

**Entropiebasiertes Data Mining
im Produktdesign**

Sandra Rudolph Wilhelm Rödder

Diskussionsbeitrag Nr. 413

Juli 2007

Diskussionsbeiträge der Fakultät Wirtschaftswissenschaft
der FernUniversität in Hagen

Herausgegeben vom Dekan der Fakultät

Alle Rechte liegen bei den Verfassern

Inhaltsverzeichnis

1	Einleitung	1
2	Lernen aus Daten	6
2.1	Mathematische Präliminarien	6
2.1.1	Die Syntax der Konditionalsprache	6
2.1.2	Bayes und Multinomial/Dirichlet auf Konfigurationen	8
2.2	Prinzipien zur Aktualisierung der Dirichlet auf Konfigurationen	9
2.2.1	Lernen aus vollständigen Daten	9
2.2.2	Lernen aus unvollständigen Daten	10
2.2.2.1	Das MinREnt-Prinzip	10
2.2.2.2	Randverteilungen und Proportional Scaling .	11
2.3	Signifikanz erlernter Abhängigkeiten	11
2.4	Mining	13
3	Produktdesign	15
3.1	Einführung	15
3.2	Ein statistischer Ansatz: Conjoint-Analyse	17
3.2.1	Idee der klassischen Conjoint-Analyse und Annahmen	17
3.2.2	OLS-Schätzung und Gütekriterien	20
3.2.2.1	Beurteilung der Güte der Schätzung	20
3.2.2.2	Beispiel zur OLS-Schätzung	24

3.2.2.3	Ergebnis der OLS-Schätzung und Interpretation	26
3.2.2.3.1	OLS-Schätzung für den Fall vollständiger Daten	26
3.2.2.3.2	OLS-Schätzung für den Fall unvollständiger Daten	28
3.2.3	Übertragung der Ergebnisse auf das Kaufverhalten	29
3.3	Ein informationstheoretischer Ansatz: MinREnt	31
3.3.1	Idee und Modellierung	31
3.3.2	'Schätzung' und Interpretation	34
3.4	Vergleich der Ergebnisse	45
4	Zusammenfassung und Ausblick	47
A	Konjugiertheit von Multinomial und Dirichlet	49
B	1. und 2. Moment der Dirichlet	51

Kapitel 1

Einleitung

Aufgrund technischer Entwicklungen sind Unternehmungen in der Lage, immer größere Mengen von konsumentenbezogenen Daten zu erfassen und zu speichern. Damit einher geht die Entwicklung von Verfahren zur Analyse dieser Datenmengen, um Zusammenhänge - Muster - zu erkennen und diese für betriebswirtschaftliche Entscheidungen nutzbar zu machen. Solche Verfahren firmieren in der Literatur unter dem Begriff Data Mining (vgl. (Kat04), S. 61 ff.). Über die Sammlung kundenbezogener Daten seien 'Beobachtungsvektoren' gegeben, deren Komponenten die Ausprägungen kundenbeschreibender Merkmale sind. Ist man an Gruppierungen der Kunden hinsichtlich ähnlicher Profile interessiert, bietet sich als Instrument die Clusteranalyse an. Ziel der Clusteranalyse ist es, Segmente von Kunden zu finden, so dass die sich darin befindlichen Kunden einander bzgl. ihrer beschreibenden Merkmale sehr ähnlich und zu Kunden in anderen Segmenten möglichst verschieden sind. Über die Maximierung der Intergruppenvarianz kann eine Zuordnung von Kunden zu Segmenten gefunden werden, die obige Forderungen erfüllt. Eine Zuordnung von neuen Objekten findet gemäß ihrer jeweiligen Merkmalsprofile statt (vgl. (Har84), S. 443 ff.).

Ein weiteres Instrument - die Diskriminanzanalyse - findet für gegebene Objekte ein sie unterscheidendes Trennkriterium, den sogenannten Diskri-

minanzwert, der abhängig von den die Objekte beschreibenden Merkmalen ist. Die Diskriminanzfunktion ist im einfachsten Fall eine lineare Funktion, deren Koeffizienten mittels geeigneter Verfahren geschätzt werden (vgl. (Har84), S. 240 ff.). Neue Objekte können wiederum mittels des Trennkriteriums bestehenden Segmenten zugeordnet werden.

Die eben dargestellten Ansätze haben zum Ziel, eine Klassifizierung von Objekten vorzunehmen. Um unter Angabe Objekte beschreibender Merkmale eine Prognose bezüglich eines weiteren Attributes, wie bspw. einer Entscheidungsvariablen, treffen zu können, sucht die folgende Verfahrensklasse nach Regeln. Gefragt sind hier also signifikante bedingte Häufigkeiten, die auf starke Abhängigkeiten der entsprechenden Variablen hinweisen. Hat man solche Abhängigkeiten identifiziert, kann bei Kenntnis einiger Variablenwerte auf das Auftreten anderer Variablenwerte geschlossen werden. Der als Assoziation bezeichnete Ansatz ist eine weitere wesentliche Aufgabe des Data Minings (vgl. (Fre98), S. 23 ff.).

Dieser zentralen Aufgabe widmen wir uns in diesem Papier, allerdings mit einem neuen Instrumentarium; die essentiellen Gedanken wollen wir kurz skizzieren.

- Für den Fall einer Datei, deren Daten Vektoren von Ausprägungen endlichwertiger Variabler sind:
 - Rekonstruiere bestmöglich die zu dieser Datei gehörige Verteilung
 - die also die Daten erzeugt haben könnte
 - Mache diese Verteilung zum Untersuchungsgegenstand des 'Minings' im Sinne von assoziativer Regelsuche.
- Für diese Rekonstruktion bediene dich des Bayes'schen Satzes und des Prinzips maximaler Entropie (MaxEnt) bzw. minimaler relativer Entropie (MinREnt).
 - Falls jedes Datum vollständig ist, aktualisiere den Kenntnisstand

über die unbekannte Verteilung jeweils mittels des Bayes'schen Satzes.

- Falls einige Daten unvollständig sind, vervollständige die Lücken mittels des MinREnt-Prinzips.
- Hat man eine Verteilung mittels der Datei erzeugt, beginnt das 'Mining'. Hier kaprizieren wir uns auf den Fall der Regelbehandlung wie oben angedeutet. So wird gezeigt werden, dass
 - auf Anfragen eines Benutzers jede beliebige bedingte Wahrscheinlichkeit ausgewiesen und mit einem Signifikanzmaß versehen werden kann.
 - die Transinformation zwischen Variablengruppen ein Maß für deren Abhängigkeitsgrad ist. Hat man dieses Maß, so unterstützt es die Auswahl hochgradig abhängiger Variabler. Somit ist die Transinformation ein Indikator für die weitere Suche nach Assoziationsregeln.

Nachdem das Instrumentarium in seiner mathematischen Konzeption entwickelt ist, werden wir für den Fall des Produktdesigns die Mächtigkeit des Instrumentariums zeigen.

- Aus einer Datei, bestehend aus Kundenbewertungen zu Produktvarianten, wird mittels der vorgestellten Verfahren 'Bayes' und 'Entropie' die erzeugende Verteilung rekonstruiert.
- Zu den den Experten interessierenden Regelwahrscheinlichkeiten können Antworten mit Signifikanzmaßen geliefert werden, was exemplarisch vorgeführt wird.
- Zu verschiedenen Merkmalsprofilen des Produktes gestattet die gefundene Wahrscheinlichkeitsverteilung die Schätzung dessen Rating-Gütemaßes. Eine vollständige Enumeration aller Merkmalsprofile und

die Berechnung ihres jeweiligen Rates ist eine einfache Vorgehensweise zum Finden optimaler Produktdesigns.

- Da die vollständige Enumeration bei großen Variablenmengen versagt, entwickeln wir eine auf dem Bayes'schen Satz basierende Heuristik 'Best Design'.
- Das Maß der Transinformation dient zur Prüfung der gefundenen Abhängigkeiten zwischen Produktmerkmalen und Rating.

Gemäß der obigen Skizzierung der Vorgehensweise werden in Abschnitt 2.1 mathematische Präliminarien aufgezählt; so werden für die Entwicklung des informationstheoretischen Ansatzes notwendige Begriffe definiert und die Syntax für konditionierte Aussagen über die Produktvarianten erläutert. Weiterhin wird der Zusammenhang der für die folgenden Überlegungen wichtigen Verteilungen Multinomial und Dirichlet dargestellt.

Daran anknüpfend werden in 2.2 Algorithmen präsentiert, die die Aktualisierung der Dirichlet aufgrund der 'gemachten' Beobachtungen erlauben.

Mit Hilfe dieser aktualisierten Verteilung ist die Signifikanz des Wahrheitsgehaltes bedingter Aussagen über die Produktvarianten bestimmbar, was in Abschnitt 2.3 dargelegt wird.

Abschnitt 2.4 widmet sich dem Mining der Datenbasis auf Grundlage der aktuellen Verteilung. Da eine vollständige Enumeration der Produktvarianten und deren aggregierte Beurteilung – wie bereits angemerkt – zu rechenaufwändig ist, wird das Maß der Transinformation eingeführt, welches zur Quantifizierung der Abhängigkeiten zwischen Variablen(-gruppen) dient und als Maß für die Reliabilität des mit der Expertensystem-Shell SPIRIT (www.xspirit.de) gefundenen Produktdesigns Verwendung findet.

In Abschnitt 3.1 wird das Problem des Produktdesigns skizziert. Es wird auf die Einordnung der Fragestellung in den marketingpolitischen Kontext sowie auf die Problematik der Verwendung von Beurteilungsskalen eingegangen.

Abschnitt 3.2 dient der Vorstellung statistischer Verfahren zur Lösung des Produktdesign-Problems. Anhand einer erzeugten Datenbasis werden die theoretischen Ausführungen zur Conjoint-Analyse veranschaulicht. Es wird exemplarisch gezeigt, wie das optimale Produktdesign auf Basis einer aggregierten Ratefunktion bestimmt werden kann. Die Schätzung der Parameter dieser Funktion findet einerseits für den Fall vollständiger Daten und andererseits für den Fall unvollständiger Daten statt.

In Abschnitt 3.3 wird in Analogie zur Vorgehensweise in Abschnitt 3.2 mit Hilfe des informationstheoretischen Ansatzes das optimale Produktdesign gesucht. Wie im vorigen Kapitel werden die Fälle 'vollständige Daten' und 'unvollständige Daten' unterschieden.

In Abschnitt 3.4 findet ein kritischer Vergleich der Ergebnisse aus den Abschnitten 3.2 und 3.3 statt mit dem Ziel, Überlegungen bezüglich der Güte der Verfahren anzustellen. Es sei schon vorweggeschickt, dass sich anhand der Ergebnisse die Überlegenheit des Prinzips MinREnt gegenüber klassischen Verfahren zeigt.

Das Papier schließt in Kapitel 4 mit einer Zusammenfassung und einem Ausblick auf zukünftige Forschung. So wird erläutert werden, wie mittels des vorgestellten Instrumentariums eine LEG-Zerlegung vorgenommen werden kann und somit Produkt-Vollprofile in hinsichtlich der Bewertung aussagekräftige Teilprofile zerlegt werden können.

Kapitel 2

Lernen aus Daten

2.1 Mathematische Präliminarien

2.1.1 Die Syntax der Konditionalsprache

Daten kommen von einer zufallsgesteuerten Quelle, die es zu rekonstruieren gilt (vgl. Kapitel 1). Ist diese Aufgabe erfüllt, ersetzen (bedingte) Wahrscheinlichkeiten die bedingten Frequenzen im ursprünglichen Datensatz. Zur Beschreibung sowohl des Vorgangs der Rekonstruktion als auch der Auswertung bedingter Wahrscheinlichkeiten bedarf es einer geeigneten Symbolik (vgl. (Röd06b)).

Die Syntax zur Beschreibung von Beobachtungen und Sachverhalten über die Individuen einer Population besteht aus einer endlichen Menge endlichwertiger Variabler $\mathcal{V} = \{V_1, \dots, V_L\}$ und deren Ausprägungen oder Werten v_l von V_l . Sie enthält ferner die Konnektive \neg (nicht), \wedge (und), \vee (oder), Klammern und Vereinbarungen über die Auswertung von aus diesen Elementen gebildeten Formeln. Mit elementaren Formeln $VARIABLE = \text{wert}$ und den Konnektiven beschreibt man Aussagen über die Individuen der Population. Vom logischen Standpunkt aus sind elementare Formeln Literale, die unter einer Interpretation w(ahr) oder f(alsch) sein können. Auch zusam-

mengesetzte Formeln sind Aussagen und werden mit $\mathbf{A}, \mathbf{B}, \mathbf{C} \dots$ bezeichnet. Konjunkte von Literalen werden kurz durch ihre Werte notiert, wie zum Beispiel $\mathbf{v} = v_1 \dots v_L$ oder \mathbf{v}_I mit $I \subset \{1, \dots, L\}$. Für Teilkonjunkte \mathbf{v}_I findet auch eine eingängige Notierung \mathbf{v}_I . Verwendung. Sie drückt aus, dass \mathbf{v}_I aus einer Menge von Vollkonjunkten \mathbf{v} besteht. \mathbf{V}_I (bzw. \mathbf{V}) ist die Menge aller (Teil-) Konjunkte und $|\mathbf{V}_I|$ (bzw. $|\mathbf{V}|$) ihre Mächtigkeit. \mathcal{V}_I bezeichnet die entsprechende Variablenmenge.

$|$ ist der gerichtete Konditionaloperator. Formeln der Art $\mathbf{B}|\mathbf{A}$ heißen Konditionale; lies: \mathbf{B} gegeben \mathbf{A} . $\mathbf{B}|\mathbf{A}$ ist w(ahr) für ein wahres \mathbf{B} und ein wahres \mathbf{A} , es ist f(alsch) für ein falsches \mathbf{B} und ein falsches \mathbf{A} , und es ist u(nbestimmt) für ein falsches \mathbf{A} (vgl. (Cal91)). $BUERGE = ja|KREDITWUERDIG = ja[w]$ ist ein wahres Konditional. Um die Nützlichkeit solcher Konditionale zu erhöhen, erlaubt man als Einschätzungen ihres Wahr-Seins Wahrscheinlichkeiten x statt w oder f . Solche konditionalen Sachverhalte schreiben wir $\mathbf{B}|\mathbf{A}[x]$.

