Inhaltsübersicht

Definitionen und Symbole

$\{a,b,c,\dots\}$	Menge, bestehend aus den Elementen a,b,c,...
$\emptyset$	leere Menge
$\mathbb{N}$	Menge der natürlichen Zahlen 1,2,...
$\min\limits_{j \in A} j$	kleinstes Element der Menge A
$\max\limits_{j \in A} j$	größtes Element der Menge A
$A \cup B$	Vereinigung der Mengen A und B
$A \cap B$	Durchschnitt der Mengen A und B
$A \subset B$	Menge A ist Teilmenge der Menge B

Einleitung

Eine Schule kann wie jeder andere organisatorische Komplex, der zur Erfüllung planvoller Entscheidungen dient, unter ökonomischen Gesichtspunkten analysiert werden. Führt man sich vor Augen, daß es bei einer Anzahl s von Schülern 2^s-1 Möglichkeiten gibt, die Gesamtmenge der Schüler zu Klassen zu gruppieren - bei 10 Schülern sind dies 1023 Möglichkeiten -, so liegt die Frage nahe, ob alle diese Möglichkeiten, selbst wenn jede von ihnen bestimmte pädagogische Ziele gleich gut erfüllt, im Hinblick auf zusätzliche ökonomische Zielsetzungen äquivalent sind. Geht man davon aus, daß die Schüler unterschiedlich begabt sind, d.h. daß ihre Fähigkeit zu lernen unterschiedlich entwickelt ist, dann ergibt sich die Frage, ob aus der stochastischen Verteilung der Lernfähigkeit bestimmte Regeln abgeleitet werden können, die zu einer Klasseneinteilung führen, bei der beispielsweise die von allen Schülern aufzuwendende Lernzeit minimal ist.

Um derartige Fragen untersuchen zu können, bedarf es einer möglichst präzisen Bestimmung des Begriffes Lernen bzw. Lernfähigkeit. Bei der Verwendung dieser Begriffe in den Wirtschaftswissenschaften, beispielsweise zur Erklärung des Phänomens des technischen Fortschritts[1], wird von

1) Arrow, K.J., "The Economic Implications of Learning by Doing", in: Review of Economic Studies, Vol. 29, 1961-62, S. 155-173; Weizsäcker, v., C.Chr., "Zur ökonomischen Theorie des technischen Fortschritts", Göt-

den Autoren auf eine Fundierung ihrer lerntheoretischen Prämissen weitgehend verzichtet. Ähnliches gilt für den Bereich der bildungsökonomischen Literatur: "... the model of student flows divided the educational system up into branches but did not try to get inside the branches and see what was going on there: in other words the educator and the psychologist were left out of the models".[1)]

Da die Relevanz der ökonomischen Deduktionen entscheidend davon abhängt, in welchem Maße die lerntheoretischen Prämissen zutreffen, wurden im ersten Kapitel dieser Arbeit zunächst die wichtigsten Bausteine der stochastischen Theorie des Lernens diskutiert. Dieser Zweig der Lerntheorie verfügt über ein gut ausgebautes formalanalytisches Instrumentarium und schien als Grundlage dieser Arbeit am besten geeignet.

Fortsetzung Fußn. 1) von der vorigen Seite:
tingen, 1966; Wright, T.P., "Factors Affecting the Cost of Airplanes", in: Journal of Aeronautical Sciences, Vol. 3, No. 4, Febr. 1936, S. 122-128; Hirsch, W.Z., "Firm Progress Ratios", in: Econometrica, Vol. 24, April 1956, S. 136-143; ders., "Manufacturing Progress Functions", in: Review of Economics and Statistics, Vol. 34, Mai 1952, S. 143-155; Gody, C.S., "Productivity Changes in Selected Wartime Shipbuilding Programs", in: Monthly Labor Review, Vol. 61, Dez. 1945, S. 1132; Asher, H., Cost-Quantity Relationship in the Airframe Industry", Project RAND Paper R - 291, RAND Corp., Santa Monica (Calif.), Juli 1956.

1) Stone, R., "A View of the Conference", in: Mathematical Models in Educational Planning, OECD: Education and Development, Technical Report, Paris 1967, S. 14.

Nach der Diskussion der lerntheoretischen Modellbausteine werden im zweiten Kapitel die wichtigsten lerntheoretischen Prämissen und die ökonomischen Zielfunktionen zusammengestellt. Dabei wird davon ausgegangen, daß die gegebene Grundgesamtheit der Schüler so in Klassen zusammengefaßt werden muß, daß (a) die pädagogische Zielsetzung, jeder Schüler müsse am Ende des Lernprozesses die gestellte Lernaufgabe mit einer bestimmten Mindest-Erfolgswahrscheinlichkeit beherrschen, erfüllt wird, und (b) darüber hinaus die jeweilige ökonomische Zielfunktion optimiert wird. In den Kapiteln 3 bis 7 werden nacheinander notwendige und hinreichende Bedingungen des optimalen Klassensystems, das die Forderungen (a) und (b) erfüllt, für folgende alternativen Zielfunktionen abgeleitet: Minimierung der von den Lehrern aufzuwendenden Zeit, Minimierung der von den Schülern aufzuwendenden Zeit, Maximierung der Ausbildungserträge, Minimierung der Ausbildungskosten, Maximierung der Nettoerträge der Ausbildung. Darüber hinaus werden einige der Zielfunktionen danach differenziert, ob (a) nur eine, (b) eine endliche Anzahl oder (c) eine unendliche Anzahl von Kohorten in jeder Klasse auszubilden ist.

Fragen, die damit zusammenhängen, ob ein Schüler ausgebildet werden soll oder nicht, bzw. in welchem Fach er ausgebildet werden soll, sind in dieser Untersuchung ausgeklammert worden, d.h. es wurde davon ausgegangen, daß über derartige Fragen bereits Entscheidungen gefallen sind.

Kapitel 1: Abriß der stochastischen Theorie des Lernens

1.1 Einleitende Bemerkungen

Ehe die Axiome der stochastischen Lerntheorie erläutert und ihre wichtigsten Hypothesen auseinandergesetzt werden, ist es angebracht, auf einige Gründe einzugehen, die zur Wahl gerade der stochastischen Lerntheorie als Ausgangspunkt dieser ökonomischen Untersuchung den Ausschlag gegeben haben.
Die stochastische Lerntheorie wurde bisher überwiegend in der experimentellen Tierpsychologie angewandt. Daraus wurde hin und wieder der Schluß gezogen, daß sie auf das Phänomen des menschlichen Lernens nicht anwendbar sei. Dieser Schluß ist vom logischen Standpunkt aus betrachtet nicht begründbar. Ob er vom sachlichen Gesichtspunkt her gerechtfertigt ist, soll in dieser ökonomischen Arbeit nicht entschieden werden. Um Mißverständnissen vorzubeugen, sei betont, daß die Ergebnisse dieser Arbeit nicht im mindesten auf der Voraussetzung basieren, daß sich beispielsweise das bei Ratten experimentell festgestellte Lernverhalten auf den Menschen übertragen läßt. Die Prämissen, die für die abgeleiteten Ergebnisse entscheidend sind, haben vielmehr die Eigenschaft, daß sie mit den unterschiedlichsten Lerntheorien verträglich sind. Die wichtigste Prämisse

lautet hierbei, daß unterschiedliche Lernsubjekte einen unterschiedlich lang bemessenen Unterricht benötigen, um das gleiche Lernproblem mit der gleichen Erfolgswahrscheinlichkeit zu lösen. Lediglich um zeigen zu können, welche Probleme sich ergeben, wenn diese individuell verschiedenen Lernzeiten als stochastische Größen angesehen werden, wurde die stochastische Lerntheorie als formales Gerüst übernommen.

