

---

<sup>11</sup> G.Klir, B.Yuan: „Fuzzy Sets and Fuzzy Logic“, Prentice Hall 1995, 88f

In Anlehnung an den sogenannten  $\lambda$ -Durchschnitt<sup>12</sup> können Durchschnittsoperatoren betrachtet werden, welche die Eigenschaften verschiedener Durchschnittsoperatoren verbinden. Wir bezeichnen sie als **Bereichsdurchschnitte**  $D_B$  und definieren:

$$D_B(a_1, \dots, a_p) := \begin{cases} D_1(a_1, \dots, a_p) & \text{sofern } a_i \in [0; \lambda_1], i = 1, \dots, p \\ \lambda \max(\lambda_1, \min(a_1, \dots, a_p)) + (1 - \lambda) \min(\lambda_2, \max(a_1, \dots, a_p)), & \text{sonst} \\ D_2(a_1, \dots, a_p) & \text{sofern } a_i \in (\lambda_2; 1], i = 1, \dots, p \end{cases}$$

mit  $0 \leq \lambda_1 \leq \lambda_2 \leq 1$ ,  $\lambda \in [0; 1]$  und  $D_1, D_2$  beliebige Durchschnittsoperatoren.

Im Spezialfall zweier Merkmale ist mit  $0 = \lambda_1 = \lambda_2$  und  $D_1 \equiv 0$  sowie  $D_2 = D_{WA}$ , dann

$$D(a, b) = \begin{cases} \frac{a + b}{2}, & \text{falls } a, b \in (0; 1] \\ 0, & \text{sonst} \end{cases}$$

ein Bereichsdurchschnitt der verdeutlicht, dass insbesondere die Exklusivität der Durchschnittsbildung – Zugehörigkeit 0 für die nicht allen betrachteten Konzepten positiv zugeordneten Objekte – eine wesentliche Eigenschaft der Bereichsdurchschnitte ist, die sich besonders zur Bildung von abstrakten Relationen also etwa auch für ein regelbasiertes Rating anbietet.

### 2.3.4 Verfahrensabgrenzung

#### 2.3.4.1 Copulas und t-Normen

Zur Abgrenzung der Durchschnittsoperatoren oben seien im Folgenden weitere Operatoren betrachtet, die insbesondere auch einen Bezug zu stochastischen Interpretationen herstellen.

#### **Definitionen**<sup>13</sup>:

1. Eine Funktion  $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$  heißt **T-Norm**, falls gilt

<sup>12</sup> Vgl. etwa Klir, Yuan: „Fuzzy Sets and Fuzzy Logic“, Prentice Hall 1995, S.93ff.

<sup>13</sup> Vgl. auch R.Holz 1998: „Rating, Ranking, Scoring, ...“

- i.  $T(a,1) = a$  (Einselement)
- ii.  $a \leq b \Rightarrow T(a,c) \leq T(b,c)$  (Monotonie)
- iii.  $T(a,b) = T(b,a)$  (Kommutativität)
- iv.  $T(a,T(b,c)) = T(T(a,b),c)$  (Assoziativität).

Wir bezeichnen  $T$  auch als verallgemeinertes Und.

2. Eine Funktion  $S: [0,1] \times [0,1] \rightarrow [0,1]$  heißt **T-Conorm** (auch **S-Norm**), falls gilt

- i.  $S(a,0) = a$  (Nullelement)
- ii.  $a \leq b \Rightarrow S(a,c) \leq S(b,c)$  (Monotonie)
- iii.  $S(a,b) = S(b,a)$  (Kommutativität)
- iv.  $S(a,S(b,c)) = S(S(a,b),c)$  (Assoziativität).

Wir bezeichnen  $S$  auch als verallgemeinertes Oder.

### Bemerkungen + Definitionen:

1. Das Einselement der T-Normen bzw. das Nullelement der S-Normen gewährleisten offensichtlich, daß T-Normen bzw. S-Normen für die Wahrheitswerte 0 und 1 die Wahrheitstabellen des logischen Und bzw. Oder wiedergeben.
2. Wichtige Beispiele für T-Normen sind<sup>14</sup>:

$$\begin{aligned}
 T_{\min}(a,b) &= \min(a,b), \\
 T_{\text{prod}}(a,b) &= ab, \\
 T_{\text{w}}(a,b) &= \begin{cases} a, & \text{falls } b = 1 \\ b, & \text{falls } a = 1 \\ 0, & \text{sonst.} \end{cases}
 \end{aligned}$$

Ferner gilt,

$$T_{\text{w}}(a,b) \leq T(a,b) \leq T_{\min}(a,b),$$

für alle T-Normen  $T$ .

---

<sup>14</sup> Vgl. [Klir 95]

3. Wichtige Beispiele für S-Normen sind<sup>15</sup>:

$$\begin{aligned} S_{\min}(a, b) &= \max(a, b), \\ S_{\text{prod}}(a, b) &= a + b - ab, \\ S_w(a, b) &= \begin{cases} a, & \text{falls } b = 0 \\ b, & \text{falls } a = 0 \\ 1, & \text{sonst.} \end{cases} \end{aligned}$$

Analog zu 2. gilt,

$$S_{\min}(a, b) \leq S(a, b) \leq S_w(a, b),$$

für alle S-Normen S.

4. Begründet auf die Assoziativität der T-Normen bzw. T-Conormen verallgemeinern wir die 2-stelligen Verknüpfungen zu r-stelligen über  $[0, 1]^r$ ,  $r = 3, 4, \dots$ :

$$T_r(a_1, \dots, a_r) := \underbrace{T(\dots T(a_{r-2}, a_{r-1}), a_r) \dots}_{(r-1) \text{ mal}}, \quad a_i \in [0, 1], \quad i = 1, \dots, r.$$

Analog für  $S_r$ . Außerdem setzen wir  $T_1(a) := S_1(a) := a$ ,  $a \in [0, 1]$ .

### Definition:<sup>16</sup>

Eine Funktion  $C: [0, 1]^r \rightarrow [0, 1]$  heißt **Copula**, falls gilt:

1.  $C(x_1, \dots, x_r) = 0$ , falls  $x_i = 0$  für mindestens ein  $i = 1, \dots, r$
2.  $C(x_1, \dots, x_r) = c$ , falls  $x_j = c$  für genau ein  $j = 1, \dots, r$  und  $x_i = 1$  für  $j \neq i = 1, \dots, r$

$$3. \quad \forall a \leq b \text{ in } [0, 1]^n \text{ ist } \sum_{(\varepsilon_1, \dots, \varepsilon_n) \in \{0, 1\}^n} (-1)^{\sum_{i=1}^n \varepsilon_i} C(\varepsilon_1 a_1 + (1 - \varepsilon_1) b_1, \dots, \varepsilon_n a_n + (1 - \varepsilon_n) b_n) \geq 0$$

### Bemerkungen + Definitionen<sup>17</sup>:

1. Für Copulas gilt:

---

<sup>15</sup> Vgl. ebenda

<sup>16</sup> Vgl. Pfeifer 2003

<sup>17</sup> Vgl. Nelsen 1999 und Pfeifer 2003

$$\max(x_1, \dots, x_n - n + 1, 0) \leq C(x_1, \dots, x_n) \leq \min(x_1, \dots, x_n)$$

und offensichtlich existieren Copulas die auch t-Normen sind, womit dann die untere Grenze aus der Verallgemeinerung von  $T_w$  auf  $[0,1]^n$  im obigen Sinne resultiert.