Ein semantisches Modell zu dieser Syntax ist eine Wahrscheinlichkeitsverteilung P auf \mathcal{V} . Ein Sachverhalt $\mathbf{B}|\mathbf{A}[x]$ ist gültig in P genau dann, wenn $p(\mathbf{B}\mathbf{A}) = x \cdot p(\mathbf{A})$. P ist der epistemische Zustand auf \mathcal{V} , in dem das Wissen über die Population ausgedrückt wird.

Für $I \subset \{1, \dots, L\}$ ist \mathbf{v}_I das Konjunkt der entsprechenden Variablenwerte, siehe weiter oben. Werden solche Variablenwerte empirisch erfahren, so wird dies Beobachtung genannt und mit \mathbf{v}_I notiert. Eine Folge solcher Beobachtungen heißt Beobachtungsdatei: $v_{I_1}, \dots, v_{I_n}, \dots, v_{I_N}$; \mathbf{v} steht für eine vollständige Beobachtung aller L Variablen. Eine einsichtige Darstellung von \mathbf{v}_I ist die vollständige Benennung aller L Variablen mit Einträgen v_i , $i \in I$, und ? bei unbekanntem Werten. Für $N=6$, die vier binären Variablen $V_l = j/n$, $l = 1, \dots, 4$, und $I_1 = \{1, 2\}$, $I_2 = \{1, 3\}$, $I_3 = \{1, 4\}$, $I_4 = \{1, 2, 3, 4\}$, $I_5 = \{4\}$ und $I_6 = \{1\}$ ist eine mögliche Beobachtungsdatei in Tabelle 2.1 dargestellt.

lfd. Nr.	V_1	V_2	V_3	V_4
1	n	j	?	?
2	j	?	j	?
3	n	?	?	n
4	n	n	n	j
5	?	?	?	n
6	j	?	?	?

Tabelle 2.1: Beobachtungsdatei

Aus Darstellungsgründen betrachten wir im Folgenden nur binäre Variable, alle Ergebnisse sind aber sofort auf Variable mit beliebig endlichem Wertebereich übertragbar (vgl. (Röd06b)).

2.1.2 Bayes und Multinomial/Dirichlet auf Konfigurationen

Der Satz von Bayes findet in unserem Kontext Anwendung zur Schätzung der unbekannt Wahrscheinlichkeiten aller Konfigurationen $\mathbf{v} \in \mathbf{V}$. Der jeweils aktuelle Informationsstand über die unbekannt Wahrscheinlichkeiten wird mittels einer Dirichlet beschrieben mit der Dichte

$$Dir(a(\mathbf{v}), \mathbf{v} \in \mathbf{V}) = \frac{\Gamma \sum_{\mathbf{v} \in \mathbf{V}} a(\mathbf{v})}{\prod_{\mathbf{v} \in \mathbf{V}} \Gamma a(\mathbf{v})} \cdot \prod_{\mathbf{v} \in \mathbf{V}} p(\mathbf{v})^{a(\mathbf{v})-1}, \quad (2.1)$$

$a(\mathbf{v}) > 0$ für alle $\mathbf{v} \in \mathbf{V}$.

Die Dirichlet ist konjugiert zur Multinomial. Bei N unabhängigen Beobachtungen vollständiger Konfigurationen und *bekannt* Wahrscheinlichkeiten $p(\mathbf{v}), \mathbf{v} \in \mathbf{V}$, folgt die Zählstatistik $\mathbf{y}(\mathbf{v}), \mathbf{v} \in \mathbf{V}$, einer Multinomial mit Wahrscheinlichkeitsfunktion

$$\frac{N!}{\prod_{\mathbf{v} \in \mathbf{V}} y(\mathbf{v})!} \prod_{\mathbf{v} \in \mathbf{V}} p(\mathbf{v})^{y(\mathbf{v})}. \quad (2.2)$$

Der im folgenden Abschnitt beschriebene Lernvorgang gestattet die Aktualisierung der Dirichlet durch Beobachtungen und mittels des Satzes von Bayes. Der mit den Zusammenhängen weniger vertraute Leser sei verwiesen auf den Anhang A.

2.2 Prinzipien zur Aktualisierung der Dirichlet auf Konfigurationen

2.2.1 Lernen aus vollständigen Daten

Eine Datei mit vollständigen empirischen Daten \mathbf{v} wird zur Verdichtung der Information über die unbekanntes $p(\mathbf{v})$, $\mathbf{v} \in \mathbf{V}$, herangezogen. Der beschriebene Lernvorgang ist dem frequentistischen Lernvorgang auf der Kontingenztafel sehr ähnlich, hat jedoch diesem gegenüber erhebliche Vorteile:

- Der bei unvollständigen Daten notwendige Abgleich – siehe Abschnitt 2.2.2 – erzeugt ggfs. nichtganzzahlige 'Häufigkeiten', die in der hiesigen Darstellung problemlos behandelt werden können.
- Die Unsicherheitsbereiche bedingter Wahrscheinlichkeiten werden elegant mittels Bayes berechenbar, siehe auch Abschnitt 2.3.

Algorithmus 'LernVollstD'

$a(\mathbf{v}) := \varepsilon \quad \forall \mathbf{v} \in \mathbf{V}$, ε eine beliebig kleine positive Zahl, vgl. auch Anhang B.

Für alle Daten \mathbf{v}

falls $\mathbf{v} = \mathbf{v}$ setze $a(\mathbf{v}) := a(\mathbf{v}) + 1$

$Dir(a(\mathbf{v}), \mathbf{v} \in \mathbf{V})$ ist die resultierende Dirichlet auf \mathcal{V}

$\hat{p}(\mathbf{v}) = \frac{a(\mathbf{v})}{\sum_{\mathbf{v}} a(\mathbf{v})}$, $\mathbf{v} \in \mathbf{V}$, ist der Erwartungswert der aktuellen Dirichlet und als solcher ein Schätzer für die unbekanntenen Wahrscheinlichkeiten $p(\mathbf{v})$, $\mathbf{v} \in \mathbf{V}$.

2.2.2 Lernen aus unvollständigen Daten

2.2.2.1 Das MinREnt-Prinzip

Es folgen allgemeine Ausführungen zum MinREnt-Prinzip (vgl. (Röd06a)). Ist die aktuelle Wahrscheinlichkeitsverteilung über die probabilistische Abhängigkeitsstruktur stochastischer Variabler \mathcal{V} in Form einer Verteilung P auf \mathcal{V} gegeben und fließt weitere Information über die Struktur zu, so adaptiert man P an die neue Situation mittels des MinREnt-Prinzips. Dazu wählt man die Verteilung minimalen relativen informationstheoretischen Abstands von P unter Wahrung der *gesamten* Information über die Abhängigkeitsstruktur.

Gegeben sei nun $P = (p(\mathbf{v}), \mathbf{v} \in \mathbf{V})$, ggfs. der Erwartungswert einer aktualisierten Dirichlet, siehe Abschnitt 2.2.1. Für festes \mathbf{I} seien die entsprechenden Wahrscheinlichkeiten $p(\mathbf{v}_{\mathbf{I}})$ auf $\mathbf{v}_{\mathbf{I}}$, $\mathbf{v}_{\mathbf{I}} \in \mathbf{V}_{\mathbf{I}}$, also eine Randverteilung auf $\mathcal{V}_{\mathbf{I}}$, gegeben. Ändern sich nun aufgrund weiterer Information die $p(\mathbf{v}_{\mathbf{I}})$ zu $p'(\mathbf{v}_{\mathbf{I}})$, besagt das MinREnt-Prinzip: $P' = (p'(\mathbf{v}), \mathbf{v} \in \mathbf{V})$ ist eine informationstreue Adaption an $p'(\mathbf{v}_{\mathbf{I}})$, falls es eine Lösung der Aufgabe (2.3) ist.

$$P' = \arg \min R(Q, P) = \sum_{\mathbf{v}} q(\mathbf{v}) \cdot ld \frac{q(\mathbf{v})}{p(\mathbf{v})} \quad (2.3)$$

$$s.d. \quad q(\mathbf{v}_{\mathbf{I}}) = p'(\mathbf{v}_{\mathbf{I}}) \quad \forall \mathbf{v}_{\mathbf{I}} \in \mathbf{V}_{\mathbf{I}}.$$

P' respektiert die Randverteilung $P'_{\mathbf{I}}$ und behält weitestgehend die Bedingungsstruktur von P bei. Die konkrete Berechnung von P' gemäß (2.3) erfolgt wegen seiner speziellen Struktur mittels des sogenannten Proportional Scaling, welches ein Spezialfall des bekannten allgemeineren Iterative Proportional Scaling ist (vgl. (Röd06a)). Zur Bedeutung des Begriffs siehe den Folgeabschnitt.

2.2.2.2 Randverteilungen und Proportional Scaling

Es liege eine Datei mit vollständigen und unvollständigen Daten vor. Die Ideen des Bayes'schen Lernens und des MinREnt-Prinzips werden jetzt in einen Algorithmus zusammengefasst.

Algorithmus 'LernUVollstD'

$a(\mathbf{v}) := \varepsilon \quad \forall \mathbf{v} \in \mathbf{V}$, ε eine beliebig kleine positive Zahl, vgl. auch Anhang B.

Für alle Daten \mathbf{v}_I

Für alle $\mathbf{v}_I \in \mathbf{V}_I$ setze $a(\mathbf{v}_I) = \bar{a}(\mathbf{v}_I) = \sum_{\mathbf{v} \in \mathbf{v}_I} a(\mathbf{v})$

Falls $\mathbf{v}_I = \mathbf{v}_I$ setze $\bar{a}(\mathbf{v}_I) = a(\mathbf{v}_I) + 1$

$Dir(\bar{a}(\mathbf{v}_I), \mathbf{v}_I \in \mathbf{V}_I)$ ist die resultierende Dirichlet auf \mathcal{V}_I

Für alle $\mathbf{v}_I \in \mathbf{V}_I$, für alle $\mathbf{v} \in \mathbf{v}_I$. setze $a(\mathbf{v}) = \frac{a(\mathbf{v})}{a(\mathbf{v}_I)} \cdot \bar{a}(\mathbf{v}_I)$

$Dir(a(\mathbf{v}), \mathbf{v} \in \mathbf{V})$ ist die resultierende Dirichlet auf \mathcal{V}

Der Leser vergegenwärtige sich, dass die $a(\mathbf{v})$ ggfs. nicht ganzzahlig sind, siehe Abschnitt 2.2.1.

$\hat{p}(\mathbf{v}) = \frac{a(\mathbf{v})}{\sum_{\mathbf{v}} a(\mathbf{v})}$, $\mathbf{v} \in \mathbf{V}$, ist der Erwartungswert der aktuellen Dirichlet und als solcher ein Schätzer für $p(\mathbf{v})$, $\mathbf{v} \in \mathbf{V}$.

2.3 Signifikanz erlernter Abhängigkeiten

Sind die – evtl. nicht-ganzzahligen – Parameter $a(\mathbf{v})$, $\mathbf{v} \in \mathbf{V}$, wie im vorigen Abschnitt berechnet, können diese natürlich für jedes $\mathbf{A} \subset \mathbf{V}$ aufaddiert werden. Der Vorgang kann zur Berechnung von Randverteilungen auf vergrößerten Merkmalsräumen Verwendung finden.

Gegeben sei also eine $Dir(a(\mathbf{v}), \mathbf{v} \in \mathbf{V})$. Ferner sei $\mathbf{A} \subset \mathbf{V}$ und $\mathcal{E} = \{\mathbf{E}\}$,

$\bigcup \mathbf{E} = \mathbf{A}$ eine Partition von \mathbf{A} . Dann ist $Dir(\sum_{\mathbf{v} \in \mathbf{E}} a(\mathbf{v}))$, alle \mathbf{E} eine Dirichlet auf der Partition.

Offensichtlich sind alle Teilkonjunkte \mathbf{v}_I zur Indexmenge $I \subset \{1, \dots, L\}$ eine solche Partition. Zu ihr gehört die Dirichlet $Dir(a(\mathbf{v}_I) = \sum_{\mathbf{v} \in \mathbf{V}_I} a(\mathbf{v}))$, alle $\mathbf{v}_I \in \mathbf{V}_I$). Dann sind $\frac{a(\mathbf{v}_I)}{\sum_{\mathbf{v} \in \mathbf{V}} a(\mathbf{v})}$, alle $\mathbf{v}_I \in \mathbf{V}_I$, die Schätzer von $p(\mathbf{v}_I)$, mithin der Randverteilung.

Es seien $\mathbf{B}, \mathbf{A} \subset \mathbf{V}$ beliebige Ereignisse, $\{\mathbf{B}\mathbf{A}, \bar{\mathbf{B}}\mathbf{A}\}$ eine Partition von \mathbf{A} . Dann ist $Beta(s = \sum_{\mathbf{v} \in \mathbf{B}\mathbf{A}} a(\mathbf{v}), t = \sum_{\mathbf{v} \in \bar{\mathbf{B}}\mathbf{A}} a(\mathbf{v}))$ eine spezielle Dirichlet darauf.

$\hat{p}(\mathbf{B}|\mathbf{A}) = \frac{\sum_{\mathbf{v} \in \mathbf{B}\mathbf{A}} a(\mathbf{v})}{\sum_{\mathbf{v} \in \mathbf{A}} a(\mathbf{v})}$ ist Schätzer der bedingten Wahrscheinlichkeit von $\mathbf{B}|\mathbf{A}$. Mittels der $Beta$ lassen sich Unsicherheitsbereiche für das unbekannte $p(\mathbf{B}|\mathbf{A})$ angeben. So ist z. B. der Bereich bei zweiseitiger symmetrischer Abgrenzung zur Sicherheit $S = 1 - \alpha$ so, dass $I_p(s, t) = \frac{\alpha}{2}$ und $I_{\bar{p}}(s, t) = 1 - \frac{\alpha}{2}$. Kalkulationsprogramme und Tabellen stellen die Werte zur Verfügung – für die geläufige Symbolik vergleiche (Gra87) oder auch (Pea56).

Tabelle 2.2 liefert zu verschiedenen $Beta(s, t)$ zur Sicherheit $S = 95\%$ und 99% die entsprechenden Unsicherheitsintervalle.

