Ben-Alexander Bohnke 23.01.1976

STATISTISCHE ASPEKTE DER VARIABLEN-REGELN IN DER LINGUISTIK

Inhalt:

X) EINFÜHRUNG

I) Vorbemerkungen

1) Ziel dieses Aufsatzes

Primäres Ziel dieses Aufsatzes ist es zu untersuchen, inwieweit sich die absoluten oder relativen Häufigkeiten einer abhängigen Variablen bei Abhängigkeit von einer Kombination von (zwei) unabhängigen Variablen aus den absoluten oder relativen Häufigkeiten bei Abhängigkeit von den einzelnen (zwei) unabhängigen Variablen berechnen lassen.
Unter Bezug auf den linguistischen Terminus der Variablenregel könnte man das Ziel folgendermaßen formulieren: zu untersuchen,inwieweit sich die Wahrscheinlichkeit einer kombinierten Variablenregel aus der von singulären Variablenregeln berechnen läßt.

2) Komplexität des Themas

Wie schon die Formulierung des Aufsatzziels andeutet, handelt es sich um ein sehr komplexes und umfangreiches Thema;aus Gründen der Kürze und Verständlichkeit,soll dieses hier nur selektiv,unter Ausklammerung vieler Probleme und mehr exemplarisch,anhand linguistischer Beispiele, als allgemein behandelt werden,wobei allerdings von den Beispielen aus relativ leicht der jeweilige allgemeine Sachverhalt erschließbar ist.

3) Bedeutung des Themas

Das Thema ist nicht nur von erheblichem wissenschaftstheoretischen Interesse,sondern seine Bedeutung liegt vorallem darin,daß sich bei einer Lösung des genannten Berechnungsproblems ein Großteil empirischer Untersuchungen einsparen ließen.

4) Statistische Problematik

Das Problem,um das es hier geht,gehört letztlich in den Bereich der Statistik.Deshalb soll am Anfang eine in diesem Rahmen eher ausführliche Einführung in die Grundlagen statistischer Methodik vorgenommen werden,da deren Kenntnis Voraussetzung für das Verständnis des speziellen Themas ist;zunächst wird dazu die linguistische Schreibweise einer Variablenregel in die statistische transformiert.

5) Form der Darstellung

Die Darstellung ist insgesamt halb systematisch gehalten, halb orientiert sie sich am zeitlichen Verlauf einer empirischen Untersuchung.

II) Transformation der allgemeinen Form einer linguistischen Variablenregel in einen logisch-statistischen Ausdruck

1) Linguistische Darstellung

a) Struktur einer linguistischen Regel

Die generelle Struktur einer linguistischen Regel ist: X → Y / A___B .Man geht also davon aus,daß ein Element X in einer bestimmten Umgebung,nämlich wenn der Ausdruck A vorausgeht und der Ausdruck B folgt, immer durch ein Element Y ersetzt wird.

b) Struktur einer linguistischen Variablenregel

Auch eine linguistische Variablenregel hat die Grundstruktur X → Y / A___B. Es werden dabei aber die Zeichen Y,A und B in (besondere) Klammern gesetzt.: X → ⦃Y⦄ / ⦃A⦄___⦃B⦄ ;damit soll gekennzeichnet werden,daß X nicht immer (d.h. nicht in 1oo% der Fälle) in der Umgebung A___B durch Y ersetzt wird,sondern in einigen Fällen (der genaue Prozentsatz wird durch diese Formel nicht angegeben) auch nicht.

c) Als Beispielregel,die durch den ganzen Aufsatz hindurch verfolgt werden soll,dient die Regel:
t → ⦃t^h⦄ / ⦃M⦄___ ⦃'V⦄ (M=$_{df}$ Morphemgrenze)
Die Regel ist zu lesen: [t] wird in einigen Fällen in der Umgebung nach Morphemgrenze vor betontem Vokal aspiriert,in anderen Fällen nicht.
Es handelt sich hierbei um ein künstliches Beispiel, das nach seinem Demonstrationswert für diesen Aufsatz konstruiert wurde.

2) Statistische Darstellung

Es soll nun die linguistische Darstellung Element für Element in die statistische übersetzt werden.
(Auf eine genaue Unterscheidung zwischen objekt- und metasprachlicher Notierung wird dabei verzichtet)

a) X

X,also das Zeichen für das bzw. die Objekte (ob X in der linguistischen Regel singulär zu interpretieren ist oder nicht,scheint mir nicht eindeutig),über das bzw. die etwas ausgesagt werden soll,könnte man logisch-statistisch wohl durch die Individuenvariable x wiedergeben.

(x soll hier aber nicht im Sinne eines formalen Objekts verstanden werden,das erst noch durch Zuweisung einer Merkmalsdimension inhaltlich interpretiert werden muß,sondern als Repräsentant eines bestimmten Objektbereiches;im Beispiel gesprochen: x repräsentiert [t].)

b) →

Den Pfeil mit der Bedeutung "wird zu" braucht man nicht zu übersetzen,denn er hat in der statistischen Beschreibung kein Äquivalent;denn hier werden keine Veränderungen dargesttelt,sondern nur Zustände, also Resultate von Veränderungen.

c) Y

Unter Y ist in der linguistischen Darstellung zu verstehen: X,mit einem bestimmten Merkmal versehen, also z.B. [t] mit dem Merkmal Aspiration,also hier $Y = t^{h}$ .In der Statistik geht man anders vor; neben dem Objektbereich (hier [t]) unterscheidet man Variablen.Variablen sind (Repräsentationen von) Merkmalsdimensionen,auf denen die Elemente des Objektbereiches Ausprägungen haben können oder nicht. Bezeichnet man Y als Variable,so schreibt man "x hat eine Ausprägung auf Y" als Y+x und "x hat keine Ausprägung auf Y" als Y-x.

(Viele Autoren schreiben die +/- - Zeichen vor den Variablenausdruck,m.E. ist die umgekehrte Schreibweise aber sinnvoller.)

Will man offenlassen,ob x eine Ausprägung auf Y hat oder nicht,schreibt man Yx.

Y in der linguistischen Regel entspricht also Yx (oder Y+x,je nach Interpretation) statistisch gesehen; allerdings gibt es auch in der Linguistik entsprechen-Notierungen.

d) /

Den Schrägstrich kann man als Zeichen übernehmen; er hat in der linguistischen Beschreibung die Bedeutung "in" im Kontext: "in der Umgebung A___B".Wie unschwer zu erkennen ist,handelt es sich hier um eine wenn-dann-Beziehung: Wenn A___B,dann X → Y. Allerdings darf diese wenn-dann-Beziehung (aus Gründen,die hier nicht aufgezeigt werden können) nicht mit der Implikation gleichgesetzt werden.

e) A___ bzw. ___B

Das Umgebungsmerkmal A___ könnte man übersetzen in: A geht x voraus.Streng genommen müßte man das z.B. folgendermaßen formalisieren: V(A)x;man wird diesen Ausdruck aber vereinfachen,z.B. V(A) als X zusammen fassen und somit insgesamt dann Xx schreiben;entsprechend könnte man ___B mit Tx wiedergeben.

Der vollständige Ausdruck A___B ist dann zu übersetzen: Xx,Tx oder Xx & Tx;die Verknüpfung von X und T kann man also durch Komma oder Und-Zeichen kennzeichnen.

f) ≠ ≠

Nun geht es hier ja aber um die Struktur einer Variablenregel,d.h. Y+x (bzw. Y-x) tritt nicht in allen (1oo %) Fällen oder keinem Fall (o %),sondern in einem Teil der Fälle auf.

In der Statistik spricht man hierbei von einer stochastischen oder statistischen im Gegensatz zu einer deterministischen Beziehung.

In der linguistischen Schreibweise wird dies -wie schon gesagt- dadurch gekennzeichnet,daß man Y, A und B in Klammern setzt.

Da sich stochastische Beziehungen nur ungenau mit logischen Quantoren beschreiben lassen,gibt man in der Statistik die Wahrscheinlichkeit p einer solchen Beziehung an.p bezeichnet hier die relative Häufigkeit,d.h. die Anzahl der tatsächlichen Fälle (z.B. von Y+x) geteilt durch die Anzahl der möglichen Fälle;dies wird später noch genauer erläutert werden.