1.2 Bausteine linearer Lernmodelle

1.2.1 Zur Definition des Begriffes Lernen

Der Begriff Lernen wird in der Psychologie auf vielerlei Art definiert. Dabei können zwei große Gruppen von Definitionen unterschieden werden: theoretische und empirische Definitionen. Psychologen, die mit theoretischen Definitionen arbeiten (z.B. Hunter und Hull), versuchen zumeist, das "wahre Wesen" dessen, was sie Lernen nennen, zu entschleiern. Ihre Theorien sind wenig operabel und schwer falsifizierbar. Operable Instrumente und exakte Methoden wurden vor allem von denjenigen Psychologen erarbeitet, die empirische Definitionen des Begriffs Lernen verwenden. Von diesen Definitionen wird im folgenden ausgegangen.

Die empirischen Definitionen werden von Hilgard und Marquis folgendermaßen gekennzeichnet: "... there has always be general agreement among authorities on the subject that learning refers to a more or less permanent change in behavior which occurs as a result of practice. In such a statement, both the dependent variable (changes in behavior) and the independent variable (practice) are reasonably objective. The term learning itself has the status of an intervening, unobserved variable linking these two sets of observals".[1] Die Gruppe der empirischen Definitionen um-

1) Hilgard und Marquis, "Conditioning and Learning", New York 1961, S. 2

faßt ihrerseits zum größten Teil sogenannte stochastische Definitionen des Lernens, die wie folgt charakterisiert werden können. (Auf andere empirische Definitionen, die beispielsweise in der Automatentheorie[1)] verwendet werden, soll im Rahmen dieser Untersuchung nicht eingegangen werden).

Geht man davon aus, daß ein bestimmter Organismus (Lernsubjekt) - sei es ein menschlicher oder tierischer Organismus - in einer bestimmten Entscheidungssituation mit einer Reihe von Verhaltensweisen reagieren kann, wobei keine dieser Verhaltensweisen mit Gewißheit, sondern jede einzelne von ihnen nur mit einer bestimmten Wahrscheinlichkeit auftritt, so kann Lernen allgemein als eine bestimmte Veränderung der Wahrscheinlichkeit definiert werden, mit der eine Verhaltensweise in einer Folge von Entscheidungssituationen auftritt, wenn diese Veränderung als eine Auswirkung kontrollierter Einflüsse interpretierbar ist. Vollzieht sich die Veränderung der Wahrscheinlichkeiten unter der Aufsicht eines Experimentators (Instruktors, Lehrers etc.), so spricht man in bezug auf die Bedingungen und Einflüsse, denen das Lernsubjekt in einer Entscheidungssituation unterworfen ist, von einem Lernexperiment.

In der Lerntheorie wird unter einem Lernexperiment

---

1) Vgl. Menzel, W., "Theorie der Lernsysteme", Berlin 1970; Feichtinger, G., "Stochastische Automaten als Grundlage linearer Lernmodelle", in: Statistische Hefte, N.F., 10. Jg., 1969, Heft 1, S. 5-21

darüber hinaus auch die Gesamtheit einer Reihe von Lernversuchen (trials) verstanden. Der Ablauf eines jeden Lernversuchs läßt sich dabei folgendermaßen schematisieren:

(a) Die Versuchsperson (Schüler) wird dazu veranlaßt, auf eine bestimmte Situation zu antworten, indem man ihr eine Reihe von Möglichkeiten, auf die Situation zu reagieren, vorgibt. Jeder dieser Möglichkeiten entspricht eine besondere Art und Weise, sich zu verhalten.

(b) Bezeichnet man mit A_1, A_2, ..., A_r die verschiedenen Verhaltensweisen bzw. Reaktionen (responses), die im Lernexperiment kontrolliert werden, dann ist die Wahrscheinlichkeit, mit der die Reaktion A_i von der Versuchsperson im ersten Lernversuch gewählt wird, p_i, $i = 1, 2, \ldots, r$. Die Wahrscheinlichkeiten p_1, p_2, ..., p_r des ersten Lernversuches werden dabei als Konstante angesehen, die vom Organismus der Versuchsperson und der Art des Entscheidungsproblems abhängen.

(c) Hat die Versuchsperson mit einer bestimmten Verhaltensweise reagiert, z.B. eine bestimmte Handlung ausgeführt oder unterlassen, dann tritt als Folge der individuellen Reaktionsweise ein bestimmtes Ereignis (event) auf. In bezug darauf, welche Wirkung ein Ereignis auf die Wahrscheinlichkeit ausübt, mit der eine bestimmte Verhaltensweise in dem darauf folgenden Lernversuch ergriffen wird, unterscheidet man folgende Ereignisgruppen:

1. Mit E_0 wird ein Ereignis symbolisiert, das die Wahrscheinlichkeiten aller alternativen Reaktionsweisen unverändert läßt.
2. Verändert ein bestimmtes Ereignis die Wahrscheinlichkeit, mit der eine bestimmte Alternative in einem Lernversuch gewählt wurde, so, daß diese Alternative im nächsten Lernversuch mit einer größeren Wahrscheinlichkeit auftritt als im vorangegangenen Versuch, so spricht man von einem verstärkenden Ereignis (reinforcing event).
3. Davon zu unterscheiden ist wiederum ein Ereignis, das nicht die Wahrscheinlichkeit derjenigen Alternative verstärkt, die zusammen mit diesem Ereignis aufgetreten ist, sondern die Wahrscheinlichkeit irgend einer anderen Alternative.