Beispiel

s	t	S	\underline{p}	\bar{p}
0,5	0,5	0,95	0,0015	0,9985
1	4	0,95	0,0063	0,6024
20	80	0,95	0,1280	0,2834
250	100	0,95	0,6659	0,7603
0,5	0,5	0,99	0,0001	0,9999
1	4	0,99	0,0013	0,7341
20	80	0,99	0,1096	0,3127
250	100	0,99	0,6500	0,7603

Tabelle 2.2: Zweiseitige symmetrische Unsicherheitsgrenzen der $Beta$ zur Sicherheit $S = 95\%$ und $S = 99\%$

Hinweis: Bei ganzzahligen s, t sind die Unsicherheitsgrenzen identisch mit den Vertrauensgrenzen für p der Binomialverteilung für die beobachtete relative Häufigkeit $\hat{p} = \frac{s}{s+t}$ in der Probe, wie der Leser sofort anhand entsprechender Tabellen oder Nomogramme (etwa in (Gra87)) nachvollzieht.

2.4 Mining

Wie in Abschnitt 2.3 ausgeführt und anhand eines Beispiels illustriert wurde, kann nach dem Lernvorgang mit der $Dir(a(\mathbf{v}), \mathbf{v} \in \mathbf{V})$ zu jedem $\mathbf{B}|\mathbf{A}$ die bedingte Wahrscheinlichkeit errechnet und die *Beta* auf $\mathbf{BA}, \overline{\mathbf{BA}}$ bestimmt werden. Mit den gegen Ende von Abschnitt 2.3 getroffenen Feststellungen kann damit $\hat{p}(\mathbf{B}|\mathbf{A})$ errechnet (– Erwartungswert der *Beta* –) und kann der symmetrische Unsicherheitsbereich zur Sicherheit $S = 1 - \alpha$ für $p(\mathbf{B}|\mathbf{A})$ angegeben werden. Das ist eine rudimentäre Form des Mining.

Oft sind jedoch die bedingten Ereignisse nicht vom Anwender vorformuliert, sondern er wünscht die Abhängigkeitsbeziehungen zwischen den Variablen $\{V_1, \dots, V_L\}$ zu entdecken und sucht nach diese Abhängigkeiten repräsentierenden signifikanten Konditionalen.

Zumindest für die Suche nach Abhängigkeitsbeziehungen zwischen Variablen (-gruppen) stellt die Expertensystem-Shell SPIRIT ein probates Mittel zur Verfügung (vgl. (Röd06a)). Zum Verständnis des Instruments bedarf es eines mathematischen Einschubs zum Konzept der **Transinformation**.

Es seien $I, J \subset \{1, \dots, L\}, I \cap J = \emptyset, I \cup J = K$ eine Partition von K und P_K eine beliebige Randverteilung auf \mathcal{V}_K ; P_I, P_J entsprechend. Die Faktorisierung

$$p(\mathbf{v}_K) = p(\mathbf{v}_I) \cdot p(\mathbf{v}_J|\mathbf{v}_I) \quad (2.4)$$

liefert sofort die Gleichung

$$\begin{aligned}
& - \sum_{\mathbf{v}_K} p(\mathbf{v}_K) \cdot \text{ld } p(\mathbf{v}_K) = \\
& - \sum_{\mathbf{v}_I} p(\mathbf{v}_I) \cdot \text{ld } p(\mathbf{v}_I) - \sum_{\mathbf{v}_I} p(\mathbf{v}_I) \cdot \sum_{\mathbf{v}_J} p(\mathbf{v}_J|\mathbf{v}_I) \cdot \text{ld } p(\mathbf{v}_J|\mathbf{v}_I).
\end{aligned} \tag{2.5}$$

Die linke Seite ist bekanntlich die Entropie von P_K auf \mathcal{V}_K , der erste Term rechts die Entropie der Randverteilung auf \mathcal{V}_I , und der zweite Term die bedingte Entropie in P_J bezüglich P_I . Diese bedingte Entropie ist die erwartete Restunsicherheit in P_J bei bekannten Konfigurationen \mathbf{v}_I . (2.5) schreibt man auch (vgl. (Mat96))

$$H(\mathcal{V}_K) = H(\mathcal{V}_I) + H(\mathcal{V}_J|\mathcal{V}_I). \tag{2.6}$$

Die Ausdrücke $H(\mathcal{V}_J)$ und $H(\mathcal{V}_I|\mathcal{V}_J)$ sind völlig analog definiert. Die Größe

$$\begin{aligned}
T(\mathcal{V}_J, \mathcal{V}_I) &= H(\mathcal{V}_J) - H(\mathcal{V}_J|\mathcal{V}_I) \\
&= H(\mathcal{V}_J) + H(\mathcal{V}_I) - H(\mathcal{V}_K) \\
&= H(\mathcal{V}_I) - H(\mathcal{V}_I|\mathcal{V}_J)
\end{aligned} \tag{2.7}$$

ist symmetrisch in $\mathcal{V}_J, \mathcal{V}_I$ und heißt Transinformation zwischen \mathcal{V}_I und \mathcal{V}_J . Sie misst die wechselseitige Information der Variablengruppen \mathcal{V}_I und \mathcal{V}_J übereinander unter P . Ist $T = 0$, so sind $\mathcal{V}_I, \mathcal{V}_J$ unabhängig unter P ; ist T maximal, determinieren sie einander.

Ein großes T ist also ein Indikator für signifikante Regeln mit den Variablen \mathcal{V}_I und \mathcal{V}_J in Prämisse bzw. Konklusion.

Die Shell SPIRIT gestattet die Berechnung von T zu beliebigen Variablengruppen $\mathcal{V}_I, \mathcal{V}_J$. Ein Beispiel erübrigt sich, da in Abschnitt 3.3 das Maß Transinformation in einer Anwendung benutzt wird. In unserem Papier wird das Transinformationsmaß Verwendung finden, nämlich dahingehend, wie stark ein Merkmal=Variable des Produktdesigns die Ergebnisvariable Kundeneinschätzung=Rate beeinflusst. Je größer dieser Einfluss, umso wichtiger die richtige Wahl der Ausprägung des Merkmals für ein gutes Rating. Mehr darüber im folgenden Kapitel.

Kapitel 3

Produktdesign

3.1 Einführung

Aufgrund der zunehmenden Anzahl von Produkten und den immer kürzer werdenden Produktlebenszyklen sehen sich Unternehmungen mit der Problematik konfrontiert, den Erwartungen der (potenziellen) Konsumenten entsprechend, innovative Produkte auf den Markt zu bringen. Mit diesem Innovationsprozess gehen recht aufwändige Produktkonzepttests einher, die die Akzeptanz der potenziellen Konsumenten bezüglich der Produktinnovation ermitteln sollen. Gerade im Konsumgüterbereich wollen Unternehmungen die Floprate von Produktinnovationen verringern, indem sie die Bedürfnisse der Konsumenten ergründen und sowohl im Rahmen der Produktentwicklung als auch bei der Konzeption von Marketingmaßnahmen berücksichtigen. Der Konsument wird in diesem Kontext als Entscheider verstanden, dem verschiedene Produktalternativen zur Wahl vorliegen. Es wird davon ausgegangen, dass der Konsument für alle Paare von Produktvarianten angeben kann, welche Variante er der anderen vorzieht, d.h. welche er präferiert. Der Konsument trägt also eine Präferenzordnung auf den Produktalternativen in sich. Besitzt diese Präferenzordnung bestimmte Eigenschaften (vgl. (Pfa62), S. 19 ff.), so kann aus ihr eine Nutzenfunktion

abgeleitet werden (siehe (Fer75), S. 41 ff.).

Die Präferenzanalyse basiert auf der Annahme, dass jede Eigenschaft (bzw. jede Eigenschaftsausprägung) eines Produktes nutzenstiftend ist und dass sich diese Teilnutzen zu einem Gesamtnutzen (der Produktvariante) aggregieren lassen (vgl. (Sch91), S. 41); die konkrete Aggregationsform ist Untersuchungsgegenstand der Präferenzanalyse. Unter diesem Begriff subsumiert man Verfahren, die sich in kompositionelle und dekompositionelle unterscheiden lassen.

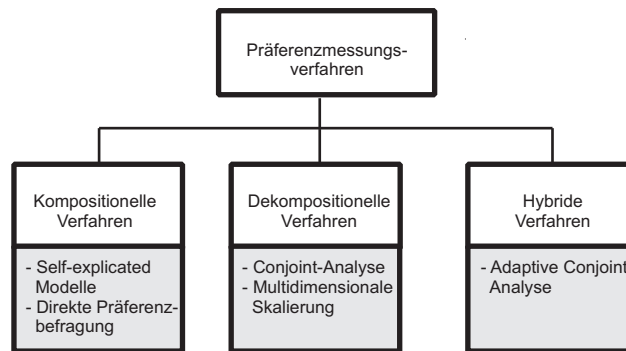


Abbildung 3.1: Präferenzmessungsverfahren

Kompositionelle Verfahren ermitteln und komponieren Präferenzurteile für einzelne Produkteigenschaften und deren Ausprägungen, während dekompositionelle Verfahren selbige aus ganzheitlichen Urteilen mittels statistischer Methoden ableiten. Anfang der 70er Jahre übertrugen GREEN/RAO (vgl. (Gre71)) den aus der Psychologie stammenden dekompositionellen Ansatz der Conjoint-Analyse auf die Analyse der Akzeptanz von Produktkonzepten. Im Folgenden werden wir uns auf diese Conjoint-Analyse kaprizieren, die breite Anerkennung gefunden hat.

3.2 Ein statistischer Ansatz: Conjoint-Analyse

3.2.1 Idee der klassischen Conjoint-Analyse und Annahmen

Die klassische Conjoint-Analyse setzt voraus, dass der Nutzen, den ein Produkt einem Konsumenten stiftet, sich als Aggregat der Teilnutzen der Eigenschaften (bzw. deren Ausprägungen) ergibt (siehe Abschnitt 3.1). Zudem wird die Annahme getroffen, die Merkmale stünden in einer kompensatorischen Beziehung und seien unabhängig (vgl. (Gen03), S. 15). Mittels der Kleinst-Quadrate Methode, englisch: Ordinary Least Squares (OLS), werden die Teilnutzen unter Verwendung der Gesamturteile geschätzt. Auf diese Weise sind Rückschlüsse auf die Wichtigkeit einzelner Eigenschaften bzw. deren Ausprägungen möglich. In der klassischen Conjoint-Analyse werden dem Probanden fiktive Produkte (bzw. Produktvarianten), sogenannte Stimuli, zur Bewertung vorgelegt. Der Proband äußert jedoch keine Auswahlentscheidung, sondern gibt nur Präferenzurteile ab. Mithin können anhand der Ergebnisse der Conjoint-Analyse nur bedingt Aussagen über die Kaufabsicht gemacht werden.

Aufgrund der obigen Annahmen bezüglich ihrer Struktur kann die Nutzenfunktion, die einen funktionalen Zusammenhang zwischen den Eigenschaften (bzw. deren Ausprägungen) und dem jeweiligen Gesamtnutzen der Produktvariante herstellt, wie folgt formuliert werden (vgl. (Voe00), S. 37) :

$$u_{k,j} = \phi(v_{k,j,l,m}(x_{j,l,m})) \quad \forall k \in \mathcal{K}, j \in \mathcal{J} \quad (3.1)$$

mit

\mathcal{K}	Indexmenge der Konsumenten; $k = 1, \dots, K$
\mathcal{J}	Indexmenge der Produktvarianten (=Stimuli); $j = 1, \dots, J$
\mathcal{L}	Indexmenge der Eigenschaften; $l = 1, \dots, L$
\mathcal{M}_l	Indexmenge der Ausprägungen der l -ten Eigenschaft; $m = 1, \dots, M_l$
$u_{k,j}$	Gesamtnutzenwert der j -ten Produktvariante für den k -ten Konsumenten
$\phi(\cdot)$	Verknüpfungsfunktion der Teilnutzen der $ \mathcal{L} $ Eigenschaften und $ \mathcal{M}_l $ Ausprägungen
$v_{k,j,l,m}$	Teilnutzen der m -ten Ausprägung der l -ten Eigenschaft der j -ten Produktvariante
$x_{j,l,m}$	Wert der m -ten Ausprägung der l -ten Eigenschaft der j -ten Produktvariante

Die Verknüpfungsfunktion gibt eine Aggregationsvorschrift für die Teilnutzen der einzelnen Eigenschaften bzw. deren Ausprägungen zu dem Gesamtnutzenwert an. Es wurde bereits erwähnt, dass in der klassischen Conjoint-Analyse die Annahme getroffen wird, die Merkmale seien unabhängig und kompensatorisch; daher wird stets vereinfachend eine linear-additive Funktion als Verknüpfungsfunktion unterstellt.

Sind Merkmale eines Produktes nominal skaliert, werden sie als $(0, 1)$ -Variable codiert und ihr Nutzenbeitrag unter Anwendung des Teilnutzenwertmodells geschätzt (vgl. (Gen03), S. 19). Auf diese Weise kann eine Bewertungsfunktion auch für die Ausprägungen von nominal skalierten Eigenschaften bestimmt werden (andere Bewertungsfunktionen wie das Idealpunkt-Modell sind bspw. in (vT01), S. 13, dargestellt).

Somit hat man:

$$u_{k,j} = \beta_{k,0} + \sum_{l \in \mathcal{L}, m \in \mathcal{M}_l} \beta_{k,j,l,m} \cdot x_{j,l,m} \quad \forall k \in \mathcal{K}, j \in \mathcal{J} \quad (3.2)$$

mit

$u_{k,j}$	Gesamtnutzenwert der j -ten Produktvariante für den k -ten Konsumenten
$\beta_{k,0}$	Regressionskonstante der Nutzenfunktion für den k -ten Konsumenten
$\beta_{k,j,l,m}$	Regressionskoeffizient der m -ten Ausprägung der l -ten Eigenschaft der j -ten Produktvariante für den k -ten Konsumenten
$x_{j,l,m}$	Wert der m -ten Ausprägung der l -ten Eigenschaft der j -ten Produktvariante

wobei

$$x_{j,l,m} = \begin{cases} 1, & \text{falls die } m\text{-te Ausprägung der } l\text{-ten Eigenschaft} \\ & \text{bei der } j\text{-ten Produktvariante vorhanden ist,} \\ 0 & \text{sonst} \end{cases}$$

Für die Bewertung der Produktvarianten (=Stimuli), also für die Spezifizierung der Nutzen $u_{k,j}$, werden in der Literatur verschiedene Ansätze diskutiert (vergleiche Abbildung 3.2). Grundsätzlich unterscheiden sich die Ansätze darin, ob die erfassten Werte der Nutzen ordinal oder kardinal skaliert sind. Während bei den nicht-metrischen Ansätzen die Probanden lediglich die Vorzugswürdigkeit eines Stimulus gegenüber einem anderen äußern müssen, fordern die metrischen Ansätze dem Probanden ab, auch die Stärke der Präferenz zu formulieren. Dem höheren Informationsgehalt der Ergebnisse der metrischen Ansätze steht die geringere Reliabilität entgegen, da Probanden häufig nicht zuverlässig die Stärke ihrer Präferenz zum Ausdruck bringen können (zur Diskussion vgl. bspw. (Gen03), S. 25 ff.).