Man gibt p in Prozentwerten oder in Werten von o bis 1; um auszudrücken, daß z.B. p(Y+x) zwischen o und 1 liegt, kann man schreiben: $p(Y+x) = r$ und $o < r < 1$.

g) Vereinfachung des statistischen Ausdrucks
Die formale linguistische Variablenregel ist $X \rightarrow$ ⟨Y⟩ / ⟨A⟩___⟨B⟩ ist jetzt also umgewandelt in den statistischen Ausdruck: $p(Xx,Tx/Yx) = r$ wobei $o < r < 1$. Wenn eindeutig ist, für welchen Objektbereich (x) die ausgesagte Beziehung gelten soll, kann man das x auch weglassen und einfach (hier) schreiben: $p(X,T/Y) = r$

h) Allgemeinere Form der statistischen Darstellung einer Variablenregel
Die Regel $p(X,T/Y) = r$ mit 2 Bedingungen, nämlich X und T, stellt aber eigentlich nur einen Sonderfall dar, der sich eben daraus ergibt, daß man in der Linguistik üblicherweise nur die sprachliche Umgebung einer Spracheinheit berücksichtigt und diese eben durch die vorausgehende(n) und nachfolgende(n) Spracheinheit(en) kennzeichnet. Gerade im Zusammenhang mit den Variablenregeln versucht man ja aber auch zusätzliche, nichtsprachliche Variablen zu berücksichtigen. Als allgemeinere Form einer Variablenregel könnte man daher statistisch etwa formulieren:
$p(X_1,X_2,....,X_n/Y) = r$ bei $o < r < 1$.
Im Rahmen dieses Aufsatzes wird aber nur auf Regeln mit 2 bedingenden Variablen, X und T, eingegangen.

i) Darstellung der Beispielregel in statistischer Schreibweise
Jetzt kann auch die linguistische Beispielregel $t \rightarrow$ ⟨t^h⟩ / ⟨M⟩___⟨'V⟩ statistisch dargestellt werden: M___ soll der Einfachheit halber mit M (=X) symbolisiert werden und ___'V mit B (=T) (B wegen betonter Vokal). Für das Merkmal Aspiration $..^h$ wird A (=Y) geschrieben. Es ergibt sich also: $p(M,B/A) = r$ bei $o < r < 1$.

j) Diagramm

Zur Übersichtlichkeit seien sich die linguistische und statistische Schreibweise einer Variablenregel noch einmal in einem Diagramm gegenübergestellt:

Variablenregel
(allgemeine Form)

linguistische Schreibweise:

$X \rightarrow \{Y\} / \{A\} \_\_ \{B\}$

künstliches Beispiel:

$t \rightarrow \{t^h\} / \{M\} \_\_ \{'V\}$

($M =_{df}$ Morphemgrenze)

statistische Schreibweise:

$p(Xx,Tx/Yx) = r$ bzw.

$p(X,T/Y) = r$

allgemeiner:

$p(X_1,X_2,\ldots,X_n/Y) = r$

bei $o < r < 1$

X	x
→	
	Y(+/-)
Y	Yx
/	/
A__	(A geht x voraus) V(A)x Xx
__B	(B folgt x) F(B)x Tx
{ }	p( ) = r o < r < 1

3) Kategorien einer statistischen Variablenregel

Wenn wir die Regel $p(Xx,Tx/Yx) = r$ analysieren, stellen wir fest, daß sie 3 Kategorien enthält:

(1) den objektbereich (x)

(2) Variablen (X,T,Y)

(3) Beziehungen zwischen Objektbereich und Variablen (/ , p( ) = r)

Es sollen nun im folgenden diese 3 Bereiche näher dargestellt werden.

Y) HAUPTTEIL

I) Objektbereich

1) Begriffsbestimmung

Objektbereich ist hier zu verstehen als der Bereich, über den Aussagen gemacht werden sollen.

(Zwar kann man auch den Objektbereich als Merkmalsdimension mit formalem Objekt x auffassen (vgl.X,II,2,a/Seite 3), diese Möglichkeit soll hier aber nicht weiter diskutiert werden.)

Man kann den Objektbereich spezifizieren in:

a) Untersuchungsbereich

b) Grundgesamtheit (Population)

c) Stichprobe (Sample)

ad a): Unter dem Untersuchungsbereich ist die zugrundeliegende Klasse zu verstehen, aus der eine Teilklasse beschrieben werden soll. Bei linguistischen Untersuchungen ist als Untersuchungsbereich also i.A. die Sprache(n) aufzufassen, aus der (den) Sprachelemente beschrieben werden sollen.

ad b) Als Grundgesamtheit bezeichnet man die Teilklasse, die beschrieben werden soll, im linguistischen Bereich also die Sprach-Elemente (in dem angeführten Beispiel der Aspiration von [t] die Klasse aller [t] -Äußerungen), die beschrieben werden sollen.

ad c) Bei der Stichprobe handelt es sich um eine Teilmenge der Grundgesamtheit (Genaueres hierüber: Y,I,4).

Zwischen a),b) und c) gilt also folgende Beziehung:
Stichprobe $\subset$ Grundgesamtheit $\subset$ Untersuchungsbereich

2) Bestimmung des Objektbereiches

a) Wie sich Untersuchungsbereich und Population genau bestimmen lassen, ist oft problematisch, gerade im linguistischen Bereich.

b) So ergibt sich etwa das Problem der Abgrenzung einer Sprache:
- räumlich (in Bezug auf Dialekte)
- zeitlich (in Bezug auf Sprachentwicklung)
- grammatisch (in Bezug auf ungrammatische Äußerungen) usw.

c) Aber auch bei Lösung des Abgrenzungsproblems blieben noch verschiedene Möglichkeiten bestehen, wie eine Sprache als Untersuchungsbereich aufzufassen ist; die adäquateste Lösung wäre wahrscheinlich, daß man von der Menge aller in der betr. Sprache getätigten Äußerungen ausgeht.

3) Größe des Objektbereiches und Allgemeinheit der Beschreibung

a) Als zentrale Unterscheidung ist hier einzuführen die zwischen (1) finitem und (2) infinitem Objektbereich Faßt man eine Sprache als Untersuchungsbereich als Menge aller in dieser Sprache getätigten Äußerungen auf (vgl. 2),dann wäre also zu fragen,ob es sich dabei um endlich oder unendlich viele Äußerungen handelt. Die Unterscheidung zwischen finitem und infinitem Objektbereich ist deshalb so wichtig,weil die Überprüfungsmöglichkeiten von Aussagen über endlich viele Objekte beträchtlich von denen von Aussagen über unendlich viele Objekte differieren;auf diesen höchst komplexen Themenbereich kann hier aber nicht eingegangen werden.

b) Weiterhin ist zu unterscheiden,ob die betreffende Grundgesamtheit (1) vollständig (Klasse),(2) partikulär (Teilmenge) oder (3) singulär (Element) erfaßt werden soll.

4) Allgemeinheit der Datenerhebung

a) Ist der Objektbereich festgelegt,seine Größe und die Allgemeinheit der Beschreibung bestimmt,so bleibt noch offen,in welchem Umfang man den betreffenden Objektbereich untersuchen muß;hierbei ist vorallem zu unterscheiden zwischen:

(A) Totalerhebung (Untersuchung aller Elemente der Population)

(B) Teilerhebung (Stichprobe)

(1) repräsentative (Zufallsstichprobe)

(2) nicht repräsentative

b) Verlangt man von einer Beschreibung 1oo%-ige Sicherheit, so muß man eine Totalerhebung vornehmen;dies ist bei einem infiniten Objektbereich theoretisch gar nicht möglich,kommt aber auch bei finitem Objektbereich auf Grund praktischer Schwierigkeiten fast nie in Frage.Deshalb behilft man sich mit Stichproben,die allerdings ihre Berechtigung nur bei finitem Objektbereich haben.Als repräsentativ können dabei nur Zufallsstichproben angesehen werden,d.h. Stichproben,bei denen jedes Element der Population die gleiche Chance hat,berücksichtigt zu werden.

II) Variablen

A) Abhängige Variable

1) Abhängige vs. unabhängige Variable(n)

Die Bestimmung,welche Variablen abhängig und welche unabhängig sind,ist keineswegs immer schon von vorneherein festzustellen,sondern oft erst nach Abschluß der Untersuchung.

2) Begriffsbestimmung

Die abhängige Variable ist diejenige Merkmalsdimension, von der ausgesagt werden soll,wieviele Elemente der Population (unter welchen Bedingungen → unabhängige Variable) darauf eine Ausprägung (welcher Größe) besitzen oder nicht.

Beim Beispiel der Aspiration von [t] ist die abhängige Variable also die Merkmalsdimension "Aspiration".