2. Seien  $X_i, i=1, \dots, n$  stetige Zufallsvariablen mit Verteilungsfunktionen  $F_i$ , dann sind die  $X_i$  stochastisch unabhängig genau dann, wenn

$$C(F_1(x_1), \dots, F_n(x_n)) = \prod_{i=1}^n F_i(x_i)$$

gilt. Die so definierte Copula wird auch als Unabhängigkeits-Copula bezeichnet.

3. Copulas sind ein noch junges Forschungsgebiet zur Modellierung von Abhängigkeiten. Wir verwenden die Clusterung der Quantilsränge von Merkmalen gemeinsam mit der Unabhängigkeits-Copula dazu ähnlich dem Konzept der Rangkorrelation, die Annahme stochastischer Unabhängigkeit zu überprüfen bzw. Einsicht in deren gemeinsame Verteilung zu erhalten. Weitere Copulas finden sich beispielsweise bei Pfeifer<sup>18</sup>.
4. Die nicht gegebene Idempotenz der t-Normen macht ausserdem bereits deutlich, dass statistische Kenngrößen die sich empirisch sowohl auf das Zählmaß also auf die Anzahl berücksichtigter Beobachtungen wie auch auf die Metrik der Beobachtungen beziehen und auf dem Produkt-Operator beruhen Ungleichgewichtungen in der Bedeutung gegenüber dem Zählmaß hervorrufen. Dies ist insbesondere beim Pearsonschen Korrelationskoeffizienten der Fall, wie auch im Vergleich zur Rangkorrelation noch deutlicher wird.

#### 2.3.4.2 Bestimmtheit

Eine dem quadrierten Pearsonschen Korrelationskoeffizienten angelehnte Prüfung der Stärke von Diskriminierung mittels Trennmaßen<sup>19</sup>, ist einerseits wegen der damit verbundenen zumeist nicht notwendigen weiteren Informationsreduzierung in der Segmentstatistik oben nicht betrachtet. Andererseits ist ein solches Maß sofern es normierte Vergleichbarkeit erzeugen soll nicht trivial naheliegend. Für eine relationale Modellauswahl etwa in Bezug auf besonders diskriminierende Merkmale ist ein solches Bestimmtheitsmaß jedoch notwendig.

<sup>18</sup> Vgl. Pfeifer 2003: „Modeling Dependence ..“

<sup>19</sup> Vgl. etwa Elpelt, Hartung: „Multivariate Statistik“, Oldenburg 1995, S. 145f sowie 251ff

Mit  $a \in s \vee m$  und (vgl. Abschnitt 2.3.1 „Segment-Schichtung“) den geschichteten Mittelwerten  $m_{kji}$ ,  $k=1,2,3$  sowie den ebensolchen Streuungen  $s_{kji}$ ,  $k=4,5$  bezüglich zweier Segmente  $j=1,2$  und Merkmalen  $i$  bietet sich für den Vergleich der Stärke von Diskriminierung in Bezug auf die einzelne Kenngröße  $a$  das Maß

$$r_{ki} = \begin{cases} \frac{|a_{k1i} - a_{k2i}|}{\max\{|a_{k1i}|, |a_{k2i}|\}}, & \text{falls nicht } a_{k1i} = 0 \text{ und } a_{k2i} = 0 \\ 0, & \text{sonst} \end{cases}$$

aber direkt an. Die Wahl des Nenners stellt dann eine Möglichkeit dar, die absolute Abweichung des Zählers nicht durch kleine Bezugsgrößen in ihrer relativen Bedeutung zu überzeichnen. Abhängig vom Ziel einer Modellbildung sind für eine Auswahl relational diskriminierender Kennzahlen  $i$  die zugehörigen Kenngrößen dann noch geeignet zu verknüpfen.

#### 2.3.4.3 Korrelation

Zur Beurteilung sachlicher Zusammenhänge zwischen Kennzahlen bzw. Merkmalen werden überwiegend Korrelationen insbesondere bei vermuteten linearen Zusammenhängen herangezogen. Der Pearsonsche Korrelationskoeffizient als Maß der Güte eines solchen Zusammenhanges wird auch Produktmomentenkorrelation genannt<sup>20</sup>, womit wesentliche Kritikpunkte bezüglich der mit der genannten Korrelation gemessenen Zusammenhänge schon deutlich werden.

Zum einen ist die Momenten-Methode eine von speziellen Verteilungen stark vereinfachende Methode der Erkenntnisgewinnung und zum anderen ist das Produkt als Aggregator von Teilinformationen ein stark verzerrender Aggregator bezüglich der Metrik der Teilräume.

Bezüglich zweier Zufallsvariablen  $X, Y$  wird die sogenannte empirische oder Pearsonsche Korrelation  $P$  wie folgt berechnet:

$$P_{X,Y} = \frac{\sum_{x \in X} \left( x - \frac{1}{|X|} \sum_{x \in X} x \right) \sum_{y \in Y} \left( y - \frac{1}{|Y|} \sum_{y \in Y} y \right)}{\sqrt{\sum_{x \in X} \left( x - \frac{1}{|X|} \sum_{x \in X} x \right)^2} \sqrt{\sum_{y \in Y} \left( y - \frac{1}{|Y|} \sum_{y \in Y} y \right)^2}}.$$

<sup>20</sup> Vgl. ebenda

P ist dann, wie man leicht nachprüft, invariant bezüglich multiplikativer Veränderungen von X und Y (Skalenparameter) wie auch gegenüber der Addition von Konstanten (Lageparameter). Erheblichen Einfluß besonders bezüglich der Skalen verzerrenden Wirkung der Multiplikation hat hingegen die Verteilung der Kennzahlenausprägungen.

Vergleichen wir P mit den im Zusammenhang mit der Clusterung betrachteten Normen, so tritt im Zähler von P an die Stelle der Addition von Abständen verschiedener Teilräume die Multiplikation dieser. Die Korrelation P ist dann bezüglich des Produktes der Standardabweichungen normiert und summiert im Zähler die multiplizierten gezählten Abstände von den nicht notwendig zentralen Durchschnitten der betrachteten Kennzahlen.

Die Multiplikation der Abstände im Vergleich zur Addition der Abstände in den Normen hat dann zwar den Effekt, dass sich einseitig gemessene geringe Abstände im Ergebnis zugunsten geringer Korrelationen auswirken. Andererseits werden aber nahe an den Mittelpunkten liegende betragsmäßig übereinstimmende Abstände im Vergleich zu entfernt liegenden betragsmäßig übereinstimmenden Abständen stark unterbewertet, womit Korrelationen fast ausschließlich von den entfernt liegenden Werten bestimmt sind, die die Anzahl der Beobachtungspunkte in ihrer Bedeutung unwesentlicher werden lassen. Dies zeigt auch der Vergleich zur Rangkorrelation<sup>21</sup> im Sinne Kendall's  $\tau$ , das sich als folgende Differenz von Wahrscheinlichkeiten ermitteln lässt:

Seien  $(X_1, Y_1)$  und  $(X_2, Y_2)$  zwei unabhängig und identisch verteilte Zufallsvektoren, dann ist

$$\tau_{X,Y} := \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

und es gelten die Beziehungen

$$\frac{3\tau - 1}{2} \leq P \leq \frac{1 + 2\tau - \tau^2}{2}, \tau \geq 0$$

$$\frac{\tau^2 + 2\tau - 1}{2} \leq P \leq \frac{1 + 3\tau}{2}, \tau \leq 0$$

sowie

$$-1 \leq 3\tau - 2P \leq 1$$

Die also in der Korrelation enthaltene wesentliche Orientierung an den Rändern der Skalen findet sich in der oben beschriebenen relationalen Diskriminanzanalyse als Verfahrens bildendes Element wieder, kann mit dem Verfahren oben aber auch auf innere Bereiche der Verteilungen angewendet werden, wobei die Betrachtung außerdem mit weniger Informationsverlust und damit exakter vorgenommen

---

<sup>21</sup> Vgl. etwa Nelsen 1999

wird. Insbesondere die bei Normalverteilungen wesentlich um den Erwartungswert konzentrierte Masse kann so auch entscheidungsrelevant verwendet werden.