In der Literatur haben sich das Ranking als nicht-metrischer Ansatz sowie das Rating als metrischer Ansatz durchgesetzt. Insbesondere das Rating besitzt den Vorteil, dass die Schätzung der Nutzenparameter mittels (linearer) Regressionsanalyse zulässig ist, da der Rate als abhängige Variable kardinal skaliert ist. Wir folgen diesem Ansatz und lassen den Probanden k ein

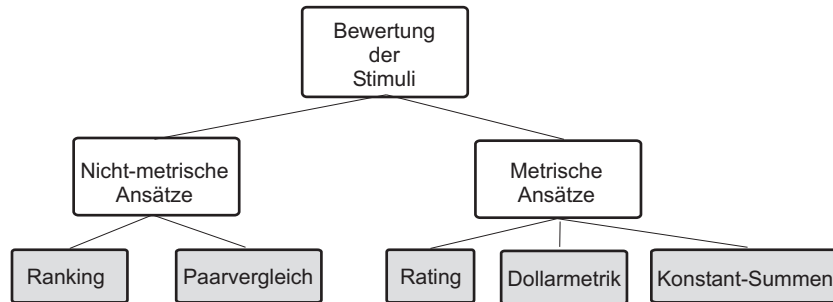


Abbildung 3.2: Bewertungsmethoden von Stimuli in der Conjoint-Analyse

Werturteil zur Produktvariante j in Form des Rates $r_{k,j}$ abgeben.

3.2.2 OLS-Schätzung und Gütekriterien

3.2.2.1 Beurteilung der Güte der Schätzung

Ist die Bewertung der Stimuli durch die Probanden abgeschlossen, erfolgt die Schätzung der Parameter, die den Beitrag einer Eigenschaft (bzw. ihrer Ausprägung) zum Rate bestimmt. Die Anwendung der OLS-Methode setzt voraus, dass sowohl die abhängige Variable wie auch die unabhängigen kardinal skaliert sind; wie schon in 3.2.1 erläutert, wird diese Voraussetzung bei der Bewertungsmethode Rating als gegeben angenommen. Der Rate r lässt sich als Nutzen mit speziellen Eigenschaften interpretieren. So ist $r \in \mathbb{N}_0^+$ und liegt im Intervall $[a, b]$ mit $a < b$. Ist $a = 1$ und $b = |\mathcal{J}|$, so entspricht die im Rating ermittelte Ordnung auf den Produktvarianten der Rangordnung. Empirisch hat sich jedoch gezeigt, dass Konsumenten für $|\mathcal{J}| > 20$ Stimuli ihre Präferenzordnung nicht mehr zum Ausdruck bringen können, sodass häufig $b \ll |\mathcal{J}|$ gewählt wird.

Aus (3.2) leitet sich demnach (3.3) ab:

$$\hat{r}_{k,j} = \hat{\beta}_{k,0} + \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}_l} \hat{\beta}_{k,j,l,m} \cdot x_{j,l,m} \quad \forall k \in \mathcal{K}, j \in \mathcal{J} \quad (3.3)$$

mit

$\hat{r}_{k,j}$	Geschätzter Rate der j -ten Produktvariante für den k -ten Konsumenten
$\hat{\beta}_{k,0}$	Geschätzte Regressionskonstante der Ratefunktion für den k -ten Konsumenten
$\hat{\beta}_{k,j,l,m}$	Geschätzter Regressionskoeffizient der m -ten Ausprägung der l -ten Eigenschaft der j -ten Produktvariante für den k -ten Konsumenten
$x_{j,l,m}$	Wert der m -ten Ausprägung der l -ten Eigenschaft der j -ten Produktvariante

wobei

$$x_{j,l,m} = \begin{cases} 1, & \text{falls die } m\text{-te Ausprägung der } l\text{-ten Eigenschaft} \\ & \text{bei der } j\text{-ten Produktvariante vorhanden ist,} \\ 0 & \text{sonst} \end{cases}$$

Die Parameter der Ratingfunktion des Konsumenten sollen geschätzt werden. Als Kriterium dient die Abweichung zwischen empirisch beobachteten und gemäß (3.3) geschätzten Rates, also der Fehler der Schätzung. Bei der OLS-Schätzung sollen die quadrierten Abweichungen zwischen den empirisch beobachteten und den geschätzten Rates der Konsumenten minimiert werden.

$$\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} (r_{k,j} - \hat{r}_{k,j})^2 \rightarrow \text{Min!} \quad (3.4)$$

Der Schätzung der Regressionskoeffizienten schließt sich die Beurteilung der Güte der Schätzung an, die anhand der folgenden Kriterien vorgenommen wird (vgl. bspw. (Dec02), S. 272 ff.):

- Inhaltsvalidität (synonym: Face Validity): Grad der inhaltlichen Genauigkeit des verwendeten Modells. Sie misst, inwieweit das geschätzte Modell die a priori Erwartungen der Experten widerspiegelt und wie signifikant die Schätzungen sind.

- **Kriteriumsvalidität:** Höhe der Korrelation der Messungen zweier Konstrukte, die theoretisch in einem engen Verhältnis stehen. Unterschieden wird nach dem Zeitpunkt der Erhebung der Konstrukte zwischen
 - **Prognosevalidität:** Grad, zu dem aufgrund der Messungen des einen Konstruktes die Messungen eines kausal abhängigen, zeitlich nachgelagerten Konstruktes prognostiziert werden können. Im Rahmen der Produktkonzeptionsforschung bestimmt man u. a. die Güte der Parameter-Schätzungen anhand der später erfolgten realen Käufe der Probanden (bspw. in einem Test-Store).
 - **Konkurrentvalidität:** Grad, zu dem aufgrund der Messungen des einen Konstruktes die Messungen eines kausal abhängigen, gleichzeitig erhobenen, Konstruktes erklärt werden können. Zu dem Zwecke der Bestimmung der Konkurrentvalidität werden sogenannte Hold-Out Samples von den Konsumenten beurteilt, die nicht in die Schätzung eingehen. Nach der Analyse kann anhand der Differenz zwischen geschätztem und 'realem' Rate der Hold-Outs die Güte der Schätzung bestimmt werden.
- **Reliabilität:** Grad der Genauigkeit der Schätzung. Die beobachteten Rates sollen durch die geschätzten möglichst gut reproduziert werden.

Die **Inhaltsvalidität** wird bei der OLS-Schätzung durch einen Linearitätstest bestimmt. Die Nullhypothese lautet, dass kein linearer Zusammenhang zwischen den Rates r und den Eigenschaftsausprägungen besteht (vgl. (Moo78), S. 68 ff.):

$$H_0 : r_{k,j} \neq \beta_{k,0} + \sum_{l \in \mathcal{L}, m \in \mathcal{M}_l} \beta_{k,j,l,m} \cdot x_{j,l,m}.$$

Entsprechend lautet die Gegenhypothese:

$$H_1 : r_{k,j} = \beta_{k,0} + \sum_{l \in \mathcal{L}, m \in \mathcal{M}_l} \beta_{k,j,l,m} \cdot x_{j,l,m}.$$

Eine Ablehnung der Nullhypothese bedeutet, dass ein linearer Zusammenhang zwischen der abhängigen und den unabhängigen Variablen besteht. Dieser Test kann auf einen Test auf Varianz zurückgeführt werden. Besteht ein perfekt-linearer Zusammenhang, so ist die erklärte Varianz gleich der Gesamtvarianz. Würde H_0 nicht abgelehnt werden können, so würden sich die unabhängigen Variablen nicht zur Erklärung der Varianz eignen; die multiple Bestimmtheit würde in diesem Modell also Null sein und das Erklärungsmodell besäße keine Gültigkeit.

Der Testwert F wird berechnet zu (vgl. (Moo02), S. 93 ff.)

$$F = \frac{Q_d/o}{Q_e/(N - o)} \quad (3.5)$$

mit

$\hat{r}_{k,j}$ Geschätzter Rate mit $\hat{r}_{k,j} = \hat{\beta}_{k,0} + \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}_l} \hat{\beta}_{k,j,l,m} \cdot x_{j,l,m}$

\bar{r} Durchschnittlicher Rate mit $\bar{r} = \frac{1}{K} \sum_k \sum_j r_{k,j}$

Q_d Erklärte Varianz, d. h. Summe der quadratischen Abweichungen der geschätzten Rates von den durchschnittlichen Rates

mit

$$Q_d = \sum_k \sum_j (\hat{r}_{k,j} - \bar{r})^2$$

o Anzahl der unabhängigen Variablen, die r erklären

Q_e Nicht erklärte Varianz, d. h. Summe der quadratischen Abweichungen der beobachteten Rates von den geschätzten Rates mit

$$Q_e = \sum_k \sum_j (r_{k,j} - \hat{r}_{k,j})^2$$

N Anzahl der Beobachtungen mit

$$N = |\mathcal{K}| \cdot |\mathcal{J}|$$

Die Nullhypothese H_0 wird abgelehnt, falls der Testwert $F > F_{f_1, f_2; 1-\alpha}$ mit $f_1 = o, f_2 = N - o$ ist; F_{f_1, f_2} ist das $(1 - \alpha)$ -Quantil der F -Verteilung mit (f_1, f_2) Freiheitsgraden. Die Werte für $F_{f_1, f_2; 1-\alpha}$ können statistischen

Tafelwerken entnommen werden (wie bspw. (Gra87)). Über diesen Hypothesentest ist es ebenfalls möglich, Aussagen bezüglich einzelner Variabler zu bestätigen bzw. abzulehnen. So kann über eine entsprechende Anpassung der Nullhypothese die Signifikanz des Einflusses einzelner Eigenschaftsausprägungen auf die abhängige Variable Rate bestimmt werden. Insbesondere ist es dadurch auch möglich, eine Modellreduktion vorzunehmen, d. h. weniger relevante Eigenschaften zu eliminieren.

Zur Bestimmung der **Kriteriumsvalidität** wird die First Choice-Regel angewendet; diese besagt, dass lediglich das Produkt mit dem höchsten Rate auch gewählt wird. Somit kann auf aggregierter Ebene die Korrelation zwischen empirisch beobachteten (durch Kauf oder Bewertung der Hold-Out Samples) und geschätzten First Choices fest gestellt werden.

Eine Aussage bezüglich der **Reliabilität** kann unter Verwendung des Bestimmtheitsmaßes gemacht werden. Das Bestimmtheitsmaß gibt den Anteil der erklärten Varianz an der Gesamtvarianz wieder.

$$R^2 = \frac{\sum_k \sum_j (\hat{r}_{k,j} - \bar{r}_k)^2}{\sum_k \sum_j (r_{k,j} - \bar{r}_k)^2} \quad (3.6)$$

Die Notation lehnt sich an die zu (3.5) an.

Die obigen Ausführungen sollen anhand eines Beispiels veranschaulicht werden.

3.2.2.2 Beispiel zur OLS-Schätzung

Ein Automobilhersteller möchte ein innovatives Fahrzeugdesign finden, Tabelle 3.1 gibt die Eigenschaften und ihre Ausprägungen wieder.

Mittels einer Regressionsfunktion kann der aggregierte Rate für das j -te Produkt geschätzt werden. Es wird die vereinfachende Annahme getroffen, die Konsumenten seien hinsichtlich ihrer soziodemographischen und anderen einstellungsrelevanten Merkmalen ähnlich, sodass keine Clusterung

l	Code Eigenschaft	Eigenschaft	Ausprägungen	Code Ausprägung
1	<i>IMAGE</i>	Image	<i>modern/ konservativ</i>	1/0
2	<i>TYPE</i>	Typ	<i>Limousine/Kombi</i>	1/0
3	<i>SHAPE</i>	Gestalt	<i>kantig/ weich</i>	1/0
4	<i>STANDEQUIP</i>	Standardausstattung	<i>umfangreich/ gering</i>	1/0
5	<i>LIGHTCONST</i>	Leichtbauweise	<i>ja/ nein</i>	1/0
6	<i>INTSPACE</i>	Innenraum	<i>großzuegig/ klein</i>	1/0
7	<i>CONTROLELE</i>	Kontrollelemente	<i>viele/ wenige</i>	1/0

Tabelle 3.1: Eigenschaften und deren Ausprägungen für das Beispiel 'Automobil'

notwendig ist. Daher können die Parameter der aggregierten Ratefunktion (3.3) auf Basis aller Konsumentenurteile geschätzt werden; somit gelten die Ausführungen des Abschnitts 3.2.1.

Für die aggregierte Ratefunktion (3.3) erhält man den Ausdruck (3.7).

$$\begin{aligned}
\hat{r}_j = & \hat{\beta}_0 + \hat{\beta}_{IMAGE} \cdot x_{j,IMAGE,modern} \\
& + \hat{\beta}_{TYPE} \cdot x_{j,TYPE,Limousine} \\
& + \hat{\beta}_{SHAPE} \cdot x_{j,SHAPE,kantig} \\
& + \hat{\beta}_{STANDEQUIP} \cdot x_{j,STANDEQUIP,umfangreich} \quad \forall j \in \mathcal{J} \quad (3.7) \\
& + \hat{\beta}_{LIGHTCONST} \cdot x_{j,LIGHTCONST,ja} \\
& + \hat{\beta}_{INTSPACE} \cdot x_{j,INTSPACE,großzuegig} \\
& + \hat{\beta}_{CONTROLELE} \cdot x_{j,CONTROLELE,viele}
\end{aligned}$$

Die Datenbasis, die der Schätzung zugrunde liegt, beinhaltet simulierte Urteile von $K = 200$ fiktiven Probanden zu den $2^7 = 128$ Produktvarianten. Das jeweilige Urteil wurde gemäß einer individuellen linear-additiven Ratefunktion, deren Parameter zufällig (und damit unabhängig) erzeugt wur-

den, berechnet. Das Ergebnis dieser fiktiven Beurteilung ist also ein zu jedem Kunden und zu jedem Stimulus mittels einer zufälligen linearen Nutzenfunktion erzeugter Rate. Diese 8-Tupel werden nun zur Schätzung der Regressionskoeffizienten unter Anwendung einer linearen Regressionsanalyse in der gängigen Software SPSS herangezogen. Die folgenden Abschnitte behandeln zwei 'Typen' von Daten, die vorliegen können: Vollständige und unvollständige.