3) Meßbarkeit der Variablen

Zunächst muß festgestellt werden,ob die Variable meßbar ist,d.h. ob Ausprägungen auf der betr. Merkmalsdimension empirisch erfaßbar sind.Das Merkmal "Aspiration" ist z.B. direkt beobachtbar.Ist dies aber nicht der Fall (z.B. beim Merkmal "Grammatikalität"),muß das Merkmal operationalisiert werden;dies geschieht meistens in zwei Schritten:

a) Es werden Indikatoren aufgestellt,das sind prinzipiell wahrnehmbare Merkmale,die eine Ausprägung der gesuchten Merkmalsdimension anzeigen.

b) Die Indikatoren werden operationalisiert,d.h. es werden Untersuchungsoperationen angegeben,durch die eindeutig das Vorhandensein oder Nichtvorhandensein der Indikatoren gemessen werden kann.

4) Meßniveau und Meßskala

Weiterhin muß entschieden werden,ob man nur bestimmen will ob die Untersuchungseinheiten eine Ausprägung auf der betr. Merkmalsdimension besitzen oder nicht,oder ob man, wenn ja,die Größe dieser Merkmalsausprägungen genauer erfassen möchte.Dies betrifft die Frage des Meßniveaus. Je nach Art der verwendeten Meßskala erzielt man unterschiedliche Meßniveaus;es lassen sich folgende 4 Meßskalas und Meßniveaus unterscheiden:

a) Nominalskala / Nominalmeßniveau
Hier wird nur gemessen,ob eine Ausprägung auf der betr. Merkmalsdimension vorliegt oder nicht.
Beispiel: t_1 ist aspiriert ($A+t_1$),t_2 ist nicht aspiriert ($A-t_2$) usw.

b) Ordinalskala / Ordinalmeßniveau
Hier werden die Untersuchungseinheiten hinsichtlich der Größe ihrer Merkmalsausprägung geordnet;dabei ergeben sich prinzipiell 3 Möglichkeiten: Eine Untersuchungseinheit (die eine Merkmalsausprägung besitzt),kann eine größere,kleinere oder gleichgroße Merkmalsausprägung wie eine andere Untersuchungseinheit haben.
Beispiel: $g(At_1) > g(At_2)$, $g(At_1) < g(At_3)$, $g(At_1) \approx g(At_4)$

($g =_{df}$ Größe = Ausprägung)

c) Intervallskala / Intervallmeßniveau
Die Intervallskala erlaubt es,die Größe der Abstände (Intervall) von Merkmalsausprägungen genau zu bestimmen;das setzt das Vorhandensein einer Maßeinheit voraus.
Beispiel: Man führt zur Messung der Merkmalsdimension "Aspiration" die Einheit "Asp" ein;es ergeben sich dann z.B. folgende Werte: $g(At_1)$ = 1o Asp, $g(At_2)$ = 5 Asp , $g(At_3)$ = o Asp.Hier sind dann Addition und Subtraktion möglich,z.B.: $g(At_1)$ - 5 Asp = $g(At_2)$;nicht möglich sind aber Multiplikation und Division,da diese einen natürlichen Nullpunkt voraussetzen.

d) Ratioskala / Ratiomeßniveau
Die Ratioskala ist eine Intervallskala mit natürlichem Nullpunkt.
Beispiel: Für das Beispiel der Aspiration ist keine Ratioskalierung möglich,weil es hier keinen natürlichen (dem Objektbereich immanenten) Nullpunkt gibt.Dagegen könnte man z.B. die Anzahl der Wörter in einem Satz ratioskalieren;hat etwa ein Satz S_1 5 Wörter und ein Satz S_2 1o,so kann man dann rechnen: (Anzahl der Wörter von S_1). 2 = (Anzahl der Wörter von S_2)

5) Diagramm

Die verschiedenen Meßniveaus mit ihren unterschiedlichen Möglichkeiten, Ausprägungen quantitativ zu erfassen, seien noch einmal in einem Diagramm in etwas allgemeinerer Form zusammengefaßt:

Meßniveau	Ausprägungen			
1) nominal-skaliert	+ $Y+x_i$			- $Y-x_i$
2) ordinal-skaliert	< $g(Yx_i) < g(Yx_j)$	> $g(Yx_i) > g(Yx_j)$	= $g(Yx_i) = g(Yx_j)$	(wie 1))
3) intervall-skaliert	+ $g(Yx_i) + a.e = g(Yx_j)$	- $g(Yx_i) - a.e = g(Yx_j)$	= (wie 2))	$g(Yx_i) = o.e$
4) ratio-skaliert	. $g(Yx_i) .a = g(Yx_j)$	: $g(Yx_i) :a = g(Yx_j)$	= (wie 2))	(wie 3))

($g =_{df}$ Größe $e =_{df}$ Einheit)

6) Wahl einer geeigneten Meßskala

Wie man am Beispiel des Merkmals "Aspiration" gesehen hat, ist die Wahl einer Meßskala

a) von dem angestrebten Quantifizierungsgrad

b) von dem Objektbereich

abhängig; nominal lassen sich alle Merkmale skalieren, je höher aber der Quantifizierungsgrad einer Skala, dest weniger Merkmale können damit skaliert werden. Im Folgenden soll es nur um nominal skalierte Variablen gehen; einmal dominieren diese in der linguistischen Arbeit, zum anderen würde der Aufsatz sonst für den gegebenen Rahmen zu umfangreich und zu kompliziert.

B) Unabhängige Variable(n)

Unabhängige Variablen sind solche, die Einfluß darauf besitzen, wieviele Elemente der Population eine Ausprägung (welcher Größe) auf der abhängigen Merkmalsdimension besitzen. Damit ist noch nicht gesagt, daß es sich hierbei um einen(teil-)kausalen Einfluß handelt; man führt allerdings primär solche unabhängigen Variablen ein, bei denen man dies vermutet; im übrigen gilt Entsprechendes wie bei der abhängigen Variablen.

III) Beziehungen zwischen Objektbereich und Variablen

A) Vorbemerkung

Man kann generell unterscheiden :

1) deskriptive
2) explikative/prediktive

Hypothesen und entsprechend

1) korrelative
2) (teil-)kausale

Beziehungen.

In diesem Rahmen soll nur auf korrelative Beziehungen eingegangen werden,und zwar auch nur auf einen Teil von ihnen und unter maximaler Berücksichtigung von 3 Variablen.

Ich bin auch hier -wie in dem ganzen Referat- davon ausgegangen,Verständlichkeit und Anschaulichkeit gegenüber Allgemeinheit,Umfassendheit und formaler Strenge höher zu bewerten.

B) Beziehungen zwischen Objektbereich und einer (abhängigen) Varieblen

1) Häufigkeitsverteilung

Nachdem nun

a) der Objektbereich bestimmt ist
b) die abhängige Variable und deren Meßniveau (hier nominal) eingeführt sind,

untersucht man,wieviele Elemente der Population (hier: [t]) eine Ausprägung auf der abhängigen Merkmalsdimension (hier: $A=_{df}$ Aspiration) besitzen.

Man nennt dies Häufigkeitsverteilung,und zwar läßt sich unterscheiden zwischen

a) absoluter Häufigkeit
b) relativer Häufigkeit

Dies läßt sich am besten an Hand einer Tabelle erklären.

2) Tabelle

		Sample: [t]
A	+	v=1200
	-	w= 800
		N=v+w= 2000

Es handelt sich hierbei um eine sog. 2-Felder-Tabelle mit den 2 Feldern v und w.
Die Tabelle gibt die Häufigkeitsverteilung einer (natürlich konstruierten) Stichprobe von 2ooo [t] -Äußerungen bezüglich der Merkmalsdimension A ($=_{df}$Aspiration) wieder.

3) Absolute Häufigkeiten f(A)
Die absoluten Häufigkeiten (sie sollen hier mit dem Symbol f(..) gekennzeichnet werden) sind einfach aus der Tabelle abzulesen.
Es gilt: f(A+) = v = 12oo
Das bedeutet: 12oo [t] der Stichprobe haben eine Ausprägung auf A,sind also aspiriert.
f(A-) = w = 8oo , 8oo [t] sind folglich nicht aspiriert.