## 2.4 Das Problem der impliziten Gewichte

In Holz 1998<sup>22</sup> ist unter anderem ausgeführt, dass nahezu alle in der Finanzpresse zu findenden Rankings bzw. Ratings direkt oder indirekt auf die Methode der gewichteten Durchschnittsbildung zurückführbar sind. Es wird dann davon ausgegangen, dass sich die Rating-Objekte  $u \in U$  durch gemeinsame Kennzahlen  $K_i: U \rightarrow \mathbb{R}$  charakterisieren lassen,  $U$  die Gesamtheit der betrachteten Objekte.

Mit transformierten Kennzahlausprägungen  $T_1(K_1(u)), \dots, T_p(K_p(u)), p \in \{1, 2, \dots\}$  ist ein auf einer gewichteten Durchschnittsbildung beruhendes Ratingverfahren  $r_p$  dann wie folgt gegeben:

$$r_p(u) = \sum_{i=1}^p g_i T_i(K_i(u)), \quad g_i \in [0; 1], \quad \sum_{i=1}^p g_i = 1.$$

Mittels  $r_p$  ließe sich auch der Pearsonsche Korrelationskoeffizient darstellen sofern wir die Summe  $\Sigma$  durch das Produkt  $\Pi$  ersetzen würden und  $r_p$  über  $U$  summierten. Offensichtlich kann eine Variation der Gewichte  $g$  dann unterschiedliche Bedeutungen der betrachteten Objekte  $u$  für statistische Kenngrößen oder einfach für das Rating<sup>23</sup> der Objekte erzeugen. Darüber hinaus wirkt die Verteilung der Kennzahlwerte selbst sowie die Art der Verknüpfung (hier die gewichtete Addition) der verschiedenen Kennzahlwerte eines Objektes auf statistische Kenngrößen und auf Ratings.

So wurde an genannter Stelle ausgeführt, dass für eine bewußte Gewichtsbildung anhand der Gewichte  $g_i$  in den Transformationen  $T_i$  eine Normierung der Kennzahlausprägungen vorzunehmen ist, um unbewußte implizite Gewichte zu vermeiden. Mit einem intuitiveren Zugang sei hier nochmals der Zusammenhang dargestellt.

### 2.4.1 Vergleichbarkeit von Kennzahlen

Betrachten wir mit  $U$  beispielshalber 5 Unternehmen, die bezüglich der gerichteten Kostenquote<sup>24</sup>  $K(u) = 100 - K^o(u)$  und bezüglich der Nettoverzinsung  $Z(u)$  die Ausprägungen

<sup>22</sup> Vgl. R. Holz 1998: „Rating, Ranking, Scoring und Fuzzy Sets“

<sup>23</sup> Wir verwenden die Begriffe Rating und Ranking synonym.

<sup>24</sup> Kostenquoten sind aus Bewertungssicht je besser je niedriger sie sind, womit originäre Kostenquoten  $K^o = \{10; 15; 20; 25; 30\}$  in % etwa in der obigen Form bezüglich der Richtung transformiert werden können.

$$K(U) = \{70; 75; 80; 85; 90\}$$

$$Z(U) = \{3,5; 3,75; 4; 4,25; 4,5\}$$

vorweisen. Da es sich bei den Angaben jeweils um Prozentwerte handelt, ist im Folgenden unwesentlich. Zunächst interessiert uns auch die Kombination der Kennzahlensausprägungen bezüglich der betrachteten Unternehmen also die Relation

$$(K \times Z)_U = \{(k(u), z(u)) : k(u) \in K, z(u) \in Z, u \in U\}$$

der 5 betrachteten Unternehmen  $u \in U$  nicht.

Wir betrachten die Gewichte  $w_K$  und  $w_Z$  der Gleichung

$$S(k, z) = w_K k + w_Z z, k \in K \text{ und } z \in Z,$$

und fragen uns, welchen Einfluß die genannten Gewichte auf die Rangfolge der Ergebnisse  $S(k, z)$  einer gewichteten additiven Aggregation von Kennzahlen bei den gegebenen Bandbreiten

$$d_K = 20 \text{ und } d_Z = 1$$

von  $K$  und  $Z$  haben. Hierzu setzen wir  $w_Z = 1$  und finden heraus, dass für

$$w_K = \frac{1}{d_K}$$

die möglichen Wertekombinationen der Ergebnisse  $S(k, z)$  in den folgenden Intervallen liegen:

$$\begin{aligned} S(70, z) \in I_1 &:= [7,0; 8,0] \\ S(75, z) \in I_2 &:= [7,25; 8,25] \\ S(80, z) \in I_3 &:= [7,5; 8,5] \quad , z \in Z. \\ S(85, z) \in I_4 &:= [7,75; 8,75] \\ S(90, z) \in I_5 &:= [8,0; 9,0] \end{aligned}$$

Hierbei ist  $w_K$  der einzige freie Parameter in der Berechnung von  $S(k,z)$  und es gibt kein  $w_K > \frac{1}{d_K}$ ,

für das der gemeinsame Schnitt  $\bigcap_{j=1}^5 I_j$  nicht von der leeren Menge verschieden wäre und kein

$w_K < \frac{1}{d_K}$ , für das die Bandbreite des Schnittes  $\bigcap_{j=1}^5 I_j \neq 0$  ist, wobei sich die genannte Bandbreite für

$\text{Max } I_1 > \text{Min } I_5$  zu

$$\begin{aligned} \text{Max } I_1 - \text{Min } I_5 &= w_K k_5 + w_{ZZ1} - w_K k_1 + w_{ZZ5} \\ &= w_K d_K - w_Z d_Z \end{aligned}$$

errechnet.  $w_K = \frac{1}{d_K}$  ist dann genau das Gewicht, welches das

Vergleichbarkeitskriterium:

$$S(\text{Min}\{k : k \in K\}, \text{Max}\{z : z \in Z\}) = S(\text{Max}\{k : k \in K\}, \text{Min}\{z : z \in Z\})$$

sicherstellt und damit, sowohl unabhängig von einer Interpretation der Kennzahlen als auch unabhängig von einer Verteilung der Kennzahlenausprägungen zwischen der minimalen und der maximalen beobachteten Ausprägung, eine (gleichgewichtige) Vergleichbarkeit der betrachteten Kennzahlen bezüglich der Rangfolge der Ergebnisse  $S(k,z)$  ermöglicht.