3.2.2.3 Ergebnis der OLS-Schätzung und Interpretation

3.2.2.3.1 OLS-Schätzung für den Fall vollständiger Daten

Es wurden die in Tabelle 3.2 notierten Regressionskoeffizienten einer aggregierten Ratefunktion mittels Regressionsanalyse geschätzt.

Anhand der geschätzten Regressionskoeffizienten können nun Aussagen bezüglich der Gestaltung des optimalen Produktdesigns gemacht werden. Geht man davon aus, dass alle Eigenschaften relevant sind, so müssen zur Bestimmung einer konkreten Produktvariante deren sämtliche Merkmalsausprägungen festgelegt werden, und zwar in der Reihenfolge ihrer Wichtigkeit. Damit ist eine Rangordnung auf den Eigenschaften gemäß ihrer Erklärung der abhängigen Variablen Rate sinnvoll. Eine Rangordnung gemäß dieses Kriteriums ist ebenfalls in Tabelle 3.2 verzeichnet.

Das Bestimmtheitsmaß gibt an, dass die Reliabilität der Schätzung sehr hoch ist. Zudem gilt $F = 15729,705 > F_{f_1=7, f_2=25593, \alpha=0,05} = 2,01$, so dass H_0 zu verwerfen ist und damit ein linearer Zusammenhang zwischen Rate und den Eigenschaftsausprägungen (statistisch) nachgewiesen ist (was den Leser nicht weiter verwundert, da die Datenbasis eben wie weiter oben beschrieben erzeugt wurde).

Variable	$\hat{\beta}$	Rang
Konstante	7,963	
<i>IMAGE</i>	1,407	2
<i>TYPE</i>	0,938	5
<i>SHAPE</i>	-0,364	7
<i>STANDEQUIP</i>	-1,021	4
<i>LIGHTCONST</i>	-1,605	1
<i>INTSPACE</i>	-1,219	3
<i>CONTROLELE</i>	-0,501	6
<hr/>		
$R_{\text{vd}}^2=0,811$		
$Q_{\text{d}} = 53415,596$		
$Q_{\text{e}} = 12415,684$		
$F = 15729,705$		

Tabelle 3.2: Ergebnis der OLS-Schätzung der Nutzenparameter für das Beispiel 'Automobil' in dem Fall vollständiger Daten

Die in Tabelle 3.2 notierten Parameter beziehen sich auf die Merkmalsausprägung, die mit '1' codiert ist. Daher wird der Designer mit dem Ziel der Maximierung des Konsumentennutzens bei negativem Vorzeichen des Parameters die Merkmalsausprägung für das absatzkräftige Produktdesign festlegen, welche mit '0' codiert ist.

Das optimale Produktdesign gemäß der Schätzung der Parameter der aggregierten Ratefunktion ist also

$$\begin{aligned}
 \text{BestDesign} = \{ & \textit{IMAGE} = \textit{modern}, \textit{TYPE} = \textit{Limousine}, \\
 & \textit{SHAPE} = \textit{weich}, \textit{STANDEQUIP} = \textit{gering}, \\
 & \textit{LIGHTCONST} = \textit{nein}, \textit{INTSPACE} = \textit{klein}, \\
 & \textit{CONTROLELE} = \textit{wenige} \}.
 \end{aligned}$$

Der folgende Abschnitt widmet sich dem Fall unvollständiger Daten, der in

der Empirie relevanter ist.

3.2.2.3.2 OLS-Schätzung für den Fall unvollständiger Daten

Im Folgenden wird erneut eine OLS-Schätzung mittels der erzeugten Datenbasis vorgenommen, allerdings ist diese Datenbasis bezogen auf die Variable 'Rate' lückenhaft, d.h. es wird davon ausgegangen, dass die Probanden gewisse Stimuli nicht bewerten konnten; 20% der Rates sind nicht gegeben. Erneut werden die Parameter der Regressionsfunktion (3.7) geschätzt.

Variable	Schätzung von $\hat{\beta}$		Rang
	mit fallweisem Ausschluss	Ersetzung der Missings durch Mittelwert	
Konstante	7,967	7,724	
<i>IMAGE</i>	1,403	1,117	2
<i>TYPE</i>	0,943	0,751	5
<i>SHAPE</i>	-0,363	-0,289	7
<i>STANDEQUIP</i>	-1,025	-0,823	4
<i>LIGHTCONST</i>	-1,603	-1,277	1
<i>INTSPACE</i>	-1,222	-0,970	3
<i>CONTROLELE</i>	-0,507	-0,411	6
R_{unvDA}^2 bzw. R_{unvDM}^2	0,812	0,647	
$Q_{\text{d}_{\text{unvDA}}}$ bzw. $Q_{\text{d}_{\text{unvDM}}}$	42654,610	34000,347	
$Q_{\text{e}_{\text{unvDA}}}$ bzw. $Q_{\text{e}_{\text{unvDM}}}$	9903,199	18557,463	
F_{unvDA} bzw. F_{unvDM}	12594,736	6698,659	

Tabelle 3.3: Ergebnis der OLS-Schätzung der Nutzenparameter für das Beispiel 'Automobil' in dem Fall unvollständiger Daten

Die Methode des fallweisen Ausschlusses führt zu einer Löschung von unvollständigen Datensätzen. Die Schätzung der Lücken, englisch: Missing

Values, durch das arithmetische Mittel der 'umliegenden' gültigen Variablenwerte führt zu einer Verzerrung der Datenlage und ist nur bei kardinal skalierten Merkmalen überhaupt zulässig. Bei nominal skalierten Variablen, beispielsweise bei soziodemographischen Merkmalen, können solche Missing Values nicht ersetzt werden. Für das vorliegende Beispiel wurden die Regressionskoeffizienten mit den für unvollständigen Daten möglichen Optionen 'Fallweiser Ausschluss' und 'Ersetzung durch Mittelwert' geschätzt; die Schätzungen erfolgten auf Basis der 80% vollständigen Daten und sind in Tabelle 3.3 notiert. Es ist erkennbar, dass erneut das optimale Produktdesign bestimmt wird (vgl. Abschnitt 3.2.2.3.1); die Güte der Schätzung im Falle unvollständiger Daten unter Verwendung der Option 'Ersetzung der Missing Values durch Mittelwert' ist schlechter als bei der Regressionsanalyse auf Basis vollständiger Daten (vgl. Tabelle 3.2 und 3.3, $R_{\text{vD}}^2 = 0,811 > R_{\text{unvDM}}^2 = 0,647$). Findet ein fallweiser Ausschluss statt, so wird die Datenbasis um die unvollständigen Datensätze reduziert; aufgrund der umfangreichen Datenbasis ändert sich in diesem Falle das Bestimmtheitsmaß kaum ($R_{\text{vD}}^2 = 0,811, R_{\text{unvDA}}^2 = 0,812$). Ein linearer Zusammenhang wird unabhängig von der gewählten Option attestiert ($F_{\text{unvDM}} = 6698,659$ bzw. $F_{\text{unvDA}} = 12594,736$, vgl. Tabelle 3.3).

3.2.3 Übertragung der Ergebnisse auf das Kaufverhalten

In den Abschnitten 3.2.2.3.1 und 3.2.2.3.2 wurden die Parameter einer aggregierten linearen Nutzenfunktion auf Basis der zufällig erzeugten Daten mittels linearer Regressionsanalyse geschätzt. Wie schon in Abschnitt 3.2.1 erwähnt, kann von diesen Ergebnissen noch nicht auf die Kaufentscheidung der Konsumenten geschlossen werden. Unterstellt man eine 'Entscheidungsregel' wie die First Choice-Regel, so kann unter bestimmten Annahmen eine Aussage bezüglich des Wahlverhaltens der Konsumenten gemacht werden.

Unterstellt man ferner die Vollständigkeit und Unverrauschtheit der über die Gesamturteile ermittelten Teilnutzenwerte, so ist der Schluss zulässig, nur das Produkt mit dem höchsten Nutzen werde auch gewählt und – unter Ausschluss weiterer kaufrelevanter Faktoren – gekauft. Daher ist die Wahrscheinlichkeit, mit der der Konsument k die Produktvariante g mit dem höchsten Nutzen auch kauft, gleich Eins und Null, sonst (vgl. (Tei01), S. 72).

$$\begin{aligned}
 p_k(g|\mathcal{J}) &= 1, & \text{genau wenn } u_{k,g} &= \max_j u_{k,j}, j \in \mathcal{J} \\
 p_k(j|\mathcal{J}) &= 0 & \forall j \in \mathcal{J} \setminus \{g\}
 \end{aligned}$$

Interpretiert man die Wahrscheinlichkeit $p_k(g|\mathcal{J}) = 1$ als sichere Entscheidung für den Kauf des Produkts, so kann für jede Produktvariante 'ausgezählt' werden, wieviele der Konsumenten diese Variante den anderen vorziehen. Auf diese Weise können Aussagen über die Absatzkraft der Produktvariante gemacht werden; für eine realistische Einschätzung müssen allerdings die Konkurrenzprodukte in die Beurteilung mit einbezogen werden. Erweitert man den Ansatz um die Konkurrenzprodukte, so liegt der in der Literatur unter dem Namen 'Share of Choices' firmierende vor.

Die Idee der First Choice-Regel und ihre theoretische Verwertung von Konsumentenurteilen wird im folgenden informationstheoretischen Ansatz wieder aufgegriffen werden und an entsprechender Stelle ein Rückvermerk gemacht.

Abschnitt 3.2 widmete sich der Darstellung der Idee statistischer Verfahren, die in der Conjoint Analyse verwendet werden. Die Beobachtungsdateien wurden unter der Annahme kompensatorischer Beziehungen und daher der Unabhängigkeit der Eigenschaften und linearem Einfluss auf den Rate erzeugt, obwohl empirisch durchaus komplexere Beziehungen nachgewiesen sind (vgl. (Voe00), S. 57 f.). Selbst in diesem linearen Fall wurde bei unvollständigen Daten die Erklärungsgüte des Modells geringer. Aus die-

sen Gründen wird in Abschnitt 3.3 ein Ansatz vorgestellt, der es gestattet, nach Abhängigkeiten innerhalb der Datenbasis zu suchen, ohne dass ein Erklärungsansatz unterstellt werden muss. Dieser – informationstheoretische – Ansatz ist in seiner technischen Umsetzung in der Expertensystem-Shell SPIRIT implementiert, die im Weiteren Verwendung findet (vgl. (Röd06a)).

3.3 Ein informationstheoretischer Ansatz: MinREnt

3.3.1 Idee und Modellierung

Gegeben seien die $N = |\mathcal{K}| \cdot |\mathcal{J}|$ Beobachtungen \mathbf{v} , die die Beurteilung der K Probanden bezüglich der J Stimuli beinhalten. In der Terminologie des Abschnitts 2.1 sei die endliche Menge endlichwertiger Variabler $\mathcal{V} = \{V_1, \dots, V_L\}$ gegeben, jetzt aber ergänzt um die Variable *RATE* mit Ausprägungen r_i : $\mathcal{V}' = \mathcal{V} \cup \text{RATE}$. Die Variable *RATE* sei 10-wertig, $\text{RATE} = (r_1 = 1, r_2 = 2, \dots, r_{10} = 10)$ (vgl. Abschnitt 3.2.2.1); die Mächtigkeit der Menge \mathbf{V}' ist somit $|\mathbf{V}'| = 10 \cdot 2^7 = 1280$.

Erhebt man die Beurteilungen von $K = 200$ fiktiven Probanden, so ergeben sich $N = 2^7 \cdot 200 = 25600$ Beobachtungen.

Zunächst gehen wir wieder wie in Abschnitt 3.2 von vollständigen Daten aus. Gemäß den Darstellungen in Abschnitt 2.1.2 folgen die Häufigkeiten auf den Konfigurationen einer Multinomial (2.2). Die dazu konjugierte Verteilung ist die Dirichlet (2.1). Der Lernvorgang aus den vollständigen Daten erfolgt gemäß dem Algorithmus 'LernVollstD', siehe Seite 9. Zur Illustration gibt Tabelle 3.4 einen Ausschnitt der Beobachtungsdatei für $K = 200$ Probanden zum Beispiel 'Automobil' wieder.

Für jede Beobachtung \mathbf{v}' werden gemäß Algorithmus 'LernVollstD' die Wahrscheinlichkeiten $\hat{p}(\mathbf{v}')$ für die Konfigurationen \mathbf{v}' über den Erwartungswert der Dirichlet gemäß $\hat{p}(\mathbf{v}') = \frac{a(\mathbf{v}')}{\sum_{\mathbf{v}'} a(\mathbf{v}')} , \mathbf{v}' \in \mathbf{V}'$, geschätzt.

Laufende Nr.	IMAGE	TYPE	SHAPE	Stimuli				RATE
				STAND- EQUIP	LIGHT- CONST	INT- SPACE	CONTROL- ELE	
1	1	1	1	0	0	0	1	10
3	1	1	1	1	0	0	1	9
4	1	1	0	1	0	0	0	8
2	1	1	0	0	0	0	1	10
5	0	1	1	0	0	1	1	7
6	1	1	0	0	0	0	0	10
⋮	⋮			⋮			⋮	⋮
25600	0	0	0	1	1	1	1	2

Tabelle 3.4: Ausschnitt Beobachtungsdatei für das Beispiel 'Automobil' in dem Fall vollständiger Daten

Mit den geschätzten Wahrscheinlichkeiten $\hat{p}(\mathbf{v}')$ können nun Konditionale der Form $RATE|\mathbf{v}$ ausgewertet werden. Es sei kurz angemerkt, dass hier nun eine empirische Wahrscheinlichkeitsverteilung vorliegt. Unter dieser Wahrscheinlichkeitsverteilung sind die Konfigurationen $\mathbf{v}' \in \mathbf{V}'$ 'auszuwerten'. Hinsichtlich des Produktdesigns muss nun geklärt werden, wann ein Produktdesign 'gut' bzw. 'optimal' ist. Wie bereits erläutert, muss also eine Entscheidungsregel wie bspw. die First Choice-Regel eingeführt werden; diese impliziert die Annahme, dass nur dann ein Produkt gekauft wird, wenn es den Rate 10 erhalten hat. Zu suchen ist also nach einer Konfiguration \mathbf{v} , für die die meisten Probanden einen $RATE$ von 10 gegeben haben. Man betrachtet somit den Wahrheitsgehalt der Aussage $RATE = 10|\mathbf{v}$, alle $\mathbf{v} \in \mathbf{V}$, für die Population; eine solche Aussage könnte bspw. $RATE = 10|(\overline{IMAGE} \wedge \overline{TYPE} \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE})$ sein.