4) Relative Häufigkeiten p(A)
Die relative Häufigkeit kann definiert werden (vgl.X,II,2,f/Seite 4) als Anzahl der tatsächlichen Fälle geteilt durch die Anzahl der möglichen Fälle.
Die relative Häufigkeit von A+ (geschrieben p(A+) berechnet sich dann wie folgt:
Die tatsächliche Anzahl von Elementen,die eine Ausprägung auf A besitzen,also f(A+),beträgt 12oo; die mögliche Anzahl beträgt dagegen 2ooo (f(A+) + f(A-)),denn es ist ja theoretisch denkbar,daß alle 2ooo [t] der Stichprobe aspiriert sind.
Um die komplizierteren f(..)-Ausdrücke zu vermeiden, kann man schreiben: $p(A+) = \frac{v}{v+w} = \frac{12oo}{2ooo} = o,6$
p(A-) berechnet sich entsprechend,also: $p(A-) = \frac{w}{v+w} = \frac{8oo}{2ooo} = o,4$. Da sich p(A+) und p(A-) zu 1 ergänzen müssen,kann man p(A-) auch folgendermaßen berechnen: p(A-) = 1 - p(A+) = 1 - o,6 = o,4;entsprechend gilt natürlich: p(A+) = 1 - p(A-).

5) Deterministische vs, statistische Häufigkeitsverteilung
Es lassen sich folgende relative Häufigkeiten (hier: = Wahrscheinlichkeiten) unterscheiden:
a) deterministische
b) statistische

ad a) : Im deterministischen Fall haben entweder alle Elemente oder kein Element eine Ausprägung auf der betr. Merkmalsdimension.
Es besteht also entweder eine Wahrscheinlichkeit von p=1 oder von p=o.
Für eine statistische Verteilung gilt: Die Wahrscheinlichkeit p,daß ein Element eine Ausprägung auf der betr. Merkmalsdimension besitzt,ist gleich r, wobei gilt : $o < r < 1$.
(Die beiden deterministischen Fälle sind also die Grenzfälle einer Folge von statistischen Verteilungen.)

Beispiel:

	2ooo [t]	
	A+	A-
1) determinist.		
a) p(A+) = 1	2ooo	o
b) p(A-) = o	o	2ooo
2) statistisch		
- p(A+) = o,6	12oo	8oo

Variablenregeln sind -wie schon gesagt (vgl.f/Seite 4) - genau solche Regeln,die statistische Verteilungen beschreiben;deshalb soll es im Folgenden auch nur um diese gehen.

C) Beziehungen zwischen Objektbereich, unabhängigen Variablen und abhängiger Variable

1) Vorbemerkung: möglichst allgemeine Gesetze als (ein) Ziel wissenschaftlicher Arbeit

a) Ziel wissenschaftlicher Arbeit ist es immer, Gesetze (Regeln, Hypothesen) von möglichst hoher Allgemeinheit (Wahrscheinlichkeit) aufzufinden. Als Ideal gelten deterministische Gesetze, die eine Wahrscheinlichkeit von p=1 besitzen (vgl. 5/Seite 13,14).

b) Wenn man deshalb eine stochastische Gesetzmäßigkeit aufgefunden hat, so wird man versuchen, deren Wahrscheinlichkeit durch Einführung zusätzlicher Bedingungen (variablen) zu erhöhen bzw. zu erniedrigen. Dies sei etwas näher erläutert: Der wissenschaftlich uninteressanteste Fall liegt dann vor, wenn sich eine Wahrscheinlichkeit von p=o,5 ergibt; man spricht dann von einer Zufallsverteilung. Angenommen im Beispiel der Aspiration von [t] sind 5o% der [t] aspiriert und 5o% nicht, dann ist dies so zu interpretieren, daß sich [t] gegenüber der Merkmalsdimension A"neutral" verhält, es besteht keine spezifische Beziehung von [t] zu A. Deshalb ist nicht nur eine Variable interessant, die -im Beispiel- die Wahrscheinlichkeit von p(A+) für [t] erhöht, sondern auch eine, die diese Wahrscheinlichkeit erniedrigt, denn dadurch erhöht sich ja die Wahrscheinlichkeit von A- für [t]; primär ist man allerdings an einer Erhöhung von p(A+) interessiert.

c) Die Einführung von Variablen läßt sich allerdings auch anders als mit dem Ziel der Erreichung möglichst hoher p-Werte begründen; darauf braucht hier aber nicht eingegangen zu werden.

2) Einführung einer unabhängigen Variablen

a) Beispielregel

Es sei erinnert an die Beispielregel
t → ₹t^h₹ / ₹M₹ ___ ₹'V₹ (vgl. c/Seite 2) bzw.
p(M,B/A) = r (vgl. i/Seite 5).
Es soll nun zunächst die Variabel M (=$_{df}$ vorausgehende Morphemgrenze) eingeführt werden.
Dazu sei wieder eine Tabelle aufgestellt.

b) Tabelle für M,A (Tabelle II)

Stichprobe: [t]				
	M			
A		+	-	
	+	a=800	b=400	a+b=1200
	-	c=700	d=100	c+d=800
		a+c=1500	b+d=500	N=2000

c) Absolute Häufigkeiten f(M,A)

Die absoluten Häufigkeiten können wieder -wie bei Tabelle I- einfach abgelesen werden;es gilt:
f(M+,A+) = a = 800
f(M+,A-) = c = 700
f(M-,A+) = b = 400
f(M-,A-) = d = 100

d) Relative Häufigkeiten p(M/A)

- Hier ist zunächst zu unterscheiden zwischen
  - absoluter Wahrscheinlichkeit p(M,A)
  - bedingter Wahrscheinlichkeit p(M/A)

  Die absolute Wahrscheinlichkeit wäre im Fall M+,A+ folgendermaßen zu berechnen:

  $$p(M+,A+) = \frac{a}{a+b+c+d} = \frac{800}{2000} = 0{,}4$$

  Die bedingte Wahrscheinlichkeit für M+,A+ würde man dagegen folgenderweise berechnen:

  $$p(M+/A+) = \frac{a}{a+c} = \frac{800}{1500} = 0{,}53$$

Die absolute Wahrscheinlichkeit besagt hier also, wie wahrscheinlich es ist,daß [t] aspiriert ist u n d [t] eine Morphemgrenze vorausgeht;dagegen besagt die bedingte Wahrscheinlichkeit,wie wahr-

scheinlich es ist,daß [t] aspiriert ist, w e n n [t] eine Morphemgrenze vorausgeht. Hier wird es im Folgenden nur um die bedingten Wahrscheinlichkeiten gehen.

- Es seien jetzt also die bedingten Wahrscheinlichkeiten der 4 möglichen Kombinationen von M und A angegeben:

$$p(M+/A+) = \frac{a}{a+c} = \frac{800}{1500} = 1 - p(M+/A-) = o,53$$

$$p(M+/A-) = \frac{c}{a+c} = \frac{700}{1500} = 1 - p(M+/A+) = o,47$$

$$p(M-/A+) = \frac{b}{b+d} = \frac{400}{500} = 1 - p(M-/A-) = o,8$$

$$p(M-/A-) = \frac{d}{b+d} = \frac{100}{500} = 1 - p(M-/A+) = o,2$$

-- Wie man sieht,berechnet sich die bedingte Wahrscheinlichkeit in Richtung der unabhängigen Variablen,d.h. man fragt: Wenn eine Morphemgrenze vorausgeht,wie hoch ist dann die Wahrscheinlichkeit,daß [t] aspiriert ist ,und nicht: Wenn [t] aspiriert ist,wie hoch ist dann die Wahrscheinlichkeit,daß eine Morphemgrenze vorausgeht.(Unter bestimmten Bedingungen berechnet man allerdings auch die umgekehrten Werte,also p(A/M)).

e) Interpretation der p-Werte

Es sei daran erinnert,daß berechnet wurde: p(A+)=o,6; und es gilt: p(M+/A+) = o,53 .M ist also eine Variable,die die Häufigkeit von A+ (in Bezug auf [t]) reduziert,und zwar um den Faktor o,o7;denn: p(A+) - p(M+/A+) = o,6 - o,53 = o,7 .Entsprechend wird p(A-) um den Faktor o,o7 von o,4 auf o,47 erhöht.

Dagegen ist bei M-,also wenn keine Morphemgrenze vorausgeht,p(A+) um den Faktor o,2 erhöht;denn es gilt: p(M-/A+) - p(A+) = o,8 - o,6 = o,2

Abschließend hierzu kann gesagt werden: Es ist also gelungen,durch Einführung von M (genauer M-) p(A+) (in Bezug auf [t]) vono,6 auf o,8 zu erhöhen und damit beträchtlich dem Idealwert von p=1 anzunähern.Es soll nun durch Einführung einer zweiten unabhängigen Variablen versucht werden,p(A+) weiter zu erhöhen.

3) Einführung einer zweiten unabhängigen Variablen

a) Beispiel

Als Beispiel sei hier die Variable B (=$_{df}$ ___'V =$_{df}$ "vor betontem Vokal" genommen.