Zur Konstruktion eines Lernmodells ist es keineswegs nötig, jeder Alternative genau ein verstärkendes Ereignis zuzuordnen. Die Anzahl der definierten Ereignisse und Alternativen muß deshalb nicht übereinstimmen. Ob einer bestimmten Alternative genau ein verstärkendes Ereignis zugeordnet werden muß oder mehrere, hängt davon ab, wie die konkreten Bedingungen des Lernexperiments beschaffen sind. Bei dem in Schulen systematisierten Lernen, bei dem es darauf ankommt, die Wahrscheinlichkeit des Auftretens einer bestimmten erwünschten Verhaltensweise zu erhöhen, kann man in der Regel von mehreren verstärkenden Ereignissen für die gleiche erwünschte Alternative ausgehen. Die verstärkenden Ereignisse können dabei verschiedene Arten der Belohnung für eine bestimmte Alternative sein, wobei diese Alternative als Erbringung einer bestimmten Leistung durch den Schüler definiert sein kann. Zur Definition eines Ereignisses läßt sich verallgemeinert sagen, daß unter diesem Begriff in der Literatur stets Sachverhalte bezeichnet sind, die sich objektiv feststellen lassen. Ein Ereignis, für das die Definition "ein Schüler findet eine Erklärung einleuchtend" gegeben wird, wäre beispielsweise in diesem Sinne nicht objektiv feststellbar, weil es allenfalls durch indirekte Beobachtung erkennbar ist. Dasselbe gilt in bezug auf die Definition des Begriffes Alternative. Auch unter diesem Begriff werden vorwiegend direkt beobachtbare Sachverhalte verstanden.

Um bestimmte Modelleigenschaften transparent zu machen, wird im folgenden auch der Spezialfall herangezogen, daß jede Alternative genau ein verstärkendes Ereignis besitzt. Es ist dann E_i, $i \in \{0,1,2,...,r\}$, dasjenige Ereignis, das die Wahrscheinlichkeit p_i der Alternative A_i, $i \in \{1,2,...,r\}$, verstärkt.

suchen abläuft, kann unter Verwendung dieser Definitionen in einer knappen Symbolik dargestellt werden. Beschreibt man jeden Lernversuch durch ein geordnetes Zahlenpaar (i, k), wobei die Zahl $i \in \{1,2,...,r\}$ die Reaktionsweise A_i und die Zahl $k \in \{0,1,2,...,q\}$ das Ereignis E_k angibt, die in einem Lernversuch festgestellt wurden, dann entspricht jedem Lernprozeß eine bestimmte Sequenz von geordneten Zahlenpaaren. Werden in einem Lernexperiment drei Reaktionsweisen und damit vier Ereignisse (E_0, E_1, E_2, E_3) unterschieden, so kann ein Lernexperiment beispielsweise zu folgendem Lernprozeß führen:

$$(1, 3), (1, 0), (3, 1), \ldots$$

Die Beispiele, mit denen die abstrakten Begriffe Lernversuch, Reaktionsweise und Ereignis in der Literatur illustriert werden, beziehen sich fast immer auf Tierexperimente. Am häufigsten findet man dabei die verschiedenen Anwendungsmöglichkeiten des sog. "T-maze experiment", durch das ein Tier gezwungen wird, nach dem Durchlaufen einer geraden Strecke, die an ihrem Ende eine T-förmige Gabelung aufweist, sich entweder nach rechts oder links zu wenden. In einem derartigen Experiment kann man beispielsweise definieren:

A_1 = Rechtswendung

A_2 = Linkswendung

E_1 = Fütterung am Ende der rechten Gabelung

E_2 = Fütterung am Ende der linken Gabelung

E_0 = Fütterung weder nach einer Rechts- noch nach einer Linkswendung

In diesem Beispiel kann ein Lernexperiment beispielsweise zu folgender Sequenz von Lernversuchen führen:

$$(1,1), (2,0), (2,0), (1,1), (2,0), (1,1), (1,1), \ldots$$

Werden die Ereignisse E_0, E_1, E_2 vom Experimentator kontrolliert (die andere Möglichkeit, daß das Lernsubjekt sie kontrolliert, wird weiter unten behandelt), und vollzieht sich diese Kontrolle so, daß die Alternative A_1 immer nur mit dem Ereignis E_1 und die Alternative A_2 immer nur mit E_0 auftritt, dann ist zu erwarten, daß der Lernprozeß nach einer ausreichend großen Anzahl von Lernversuchen mit immer größerer Wahrscheinlichkeit zur Wahl der Alternative A_1 führt. Dieses Experiment kann in allen möglichen Abwandlungen angewandt werden. So läßt sich beispielsweise bestimmen, daß in allen Versuchen mit der Alternative A_1 das Ereignis E_1 in 60% der Fälle und das Ereignis E_2 in 40% der Fälle auftritt, während in allen Versuchen mit A_2 das Ereignis E_1 und E_2 zu je 50% erscheinen. Bei all diesen Anwendungsmöglichkeiten wird dann die Frage gestellt, wie sich aus der Kenntnis einer endlichen Sequenz von Versuchen der Ablauf des Prozesses voraussagen läßt. Besonders interessant ist u.a. die Frage, nach wieviel Versuchen eine bestimmte Alternative A_i mit einer Mindestwahrscheinlichkeit p von beispielsweise $0{,}8 \leqq p \leqq 1$ auftreten wird.

Bei der Anwendung der Lerntheorie in den Wirtschaftswissenschaften wurde die Frage, inwieweit das menschliche Lernen aus einer Sequenz identischer Entscheidungssituationen besteht, so gut wie überhaupt nicht überprüft. In dem oben dargestellten Beispiel steht das Lernsubjekt in jedem Lernversuch vor identischen Reaktionsmöglichkeiten A_1, A_2, ..., A_r. Diese Prämisse ist sicherlich für viele Phänomene des menschlichen Lernens gerechtfertigt, für andere nicht. Da die Ergebnisse dieser Untersuchung nicht davon abhängen, ob diese Prämisse zutrifft (- zur Ableitung der Untersuchungsergebnisse wird ein einfacheres Lernmodell zugrunde gelegt, für das diese Prämisse nicht konstitutiv ist -), soll diese primär psychologische Fragestellung hier nicht diskutiert werden. Zur Illustration des Problems mag es genügen, darauf hinzuweisen, daß z.B. die Alternative eines Schülers, in einer Klausur eine schriftliche Leistung zu erbringen, nicht ohne weiteres mit einer Alternative bzw. Leistung gleich gesetzt werden kann, die aus einer mündlichen Prüfung resultiert.

1.2.2 Axiome linearer Lernmodelle - Formulierung des allgemeinen linearen Modells

Mit den bisher definierten Größen A_i, E_i und p_i kann zwar ein konkretes Lernexperiment beschrieben werden, aber eine Theorie über den zu erwartenden Ablauf solcher Experimente ist daraus nicht ableitbar. Die bloß beschreibende Funktion dieser Begriffe muß daher zu einer erklärenden erweitert werden. Dafür sind indessen eine Reihe weiterer Definitionen erforderlich.

Jedes Lernexperiment besteht aus einer Sequenz von diskreten Lernversuchen, wobei jeder Lernversuch durch ein bestimmtes Zahlentupel (i, k), $i = 1,2,\dots,r$, $k = 0,1,2,\dots,r$, repräsentiert werden kann. Bezeichnet man mit X die Menge all dieser geordneten Sequenzen mit einer unendlichen Anzahl von Zahlentupeln und mit x ein Element dieser Menge, z.B. die Sequenz

$$(1, 2), (1, 1), (1, 0), \dots,$$

dann ist $X = \{x\}$.