Man fokussiere auf eine Konfiguration \mathbf{v} , $\mathbf{v} \in \mathbf{V}$. Es seien \mathbf{A} , \mathbf{B} die zwei Ereignisse mit $\mathbf{A} = \{\mathbf{v}'|\mathbf{v}' \text{ hat Merkmale } \mathbf{v}\}$ sowie $\mathbf{B} = \{\mathbf{v}'|\mathbf{v}' \text{ hat RATE}=10\}$.

Laufende Nr.	IMAGE	TYPE	SHAPE	Konfiguration \mathbf{v}'				RATE	$a(\mathbf{v}')$	$\frac{a(\mathbf{v}')}{\sum_{\mathbf{v}'} a(\mathbf{v}')}$
				STAND- EQUIP	LIGHT- CONST	INT- SPACE	CONTROL- ELE			
1	0	0	0	0	0	0	0	10	$\frac{\varepsilon}{1280 \cdot \varepsilon + 25600}$	0,00000
2	0	0	0	0	0	0	0	9	$\frac{43 + \varepsilon}{1280 \cdot \varepsilon + 25600}$	0,00168
3	0	0	0	0	0	0	0	8	$\frac{112 + \varepsilon}{1280 \cdot \varepsilon + 25600}$	0,00437
4	0	0	0	0	0	0	0	7	$\frac{44 + \varepsilon}{1280 \cdot \varepsilon + 25600}$	0,00172
\vdots	\vdots			\vdots			\vdots	\vdots	\vdots	\vdots
970	1	1	0	0	0	0	0	10	$\frac{200 + \varepsilon}{1280 \cdot \varepsilon + 25600}$	0,00781
\vdots	\vdots			\vdots			\vdots	\vdots	\vdots	\vdots
1280	1	1	1	1	1	1	1	1	$\frac{\varepsilon}{1280 \cdot \varepsilon + 25600}$	0,00000

Tabelle 3.5: Ausschnitt Konfigurationen mit Zählstatistik für das Beispiel 'Automobil' für $\varepsilon = 0,0001$ bei vollständigen Daten

Somit kann gemäß den Ausführungen in Abschnitt 2.3 $\hat{p}(\mathbf{B}|\mathbf{A}) = (RATE = 10|\mathbf{v})$ bestimmt werden zu $\hat{p}(\mathbf{B}|\mathbf{A}) = \frac{\hat{p}(\mathbf{AB})}{\hat{p}(\mathbf{A})} = \frac{\sum_{\mathbf{v}' \in \mathbf{AB}} a(\mathbf{v}')}{\sum_{\mathbf{v}' \in \mathbf{A}} a(\mathbf{v}')}$.

Das optimale Produktdesign \mathbf{v}^* gemäß der Entscheidungsregel First Choice kann über eine vollständige Enumeration identifiziert werden, indem die bedingten Wahrscheinlichkeiten $\hat{p}(RATE = 10|\mathbf{v})$, für alle $\mathbf{v} \in \mathbf{V}$, ermittelt werden und die Produktvariante mit der höchsten Wahrscheinlichkeit für $RATE = 10$ identifiziert wird; dieser Zusammenhang wird in (3.8) angegeben.

$$\mathbf{v}^* = \arg \max_{\mathbf{v}} p(RATE = 10|\mathbf{v}), \mathbf{v} \in \mathbf{V} \quad (3.8)$$

Diese Vorgehensweise ist jedoch schon bei diesen $2^7 = 128$ Konfigurationen aufwändig.

Die folgende Beziehung zwischen bedingten Wahrscheinlichkeiten ist eine Konsequenz des Satzes von Bayes, die uns im Weiteren zur Bestimmung eines optimalen Produktdesigns mit geringem Rechenaufwand weiterhilft. $V_l = v_l$ sei die Realisierung einer Merkmalsvariablen l . Unter Annahme der Gültigkeit der First Choice-Regel ist es zulässig, hinsichtlich der Bewertung

einer Produktvariante nur zwischen $RATE = 10$ und $\overline{RATE = 10}$ zu differenzieren. Die bereits erwähnte Konsequenz lautet nun:

$$\begin{aligned} &\text{wenn } p(v_l | RATE = 10) > p(v_l | \overline{RATE = 10}), \\ &\text{dann gilt auch } p(RATE = 10 | v_l) > p(RATE = 10). \end{aligned} \tag{3.9}$$

Ist also in der Verteilung P die Ausprägung v_l einer Eigenschaft $l, l \in \mathcal{L}$, wahrscheinlicher unter $RATE = 10$ als unter $\overline{RATE = 10}$, so erhöht die Bedingung v_l die Wahrscheinlichkeit von $RATE = 10$ über ihre a priori-Wahrscheinlichkeit hinaus. Diese Tatsache ist intuitiv sofort nachvollziehbar; den einfachen Beweis überlassen wir dem Leser.

Gleichung (3.9) nutzen wir in der Heuristik 'Best Design' im folgenden Abschnitt, welche *sukzessive* mittels Einsatz der Shell SPIRIT ein im Sinne der Entscheidungsregel sehr gutes Produktdesign findet. Es sei angemerkt, dass in unseren bisherigen Testläufen mittels der Heuristik stets das optimale Produktdesign gefunden wurde.

3.3.2 'Schätzung' und Interpretation

Wie im letzten Abschnitt bereits erwähnt, ist eine vollständige Enumeration zur Bestimmung des optimalen Produktdesigns zu aufwändig. Wir stellen nun den Algorithmus 'Best Design' vor, der sich der Gleichung (3.9) im Umkehrschluss bedient, um ein gutes Produktdesign \mathbf{v}^* gemäß (3.8) zu bestimmen, für das also die Aussage $\mathbf{v} | RATE = 10$ einen hohen Wahrheitsgrad (=Wahrscheinlichkeit) aufweist.

Algorithmus 'Best Design'

$P'(\mathbf{v}_I), \mathbf{v}_I \in \mathbf{V}_I$, sei die bedingte Wahrscheinlichkeitsverteilung unter dem bedingenden Ereignis $RATE = 10$

Setze $I \equiv \mathcal{L}$

Initialisiere $\nu = 1$

Für alle V_l , $l \in \mathbf{I}$

M1 Bestimme $v_{l^*}^* = \arg \max_l \max_{v_l} p'(V_l = v_l)$

Setze $\mathbf{I} := \mathcal{L} \setminus \{l^*\}$

Setze $\nu := \nu + 1$

M2 Bestimme $P'(\mathbf{v}_{\mathbf{I}})$ unter $RATE = 10 \bigwedge_{\omega=1}^{\nu-1} v_{\omega}^*$

Gehe zu M1

Terminiere, falls $\nu = |\mathcal{L}|$

Der Algorithmus ermittelt ein gutes Produktdesign $\bigwedge_{l=1}^L V_l = v_{l^*}^*$ unter der Bedingung, dass der höchste Rate vergeben wurde; die Zulässigkeit dieses Vorgehens basiert auf der Annahme der Gültigkeit der First Choice-Regel und Gleichung (3.9).

In diesem Abschnitt wird die Vorgehensweise des Algorithmus anhand des Beispiels 'Automobil' zu einem späteren Zeitpunkt verdeutlicht.

Hat der Algorithmus 'Best Design' ein gutes Produktdesign gefunden, so ist nun von Interesse, inwiefern das Ergebnis valide und reliabel ist (vgl. dazu die in der linearen Regressionsanalyse vorgestellten Maße, siehe Abschnitt 3.2.2.1).

Die Prüfung der **Inhaltsvalidität** entfällt, da im Gegensatz zum statistischen Ansatz hier eben keine Annahmen bezüglich der strukturellen Abhängigkeiten zwischen $RATE$ und den Produkteigenschaften zu treffen sind; somit existiert auch kein zum F-Wert analoges Gütekriterium.

Interpretiert man wie in Abschnitt 3.2.3 dargestellt die Wahrscheinlichkeit $p(RATE = 10 | \mathbf{v})$ als Absatzkraft einer Produktvariante, so kann die **Krite-**

riumsvalidität über die Korrelation zwischen prognostizierter und (durch Kauf oder über Hold-Outs) beobachteter Absatzkraft ermittelt werden.

Zur Feststellung der **Reliabilität** dient im Folgenden das Maß der Transinformation (siehe Abschnitt 2.4). Eine hohe Transinformation $T(V_l, RATE)$ lässt auf eine starke Abhängigkeit zwischen der 'Zielvariablen' RATE und den Produkteigenschaften schließen.

Die obigen Ausführungen zur Ermittlung eines guten Produktdesigns und zur Beurteilung der Güte der Prognose werden nun an dem Beispiel – zunächst für den Fall vollständiger Daten, anschließend für den Fall unvollständiger – nachvollzogen.

Produktdesign bei Vorliegen vollständiger Daten

Ergebnis des Verfahrens 'Best Design'

Wenden wir den Algorithmus 'LernVollstD' (siehe Abschnitt 2.2.1, Seite 9) auf die Beobachtungsdatei (siehe Tabelle 3.4) an, so erhalten wir die Wahrscheinlichkeitsverteilung P auf dem Grundraum \mathcal{V} . Es ist nun möglich, $P'(\mathbf{v})$ unter Anwendung der Shell SPIRIT durch 'Klicken' auf den Wert $RATE = 10$ (auch Evidenzierung genannt) zu bestimmen. Die resultierenden Wahrscheinlichkeiten auf den Produkteigenschaften können abgelesen werden.

Wie der Leser sofort bemerkt haben wird, ist es nun denkbar einfach, den Algorithmus 'Best Design' fortzusetzen und die $v_{l^*}^*, \forall l = 1, \dots, L$, zu ermitteln. In Abbildung 3.3 wurde bereits $RATE = 10$ evidenziiert und die Wahrscheinlichkeitsverteilung $P'(\mathbf{v}|RATE = 10)$ von SPIRIT erzeugt. In der Abbildung sind die Randverteilungen auf den Produktmerkmalen V_l , $l = 1, \dots, L$, gut zu erkennen. Betrachtet man die Wahrscheinlichkeiten der

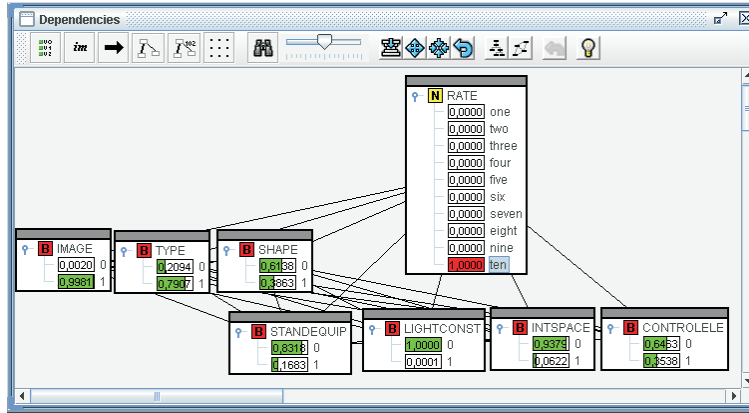


Abbildung 3.3: Evidenzierung von $RATE = 10$

Realisierungen v_l der Variable V_l , für alle l , so erkennt man, dass die Realisierung mit der höchsten Wahrscheinlichkeit $LIGHTCONST = 0$ ist, d. h. $v_{l^*} = \arg \max_l \max_{v_l} (V_l = v_l) = (LIGHTCONST = 0)$.

Im zweiten Iterationsschritt evidenzieren wir $RATE = 10$ und $LIGHTCONST = 0$, und bestimmen erneut die Realisierung mit der höchsten Wahrscheinlichkeit auf Basis der mit $RATE = 10$ und $LIGHTCONST = 0$ bedingten Wahrscheinlichkeitsverteilung. Wir berichten an dieser Stelle nur das Ergebnis des Iterationsschrittes: $IMAGE = 1$.

Die Schritte M1 und M2 werden durchlaufen, bis das Abbruchkriterium $\nu = 7$ greift. Die gefundene Lösung lautet nun:

$$\mathbf{v}^* = (IMAGE \wedge TYPE \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE}).$$

Nun soll die auf Seite 35 angesprochene Reliabilität des ermittelten Ergebnisses ausgewiesen werden.

Reliabilität

Um die Reliabilität zu prüfen, wird die Transinformation zwischen der Variablen $RATE$ und den Eigenschaften V_l , $l = 1, \dots, L$, berechnet. Die Transin-

formation gibt, wie in Abschnitt 2.4 dargestellt, einen Hinweis auf die Stärke der Beziehungen zwischen Rate und Produktmerkmalen.