Streng genommen handelt es sich hierbei um eine zusammengesetzte Variable, die hier aber als einfache Variable behandelt werden soll.

b) Tabelle B/A (Tabelle III)

Stichprobe: [t]				
		B(V) =B		
		+	-	
A	+	a'=900	b'=300	a'+b'= 1200
	-	c'=450	d'=350	c'+d'= 800
		a'+c'= 1350	b'+d'= 650	N= 2000

c) Absolute Häufigkeiten

f(B+,A+) = a' = 900

f(B+,A-) = c' = 450

f(B-,A+) = b' = 300

f(B-,A-) = d' = 350

d) Relative Häufigkeiten:

p(B+/A+) = 0,67

p(B+/A-) = 0,33

p(B-/A+) = 0,46

p(B-/A-) = 0,54

e) Interpretation der p-Werte

Bei der Variablen B ist es umgekehrt wie bei der Variablem M; B+ erhöht die Wahrscheinlichkeit von A+ um den Faktor 0,07 von 0,6 auf 0,07; dagegen erniedrigt B- p(A+) (in Bezug auf [t]) um den Faktor 0,06 von 0,6 auf 0,54.

4) Modelle zur Berechnung von p(X,T/Y) aus p(X/Y) und p(T/Y)

a) Vorbemerkung

Man ist nun bei dem wohl zentralen statistischen Problem im Zusammenhang mit Variablenregeln angelangt,nämlich: Gelingt es,aus der Häufigkeitsverteilung der Untersuchungselemente auf der abhängigen Merkmalsdimension bei Abhängigkeit von einzelnen unabhängigen Variablen die Häufigkeitsverteilung bei Abhängigkeit von Kombinationen von (2) unabhängigen Variablen,sog. Umgebungen,zuberechnen ?

Wir können das jetzt in der statistischen Schreibweise sehr viel eleganter und verständlicher formulieren: Gelingt es,aus p(X/Y) und p(T/Y) den Ausdruck p(X,T/Y) zu berechnen ?

Unter Verwendung der Beispielvariablen M,B und A wäre zu formulieren: Gelingt es p(M,B/A) aus p(M/A) und p(B/A) zu berechnen ?

Beziehungsweise im Einzelfall: Gelingt es, p(M+,B+/A+) aus p(M+/A+) und p(B+/A+) zu berechnen ? usw. Dies sei noch einmal veranschaulicht:

Wenn wir wissen:

(1) wenn eine Morphemgrenze vorausgeht,sind 53% aller [t] aspiriert und

(2) wenn ein betonter Vokal folgt,sind 67% aller [t] aspiriert

Können wir dann daraus berechnen,wieviel [t] aspiriert sind,wenn eine Morphemgrenze vorausgeht u n d ein betonter Vokal folgt ?

b) Linguistische Untersuchungen

Bekanntlich hat man sich gerade in der linguistischen Literatur um eine Lösung dieses Problems bemüht.

Es sollen im Folgenden die wichtigsten der dort unternommenen mathematischen Lösungsversuche systematisiert dargestellt werden.Dabei sollen die entsprechenden Gleichungen aber nicht in allgemeiner Form,sondern wieder für das Beispiel der Aspiration von [t] formuliert werden.

c) Formeln zur Berechnung von p(M,B/A) aus p(M/A) und p(B/A)

Da man ja-wie gesagt- primär an der Kombination von M und B interessiert ist,bei der sich der höchste Wert für A+ erwarten läßt,soll folglich von den höchsten p(A+)-Werten ausgegangen werden,nämlich p(M-/A+) = o,8 und p(B+/A+) = o,67.

	p(M-/A+)		p(B+/A+)		p(M-,B+/A+)
(I) Anwendungsmodelle					
(A) additive					
(1) mit M-Faktor					
(a) konstant (k=o,5)	o,8	.o,5+	o,67	.o,5=	o,74
(b) variabel		.a +		.b =	
(2) ohne M-Faktor		+		=	1,47
(B) multiplikative					
(1) mit M-Faktor					
(a) konstant		.		.o,5=	o,27
(b) variabel		.a .		.b =	
(2) ohne M-Faktor		.		=	o,54

	1-p(M-/A+)		1-p(B+/A+)		1-p(.)
(II) Nichtanwendungsmodelle					
(A) additive					
(1) mit M-Faktor					
(a) konstant	o,2	.o,5+	o,33	.o,5=	o,26
(b) variabel		.a +		.b =	
(2) ohne M-Faktor		+		=	o,53
(B) multiplikative					
(1) mit M-Faktor					
(a) konstant		.		.o,5=	o,o3
(b) variabel		.a .		.b =	
(2) ohne M-Faktor		.		=	o,o7

M-Faktor $=_{df}$ Multiplikationsfaktor

(Die Bezeichnungen Anwendungsmodell und Nichtanwendungsmodell ergeben sich daher,daß man in der linguistischen Beschreibung von der Wahrscheinlichkeit der Regelanwendung bzw. -nichtanwendung ausgeht,d.h. in der Linguistik faßt man eine Regel meistens dynamisch,als eine Operation auf,und nicht wie in der Statistik,statisch,im Sinne einer Regelmäßigkeit. (vgl. auch b/Seite 3)

d) Besprechung der verschiedenen Formeln

- Als erste Bedingung muß man fordern,daß nicht Lösungen entstehen können,die mathematisch unmöglich sind,vorallem keine p-Werte größer oder kleiner 1;von daher kommen die beiden additiven Modelle ohne Multiplikationsfaktor nicht in Frage,denn wie man sieht,kommt etwa beim vorliegenden Beispiel beim additiven Anwendungsmodell (ohne M-Faktor) p(M-,B+/A+) = 1,47 heraus;ebenso können beim

Nichtanwendungsmodell mathematisch unmögliche Werte herauskommen.

- Auch die Modelle mit variablem Multiplikationsfaktor (also die Gleichungen,in denen die Variablen a und b vorkommen) sind zunächst einmal uninteressant;sie könnten zwar bei Einsetzung geeigneter Faktoren zu richtigen Lösungen führen,es ist aber bis jetzt nicht zu sehen,wie man im Einzelfall die Größe dieser variablen Faktoren bestimmen könnte.
- Schließlich läßt sich zu den Multiplikationsmodellen allgemein sagen,daß sie kaum zu adäquaten Lösungen führen dürften.Denn außer in dem Extremfall,daß beide singuläre Wahrscheinlichkeiten gleich 1 sind,ergibt sich für die kombinierte Wahrscheinlichkeit immer ein geringerer Wert als jeder der beiden Einzelwahrscheinlichkeiten. Das würde bedeuten (im Beispiel),es wäre gar nicht möglich,durch Einführung zusätzlicher Variablen p(M-,B+/A+) zu erhöhen,ein kaum denkbares Ergebnis. Übrigens ergeben bei den multiplikativen Modellen (im Gegensatz zu den additiven Modellen) die Anwendungs- und Nichtanwendungsmodelle nicht einander entsprechende Werte.
- Es bleiben übrig: Das additive Anwendungsmodell mit konstantem Multiplikationsfaktor von o,5 und das entsprechende Nichtanwendungsmodell.Hier soll nur das erste herangezogen werden,das zweite führt ja -wie gesagt- zu entsprechenden Ergebnissen.Das additive Anwendungsmodell mit konstantem Multiplikationsfaktor von o,5 : $p(M/A) \cdot o,5 + p(B/A) \cdot o,5 = p(M,B/A)$ ist natürlich nichts anderes als die Berechnung des arithmetischen Mittels,ein Verfahren,das also auch intuitiv einleuchtet.

e) Berechnung aller Kombinationsmöglichkeiten mit der Formel: $p(M/A).o,5 + p(B/A).o,5 = p(M,B/A)$

Es soll nun mit dieser Formel bzw. der äquivalenten Formel $(p(M/A) + p(B/A)) \cdot o,5 = p(M,B/A)$ die 8 möglichen kombinierten Wahrscheinlichkeiten berechnet werden, die sich beim Beispiel der Aspiration von [t] in Abhängigkeit von M und K ergeben.