Anschaulich stellt das Vergleichbarkeitskriterium sicher, dass die Veränderung einer der beiden Kennzahlen sich auch auf die Rangfolge der Ergebnisse  $S(.,.)$  auswirken kann. Das heißt das Vergleichbarkeitskriterium gewährleistet den technisch gleichgewichtigen Einfluß der betrachteten Kennzahlen auf die Rangfolge der Ergebnisse  $S(.,.)$ .

Ist die Bandbreite  $d_Z$  von 1 verschieden so wird das Vergleichbarkeitskriterium bezogen auf die Bandbreite der Intervalle  $I$  entsprechend der Bandbreite von  $Z$  durch die Gewichte

$$w_K = \frac{d_Z}{d_K} \text{ und } w_Z = 1 \text{ fest}$$

gewährleistet oder für  $w_Z \neq 1$  auch durch die von der jeweils anderen Kennzahl unabhängigen Vergleichbarkeitsgewichte

$$w_K = \frac{1}{d_K} \text{ und } w_Z = \frac{1}{d_Z} .$$

Wir betrachten im folgenden den letzteren Fall, da er die Kennzahlen für sich, unabhängig vom Zusammenkommen mit speziellen anderen Kennzahlen vergleichbar macht.

Damit ist auch die Verallgemeinerung des Vergleichbarkeitskriteriums auf beliebige Anzahlen von Kennzahlen gegeben.

***Bei mehreren Kennzahlen soll jede Kennzahl für sich mit jeder anderen Kennzahl vergleichbar sein.***

Das Vergleichbarkeitskriterium erfordert also, dass wir die Kennzahlenausprägungen jeweils mittels der Vergleichbarkeitsgewichte durch ihre Bandbreiten normieren bevor inhaltliche Gewichte - aus einer bewußten Interpretation heraus - unabhängig von technischen impliziten Gewichten für die Bildung einer Rangfolge gewählt werden.

#### 2.4.2 A-priori Bewertung von Kennzahlen

Ohne die direkte Berücksichtigung von Beziehungen zwischen den Kennzahlen, wie sie mit einer Relation des Typs

$$(K \times Z)_U = \{(k(u), z(u)) : k(u) \in K, z(u) \in Z, u \in U\}$$

gegeben sein können und etwa durch Clusterungen abstrahierend betrachtet werden können, stellt sich das Problem einer objektiveren Festlegung der Bandbreiten der einzelnen Kennzahlen. Die naheliegende unmittelbar aus dem Vergleichbarkeitskriterium resultierende Festlegung der Bandbreiten bezüglich des minimalen und maximalen Wertes der jeweilig betrachteten Kennzahl zeigt sich im allgemeinen als weniger sinnvoll, da das Minimum und das Maximum von Kennzahlenausprägungen vereinzelter, eher zufälliger Beobachtungen geprägt sein kann.

Für die a-priori Bewertung der Kennzahlen, d.h. für die objektive Festlegung der Bandbreite der Kennzahlen, kommt ohne weiteres Wissen als Basis dennoch nur die beobachtete Bandbreite der minimalen zur maximalen Kennzahlenausprägung in Frage. Als etablierte Verfahren zur Berücksichtigung der Sondersituationen im Sinne einer Transformation der Kennzahlenausprägungen

und damit auch einer Transformation der Bandbreite der Kennzahlenausprägungen sind zunächst verteilungsabhängige Ansätze und auch die Transformation mittels Präferenzfunktionen von Bedeutung. Letztere betrachten wir nicht, da es sich um subjektive Ansätze handelt.

Verteilungsabhängige Transformationen sind, wenn nicht über die Verteilungsfunktion selbst, über Momente dieser wie etwa der empirischen Standardabweichung  $\sigma$  gebildet<sup>25</sup>. Mit Kennzahlenausprägungen  $k_i, i=1, \dots, n$ , kann dann die Bandbreite der Kennzahlenausprägungen verzerrt etwa bezüglich eines vielfachen von  $\sigma$  mit

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (k_i - \frac{1}{n} \sum k_i)^2}$$

festgelegt werden.

Die lineare Transformation ordnet sich über die Gleichverteilung den verteilungsabhängigen Verfahren unter.

Als weiteres Verfahren zur Festlegung objektiver Bandbreiten der Kennzahlen kommen Vergrößerungen der Kennzahlenausprägungen durch Clusterungen in Frage, was insbesondere bei nicht gegebener Stetigkeit und damit in Anwendungen der Praxis geringeren Informationsverlust ermöglicht.

Es zeigt sich jedoch bei allen Verfahren, dass die für Ratingverfahren und auch für statistische Kenn- und Prüfgrößen bedeutsamen impliziten Gewichte regelmäßig nur an wenigen Eckwerten festgemacht werden bzw. nur mittels einer geringen Ausschöpfung der mit der Verteilung tatsächlich zur Verfügung stehenden Information berücksichtigt werden. Genauere Betrachtungen ermöglichen hier die im Folgenden formalisierten evolutorischen Skalenbetrachtungen, die dann über die allgemeinste Form der Aggregation, der Relation, auch in Modellbildungen einfließen können.

### *2.5 Evolutorische Skalenbetrachtungen*

Unter **Evolution** verstehen wir eine fortschreitende Entwicklung von beispielsweise Arten, Gesellschaften, Wirtschaftszweigen, Controlling-Segmenten und ähnlichem, im Folgendem auch als **Systeme** bezeichnet.

Über die Veränderung der Verteilung von Kennzahlen der Organismen, der Individuen, der Unternehmen oder beispielsweise der Kundengruppen, im Folgenden auch als (System-) **Objekte**

---

<sup>25</sup> Vgl auch die Konstruktion des Pearsonschen Korrelationskoeffizienten oben.

bezeichnet, lassen sich dann Aussagen über die Art der Evolution bzw. der fortschreitenden Entwicklung machen.

Insbesondere wenn auf evolutionäre Entwicklungen Einfluß genommen werden kann, besonders aber wenn evolutionäre Entwicklungen gesteuert sind, ist sowohl für eine eventuell notwendige Einflußnahme wie auch für ein Controlling von Steuerungsmaßnahmen oder einfach des schaffens von Wissen halber eine zeitnahe kontinuierliche und detaillierte Betrachtung von Veränderungen von Interesse. Die Betrachtung allein von Verteilungsmomenten wie etwa des Erwartungswertes und der Streuung ist hierbei zumeist wenig aussagekräftig.

Wir betrachten daher im Folgenden bezogen auf eine einzelne gemeinsame Kennzahl  $K_i : U \rightarrow \mathbb{R}$  einer Objektmenge  $U$  sowohl den Status der zeitlichen Veränderung der Kennzahlwerte  $K_i(u)$  wie auch die Dynamik der Veränderung bezogen auf ein Zeitintervall für das einzelne Objekt und für das System.

Sei also  $K_{i,t}(u)$  der Kennzahlwert zur Kennzahl  $i$  des Objektes  $u$  zum Zeitpunkt  $t$  und

$${}_d K_{i,t}(u) = \begin{cases} \frac{K_{i,t}(u) - K_{i,t-d}(u)}{\text{abs}(K_{i,t-d}(u))}, & \text{falls } K_{i,t-d}(u) \neq 0 \\ 1, & \text{falls } K_{i,t-d}(u) = 0, K_{i,t}(u) > 0 \\ -1, & \text{falls } K_{i,t-d}(u) = 0, K_{i,t}(u) < 0 \\ 0, & \text{sonst} \end{cases}, \quad d > 0$$

die relative Veränderung oder auch **Dynamik der Kennzahlwerte  $K_i(u)$  vom Zeitpunkt  $t-d$  zum Zeitpunkt  $t$** <sup>26</sup>. Seien ferner  $q_{i,t}$  und  ${}_d q_{i,t}$  die zugehörigen Quantilsfunktionen und weiter

$$r_{i,t-d} = \frac{\sum_u K_{i,t}(u) - \sum_u K_{i,t-d}(u)}{\text{abs}(\sum_u K_{i,t-d}(u))}$$

die relative Veränderung der Durchschnitte der Kennzahlwerte der Kennzahl  $i$  von  $t-d$  nach  $t$ .