Außerdem sollen folgende Mengendefinitionen verwendet werden:

$\{A_{i,n}\}$ = Menge aller Sequenzen, bei denen im n-ten Lernversuch die Alternative A_i erscheint.

$\{E_{k,n}\}$ = Menge aller Sequenzen, bei denen im n-ten Lernversuch das Ereignis E_k auftritt.

$\{x_n\}$ = Menge aller Sequenzen, die bei den ersten n Lernversuchen dieselben Alternativen und Ereignisse aufweisen wie eine bestimmte vorgegebene Sequenz x.

Die Wahrscheinlichkeiten für das Auftreten bestimmter Alternativen und Ereignisse in bestimmten Lernversuchen werden wie folgt definiert:

$p_{i,n} = P(A_{i,n})$ = Wahrscheinlichkeit, daß im n-ten Lernversuch die Alternative A_i auftritt.

$p_{xi,n} = P(A_{i,n} | x_{n-1})$ = bedingte Wahrscheinlichkeit, daß im n-ten Lernversuch die Alternative A_i auftritt, wenn die vorangegangene Sequenz von Alternativen und Ereignissen gegeben ist.

$P_{x,n} = P(x_n)$ = Wahrscheinlichkeit, mit der eine Sequenz zu erwarten ist, die für die ersten n Lernversuche mit einer gegebenen Sequenz x übereinstimmt.

$\pi_{k,n} = P(E_{k,n})$ = Wahrscheinlichkeit, daß im n-ten Lernversuch das k-te Ereignis auftritt.

$\pi_{ik,n} = P(E_{k,n} | A_{i,n})$ = Wahrscheinlichkeit, daß im n-ten Lernversuch das Ereignis E_k auftritt, wenn im gleichen Lernversuch die Alternative A_i erschienen ist.

Anhand dieser Definitionen ist es möglich, durch Postulierung bestimmter Zusammenhänge zwischen den Wahrscheinlichkeitsvariablen ein Lernmodell zu definieren. Den Zusammenhängen selbst wird der Rang von Axiomen zugeschrieben, deren Gültigkeit als gegeben vorausgesetzt werden muß, (was nicht hindert, daß sie erforderlichenfalls durch andere Axiome ersetzt werden können).

<u>Definition</u>[1]: Ein Ein-Parameter-Lernmodell wird durch die Größen X, $p_{i,n}$, $p_{xi,n}$, $P_{x,n}$, $\pi_{k,n}$, $\pi_{ik,n}$ und θ, $0 < \theta \leq 1$, konstituiert, wenn für jedes $x \in X$ mit der Eigenschaft $P_{x,n} > 0$ und für jedes i, $1 \leq i \leq r$,

1) Vgl. auch Estes, W.K. und Suppes, P., "Foundations of Linear Models", in: Studies in Mathematical Learning Theory, Hrsg.: Bush, R.R. und Estes, W.K., Stanford (Calif.): University Press, 1959, S. 137-179

und jedes k, $0 \leqq k \leqq r$, die folgenden Axiome gelten:

<u>Axiom 1</u>. Für $x \in \{E_{i,n}\}$ gilt:

(1.1) $\quad p_{xi,n+1} = (1-\Theta)p_{xi,n} + \Theta$

<u>Axiom 2</u>. Für $x \in \{E_{k,n}\}$, $i \neq k$, $k \neq 0$ gilt:

(1.2) $\quad p_{xi,n+1} = (1-\Theta)p_{xi,n}$

<u>Axiom 3.</u> Für $x \in \{E_{o,n}\}$ gilt:

(1.3) $\quad p_{xi,n+1} = p_{xi,n}$

Wie eine Interpretation der Axiome ergibt, sind die hypostasierten Zusammenhänge nicht unplausibel. Axiom 1 besagt: Wenn innerhalb einer bestimmten Sequenz in einem bestimmten Lernversuch die Alternative A_i und im selben Lernversuch das verstärkende Ereignis E_i erschien, dann wird die Alternative A_i im darauffolgenden Lernversuch mit einer größeren Wahrscheinlichkeit auftreten als für ihr Auftreten im vorangegangenen Lernversuch zu erwarten war. Ist die Alternative A_i nicht zusammen mit Ereignis E_i, sondern mit Ereignis E_k, $k \neq i$, $k \neq 0$, aufgetreten, das die Wahrscheinlichkeit einer anderen Alternative A_k vergrößert, dann folgt aus Axiom 2, daß im darauffolgenden Lernversuch die Alternative A_i mit geringerer Wahrscheinlichkeit auftreten wird als für ihr Auftreten im vorangegangenen Lernversuch zu erwarten war. Axiom 3 schließlich nimmt bezug auf den Fall, daß im n-ten Versuch einer bestimmten Sequenz die Alternative A_i zusammen mit dem neutralen Ereignis E_o

erschien. Die bedingte Wahrscheinlichkeit, mit der A_i in diesem Versuch auftrat, ist dabei gleich groß wie die bedingte Wahrscheinlichkeit, daß A_i auch im (n+1)-ten Versuch auftritt.

In der Formulierung dieses Lernmodells sind eine Reihe wichtiger Prämissen enthalten, die kurz zusammengefaßt werden sollen:

(1) Das Spektrum aller Verhaltensweisen, die vom Lernsubjekt in einer durch das Lernexperiment gegebenen Entscheidungssituation ergriffen werden können, wird durch die Definition von r Entscheidungsalternativen $A_1, A_2, \ldots, A_r$ ausgeschöpft. Diese Prämisse ist in bezug auf die Anwendung des Modells weniger restriktiv, als es den Anschein hat. Denn es ist immer möglich, eine der r Alternativen so zu definieren, daß sie die Menge "alle übrigen Entscheidungsalternativen" enthält.

(2) Aus der Prämisse, daß die r Entscheidungsalternativen das Spektrum aller möglichen Entscheidungen ausschöpfen, läßt sich folgende Bedingung ableiten, die in jedem Lernversuch erfüllt sein muß

(1.4) $$\sum_{i=1}^{i=r} p_{i,n} = 1 \; ; \quad n = 1,2,\ldots$$

In bezug auf die Ereignisse $E_0, E_1, E_2, \ldots, E_r$ gilt die entsprechende Bedingung

(1.4.1) $$\sum_{k=0}^{k=r} \pi_{k,n} = 1; \; n = 1,2,\ldots$$

(3) Aus Prämisse (2) folgt wiederum, daß immer dann, wenn die Wahrscheinlichkeit einer Alternative A_i sich erhöht

$$p_{i,n+1} > p_{i,n} \quad ,$$

die Wahrscheinlichkeit mindestens einer anderen Alternative A_k sich vermindert:

$$p_{k,n+1} < p_{k,n} \quad .$$

(4) Jede Alternative A_i und jedes Ereignis E_k besitzt in jedem Lernversuch einer Sequenz inhaltlich die gleiche Bedeutung.