M	B	A			
+	+	+	(o,53 + o,67).o,5	=	o,6
+	+	-	(o,47 + o,33)	=	o,4
+	-	+	(o,53 + o,46)	=	o,5
+	-	-	(o,47 + o,54)	=	o,5
-	+	+	(o,8 + o,67)	=	o,74 (max.)
-	+	-	(o,2 + o,33)	=	o,26 (min.)
-	-	+	(o,8 + o,46)	=	o,63
-	-	-	(o,2 + o,54)	=	o,37

Die erste Zeile ist folgendermaßen zu lesen: $(p(M+/A+) + p(B+/A+)).o,5 = p(M+,B+/A+) = (o,53 + o,67).o,5 = o,6$; die anderen Zeilen sind entsprechend zu lesen.

f) Bewertung des Berechnungsmodells $(p(M/A) + p(B/A)).o,5 = p(M,B/A)$ anhand der Beispielwerte

Wie man sieht beträgt der Maximalwert für $p(M,B/A)$ $p=o,74$, und zwar in dem Fall M-, B+, dies ist natürlich genau der Fall, in dem auch $p(M/A)$ und $p(B/A)$ maximal groß sind, nämlich bei $p(M-/A+) = o,8$ und $p(B+/A+) = o,67$. Logischerweise kann $p(M,B/A)$ bei diesem Berechnungsmodell nicht größer als die größere der beiden Einzelwahrscheinlichkeiten bzw. nicht kleiner als die kleinere der beiden Einzelwahrscheinlichkeiten sein, sondern liegt eben immer genau zwischen $p(M/A)$ und $p(B/A)$. Das ist ohne Zweifel ein Resultat, daß den Wert dieses Berechnungsmodells sehr fragwürdig erscheinen läßt.

Um aber in allgemeinerer Form darzustellen, ob eine Berechnung von $p(X,T/Y)$ aus $p(X/Y)$ und $p(T/Y)$ überhaupt denkbar ist und welche Bedingungen ein entsprechendes Rechenmodell erfüllen müßte, muß eine genauere mathematische Analyse unternommen werden.

5) Versuch der Berechnung von X,T/Y aus X/Y und T/Y, ausgehend von absoluten Häufigkeiten

A) Absolute Häufigkeiten als primäre Häufigkeiten

Wie gezeigt wurde,führt von den verschiedenen Modellen zur Berechnung der relativen Häufigkeit von X,T/Y (im Beispiel: M,B/A) aus den relativen Häufigkeiten von X/Y (M/A) und T/Y (B/A) nur eins, nämlich: $p(X/Y).o,5 + p(T/Y).o,5 = p(X,T/Y)$ zu einer überhaupt denkbaren und verwendbaren Lösung; aber auch diese Lösung ist -wie ja erläutert wurde- recht fragwürdig.

Um genau zu überprüfen,ob dieses Rechenmodell zu richtigen Lösungen führt,gibt es zwei Möglichkeiten:

a) Man kann empirische Untersuchungen vornehmen. Dies ist hier natürlich nicht möglich und könnte zunächst auch allenfalls die Brauchbarkeit der Formel für den speziellen Fall M,B/A beweisen.

b) Man kann mathematisch-theoretisch analysieren, ob die errechneten Werte richtig sein können oder müssen.

Ein solcher Versuch soll hier -allerdings nur ansatzweise und ohne großen mathematischen Aufwand- unternommen werden.

Dabei soll auf absolute Häufigkeiten zurückgegriffen werden,denn diese sind ja die primären Häufigkeiten.Während sich relative Häufigkeiten unproblematisch aus absoluten herleiten lassen,ist das umgekehrt nicht möglich.

B) Exemplarische Darstellung des Versuchs der Berechnung von $f(X,T,Y)$ aus $f(X,Y)$ und $f(T,Y)$

a) Vorbemerkungen

Auch diese Berechnung soll zunächst wieder am Beispiel M,B,A durchgeführt werden.

Um die Rechnung optisch möglichst einfach zu halten,werden für die absoluten Häufigkeiten die Bezeichnungen der entsprechenden Tabellenfelder (vgl. Tabelle II/Seite 16,Tabelle III/Seite 18) eingesetzt,also : $f(M+,A+) = a$, $f(B+,A+) = a'$ usw. Für die 8 kombinierten Werte werden die Buchstaben a,b,c,d mit den Indizes

"1" und "2" eingesetzt,also: $f(M+,B+,A+) = a_1$, $f(M+,B-,A+) = a_2$, $f(M-,B+,A+) = b_1$, $f(M-,B-,A+) = b_2$ usw.

b) Tabellengleichungen

Man kann nun aus den beiden Einzeltabellen folgende Gleichungen aufstellen:

Tabelle M/A	Tabelle B/A
$a_1 + a_2 = a = 800$	$a_1 + b_1 = a' = 900$
$b_1 + b_2 = b = 400$	$a_2 + b_2 = b' = 300$
$c_1 + c_2 = c = 700$	$c_1 + d_1 = c' = 450$
$d_1 + d_2 = d = 100$	$c_2 + d_2 = d' = 350$

Dies sei noch einmal erläutert:

$a_1 + a_2 = a = 800$ entspricht ja:

$f(M+,B+,A+) + f(M+,B-,A+) = f(M+,A+) = 800$

Man unterteilt also die 800 Fälle,in denen M+,A+ gilt einfach in die,in denen außerdem auch B+ gilt (a_1) und die,in denen außerdem B- gilt (a_2).Und es geht dann eben (für diesen Fall) darum,a_1 und a_2 zu berechnen.

c) Maximale und minimale absolute Häufigkeiten

Aus den obigen Gleichungen lassen sich ohne weiteres die Maximal- bzw. Minimalwerte berechnen,die a_1,a_2,b_1,b_2 usw. annehmen können.

- Das sei für a_1 demonstriert;Voraussetzung ist, daß als Lösungsmenge (natürlich) nur die Menge der natürlichen Zahlen (einschließlich o) in Frage kommt.

  Aus $a_1 + a_2 = 800$ folgt dann: a_1 max = 800;

  Aus $b_1 + b_2 = 400$ folgt: b_1 max = 400

  Wenn b_1 max = 400,dann folgt aber aus

  $a_1 + b_1 = 900$: a_1 min = 500

Tabelle der Maximal- und Minimalwerte:

M	B	A		max:	Min:
+	+	+	a_1:	800	500
+	-	+	a_2:	300	0
-	+	+	b_1:	400	100
-	-	+	b_2:	300	0
+	+	-	c_1:	450	350
+	-	-	c_2:	350	250
-	+	-	d_1:	100	0
-	-	-	d_2:	100	0

d) Umformungen der Tabellengleichungen

Die in b) genannten Gleichungen lassen sich in folgender weise umformen:

$a_2 = 800 - a_1$

$b_1 = 900 - a_1$

$b_2 = 400 - b_1 = 400 - (900 - a_1) = -500 + a_1$

$c_2 = 700 - c_1$

$d_1 = 450 - c_1$

$d_2 = 100 - d_1 = 100 - (450 - c_1) = -350 + c_1$

Wie man sieht, sind also stets 3 Werte von einem vierten abhängig; wenn also z.B. a_1 festliegt, dann sind dadurch auch a_2, b_1 und b_2 bestimmt; dagegen sind c_1, c_2, d_1 und d_2 von a_1 völlig unabhängig; zwischen ihnen herrscht aber wieder die gleiche Abhängigkeit.

Diese Abhängigkeiten seien in den folgenden zwei Tabellen veranschaulicht:

e) Tabellen

- Abhängigkeit von a_1, a_2, b_1 und b_2

a_1	a_2	b_1	b_2
500	300	400	0
501	299	399	1
502	298	398	2
...	...	...	...
800	0	100	300

- Abhängigkeit von c_1,c_2;d_1 und d_2

c_1	c_2	d_1	d_2
35o	35o	1oo	o
351	349	99	1
352	348	98	2
...	...	...	...
45o	25o	o	1oo

f) Anzahl der Kombinationsmöglichkeiten
Für die Kombination von a_1,a_2,b_1 und b_2 gibt es also 3o1 (8oo-5oo+1=3o1 , 3oo-o+1= 3o1 usw.) Möglichkeiten.Für die Kombination von c_1,c_2,d_1 und d_2 gibt es 1o1 (45o-35o+1= 1o1,35o-25o+1=1o1 usw.) Möglichkeiten.
Das bedeutet konkret: bei vorgegebenem f(M+,A+) = a = 8oo und f(B+,A+) = a' = 9oo kann f(M+,B+,A+) 3o1 verschiedene Werte annehmen,nämlich von 5oo (einschließlich) bis 8oo (einschließlich);für die anderen 2er-Kombinationen gilt Entsprechendes; will man dagegen berechnen,wieviel Wertkombinationen sich bei vorgegebenem a,a',c und c' für z.B. a_1 & c_1 ergeben,so kommt man auf die Anzahl 3o1 . 1o1 = 3o4o1.