Dann bezeichnen wir

<sup>26</sup> Die im Abschnitt „Bestimmtheit“ angesprochene und mit Bezug auf Vergleichbarkeit angemessenere Konstruktion des Nenners, ist hier nicht gewählt, da hier eine breitere Information angestrebt wird, die sich auf den Vergleich auch zur Bezugsgröße stützt.

- die  $x < q(1/3)$  als den **unteren Skalenbereich**,
- die  $x \in [q(1/3); q(2/3)]$  als den **zentralen Skalenbereich**
- und die  $x > q(2/3)$  als den **oberen Skalenbereich**

einer Quantilsfunktion  $q$  und nennen die **Dynamik einer Kennzahl  $K_{i,t}$**

- **normal**, falls der untere Skalenbereich von  ${}_d K_{i,t}$  nur Werte kleiner 0 enthält und der obere Skalenbereich nur Werte größer 0,
- **inflationär**, falls der obere und der zentrale Skalenbereich  ${}_d K_{i,t}$  nur Werte größer 0 enthält und
- **reaktionär**, falls der untere und der zentrale Skalenbereich  ${}_d K_{i,t}$  nur Werte kleiner 0 enthält,

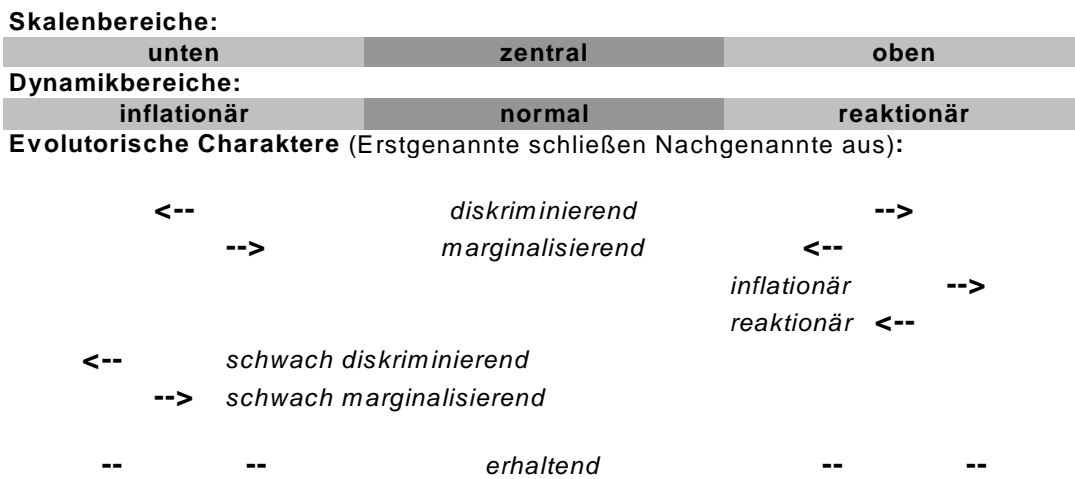
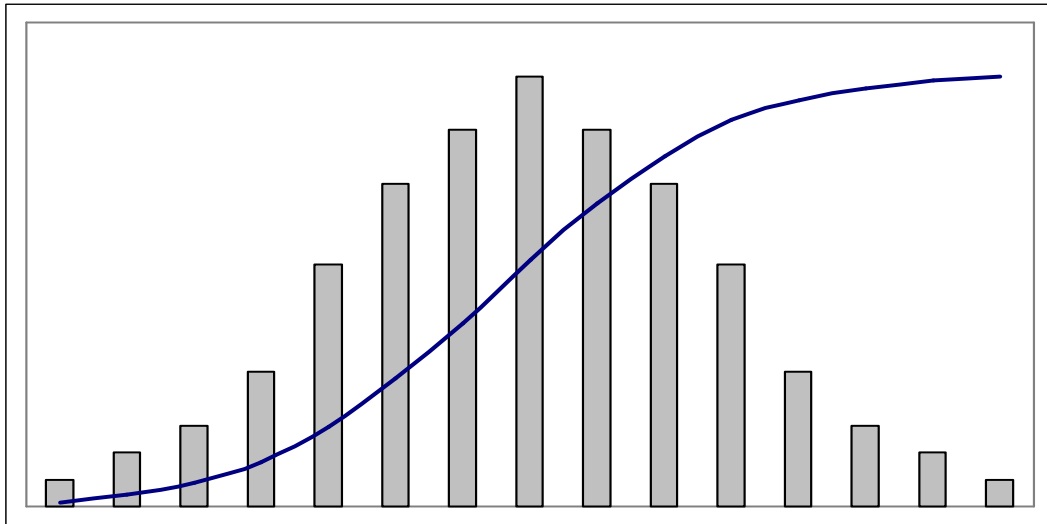
wobei die Begriffsbildung - wie auch im Folgenden - der einfacheren Verständigung und weniger der in konkreten Anwendungen immer auch intuitiv treffenden Bezeichnung dient.

Des Weiteren ist dann von Bedeutung wie - insbesondere auch mit welcher Intensität - sich die Charaktere normal, inflationär und reaktionär der Dynamiken auf die Verteilung der Kennzahlwerte  $K_{i,t}(u)$  auswirken, inwieweit sich also durch die Dynamik auch charakteristische interpretierbare Veränderungen der Gestalt der Verteilung der Kennzahlwerte im Zeitablauf ergeben.

Auch hier ist im Allgemeinen die zusätzliche Betrachtung allein der durchschnittlichen Veränderung der Kennzahlwerte in den definierten Skalenbereichen (unten, zentral und oben) weniger aussagekräftig, zumal sich das Zufallsrisiko dann auf eine geringere Anzahl zu Grunde gelegter Daten auswirkt, weswegen wir auch in den genannten Skalenbereichen der Kennzahlwerte  $K_{i,t}(u)$ , die Dreiteilung der Betrachtung beibehaltend, jeweils drei Stützstellen betrachten wollen, um die Veränderung der Form der Verteilung der Kennzahlwerte über den Drift der Stützstellen beurteilen zu können.

Wir legen daher unserer Betrachtung jeweils die Marktskala der Feinheit  $\delta=0,1$  der Kennzahlwerte  $K_i(u)$  zu Grunde. Verallgemeinerungen für eine größere Exaktheit sind naheliegend aber besonders in der Umsetzung auch aufwendiger.