(5) Die Veränderung der Wahrscheinlichkeit, mit der eine Alternative im n-ten Lernversuch auftrat, gegenüber der Wahrscheinlichkeit, mit der dieselbe Alternative auch im (n+1)-ten Versuch auftritt, läßt sich durch eine lineare Funktion beschreiben (Axiome 1 und 2). Für die Wahl dieser restriktiven Prämisse (5) gelten innerhalb der Psychometrie und Lerntheorie dieselben Gründe, die auch in der Ökonometrie zur Verwendung von zumeist linearen Beziehungen geführt haben: Bei nicht-linearen Funktionen könnte die Mehrzahl der Theoreme nicht abgeleitet werden.

1.2.3 Die Lernfunktion nach Estes und Suppes und die Funktion nach Bush und Mosteller - ein Vergleich

Der Parameter θ in der Gleichung

$$p_{xi,n+1} = (1-\theta)p_{xi,n} + \theta$$

des Axioms 1 wird in der Literatur als Lernrate (rate of learning) bezeichnet, weil die Wahrscheinlichkeit, mit der eine bestimmte Alternative A_i nach einer Reihe von Lernversuchen auftritt, sich als Folgewirkung des verstärkenden Ereignisses E_i erhöht. Für die Funktion selbst, die das Ausmaß der Erhöhung bestimmt, wird der Begriff Lernfunktion verwendet. Beide Begriffe sind indessen nicht fest fixiert; ihre jeweilige Bedeutung ist davon abhängig, auf welches spezielle Lernmodell sie sich beziehen.

Die von Bush und Mosteller[1)] einerseits und Estes und Suppes andererseits verwendeten Modelle, auf denen die Mehrzahl der lerntheoretischen Arbeiten basiert, sind weitgehend identisch. Abweichungen ergeben sich nur in den Fällen, in denen Bush und Mosteller statt einer Ereignisgruppe E_j zwei Gruppen von Ereignissen unterscheiden: (a) eine Gruppe, die die Klassen O_j (outcomes) enthält, und die inhaltlich mit den bisher verwendeten Ereignisklassen E_j identisch sind und (b) eine Gruppe von zusätzlichen Ereignissen, die aus abstrakten, inhaltlich nicht fixierten Ereignisklassen E'_k bestehen. Dabei wird angenommen, daß nicht nur die verschiedenen Ergebnisse (outcomes) O_j eines Lernversuchs auf die Wahrscheinlichkeiten der Alternativen einwirken, sondern zusätzlich auch die

1) Bush, R.R. u. Mosteller, F., "Stochastic Models for Learning", New York 1955

Ereignisklassen E'_k. Jeder Ereignisklasse E'_k wird eine stochastische Matrix

$$\begin{pmatrix} u_{11,k} & u_{12,k} & \cdots & u_{1r,k} \\ \vdots & \vdots & & \vdots \\ u_{r1,k} & u_{r2,k} & \cdots & u_{rr,k} \end{pmatrix}$$

zugeordnet, aus deren Produkt mit dem Wahrscheinlichkeitsvektor

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_r \end{pmatrix}$$

der die Wahrscheinlichkeit der einzelnen Alternativen angibt, die Veränderung resultiert, die dieser Vektor durch das Auftreten des Ereignisses E'_k erfährt.

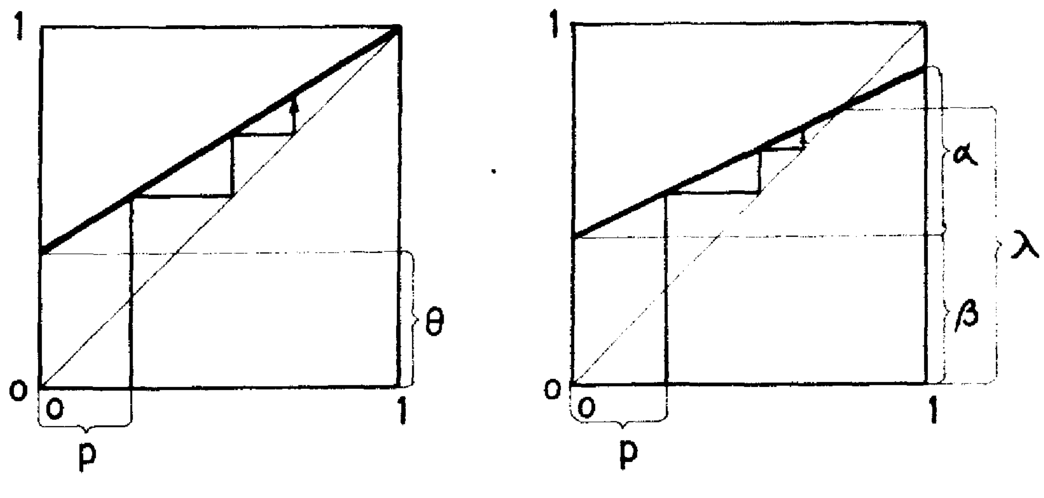
Den überwiegenden Teil ihrer Analyse widmen Bush und Mosteller indessen denjenigen Modellen, in denen zwischen den Ereignisgruppen O_j und E'_k eine eindeutige Zuordnung besteht: O_j tritt immer zugleich mit E'_j auf, so daß es genügt, mit einer einzigen Ereignisgruppe E_j zu arbeiten. In den Fällen, in denen diese Bedingung zutrifft, sind die von Bush und Mosteller und Estes und Suppes verwendeten Modelle bis auf die Hypothese über die Form der Lernfunktion identisch.

Der Unterschied der beiden Lernmodelle hinsichtlich der verwendeten Lernfunktion wird aus Fig. 1 deutlich, wobei der Fall von zwei Alternativen A_1 und A_2 und zwei verstärkenden Ereignissen E_1 und E_2 zum Vergleich herangezogen wurde.

In Fig. 1a) wurde die Lernfunktion nach Estes und Suppes zugrunde gelegt, in Fig. 1b) die von Bush und Mosteller. Während sich im ersten Fall die Wahrscheinlichkeit, mit der ein Schüler das Lernproblem löst, beliebig nahe der Wahrscheinlichkeit 1 annähern läßt, kann die maximale Wahrscheinlichkeit in Fig. 1b) nicht über den Fixpunkt λ hinaus erhöht werden.

Fig. 1

a) Estes u. Suppes b) Bush und Mosteller



Um mit einer bestimmten endlichen Folge von Lernversuchen eine maximale Erhöhung von p_1 zu erreichen, ist es nötig, daß in jedem Lernversuch das verstärkende Ereignis E_1 auftritt. Unter dieser Voraussetzung ist die

Form der Treppenkurve, mit der sich der Lernprozeß darstellen läßt (Fig. 1), stets dieselbe, ganz gleich in welcher Reihenfolge die Alternativen A_1 und A_2 in den einzelnen Versuchen auftreten.
Aus Fig. 1 ist ersichtlich, daß das Modell von Estes und Suppes als Spezialfall ($\lambda = 1$) im Modell von Bush und Mosteller enthalten ist. Die Größe λ, über die hinaus die Wahrscheinlichkeit, eine bestimmte Alternative zu wählen, nicht erhöht werden kann, ist einer der wichtigsten Parameter in derartigen Lernmodellen.