Variable V_l	$T(RATE, V_l)$	Rang
IMAGE	0,159890	2
TYPE	0,065605	5
SHAPE	0,009558	7
STANDEQUIP	0,078718	4
LIGHTCONST	0,209952	1
INTSPACE	0,114735	3
CONTROLELE	0,018236	6

Tabelle 3.6: Transinformation zwischen $RATE$ und Eigenschaften im Fall vollständiger Daten

Für die ersten beiden Iterationsschritte wurde anhand der in Abbildung 3.3 dargestellten Randverteilungen auf den Produktmerkmalen $V_l, l = 1, \dots, L$, verdeutlicht, welche Variablen in den ersten beiden Schritten sukzessiv in die Bedingung aufgenommen (sprich: evidenziiert) wurden. Erinnern wir uns: In $\nu = 1$ war dies $LIGHTCONST = 0$ und in $\nu = 2$ $IMAGE = 1$. Tabelle 3.6 gibt die Transinformation zwischen $RATE$ und den Produktmerkmalen wieder. Die Rangordnung auf den Produktmerkmalen $V_l, l = 1, \dots, L$, wurde gemäß der Höhe der Transinformation in absteigender Reihenfolge vorgenommen. $LIGHTCONST$ weist die höchste Transinformation auf, $IMAGE$ die zweithöchste. Die Reihenfolge der Variablen, die die Bedingungsstruktur sukzessive anreichert haben, ist damit plausibel, da sie der Rangordnung gemäß der Transinformation entspricht. Damit wurden die Realisierungen der Produktmerkmale in der Reihenfolge der Stärke ihrer Schlusskraft auf $RATE$ festgelegt; diese Vorgehensweise ist schlüssig und das Ergebnis damit reliabel. Somit gilt also für die Produktvariante \mathbf{v}^* Gleichung (3.8) (siehe Seite 33). Um beurteilen zu können, wie signifikant nun dieses Ergebnis ist, wird für die Wahrscheinlichkeit $p(\mathbf{v}^* | RATE = 10)$ der

Unsicherheitsbereich angegeben (vgl. Abschnitt 2.3). Der Leser hat noch die Ausführungen in Abschnitt 3.2.3 in Erinnerung: Die Wahrscheinlichkeit $p(\mathbf{v}^*|RATE = 10)$ lässt sich ökonomisch als Anteil der Konsumenten, die die Produktvariante \mathbf{v}^* präferieren, interpretieren.

Signifikanz der prognostizierten Absatzkraft

Wertet man mit der Verteilung P die Aussage $\mathbf{v}^*|RATE = 10$ aus, so erhält man die bedingte Wahrscheinlichkeit:

$$\begin{aligned} p((IMAGE \wedge TYPE \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \\ \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE})|RATE = 10) \\ = 0,19114. \end{aligned}$$

Das mit dem Wahrheitsgrad p versehene Konditional lässt sich wie folgt ökonomisch interpretieren: Der Anteil der Konsumenten, die der Produktvariante \mathbf{v}^* den höchsten $RATE$ gegeben haben, liegt bei 19,114%. Damit kann der Designer auf die Absatzkraft des Produktes schließen: Alle anderen Produktvarianten haben aufgrund des in Gleichung (3.9) dargestellten Zusammenhangs einen geringeren (geschätzten) Anteil. Um diese geschätzten 19,114% statistisch beurteilen zu können, knüpfen wir an die Ausführungen in Abschnitt 2.3 zur Bestimmung der Unsicherheitsbereiche für bedingte Wahrscheinlichkeiten an. Für $p(\mathbf{v}^*|RATE = 10) = 0,19114$ ist in der Nomenklatur des Abschnitts 2.3 das Ereignis \mathbf{A} mit $\mathbf{A} \equiv RATE = 10$ und \mathbf{B} mit $\mathbf{B} = \mathbf{v}^*$ bzw. $\overline{\mathbf{B}} = \overline{\mathbf{v}^*}$ definiert. Über die $Beta(s = \sum_{\mathbf{v} \in \mathbf{B}\mathbf{A}} a(\mathbf{v}), t = \sum_{\mathbf{v} \in \overline{\mathbf{B}}\mathbf{A}} a(\mathbf{v}))$ können die Unsicherheitsbereiche angegeben werden; hier lautet die Verteilung $Beta(s = 199,936, t = 846,08)$ und die Unsicherheitsbereiche zur Sicherheit $S = 95\%$ damit $[\underline{p} = 0,1679, \overline{p} = 0,2155]$. Somit ist die statistische Signifikanz des Ergebnisses bestätigt.

Signifikanz der geschätzten Konsumentenpräferenzen

Oben konnte aufgrund der Wahrscheinlichkeitsverteilung P die Absatzkraft der Produktvariante \mathbf{v}^* prognostiziert und mittels der $Beta$ die Signifikanz dieser Aussage ermittelt werden. Nun ist aber noch von Interesse, wie homogen die Konsumentenpräferenzen hinsichtlich der gefundenen Produktvariante sind. Aufgrund Gleichung (3.9) muss für die gefundene Produktvariante auch Gleichung (3.8) gelten, siehe Seite 34. Evidentiert man das Produktdesign, erhält man die Randverteilung auf $RATE$ – wie in Abbildung 3.4 erkennbar.

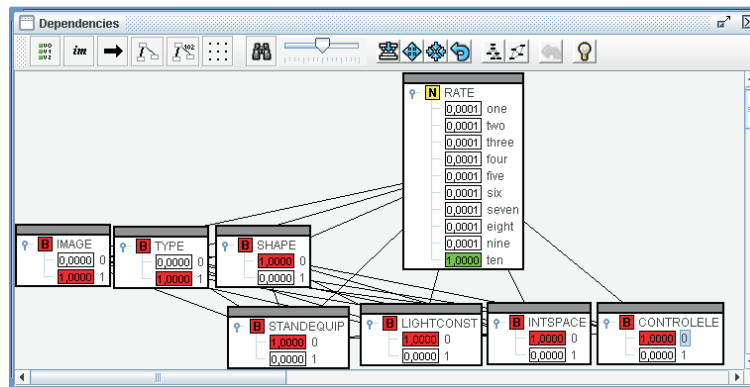


Abbildung 3.4: Verteilung auf $RATE$ unter optimalem Produktdesign

Die Graphik 3.4 zeigt, dass unter Vorliegen des obigen Produktprofils einheitlich der höchste $RATE$ vergeben wurde; es gilt also

$$\begin{aligned}
 p(RATE = 10 | (IMAGE \wedge TYPE \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \\
 \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE})) \\
 = 1.
 \end{aligned}$$

Die Gültigkeit des oben beschriebenen Zusammenhangs ist somit exemplarisch nachgewiesen.

Produktdesign bei Vorliegen unvollständiger Daten

Ergebnis des Verfahrens 'Best Design'

Auch für den Fall unvollständiger Daten \mathbf{v} , $\mathbf{v} \in V$, können mittels des Algorithmus 'LernUVollstD' (siehe Abschnitt 2.2.2) die Wahrscheinlichkeiten $\hat{p}(\mathbf{v})$ geschätzt werden. Erneut wird die Datenbasis, bei der 20% der Rates nicht angegeben wurden, untersucht. Tabelle 3.7 gibt für einige Konfigurationen die geschätzte Wahrscheinlichkeit $p(a(\mathbf{v}')) = \frac{a(\mathbf{v}')}{\sum_{\mathbf{v}'} a(\mathbf{v}')}$ wieder.

Laufende Nr.	IMAGE	TYPE	SHAPE	Konfiguration \mathbf{v}'				Rate	$\frac{a(\mathbf{v}')}{\sum_{\mathbf{v}'} a(\mathbf{v}')}$
				STAND- EQUIP	LIGHT- CONST	INT- SPACE	CONTROL- ELE		
1	0	0	0	0	0	0	0	10	0,00014
2	0	0	0	0	0	0	0	9	0,00162
3	0	0	0	0	0	0	0	8	0,00369
4	0	0	0	0	0	0	0	7	0,00150
⋮	⋮			⋮			⋮	⋮	⋮
970	1	1	0	0	0	0	0	10	0,00627
⋮	⋮			⋮			⋮	⋮	⋮
1280	1	1	1	1	1	1	1	1	0,00019

Tabelle 3.7: Ausschnitt Konfigurationen mit Zählstatistik für das Beispiel 'Automobil' für $\varepsilon = 0,0001$ bei unvollständigen Daten

Der Leser vergleiche das in Tabelle 3.7 dargestellte Ergebnis der Schätzung mit der des Falls vollständiger Daten (auszugsweise notiert in Tabelle 3.5 auf Seite 33) und bemerke, dass die $\hat{p}(\mathbf{v}')$ (aufgrund des notwendigen 'proportionalen' Abgleichs) voneinander abweichen und damit auch die Wahrscheinlichkeitsverteilung P auf \mathcal{V}' .

Nach Einlernen der Daten kann erneut das Verfahren 'Best Design' zur Suche eines guten Produktdesigns eingesetzt werden. Dem Leser werden nur der Screenshot (Abbildung 3.5) aus SPIRIT, der die Randverteilungen auf den Eigenschaftsvariablen unter Evidentiierung von $RATE = 10$ zeigt, sowie

anschließend die gefundene Lösung präsentiert.

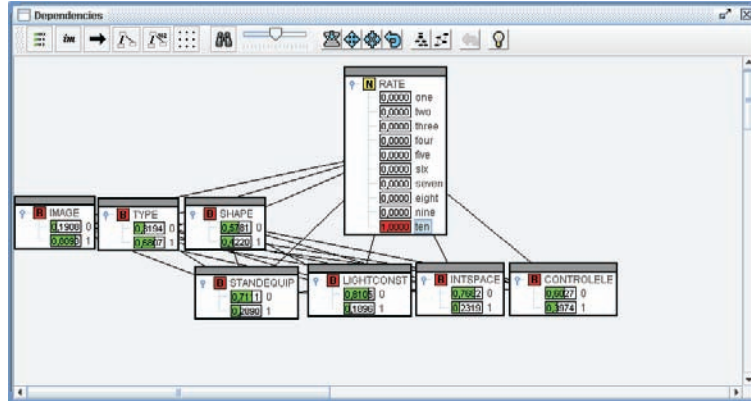


Abbildung 3.5: Evidenzierung von $RATE = 10$

Die gefundene Lösung lautet:

$$\mathbf{v}^* = (IMAGE \wedge TYPE \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE}).$$

Der Algorithmus 'Best Design' findet – trotz unvollständiger Daten – erneut das beste Produktdesign (vergleiche auch die Ausführungen auf Seite 37). Es stellt sich die Frage nach der Reliabilität des Ergebnisses.

Reliabilität

Wie bereits erwähnt liegt nun eine andere Wahrscheinlichkeitsverteilung P auf \mathcal{V}' vor; aus diesem Grunde weichen die Transinformationen $T(RATE, V_l)$, $l = 1, \dots, L$, natürlich ebenfalls ab; diese sind in Tabelle 3.8 notiert.

Ein Vergleich der Rangordnung auf den Eigenschaftsvariablen gemäß der Höhe der Transinformation (siehe Tabellen 3.6 und 3.8) zeigt, dass trotz der hohen Anzahl von Missing Values in den Rates die Stärke der Abhängigkeit zwischen $RATE$ und den einzelnen Eigenschaften korrekt bestimmt wurde. Die Reliabilität der 'Schätzung' wird aufgrund der Kohärenz von Produktdesign und Transinformation bestätigt.

Variable V_l	$T(RATE, V_l)$	Rang
IMAGE	0,083932	2
TYPE	0,035988	5
SHAPE	0,005194	7
STANDEQUIP	0,043610	4
LIGHTCONST	0,114465	1
INTSPACE	0,062038	3
CONTROLELE	0,010434	6

Tabelle 3.8: Transinformation zwischen $RATE$ und Eigenschaften im Fall unvollständiger Daten

Signifikanz der prognostizierten Absatzkraft

Mittels der Wahrscheinlichkeitsverteilung P lässt sich wieder die Aussage $\mathbf{v}^*|RATE = 10$ bewerten.

$$\begin{aligned}
 &P((IMAGE \wedge TYPE \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \\
 &\quad \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE})|RATE = 10) \\
 &= 0,11895
 \end{aligned}$$

Vergleicht der Leser dieses Ergebnis mit dem auf Seite 39 für den Fall vollständiger Daten, so erkennt er, dass die bedingte Wahrscheinlichkeit – trotz der über 5120 fehlenden Beurteilungen – nur wenig abweicht. Der Unsicherheitsbereich zur Sicherheit $S = 95\%$ mit $Beta(s = 160, 512, t = 1188, 096)$ ist mit $[p = 0,1023, \bar{p} = 0,1368]$ vergleichsweise klein geblieben.

Signifikanz der geschätzten Konsumentenpräferenzen

Die fehlenden Daten führen wie oben gezeigt zwar zu einer 'vorsichtigeren' Prognose der Absatzkraft, jedoch konnte mittels der $Beta$ die Signifikanz des Ergebnisses gezeigt werden. Erneut stellt sich die Frage, wie homogen die Konsumentenpräferenzen hinsichtlich der gefundenen Produktvariante sind.

Evidenziert man das Produktdesign, erhält man die Randverteilung auf *RATE* – wie in Abbildung 3.6 dargestellt.

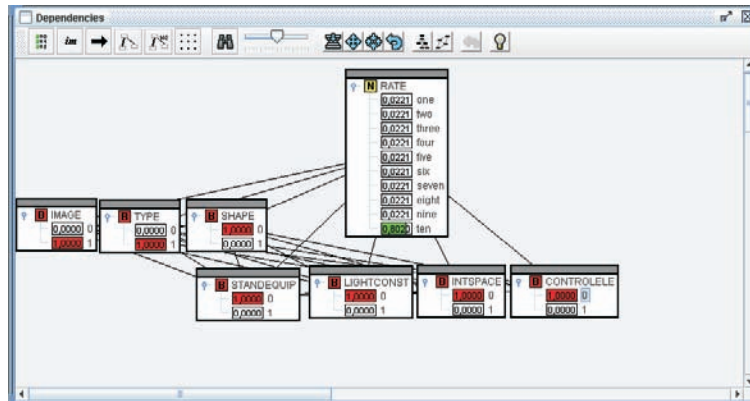


Abbildung 3.6: Verteilung auf *RATE* unter optimalem Produktdesign

Die Graphik 3.6 zeigt, dass das System einen Anteil von über 80,281% der Konsumenten, die dieser Produktvariante den höchsten Rate geben würde, prognostiziert. Diese Schätzung weicht zwar von derjenigen im Fall vollständiger Daten ab, jedoch würde auch in diesem Fall die 'richtige' Entscheidung, nämlich die Realisierung dieser Produktvariante, unterstützt werden.

$$\begin{aligned}
 p(RATE = 10 | (IMAGE \wedge TYPE \wedge \overline{SHAPE} \wedge \overline{STANDEQUIP} \wedge \\
 \overline{LIGHTCONST} \wedge \overline{INTSPACE} \wedge \overline{CONTROLELE})) \\
 = 0,80281
 \end{aligned}$$

Der Unsicherheitsbereich bestimmt sich zur Sicherheit $S = 95\%$ mit $Beta(s = 160,512, t = 38,4)$ zu $[p = 0,7494, \bar{p} = 0,8586]$. Dieses Ergebnis ist signifikant und weist somit auf den 'wahren' Sachverhalt, der Vorzugswürdigkeit der Produktvariante \mathbf{v}^* , hin.