Sei ferner die relative Veränderung der Marktskala von  $t-d$  nach  $t$  in den Stützstellen in zu  ${}_d K_{i,t}$  analoger Weise definiert. Dann nennen wir eine Kennzahl  $K_i$  von  $t-d$  nach  $t$  bezüglich einer unveränderten Objektmenge  $U$ :



**Abbildung:** Evolutorische Skalen-Charaktere

- **inflationär zur Intensität I**, wenn die zugehörige Dynamik inflationär ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=7,8,9$ , im oberen Skalenbereich von  $K_i$  von t-d nach t eine positive relative Veränderung größer I aufweisen.
- **reaktionär zur Intensität I**, wenn die zugehörige Dynamik reaktionär ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=7,8,9$ , im oberen Skalenbereich von  $K_i$  von t-d nach t eine negative relative Veränderung kleiner -I aufweisen.
- **diskriminierend zur Intensität I**, wenn die zugehörige Dynamik normal ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=1,2,3$ , im unteren Skalenbereich von  $K_i$  von t-d nach t eine negative relative Veränderung kleiner -I aufweisen und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=7,8,9$ , im oberen Skalenbereich von  $K_i$  von t-d nach t eine positive relative Veränderung größer I aufweisen.
- **marginalisierend zur Intensität I**, wenn die zugehörige Dynamik normal ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=1,2,3$ , im unteren Skalenbereich von  $K_i$  von t-d nach t eine positive

relative Veränderung größer I aufweisen und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=7,8,9$ , im oberen Skalenbereich von  $K_i$  von t-d nach t eine negative relative Veränderung kleiner -I aufweisen.

- **schwach diskriminierend zur Intensität I**, wenn die zugehörige Dynamik reaktionär ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=1,2,3$ , im unteren Skalenbereich von  $K_i$  von t-d nach t eine negative relative Veränderung kleiner -I aufweisen und  $K_i$  nicht reaktionär ist.
- **schwach marginalisierend zur Intensität I**, wenn die zugehörige Dynamik inflationär ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=1,2,3$ , im unteren Skalenbereich von  $K_i$  von t-d nach t eine positive relative Veränderung größer I aufweisen und  $K_i$  nicht inflationär ist.
- **erhaltend zum Niveau I**, wenn die zugehörige Dynamik normal ist,  $K_i$  nicht diskriminierend oder marginalisierend ist und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=1,2,3$ , im unteren Skalenbereich und mindestens 2 der 3 Skalenpunkte  $q_i(r^*0,1)$ ,  $r=7,8,9$ , im oberen Skalenbereich von  $K_i$  von t-d nach t keine betragsmäßige relative Veränderung größer I aufweisen.
- **undefiniert entwickelt**, in den sonstigen Fällen.

Von besonderer Bedeutung sind dann jeweils die maximalen Intensitäten bzw. das minimale Niveau. Womit sich über den Zeitspannenparameter d auch eine im Sinne von Frühwarnfunktionen interpretierbare Beobachtung von Systemen gestalten läßt.

Da Objektzu- und abwanderungen erhebliche von den aufgezeigten Entwicklungen unabhängige Verteilungsveränderungen zur Folge haben können, insbesondere wenn die statistische Basis eher geringeren Umfangs ist, ist es zumeist sinnvoller die Betrachtungen für sogenannte geschlossene in den Objekten unveränderte Systeme vorzunehmen, was im konkreten Fall zu einer Einschränkung zur Verfügung stehender Daten führen kann. Objekte mit fehlenden Werten bringen ähnliche Probleme mit sich wie Zu- und Abwanderungen.

Wie auch Erwartungswerte und andere Momente von Verteilungen reduzieren die definierten Entwicklungen mit der (maximalen) Intensität bzw. mit dem (minimalen) Niveau die wertmäßige Information auf einen einzelnen Wert. Auch die Intensität bzw. das Niveau sind dann direkt abhängig von der gesamten Verteilung zum Zeitpunkt t-d und t. Da wie beim Median im Vergleich zum Erwartungswert deutlicher die Anzahl der Daten als die Lage der Daten zu Bedeutung gelangt, sind die Intensitäten und das Niveau jedoch weniger durch einzelne „Ausreißer“ verfälscht.

Vom gesagten unabhängig ist die Veränderung beispielsweise des Erwartungswertes der Kennzahlen von t-d nach t eine zusätzliche Information, die jedoch keine Aussage über die innere Strukturveränderung betrachteter Systeme zuläßt. Die zusätzliche Betrachtung der Veränderung

beispielsweise der Streuung und der Schiefe der Verteilungen können hingegen insbesondere bei geringer Datenbasis sehr stark durch Ausreißer verfälscht sein.

Für die Entwicklungen inflationär, reaktionär, schwach diskriminierend und schwach marginalisierend ist die relative Veränderung der Kennzahldurchschnitte  $r_{i,t,d}$  offensichtlich eine wichtige interpretatorische Bezugsgröße.

Die Charaktere hier sind an möglichen Umsetzungen wie auch an mathematischen Interpretationen orientiert, besonders aber auch an der Beobachtung nicht rein zufälliger Entwicklungen, weswegen die Definitionen insbesondere mittels auch der Dynamik-Skalen eine Tragfähigkeit der Entwicklungen im Sinne einer ausreichenden Stützung gewährleisten. Die Evolutions-Theorie<sup>27</sup> kennt im allgemeinen andere Charaktere mit denen teilweise Überschneidungen gegeben sind.

## *2.6. Zur Technik der multivariaten Analyse mittels Bayes-Regeln*

Die Darstellung ist an das versicherungsmathematische Problem der Prämienfindung für ein Kollektiv von Versicherten mit als zufällig erachteten Schadenforderungen angelehnt.

### *2.6.1 Bayes-Regeln*

Im Folgenden sei die Nicht-Entartetheit der Probleme vorausgesetzt. Seien also Zufallsvariablen (teilweise auch Risiken genannt)  $X_1, X_2, \dots, X_n, Y$  und  $\Theta$  über einem gemeinsamen Wahrscheinlichkeitsraum gegeben. Ferner seien die Zufallsvariablen  $X_i$  gegeben  $\Theta$  stochastisch unabhängig ( $i=1, \dots, n$ ). Sei außerdem  $L: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  eine Verlustfunktion, deren Wert an der Stelle  $(x,d)$  als der Verlust interpretiert werde, den wir erleiden, wenn  $x$  die Realisierung des Risikos  $X$  und  $d=\mathbf{H}(X)$  die für  $X$  kalkulierte Prämie ist.

Eine Möglichkeit der Prämienfestlegung  $d=\mathbf{H}(X)$  ist dann über die Minimierung des erwarteten Verlustes

$$\int L(x,d) \mathbb{P}^X(dx) \quad (1)$$

bezüglich  $d$  gegeben. Mittels der folgend ausschließlich verwendeten Verlustfunktion

$$L(x,d) := (x - d)^2$$

---

<sup>27</sup> Vgl etwa Allen 1998, S. 61ff

erhält man so bekanntermaßen die Prämie  $\mathbf{H}(X)=E(X)$ .

In Erweiterung kann ein Entscheidungsmodell<sup>28</sup> betrachtet werden, mittels dessen Prämien ermittelt werden, die sich als Bayes-Regeln interpretieren lassen. Hierzu wird das Kollektiv der Risiken als a-priori mittels eines Parameters  $\vartheta$ , dem sogenannten Strukturparameter, differenziert betrachtet, der als Realisierung einer Zufallsvariable  $\Theta$  angesehen wird. Die Verteilung von  $\Theta$  wird dann auch A-priori bzw. Strukturverteilung genannt.