1.2.4 Spezifizierung des allgemeinen linearen Modells

Das allgemeine Modell kann in verschiedener Hinsicht abgewandelt, verfeinert und erweitert werden. So ist es z.B. möglich, den Parameter Θ als Variable zu behandeln[1], oder statt der linearen eine allgemeinere Lernfunktion einzuführen. Die von Bush und Mosteller aufgegriffene Möglichkeit, die verschiedenen Ereignisse in mehrere Gruppen einzuteilen, wurde bereits erwähnt.

Eine allgemeine Einteilung der Ansätze, die sich der Forschung hier bieten, hat zur Klassifizierung der Lernmodelle in "statistische Theorien" oder "stimulus-sampling-Modelle" einerseits und "stochastische Theorien des Lernens" andererseits geführt, wenngleich auch beide Forschungsgebiete eng miteinander verquickt sind.[2]

Während sich bei den stimulus-sampling-Modellen das Hauptinteresse der Forschung darauf konzentriert, durch welche Kausalgesetze Stimulus und Reaktion miteinander verbunden sind, wird in den stochastischen Theorien die aus diesen Gesetzmäßigkeiten resultierende Form des Lernprozesses analysiert. Im Rahmen dieser Untersuchung sind vor allem die Fragen des Prozeßab-

---

1) Vgl. Restle, F., "A Theory of Discrimination Learning", in: Psychological Review, No. 62, 1955, S.11-19; La Berge, D.L., "The Effect of Preliminary Trials on Rate of Conditioning in a Simple Prediction Situation", in: Journal of Experimental Psychology, No. 57, 1959, S. 20-24
2) Vgl. Bush, R.R. und Estes, W.K., "Studies in Mathematical Learning Theory", op.cit., S. 3-5

laufs von besonderem Interesse. Die Probleme des "conditioning" - des Zusammenhangs zwischen Stimulus und Reaktion - beziehen sich auf primär psychologische Tatbestände; sie sollen daher hier nicht erörtert werden.

In bezug auf den Ablauf von Lernprozessen ist die Frage von größter Bedeutung, w e r aus der als gegeben betrachteten Ereignismenge in einem bestimmten Versuch nach dem Auftreten einer Alternative ein bestimmtes Ereignis auswählt. Drei Möglichkeiten sind denkbar: (a) der Experimentator (Lehrer) kontrolliert die Ereignisse, (b) die Versuchsperson (Schüler) kontrolliert die Ereignisse, (c) Lehrer und Schüler üben die Kontrolle gemeinsam aus. Der Begriff Kontrolle wird in der Literatur zur stochastischen Lerntheorie stets anhand der bereits eingeführten Definitionen für die verschiedenen stochastischen Variablen präzisiert.

(a) <u>Vom Experimentator kontrollierte Ereignisse</u>

Mit $\pi_{k,n}$ wurde die Wahrscheinlichkeit bezeichnet, daß im n-ten Lernversuch das k-te Ereignis eintrifft. Wird das Auftreten dieses Ereignisses dadurch bestimmt, daß der Experimentator mittels der Versuchsbedingungen die Größe $\pi_{k,n}$ so beeinflußt, daß sie in allen Versuchen konstant ist, dann nennt man diese Ereignisse "vom Experimentator kontrolliert".

Dafür ein Beispiel. Einer Versuchsperson werden nacheinander 100 verschiedene Rechenaufgaben gestellt. Diese Aufgaben seien inhaltlich von der gleichen Art, es sollen beispielsweise jedes mal zwei Brüche durch einander dividiert werden. Die Lösung kann richtig sein (A_1) oder falsch (A_2), und der Experimentator kann eine Belohnung für die Lösung erteilen oder eine Strafe. Erteilt er in v_1-Prozent aller Versuche, in denen die Lösung richtig war, und in v_2-Prozent aller Versuche, in denen die Lösung falsch war, eine Belohnung ($v_1 + v_2 = 1$), so tritt in jedem Lernversuch einer der folgenden vier Fälle auf:

(i) richtige Antwort - Belohnung
(ii) richtige Antwort - Bestrafung
(iii) falsche Antwort - Belohnung
(iv) falsche Antwort - Bestrafung

Nimmt man an[1], daß die Fälle (i) und (iv) die Wahrscheinlichkeit, mit der im nächsten Lernversuch eine richtige Lösung gefunden wird, in gleichem Maße erhöhen und die Fälle (ii) und (iii) diese Wahrscheinlichkeit in gleichem Maße vermindern, dann bilden (i) und (iv) das Ereignis E_1 und (ii) und (iii) das Ereignis E_2. Unter diesen Voraussetzungen ist die Wahrscheinlichkeit, daß das Ereignis E_1 eintrifft, in jedem Versuch v_1: Ist die Wahrscheinlichkeit, eine richtige Lösung zu finden, p, dann ist im nächsten Lernversuch die Wahrscheinlich-

1) Diese Annahme ist nicht unbedingt erfüllt. In bezug darauf, was mit diesem Beispiel demonstriert werden soll, ist es jedoch unwichtig, ob die Annahme gänzlich zutrifft oder nicht.

keit, daß eine richtige Lösung gefunden und belohnt wird pv_1. Die Wahrscheinlichkeit, daß in diesem Versuch eine falsche Lösung gefunden und nicht belohnt wird, ist $(1-p)(1-v_2)$. Das Ereignis E_1 tritt folglich in jedem Versuch mit der Wahrscheinlichkeit $pv_1 + (1-p)(1-v_2) = pv_1 + (1-p)v_1 = v_1$ auf. Entsprechend ist die Wahrscheinlichkeit des Ereignisses E_2 in jedem Versuch konstant v_2.

(b) Vom Lernsubjekt kontrollierte Ereignisse

Die Ereignisse eines Lernexperiments, für das

$$\pi_{ii,n} = 1 \quad , \qquad i = 1,2,\ldots,r$$

gilt, nennt man "vom Lernsubjekt kontrolliert". Diese Bezeichnungsweise ist wenig einleuchtend, denn es lassen sich immer Beispiele für Lernexperimente finden, in denen $\pi_{ii,n} = 1$ dadurch erfüllt wird, daß der Experimentator die Ereignisse kontrolliert. Die Bezeichnung hat sich jedoch durchgesetzt[1]. Mit dem obigen

---

1) In den Fällen, in denen zwei Ereignisgruppen O_j und E'_j unterschieden werden und in denen $\pi_{ij,n}$ die bedingte Wahrscheinlichkeit zwischen der Alternative A_i und dem Ereignis O_j bedeutet, wird die Bezeichnung einleuchtender, wenn O_j eine Menge von Ereignissen beschreibt, die ausschließlich vom Lernsubjekt hervorgerufen werden können. Beispiel: O_1 = die Versuchsperson folgt einer bestimmten Aufforderung, O_2 = sie folgt der Aufforderung nicht. Ordnet man die Ereignisse der zweiten Gruppe denen der ersten Gruppe eindeutig zu, indem z.B. der Experimentator bei jedem Ereignis O_j ein bestimmtes Ereignis E'_j auftreten

Beispiel läßt sich auch dieser Fall illustrieren, wenn beispielsweise angenommen wird, daß für jede richtige Lösung eine Belohnung erteilt wird und bei jeder falschen Lösung nichts weiter geschieht.