3.4 Vergleich der Ergebnisse

In den Abschnitten 3.2 und 3.3 wurden zwei Ansätze zur Erkennung von Zusammenhängen zwischen Produkteigenschaften und Bewertung in Form eines *RATE* vorgestellt. Ersterer bedient sich des statistischen Verfahrens der Regressionsanalyse, um die Stärke des Einflusses der jeweiligen Produkteigenschaft auf die 'Zielvariable' *RATE* zu quantifizieren. Letzterer nutzt einen informationstheoretischen Ansatz, um die Abhängigkeiten zu messen. Um die Güte beider Verfahren zu testen, wurde eine Datenbasis simuliert; für $K = 200$ Probanden wurde eine linear-additive Ratefunktion zufällig erzeugt. Im Fall vollständiger Daten war auf Basis der erzeugten $2^7 \cdot 10 \cdot 200 = 25600$ Beobachtungen das optimale Produktdesign zu bestimmen. Im Fall unvollständiger Daten wurde die empirisch relevante Annahme getroffen, Missing Values seien lediglich in den Rates vorhanden; 20% der Rates sind somit nicht bekannt.

Die Analyseergebnisse zeigen, dass sowohl mittels OLS als auch mit MinREnt das optimale Produktdesign, sowohl bei vollständigen als auch bei unvollständigen Daten, gefunden wurde. Jedoch wies das Ergebnis im letzteren Fall unter Anwendung von OLS eine geringere Reliabilität auf (vergleiche Tabellen 3.2 und 3.3). Die Reliabilität des mittels MinREnt gefundenen Ergebnisses im Fall unvollständiger Daten bleibt hoch, da die Transinformationen zwischen Produkteigenschaften und Rate nicht signifikant von denen im Fall vollständiger Daten verschieden sind (vergleiche Tabellen 3.8 und 3.6).

Es bleibt also Folgendes festzuhalten: Selbst wenn die intrinsische Präferenzfunktion eine lineare Struktur aufweist und damit eine lineare Regressionsanalyse zulässig ist, so findet ein System, welches keine Präferenzstruktur unterstellt, allein aufgrund der Auswertung der Beobachtungen nach

einem informationstheoretischen Prinzip ebenfalls das optimale Produktdesign. Simulationsstudien, die komplexere Funktionen unterstellten, zeigten, dass MinREnt der Methode OLS deutlich überlegen ist, obwohl in der Literatur oftmals OLS attestiert wird, es erbringe hinreichend gute Ergebnisse auch bei nichtlinearen Funktionen (vgl. (Tei01), S. 62 f.). Die Diskussion dieser Ergebnisse sei weiteren Berichten vorbehalten.

Insbesondere aber das Ergebnis der Handhabung von Missing Values bei unvollständigen Daten in SPIRIT zeigt eine Überlegenheit gegenüber den Optionen, die bei der linearen Regressionsanalyse bereit stehen. Somit weist der informationstheoretische Ansatz auch eine geringere Sensitivität bezüglich fehlender Werte auf.

Kapitel 4

Zusammenfassung und Ausblick

In diesem Papier wurde entropiebasiertes Data Mining an dem Beispiel des Produktdesigns demonstriert. So wurde gezeigt, dass statistische Verfahren wie die OLS-Methode in der Conjoint-Analyse Annahmen bezüglich der Abhängigkeiten von Variablen treffen müssen; sie untersuchen damit Beobachtungsdateien unter einer bereits gegebenen Struktur und sind daher besonders anfällig für irreguläre Antworten oder fehlende Angaben. Die Autoren haben diesen statistischen Verfahren einen informationstheoretischen Ansatz gegenüber gestellt, der es gestattet, nach Abhängigkeiten innerhalb der Datenbasis zu suchen, ohne dass eine spezifische Struktur in den Daten vorgegeben werden muss. Da dieser Ansatz nicht nur eine fallweise ('zeilenweise') Betrachtung vornimmt, sondern auch variablenweise ('spaltenweise') untersucht, können Muster identifiziert werden.

Die Datenbasis ist so erzeugt worden, dass ein linearer Zusammenhang zwischen *RATE* und den Merkmalsausprägungen besteht. Wie zu erwarten war, liefert die OLS-Methode sehr gute Ergebnisse; der informationstheoretische Ansatz ist dem statistischen Verfahren jedoch nicht unterlegen, da ebenfalls das optimale Produktdesign gefunden wird. Zudem konnten über den

MinREnt-Ansatz auch Aussagen über die Absatzkraft der Produktdesigns gemacht werden, deren Signifikanz mittels der Unsicherheitsbereiche nachgewiesen wurde.

Im Weiteren besteht ein Forschungsbedarf hinsichtlich folgender Fragen:

- Welche Eigenschaften sind von besonderer Relevanz für die Beurteilung und wirken zudem noch stark auf andere Eigenschaftsvariable? Identifiziert man solche Variable, müssen in zukünftigen Produktdesigntests den Probanden weniger Stimuli vorgelegt werden. Das für die Messung der gegenseitigen Abhängigkeit probate Mittel wurde bereits genannt: Die Transinformation.
- Gibt es Variablenklassen, die aufgrund eines Sachbezugs gebildet werden können? So könnten Merkmale, die eher die Ästhetik, die Funktionalität oder das Image beschreiben, in Klassen zusammengefasst werden. Innerhalb solcher gibt es Merkmale, die stärker auf die Variable *RATE* schließen als andere, so dass auch über eine derartige Zerlegung der Variablenmenge Umfang und Aufwand von Befragungen reduziert werden können.

Anhang A

Konjugiertheit von Multinomial und Dirichlet

Zu M Ausprägungen einer Zufallsvariablen, die mit Wahrscheinlichkeiten (p_1, \dots, p_M) , $\sum p_m = 1$, realisieren, ist bei N unabhängigen Versuchen die Zählstatistik (y_1, \dots, y_M) multinomial verteilt nach

$$p(\mathbf{y}_1, \dots, \mathbf{y}_M | p_1, \dots, p_M) = \frac{N!}{\prod_{m=1}^M y_m!} \cdot \prod_{m=1}^M p_m^{y_m} \quad (\text{A.1})$$

Herrscht über (p_1, \dots, p_M) Unsicherheit, beschreibt man diese sinnvollerweise durch eine Dirichlet $Dir(a_1, \dots, a_M)$, deren allgemeine Dichte lautet:

$$f(p_1, \dots, p_M | a_1, \dots, a_M) = \frac{\Gamma(\sum_{m=1}^M a_m)}{\prod_{m=1}^M \Gamma(a_m)} \cdot \prod_{m=1}^M p_m^{a_m - 1} \quad (\text{A.2})$$

alle $a_m > 0$. Mit $a_0 = \sum a_m$ hat (A.2) den Erwartungswert $(\frac{a_1}{a_0}, \dots, \frac{a_M}{a_0})$ und die Varianz $(\frac{a_1(a_0 - a_1)}{a_0^2(a_0 + 1)}, \dots, \frac{a_M(a_0 - a_M)}{a_0^2(a_0 + 1)})$.

Multinomial und Dirichlet sind konjugiert. Beobachtet man $(\mathbf{y}_1, \dots, \mathbf{y}_M)$ mit $\sum y_M = N$ Realisationen der multinomialverteilten Zufallsvariablen, so ist nach dem Satz von Bayes (vgl. (Bay63)) mit der a priori-Dichte (A.2) und $a_m = a_m + y_m$, $a_0 = a_0 + N$ die a posteriori-Dichte wieder Dirichlet nach (A.2) mit aktualisierten Parametern. Statt N Beobachtungen führen N mal 1 Beobachtung zum gleichen Ergebnis.

Sind die $M = 2^7$ Ausprägungen gerade die Konfiguration $\{\mathbf{v}\}$ des Abschnitts 2.1.2, übertragen sich die Gleichungen (A.1) und (A.2) unmittelbar auf (2.1) mit $a(\mathbf{v})$ statt a_m .

Anhang B

1. und 2. Moment der Dirichlet

Für eine $Dir(\varepsilon, \dots, \varepsilon)$ mit einem kleinen $\varepsilon > 0$ kann der gesamte Lernvorgang analog zu Anhang A erfolgen. Alle Wahrscheinlichkeiten, Erwartungswerte und Varianzen werden wahlweise für $\varepsilon > 0$ oder dann, wenn es zu erlaubten Ausdrücken führt, mit $\lim \varepsilon \rightarrow 0$ hingeschrieben.

Der Leser vollziehe die folgenden Beispiele nach.

Beispiele

1. Der Erwartungswert zur $Dir(\varepsilon, \dots, \varepsilon)$ ist $(\frac{1}{M}, \dots, \frac{1}{M})$.
2. Der Erwartungswert zur $Dir(\varepsilon, \varepsilon, \varepsilon + 1, \varepsilon + 2)$ auf $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ ist $(\frac{\varepsilon}{4\varepsilon+3}, \frac{\varepsilon}{4\varepsilon+3}, \frac{\varepsilon+1}{4\varepsilon+3}, \frac{\varepsilon+2}{4\varepsilon+3})$. Der $\lim_{\varepsilon \rightarrow 0} (\cdot)$ ist $(0, 0, \frac{1}{3}, \frac{2}{3})$, stellt aber keinen erlaubten Parameter der Dirichlet dar.
3. Zu 2.: Berechnen wir nun die bedingte Wahrscheinlichkeit $p(\mathbf{v}_2 | (\mathbf{v}_1 \vee \mathbf{v}_2)) = \frac{\frac{\varepsilon}{4\varepsilon+3}}{\frac{2\varepsilon}{4\varepsilon+3}} = \frac{\varepsilon}{2\varepsilon} = \frac{1}{2}$; analog gilt: $P(\mathbf{v}_4 | (\mathbf{v}_3 \vee \mathbf{v}_4)) = \frac{\frac{\varepsilon+2}{4\varepsilon+3}}{\frac{\varepsilon+1}{4\varepsilon+3}} \stackrel{\lim_{\varepsilon \rightarrow 0}}{=} \frac{2}{3}$. Damit entspricht für kleine ε das Lernen auf Konfigurationen dem frequentistischen Lernen über Multinomial/Dirichlet. Bei der Berechnung

der Wahrscheinlichkeiten muss nur im Bedarfsfall eine Limesbetrachtung für $\varepsilon \rightarrow 0$ erfolgen.

Für $M = 2$ liegt der Spezialfall Binomial-/Betaverteilung vor. Alle angestellten Überlegungen gelten entsprechend. Die Summenfunktion $\phi(x|a_1, a_2) = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\cdot\Gamma(a_2)} \int_0^x p^{a_1-1}(1-p)^{a_2-1} = I_x(a_1, a_2)$ liegt tabellarisch vor oder ist in guten Tabellenkalkulationsprogrammen berechenbar. Damit ist auch die zweiseitige symmetrische Abgrenzung $[\underline{p}, \bar{p}]$ zur Sicherheit $S = 1 - \alpha$ so, dass $I_{\underline{p}}(a_1, a_2) = \frac{\alpha}{2}$ und $I_{\bar{p}}(a_1, a_2) = 1 - \frac{\alpha}{2}$.

Literaturverzeichnis

- [Bay63] BAYES, T.: *An Essay towards solving a Problem in the Doctrine of Chances*. In: Philosophical Transactions of the Royal Society of London, 1763.
- [Cal91] CALABRESE, P. M.: *Deduction and Inference Using Conditional Logic and Probability*. In: Conditional Logic in Expert Systems I, 1991.
- [Dec02] DECKER, R.: *Marketingforschung: Methoden und Modelle zur Bestimmung des Käuferverhaltens*. Verlag Moderne Industrie, München, 2002.
- [Fer75] FERSCHL, F.: *Nutzen- und Entscheidungstheorie*. Westdeutscher Verlag, Opladen, 1975.
- [Fre98] FREITAS, A.: *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer Verlag, Berlin et al., 1998.
- [Gen03] GENSLER, S.: *Heterogenität in der Präferenzanalyse*. Dissertation, Deutscher Universitätsverlag, Wiesbaden, 2003.
- [Gra87] GRAF, U./ HENNING, H.-J./ STANGE K./ WILRICH P.-T.: *Formeln und Tabellen der angewandten Statistik*. Springer Verlag, Berlin et al., 1987.

- [Gre71] GREEN, P.E./ RAO, V.: *Conjoint Measurement for Quantifying Judgemental Data*. In: Journal of Marketing Research, Vol. 8, 1971.
- [Har84] HARTUNG, J./ EPELT, B.: *Multivariate Statistik*. Oldenbourg Verlag, München, 1984.
- [Kat04] KATHÖFER, U.: *Data Mining mit genetischen Algorithmen*. Dissertation, Logos Verlag, Berlin, 2004.
- [Mat96] MATHAR, R.: *Informationstheorie: Diskrete Modelle und Verfahren*. Teubner Verlag, Stuttgart, 1996.
- [Moo78] MOOSBRUGGER, H.: *Multivariate statistische Analyseverfahren*. Kohlhammer Verlag, Stuttgart et al., 1978.
- [Moo02] MOOSBRUGGER, H.: *Lineare Modelle*. Verlag Hans Huber, Bern et al., 2002.
- [Pea56] PEARSON, K.: *Tables of the Incomplete Beta-Function*. Cambridge University Press, Cambridge, 1956.
- [Pfa62] PFANZAGL, J.: *Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens*. Physica-Verlag, Würzburg, 1962.
- [Röd06a] RÖDDER, W./ REUCHER, E./ KULMANN F.: *Features of the Expert-System Shell SPIRIT*. In: Logic Journal of the IGPL, 2006.
- [Röd06b] RÖDDER, W.: *Induktives Lernen bei unvollständigen Daten unter Wahrung des Entropieprinzips*. Diskussionspapier Nr. 388, FernUniversität in Hagen, 2006.
- [Sch91] SCHUBERT, B.: *Entwicklung von Konzepten für Produktinnovationen mittels Conjoint-Analyse*. Poeschel Verlag, Stuttgart, 1991.

- [Tei01] TEICHERT, T.: *Nutzenschätzung in Conjoint-Analysen*. Dissertation, Deutscher Universitätsverlag, Wiesbaden, 2001.
- [Voe00] VOETH, K.: *Nutzenmessung in der Kaufverhaltensforschung*. Deutscher Universitätsverlag, Wiesbaden, 2000.
- [vT01] THADEN, CHR. VON: *Conjoint-Analyse mit vielen Merkmalen*. Dissertation, Peter Lang, Frankfurt et al., 2001.