Bezeichnen wir die gemäß (1) analog bestimmte Prämie bezüglich  $\Theta$  gegeben  $X$  mit  $d(\vartheta)$ , so erhalten wir mit der Minimierung des erwarteten Verlustes

$$\int L(d(\vartheta), f) \mathbb{P}^{\Theta|X=x}(d\vartheta) \quad (2)$$

die auch als erfahrungstarifizierte Prämie bezeichnete reelle Zahl  $f^*(x)$ , die sich hier als  $E[\Theta|X=x]$  ergibt und wobei wir in  $X$  dann die Schadenerfahrung abbilden.  $f^*$  ist eine Entscheidungsfunktion die auch den erwarteten Verlust

$$\begin{aligned} EL(d(\Theta), f(X)) &= \int L(d(\vartheta), f(x)) \mathbb{P}^{(\Theta, X)}(d(\vartheta, x)) \\ &= \iint L(d(\vartheta), f(x)) \mathbb{P}^{X|\Theta}(dx) \mathbb{P}^{\Theta}(d\vartheta) \\ &= \iint L(d(\vartheta), f(x)) \mathbb{P}^{\Theta|X}(d\vartheta) \mathbb{P}^X(dx) \end{aligned}$$

unter allen Entscheidungsfunktionen  $f$  minimiert und somit auch als Bayes-Regel bzgl.  $\mathbb{P}^{\Theta}$  interpretiert werden kann. Die Bayes-Regel ist von der Entscheidungsfunktion, die man mit Hilfe des Minimierungsproblems

$$\min_a \iint L(y, a) \mathbb{P}^{Y|\Theta}(dy) \mathbb{P}^{\Theta|X}(d\vartheta) \quad (3)$$

erhält, der sogenannten empirischen Bayes-Regel, zu unterscheiden, die sich hier zu

$$E[Y|X=x] = \int E[Y|\Theta=\vartheta] \mathbb{P}^{\Theta|X=x}(d\vartheta)$$

---

<sup>28</sup>Vgl Heilmann, Kapitel 5.2

ergibt. Bayes- und empirische Bayes-Regeln stimmen nur in Spezialfällen von  $L$  überein, so beispielsweise, wie man leicht nachweist, insbesondere bei der oben festgelegten quadratischen Verlustfunktion.

Bei Heilmann<sup>29</sup> wird in

$$\mathbb{P}^{\Theta|X} := \mathbb{P}^{\Theta|X_1, X_2, \dots, X_n},$$

den sogenannten a-posteriori Verteilungen, ein mittels der individuellen Schadenerfahrung aufeinander folgender Perioden verbessertes a-priori Wissen gesehen. Abweichend hiervon verwenden wir  $X$  zur Modellierung zufälliger objektiver Schadenmerkmalsausprägungen, den Ausprägungen der Tarifmerkmale, um hieraus ein a-posteriori Wissen zu erhalten.

Wir befassen uns im Folgenden mit der empirischen Bayes-Regel, die, wenn wir mittels  $Y$  die antizipierte zukünftige Schadenverteilung und mittels  $X=(X_1, X_2, \dots, X_n)$  die Verteilung(en) der differenzierenden Tarifmerkmalsausprägungen eines Kollektivs sowie mittels  $\Theta$  die Strukturverteilung modellieren, eine Prognose- und eine Diagnose-Komponente enthält:

$$E[Y|X=x] = \int \underbrace{E[Y|\Theta=\vartheta]}_{\text{Prognose}} \underbrace{\mathbb{P}^{\Theta|X=x}}_{\text{Diagnose}}(d\vartheta) \quad (4)$$

### 3.2 Dempster-Shafer Updating

Von Bedeutung sei nun die Ermittlung der a-posteriori Verteilung im Sinne der Diagnose-Komponente in (4)

$$\mathbb{P}^{\Theta|X} = \mathbb{P}^{\Theta|X_1, X_2, \dots, X_n}.$$

Als Rechenregeln gelten:

- Bedingte Wahrscheinlichkeit  $\mathbb{P}^{\Theta|X} = \frac{\mathbb{P}^{\Theta, X}}{\mathbb{P}^X}$

---

<sup>29</sup> Vgl. ebenda. Es ist dann zusätzlich zur bedingten Unabhängigkeit die Verteilungsideutität über die Perioden angenommen, was für die Betrachtung der Klassen konjugierter Verteilungen von Bedeutung ist.

- Totale Wahrscheinlichkeit 
$$\mathbb{P}^X = \sum_{\vartheta} \mathbb{P}^{X|\Theta} \mathbb{P}^{\Theta} \quad (5)$$

- Bayes-Updating 
$$\mathbb{P}^{\Theta|X} = \frac{\mathbb{P}^{X|\Theta} \mathbb{P}^{\Theta}}{\sum_{\vartheta} \mathbb{P}^{X|\Theta} \mathbb{P}^{\Theta}} \stackrel{\text{stoch.}}{=} \stackrel{\text{Unabh}}{=} \frac{\mathbb{P}^{X_1|\Theta} \dots \mathbb{P}^{X_n|\Theta} \mathbb{P}^{\Theta}}{\sum_{\vartheta} \mathbb{P}^{X_1|\Theta} \dots \mathbb{P}^{X_n|\Theta} \mathbb{P}^{\Theta}} \quad (6)$$

Letzteres lässt - bei gegebener stochastischer Unabhängigkeit der durch  $X_i$  gegebenen  $\Theta$  gegebenen Verteilungen - die Ermittlung der a-posteriori Verteilung aus den einzelnen Verteilungen der Tarifmerkmale in dem bezüglich  $\Theta$  strukturierten Kollektiv zu. Die empirische Ermittlung von  $\mathbb{P}^{X|\Theta}$  anhand der Verhältnisse eines gegebenen Kollektivs hat dann aber zur Folge, dass sie sich auf das Datenvolumen des Teilkollektivs gegeben  $\Theta=\vartheta$  beschränken muß, was immer dann ein Nachteil ist, wenn  $\Theta$  stärker differenziert als die einzelnen Tarifmerkmale, was wiederum in praktischen Anwendungen nahezu immer der Fall sein wird. Besserung schafft hier das Updating nach Dempster und Shafer<sup>30</sup>, das auch als Kombinations-Regel bekannt ist.

- Kombinations-Regel 
$$\mathbb{P}^{\Theta|X_1, X_2} := \frac{\mathbb{P}^{\Theta|X_1} \mathbb{P}^{\Theta|X_2}}{\sum_{\vartheta} \mathbb{P}^{\Theta|X_1} \mathbb{P}^{\Theta|X_2}} \quad (7)$$

Die Kombinations-Regel ist assoziativ, das heißt es gilt:

- Assoziativität

$$\mathbb{P}^{\Theta|(X_1, X_2), X_3} = \frac{\mathbb{P}^{\Theta|X_1} \dots \mathbb{P}^{\Theta|X_3}}{\sum_{\vartheta} \mathbb{P}^{\Theta|X_1} \dots \mathbb{P}^{\Theta|X_3}} = \mathbb{P}^{\Theta|X_1, X_2, X_3} = \mathbb{P}^{\Theta|X_1, (X_2, X_3)} \quad (8)$$

Und wie man mittels obiger Rechenregeln leicht nachweist gilt unter Ausnutzung der stochastischen Unabhängigkeit von  $X_i$  gegeben  $\Theta$

$$\mathbb{P}^{\Theta|X_1, \dots, X_n} = \frac{\mathbb{P}^{X_1|\Theta} \dots \mathbb{P}^{X_n|\Theta} (\mathbb{P}^{\Theta})^n}{\sum_{\vartheta} \mathbb{P}^{X_1|\Theta} \dots \mathbb{P}^{X_n|\Theta} (\mathbb{P}^{\Theta})^n}, \quad (9)$$