(c) Vom Lernsubjekt und vom Experimentator kontrollierte Ereignisse

Mit $\pi_{ik,n}$ wurde die bedingte Wahrscheinlichkeit bezeichnet, daß im n-ten Lernversuch bei Auftreten der i-ten Alternative das k-te Ereignis eintritt. Bei r Alternativen und r Ereignissen erhält man für jeden Lernversuch eine Anzahl r^2 solcher Größen. Ist jede dieser Größen in allen Lernversuchen konstant, so nennt man die Ereignisse "vom Lernsubjekt und vom Experimentator kontrolliert" (experimenter-subject-controlled events). Im Unterschied zu dem Fall, daß die Wahrscheinlichkeit, mit der eine Alternative auftritt, in jedem Lernversuch konstant ist ("vom Experimentator kontrolliert"), hängt diese Wahrscheinlichkeit nun auch davon ab, welche Alternative das Lernsubjekt gewählt hat.

Fortsetzung der Fußnote der vorherigen Seite:
läßt, dann sind alle Ereignisse vom Lernsubjekt kontrolliert. Die Tatsache, daß die Tierexperimente, die zur Entwicklung der stochastischen Lerntheorie in starkem Maße beigetragen haben, in der Regel solche Ereignismengen enthielten, mag dazu geführt haben, daß die Bezeichnung allgemein für jede Ereignismenge übernommen wurde, für die $\pi_{ii,n} = 1$ gilt.

Im Beispiel zu (a) wurde angenommen, daß $v_1 + v_2 = 1$ ist. Für jeden anderen Fall $v_1 + v_2 \neq 1$ ist die Wahrscheinlichkeit des Auftretens des Ereignisses E_1 nicht mehr in jedem Versuch konstant und gleich v_1, sondern es ergibt sich als Wahrscheinlichkeit für E_1 nach Auftreten von A_1 die Größe $pv_1 + (1-p)(1-v_2)$. Da $1-v_2 \neq v_1$ angenommen wurde, ist $pv_1 + (1-p)(1-v_2) \neq v_1$. Die Wahrscheinlichkeit, mit der das Ereignis eintritt, ist also sowohl von p als auch von v_1 und v_2 abhängig. Da sich die Wahrscheinlichkeit p wiederum während des Lernprozesses ändert, ist $\pi_{1,k}$ keine Konstante mehr. Entsprechendes gilt in bezug auf das Ereignis E_2.

1.3 Der Begriff Lernen im engeren Sinn - Der Ablauf von Lernprozessen

Der Ablauf eines Lernprozesses hängt bei sonst gleichen Bedingungen entscheidend davon ab, wer das Auftreten der reaktionsverstärkenden Ereignisse kontrolliert. Hinsichtlich der Kontrollfunktion wurden die drei Fälle (a), (b) und (c) unterschieden. Zur Darstellung des Prozeßablaufs soll hier jedoch nur der Fall (b) herangezogen werden. Diese Beschränkung ist sinnvoll, weil sich in bezug auf diejenige Eigentümlichkeit der Ablaufprozesse, die in dieser Untersuchung von Interesse ist, die drei Fälle nicht unterscheiden: In allen drei Fällen ergibt sich mit steigender Anzahl von Lernversuchen ein immer breiteres Verzweigungsschema, innerhalb dessen ein bestimmter Ablaufprozeß wie eine Route verfolgt werden kann.

Ausgangspunkt ist ein Lernmodell mit den Alternativen A_1, A_2, A_3 und den verstärkenden Ereignissen E_1, E_2, E_3. Die Wahrscheinlichkeit, mit der vor Beginn des ersten Lernversuchs die Alternative A_1 gewählt wird, sei p_1, die der Alternative A_2 und A_3 entsprechend p_2 und p_3, wobei $p_1 + p_2 + p_3 = 1$. Ferner sei

$$\pi_{ii,n} = 1 \,, \qquad \begin{array}{l} i = 1,2,3. \\ n = 1,2,\ldots \end{array} .$$

Mit L_1 wird ein linearer Operator definiert, der das Ausmaß der Erhöhung der Variablen p_1 beschreibt:

(1.5) $\quad L_1 p_1 := \alpha_1 p_1 + \beta_1 \,.$

Als $L_1^2 p_1$ bzw. $L_1^m p_1$ wird der Fall bezeichnet, daß der Operator L_1 zwei- bzw. m-mal hintereinander angewandt wird (die hochgestellten Symbole sind also keine Exponenten, die eine Potenzierung ausdrücken):

$$L_1^2 p_1 = L_1(L_1 p_1) = \beta_1 + \alpha_1(\alpha_1 p_1 + \beta_1).$$

In Abweichung von den bisherigen Beispielen sei angenommen, daß nicht nur das Ereignis E_1, sondern auch E_2 und E_3 eine Erhöhung von p_1 bewirken. Diese Annahme bezieht sich auf ein Lernexperiment, in dem es dem Experimentator darauf ankommt, die Wahrscheinlichkeit einer bestimmten Alternative (hier von A_1) auf Kosten der Wahrscheinlichkeiten aller übrigen Alternativen zu erhöhen. Dies kann z.B. dann der Fall sein, wenn A_1 bedeutet, eine Aufgabe richtig zu lösen, während

A_2 und A_3 die Menge aller übrigen Reaktionsweisen, die denkbar sind, beinhalten. So kann beispielsweise A_2 bedeuten, daß die Aufgabe nur bis zu einem bestimmten, genau festgelegten Maß gelöst wird, während A_3 so viel wie "nicht gelöst" bedeutet. Wird die Aufgabe vollkommen richtig gelöst, d.h. tritt A_1 auf, dann läßt sich auch sagen, daß das Lernsubjekt in diesem Lernversuch "gelernt" hat, etwas Bestimmtes zu tun. Allerdings wird der Begriff "Lernen" in diesem Beispiel dann auf zwei verschiedene Weisen gebraucht: zum einen bedeutet er in der bisherigen Terminologie jede Veränderung irgendeiner der drei Ausgangswahrscheinlichkeiten p_1, p_2 und p_3, zum anderen bezeichnet er den davon völlig verschiedenen Tatbestand, daß eine bestimmte Alternative gewählt wurde ("Lernen im e.S."). Beim menschlichen Lernen, insbesondere wenn es in Schulen systematisiert wird, kommt es in der Regel darauf an, die Wahrscheinlichkeit des Auftretens bestimmter Verhaltensweisen auf Kosten anderer zu erhöhen. Da in dieser Untersuchung allein solche Lernexperimente interessieren, die dieser Aufgabenstellung entsprechen, wird im folgenden der Begriff Lernen auch im engeren Sinne gebraucht werden.