g) Interpretation
Wie man sieht,ist es also in diesem Fall, (bei dem aber ganz bestimmte,stark einschränkende Bedingungen vorausgesetzt wurden, auf die noch eingegan gen werden wird,) möglich,aus den absoluten Häufigkeiten f(M,A) und f(B,A) genau ein Intervall von Werten zu bestimmen,die f(M,B,A) annehmen kann.Und zwar ergibt sich für f(M,B,A+) ein Intervall von 3o1,für f(M,B,A-) ein Intervall von 1o1 Werten Es ist also (in diesem Fall) nicht möglich, eindeutig von f(M,A) und f(B,A) auf F(M,B,A) zu schließen,aber doch innerhalb gewisser Grenzen.Dieses Berechnungsmodell wird noch allgemeiner dargestellt werden,vorher soll aber,ausgehend von den hier errechneten Werten noch einmal auf Wahrscheinl. eingegangen werden

C) Berechnung von p(M,B/A) aus f(M,B,A)

a) Maximal- und Minimalwerte

Aus den Maximal- bzw. Minimalwerten von f(M,B,A) kann man leicht die Maximal- bzw. Minimalwerte von p(M,B/A) berechnen.
Will man etwa den Maximalwert von p(M+,B+/A+) berechnen,so teilt man den Maximalwert von $f(M+,B+,A+) = a_1 \max = 800$ durch $a_1 \max + c_1 \min$ ($c_1 \min$ ist entsprechend der Minimalwert von f(M+,B+,A-));man erhält also: $\frac{a_1 max}{a_1 max + c_1 min} = \frac{800}{800+350} = 0{,}7$

Den Minimalwert von p(M+,B+/A+) berechnet man umgekehrt durch folgende Formel:

$$\frac{a_1 min}{a_1 min + c_1 max} = \frac{500}{500+450} = 0{,}53$$

b) Tabelle der Maximal- und Minimalwerte von p(M,B/A)

	max	min	p(M/A).0,5 + p(B/A).0,5
p(M,B/A)			
+ + +	0,7	0,53	0,6
+ - +	0,55	0	0,5
- + +	1	0,5	0,74
- - +	1	0	0,63
+ + -	0,47	0,3	0,4
+ - -	1	0,45	0,5
- + -	0,5	0	0,26
- - -	1	0	0,37

c) Überprüfung der Formel:
p(M/A).0,5 + p(B/A).0,5 = p(M,B/A)
Es können nun die mittels der oben genannten Formel errechneten Werte überprüft werden, genauer gesagt,ob diese Werte eine mögliche Lösung darstellen,denn daß es sich dabei nur um eine von vielen Lösungen handeln kann,ist ja inzwischen demonstriert worden.Wie man sieht (vgl. b)) stellen diese Werte tatsächlich eine mögliche Lösung dar;inwieweit das allerdings allgemein gilt,bliebe noch zu prüfen.

d) Es sei noch Folgendes angemerkt:
Wie man der Tabelle (auf Seite 27) entnehmen kann, kann $p(M,B,/A)$ in vier Fällen den Wert 1 erreichen und in vier Fällen den Wert o. In diesen Fällen liegt dann keine Variablenregel mehr vor, denn eine Variablenregel ist ja dadurch definiert, daß gilt:
$p(...) = r$, $o \leq r < 1$.

D) Allgemeine Darstellung des Versuchs der Berechnung von f(X,T,Y) aus f(X,Y) und f(T,Y)

a) Einschränkende Bedingungen

Am Beispiel M,B,A wurde gezeigt,daß sich aus f(M,A) und f(B,A) genau ein Intervall von Werten berechnen läßt,die f(M,B,A) und daraus abgeleitet p(M,B/A) annehmen können.

Für das Verhältnis von M und B gelten aber ganz bestimmte Bedingungen,nämlich:

$a + b = a' + b' \;\&\; c + d = c' + d'$

Die aufgezeigte Möglichkeit der partiellen Berechnung von f(M,B,A) aus f(M,A) und f(B,A) gilt nur für Tabellen ,die diese Bedingungen erfüllen (bzw. sich bei Konstanthaltung der relativen Häufigkeiten so umformen lassen,daß sie diese Bedingungen erfüllen.)

Da diese Bedingungen sehr stark einschränkend sind,verliert so das aufgezeigte Berechnungsverfahren erheblich an Bedeutung. Es wäre zu untersuchen,inwieweit sich durch Modifikation des Berechnungsverfahren auch Häufigkeitsverteilungen erfassen ließen,die nicht die Bedingungen

$a + b = a' + b' \;\&\; c + d = c' + d'$

erfüllen.Die in b)/Seite 24 aufgestellten Tabellengleichungen sind jedenfalls nicht lösbar,wenn die genannten Bedingungen nicht erfüllt sind.

Trotzdem soll das betreffende Berechnungsverfahren im Folgenden kurz allgemein beschrieben werden.

b) Allgemeine Feststellungen

- Die Differenz zwischen Maximal- und Minimalwert ist für alle (4) absoluten Häufigkeiten einer Gruppe (für a_1, a_2, b_1, b_2 bzw. für c_1, c_2, d_1, d_2) gleich, nicht aber notwendig gleich für Häufigkeiten verschiedener Gruppen.
- Die Anzahl der Kombinationsmöglichkeiten innerhalb einer Gruppe ist gleich der der Differenz zwischen Maximalwert und Minimalwert + 1.

$A_{k_i} = a_1 \max - a_1 \min + 1 =$ (z.B.)
$b_1 \max - b_1 \min + 1 =$
$a_2 \max - a_2 \min + 1 =$
$b_2 \max - b_2 \min + 1 = 3o1$

- Die Gesamtanzahl der Kombinationsmöglichkeiten ergibt sich durch Multiplikation der singulären Kombinationsmöglichkeiten.

Beispiel: $A_{k_i} = 3o1 \quad A_{k_j} = 1o1$

$A_k = A_{k_i} \cdot A_{k_j} = 3o1 \cdot 1o1 = 3o4o1$

- Es gilt folglich: Je größer die Differenzen zwischen Maximal- und Minimalwerten, desto größer die Anzahl der Kombinationsmöglichkeiten.
- Ist die Bedingung $a+b = a'+b'$ & $c+d = c'+d'$ erfüllt, so gibt es immer endlich viele Lösungen für a_1, a_2 usw.; ist die Bedingung nicht erfüllt, so gibt es -wie gesagt- keine Lösung.
- Fallen Maximalwert und Minimalwert in beiden Gruppen zusammen, dann gibt es jeweils nur eine Kombinationsmöglichkeit.

Beispiel:

$a_1 + a_2 = a = 12oo \quad c_1 + c_2 = c = o$

$a_1 + b_1 = a' = 6oo \quad c_1 + d_1 = c' = 4oo$

$b_1 + b_2 = b = o \quad d_1 + d_2 = d = 8oo$

$a_2 + b_2 = b' = 6oo \quad c_2 + d_2 = d' = 4oo$

In diesem Fall gibt es nur jeweils eine Lösung für die absoluten Häufigkeiten:
$a_1 = 600$ $a_2 = 600$ $b_1 = 0$ $b_2 = 0$
$c_1 = 0$ $c_2 = 0$ $d_1 = 400$ $d_2 = 400$
Entsprechend ergeben sich auch für die relativen Häufigkeiten eindeutige Lösungen, z.B.: $p(X+,T+/Y+) \hat{=} \frac{a_1}{a_1 + c_1} = 1.$

Es sei noch angemerkt, daß es -wie man sieht- auch in diesem extremen Fall nicht gelingt, z.B. $f(X+,T+,Y+) = a_1$ allein aus $f(X+,Y+) = a$ und $f(T+,Y+) = a'$ zu berechnen, sondern daß man (außer in Ausnahmefällen) noch einen dritten Wert benötigt, entweder $f(X-,Y+) = b$ oder $f(T-,Y+) = b'$; dies ist allerdings kein großer Nachteil.(+)
In solchen extremen Fällen ist es also tatsächlich möglich, aus $f(X,Y)$ und $f(T,Y)$ eindeutig $f(X,T,Y)$ und somit auch $p(X,T/Y)$ zu berechnen. Derartige Häufigkeitsverteilungen dürften aber bei der praktischen empirischen Arbeit so selten vorkommen, daß ihre Berechnung nicht von Bedeutung ist.

(+) Dieser Sachverhalt wurde an einer früheren Stelle im Aufsatz (Seite 26) nicht klar genug herausgestellt.

6) Beziehungen zwischen den unabhängigen Variablen

a) Mögliche Bedeutung von X,T für X,T/Y

Wie gezeigt wurde,läßt sich aus f(X,Y) und f(T,Y) auf f(X,T,Y) und p(X,T/Y) nur unter bestimmten Bedingungen und auch dann (i.A.) nur innerhalb eines bestimmten Intervalls schließen.