<sup>30</sup> Vgl bezüglich des folgend aufgeführten Zusammenhanges auch Holz: „Fuzzy Sets in der Tarifierung“, Kapitel 6. Für die Dempster-Shafer-Theorie vgl. auch Kohlas: „A Mathematical Theory of Hints“.

was bei zusätzlicher Annahme einer Gleichverteilung für  $\Theta$  offensichtlich exakt dem Bayes-Updating entspricht. Ferner gilt unter Ausnutzung von (6) und (5) in (9) mit beliebigen Verteilungen für  $\Theta$

$$\begin{aligned}
 \mathbb{P}^{\Theta|X_1, \dots, X_n} &\stackrel{(9), \text{stoch.}}{=} \frac{\mathbb{P}^{X_1, \dots, X_n|\Theta} (\mathbb{P}^{\Theta})^n}{\sum_{\vartheta} \mathbb{P}^{X_1, \dots, X_n|\Theta} (\mathbb{P}^{\Theta})^n} \\
 &\stackrel{\text{Unabh}}{=} \\
 &\stackrel{(6), (5)}{=} \mathbb{P}^{\Theta|X_1, \dots, X_n} \frac{(\mathbb{P}^{\Theta})^{n-1}}{\sum_{\vartheta} \mathbb{P}^{\Theta|X_1, \dots, X_n} (\mathbb{P}^{\Theta})^{n-1}}
 \end{aligned} \tag{10}$$

sowie auch

$$\mathbb{P}^{\Theta|X_1, \dots, X_n} \stackrel{(9), (6), (5)}{=} \frac{\mathbb{P}^{\Theta|X_1} \dots \mathbb{P}^{\Theta|X_n}}{\sum_{\vartheta} \mathbb{P}^{\Theta|X_1} \dots \mathbb{P}^{\Theta|X_n}} \tag{11}$$

Die Verwendung der Kombinations-Regel an Stelle des Bayes-Updateings in der Diagnose-Komponente von (4) hat dann also die Verzerrung

$$z(\vartheta, x_1, \dots, x_n) = \frac{(\mathbb{P}^{\Theta})^{n-1}}{\sum_{\vartheta} \mathbb{P}^{\Theta|X_1, \dots, X_n} (\mathbb{P}^{\Theta})^{n-1}} \tag{12}$$

zur Folge, die in Anwendungen zur Gewährleistung der Bayes'schen Verhältnisse unter Verwendung von (11) herausgerechnet werden kann und im Fall der Gleichverteilung für  $\Theta$  offensichtlich gleich 1 ist.

Die Formel (11) stellt dann eine gegenüber der stochastischen Unabhängigkeit allgemeinere Form der Zerlegung von a-posteriori Verteilungen dar, die aber zur Gewährleistung der stochastischen Verhältnisse der genannten Voraussetzungen bedarf.

- Spezifität

Eine weitere, empirisch relevante Eigenschaft der Kombinations-Regel ist deren Spezifizierungs-Eigenschaft. Beschreibe

$$\delta_{\theta}(\vartheta) = \begin{cases} 1, & \text{falls } \vartheta = \theta \\ 0, & \text{sonst} \end{cases},$$

die sogenannte Dirac-Funktion für  $\theta$ , die Konzentration der gesamten Masse einer Verteilung in einem Punkt  $\theta$ . Dann ist sofern in der Kombinations-Regel nur eine der beteiligten Verteilungen durch eine Dirac-Funktion beschrieben ist, die mittels der Kombinations-Regel ermittelte a-posteriori Verteilung ebenfalls auf  $\theta$  konzentriert also auch insgesamt eine Dirac-Funktion für  $\theta$  sofern insgesamt Evidenz für  $\theta$  gegeben ist.

Die genannte Eigenschaft stellt einen Spezial-Fall der Spezifizierungs-Eigenschaft der Kombinations-Regel dar. Allgemein betont die Normierung des Produktes der a-posteriori Verteilungen gegeben  $X_i$  ( $i=1,\dots,n$ ) auf die Summe 1 bezüglich  $\Theta$ , die gemeinsam hohen a-posteriori Wahrscheinlichkeiten zu Lasten der unterschiedlichen a-posteriori Wahrscheinlichkeiten, was allein aus der Krümmung des Graphen der Verknüpfung von Verteilungen mittels der Multiplikation resultiert und womit besonders übereinstimmende Evidenz hervorgehoben wird.

- Konträre Evidenz

Verknüpfen wir andererseits sich gegenseitig ausschließende Evidenz etwa in Form zweier Dirac-Funktionen

$$\delta_{\theta_1}(\vartheta) \quad \text{und} \quad \delta_{\theta_2}(\vartheta), \quad \theta_1 \neq \theta_2$$

in der Kombinationsregel, dann erhält das so charakterisierte Objekt unabhängig von weiteren Merkmalen die Evidenz 0 zu jedem  $\vartheta$  der a-priori Struktur.

Letzteres ist bei der empirischen Ermittlung von a-posteriori Verteilungen, wie dies im Weiteren beschrieben ist, ausgeschlossen, bei einer subjektiven Festlegung von Evidenz aber zu beachten.

## **Literatur:**

- P.M. Allen:** „Modelling complex economic evolution“, in L. Pohlmann et.al. (Hrsg.): „Selbstorganisation – Jahrbuch für Komplexität in den Natur-, Sozial- und Geisteswissenschaften“, Band 9 1998, Duncker & Humblot, 47-75.
- J.C. Bezdek:** „Pattern recognition with fuzzy objective function algorithms“, Plenum Press New-York 1981.
- B. Elpelt, J. Hartung:** „Multivariate Statistik“, Oldenburg 1995.
- W.-R. Heilmann:** „Grundbegriffe der Risikotheorie“, VVW Karlsruhe 1987.
- R. Holz:** „Fuzzy Sets in der Tarifierung“, Shaker 1996.
- ... „Rating, Ranking, Scoring und Fuzzy Sets“, Blätter der DGVM, Heft 3/1998, 363-384.
- ... „Großstädte-Ranking 2000 – Informationsgewinnung unter Berücksichtigung dynamischer Kausalitäten“, Shaker-Verlag Aachen 2000.
- ... „Die private Versicherungswirtschaft 2003 – Rankings, Benchmarks, Scorecards und Marktentwicklungen zur Kranken-, Lebens- und Schaden- und Unfallversicherungswirtschaft“, Shaker-Verlag Aachen 2003.
- G.J.Klir, B.Yuan:** „Fuzzy Sets and Fuzzy Logic“, Prentice Hall 1995.
- J. Kohlas, P.-A. Monney:** „A Mathematical Theory of Hints“, Springer 1995.
- R.B. Nelsen:** An Introduction to Copulas, Lecture Notes in Statistics 139, Springer 1999.
- D. Pfeifer, J. Neslehova:** „Modelling Dependence in Finance and Insurance: the Copula Approach, Blätter der DGVM, Heft 2/2003, 177-191.
- R. Seising (Hrsg.):** „Fuzzy Theorie und Stochastik“, Gabler 1999.

Gez. Robert Holz <http://www.rankingweb.de>

Weitere Diskussion Paper und Anwendungen unter:

Applications <http://www.t-online.de/home/robert.holz/page28.html>