Es fragt sich nun,wovon es abhängt,welcher mögliche Wert von f(X,T,Y) bzw. p(X,T/Y) im konkreten Einzelfall zutrifft.Naheliegend ist es,daran zu denken,daß die jeweilige Beziehung zwischen den unabhängigen Variablen X und T dafür verantwortlich sein könnte.

b) Kontrolle von X,T

Deshalb soll hier -auch wieder nur ansatzweise und exemplarisch- untersucht werden,ob durch Kontrolle der Beziehung von X und T sich in den angegebenen Fällen nicht nur ein Intervall von Werten für X,T,Y berechnen läßt,sondern ein eindeutiger Wert. Ob sich vielleicht sogar unabhängig von der Bedingung a + b = a' + b' & c + d = c' + d' direkt aus p(X/Y) und p(T/Y) p(X,T/Y) berechnen läßt.

c) Durchrechnung von fünf Beispielen

Es seien fünf Beispielfälle von Häufigkeitsverteilungen von M,B,A unterschieden (bei gleichen Werten von M,A und B,A),für die bestimmte Parameter berechnet werden sollen,die die jeweilige Beziehung von M und B beschreiben.

Zunächst seien dafür die fünf Häufigkeitsverteilungen in eine sog. 8-Felder Tabelle eingetragen.In der Tabelle ist jeweils zunächst der f(M,B,A)-Wert eingetragen,darunter dann der p(M,B/A)-Wert.

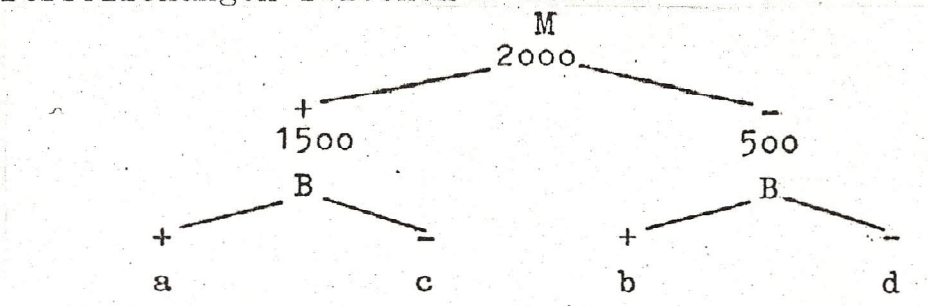
d) Tabelle M,B,A (Tabelle IV)

Stichprobe: [t]							
			M				
			+		-		
			B		B		
			+	-	+	-	
A	+		a_1	a_2	b_1	b_2	
		(1)	800 0,7	0 0	100 0,5	300 1	1200
		(2)	500 0,53	300 0,55	400 1	0 0	
		(3)	540 0,6	260 0,43	360 0,8	40 0,8	
		(4)	600 0,6	200 0,4	300 0,86	100 0,67	
		(5)	550 0,55	250 0,5	350 1	50 0,33	
	-		c_1	c_2	d_1	d_2	
		(1)	350 0,3	350 1	100 0,5	0 0	800
		(2)	450 0,47	250 0,45	0 0	100 1	
		(3)	360 0,4	340 0,57	90 0,2	10 0,2	
		(4)	400 0,4	300 0,6	50 0,14	50 0,33	
		(5)	450 0,45	250 0,5	0 0	100 0,67	
		(1)	1150	350	200	300	2000
		(2)	950	550	400	100	
		(3)	900	600	450	50	
		(4)	1000	500	350	150	
		(5)	1000	500	350	150	

e) Beziehungen zwischen M und B

Aus dieser Tabelle lassen sich auch die jeweiligen Häufigkeitsbeziehungen von M und B (in Bezug auf [t]) ersehen: So gilt z.B. bei (1) : $f(M+,B+) = 1150$, $p(M+/B+) = \frac{1150}{1150 + 350} = 0,77$ usw. Im Folgenden sollen zunächst die absoluten Häufigkeitsbeziehungen zwischen M und B isoliert in einem Diagramm (das einer 4-Felder-Tabelle entspricht) dargestellt werden.

f) Diagramm zur Darstellung der absoluten Häufigkeitsbeziehungen zwischen M und B



	a	c	b	d
(1)	1150	350	200	300
(2)	950	550	400	100
(3)	900	600	450	50
(4)	1000	500	350	150
(5)	1000	500	350	150

g) Korrelative Beziehungen zwischen M und B
Aus den in f) angegebenen absoluten Häufigkeiten sollen verschiedene Parameter berechnet werden, die die Abhängigkeit von M und B beschreiben:
- die schon bekannten relativen Häufigkeiten wie p(M+/B+)
- ein Wahrscheinlichkeitswert, der nicht nur eine singuläre relative Häufigkeit ausdrückt wie z.B. p(M+/B+), sondern die Wahrscheinlichkeiten der 8 Kombinationsmöglichkeiten von p(M|B) zusammenfaßt, nämlich: p(M $\leftrightarrow$ B)
- Korrelationskoeffizienten, $K_{\ddot{A}}$ und phi, die wie p(M $\leftrightarrow$ B) die Werte von p(M+/B+), p(M+/B-), p(M-/B+), p(M-/B-), p(B+/M+), p(B+/M-), p(B-/M+) und p(B-/M-) zusammenfassen. Die Korrelation ist allerdings ein Parameter, der nicht -wie die Wahrscheinlichkeit- von o - 1, sondern von -1 bis 1 geht. Wie man unten sieht, läßt sich der Korrelationskoeffizient $K_{\ddot{A}}(X,Y)$ leicht aus p(X $\leftrightarrow$ Y) gewinnen, denn es gilt: $K_{\ddot{A}}(X,Y) = (p(X \leftrightarrow Y)).2-1$; es bes teht also ein enger Zusammenhang zwischen Korrelation und Wahrscheinlichkeit; allerdings gibt es verschiedene Korrelationskoeffizienten, die auch zu (in begrenztem Ausmaß) differierenden Ergebnissen führen (vgl. Ergebnisse von $K_{\ddot{A}}(M,B)$ und p(M$\leftrightarrow$B)).

h) Tabelle mit Wahrscheinlichkeits- und Korrelations-Parametern für M,B

	p(M+/B+) $\frac{a}{a+c}$	p(B+/M+) $\frac{a}{a+b}$	p(M↔B) $\frac{a+d}{a+b+c+d}$	$K_{\ddot{A}}$ $\frac{2(a+d)}{a+b+c+d}-1$	phi+	p(M+B+/A+)
(1)			o,73	o,46	o,42	o,7
(2)			o,53	o,o6	o,15	o,53
(3)	o,6	o,67	o,48	-o,o4	-o,27	o,6
(4)	o,67	o,74	o,58	o,16	o,o3	o,6
(5)	o,67	o,74	o,58	o,16	o,o3	o,55

+: $\text{phi} = \frac{a.d - b.c}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$

i) Interpretation der Werte für die fünf Beispiele

- (1)/(2) : Hier geht es um die Maximal- bzw. Minimalwerte von f(M,B,A) bzw. p(M,B/A);wie man sieht, differiert die Korrelation von M und B,je nachdem, ob z.B. p(M+,B+/A+) o,7 oder o,53 ist,also maximal oder minimal ist,erheblich.$K_{\ddot{A}}$ beträgt z.B. im Fall (1) o,46,im Fall (2) o,o6.Dies ist aber nicht die maximale Korrelationsdifferenz.
  (Zur Erklärung der Eintragung der Maximal- und Minimalwerte in Tabelle IV sei noch angemerkt:
  Es gelten folgende Abhängigkeiten:
  a_1=max → a_2=min & b_1= min & b_2=max
  a_1=min → a_2=max & b_1= max & b_2=min
  Es können also nicht a_1, a_2, b_1, b_2 zusammen maximal oder minimal seien;
  für c_1, c_2, d_1, d_2 gilt Entsprechendes.)
- (3)/(4) : Im Fall (3)/(4) stimmt p(M+,B+/A+) überein,beträgt nämlich jeweils o,6.Dennoch stimmt keiner der angeführten Parameter,die die Abhängigkeit von M und B ausdrücken,überein. Wie man sieht,kann man also nicht von einem gleichen Wert von p(M+,B+/A+) auf eine gleiche Beziehung zwischen M und B schließen;für die anderen Werte gilt natürlich Entsprechendes.
- (4)/(5) : Bei (4)/(5) umgekehrt stimmen alle Parameter für M,B überein,dennoch varriert p(M+,B+/A+) geringfügig um o,o5.