Die Wirkung der drei Ereignisse auf p_1, d.h. ihr Einfluß darauf, mit welcher Wahrscheinlichkeit das Lernsubjekt im nächsten Lernversuch lernen wird (= die Alternative A_1 wählen wird), möge sich durch die folgenden drei Operatoren beschreiben lassen:

$$(1.6)\quad \begin{array}{l} L_1p_1=\alpha_1 p_1 + \beta_1 \\ L_2p_1=\alpha_2 p_1 + \beta_2 \\ L_3p_1=\alpha_3 p_1 + \beta_3 \end{array} \left\{ \begin{array}{l} 0 \leqq \alpha_i < 1 \\ 0 < \beta_i \leqq 1 \\ \text{für } \alpha_i=0 \text{ ist } \beta_i=1 \\ \alpha_i + \beta_i \leqq 1 \end{array} \right\} \; i=1,2,3$$

Unter diesen Voraussetzungen ergeben sich folgende Möglichkeiten für den Ablauf eines Lernprozesses. Das Resultat des ersten Lernversuches ist eine der drei folgenden Transformationen:

$$L_1p_1, \qquad L_2p_1, \qquad L_3p_1$$

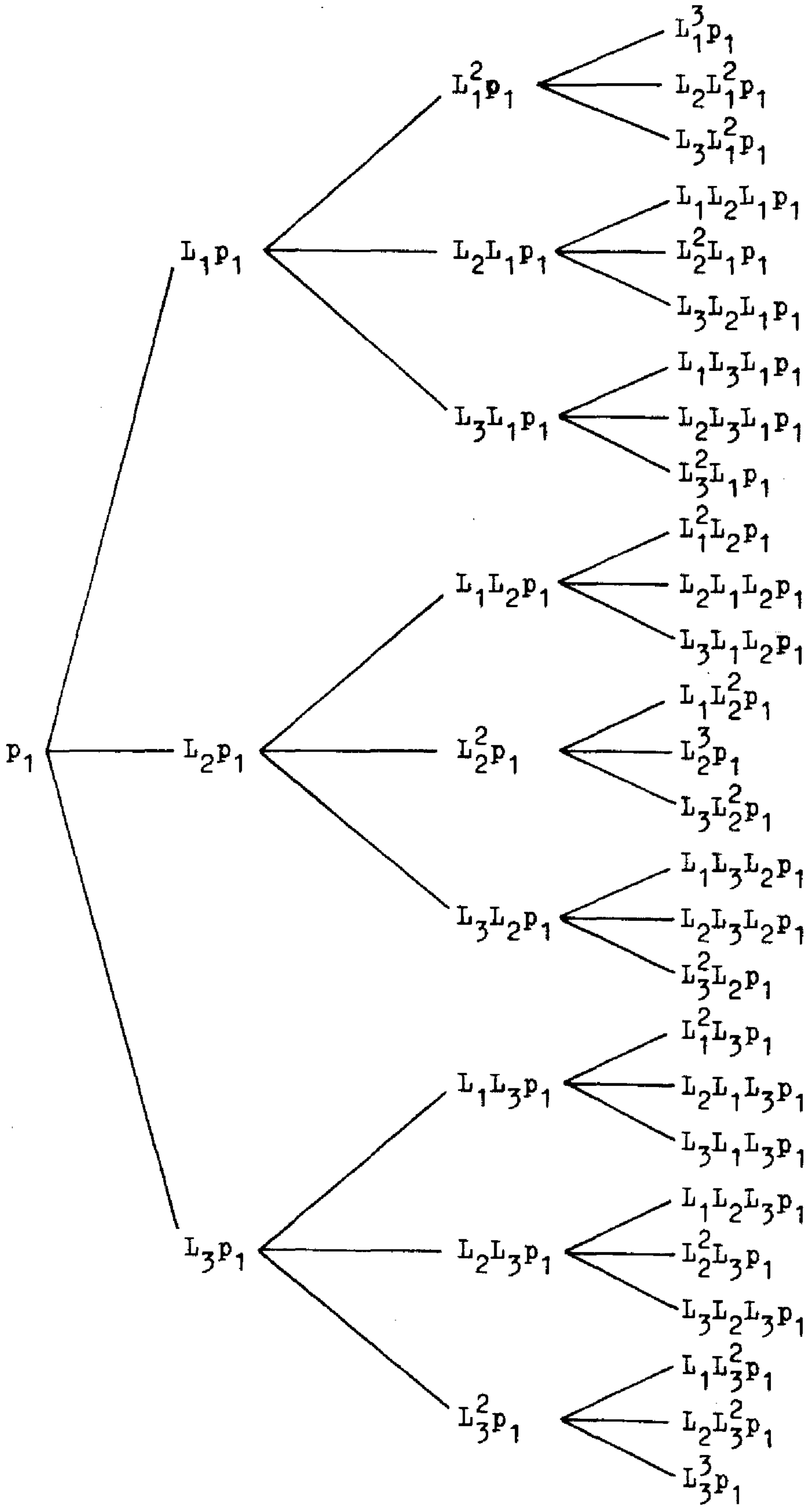
Nach dem zweiten Lernversuch ergibt sich eine der folgenden neun Möglichkeiten:

$L_1^2p_1, L_2L_1p_1, L_3L_1p_1, L_1L_2p_1, L_2^2p_1, L_3L_2p_1, L_1L_3p_1, L_2L_3p_1, L_3^2p_1$

Der Sachverhalt läßt sich anschaulich in Form eines Verzweigungsschemas darstellen (vgl. Übersichtsschema 1).

Gemäß der Bedingung, daß sich die Wahrscheinlichkeiten aller drei Alternativen in jedem Lernversuch zu 1 addieren, muß die Erhöhung von p_1 auf beispielsweise L_1p_1 von einer Verminderung entweder von p_2 oder von p_3 oder von p_2 und p_3 begleitet werden. Für die Zwecke der Darstellung genügt es, in bezug auf diese Verminderung von der einfachen Annahme auszugehen, daß die Wahrscheinlichkeit von p_2 bzw. p_3 nach einer Erhöhung von p_1 auf Lp_1 sich auf

Übersichtsschema 1



$$(1.\ 7) \qquad a_2 \frac{1 - L_1 p_1}{a_2 + a_3} \qquad \text{bzw.} \qquad a_3 \frac{1 - L_1 p_1}{a_2 + a_3}$$

vermindern, d.h. daß die Wahrscheinlichkeiten der Alternativen A_2 und A_3 stets im Verhältnis $a_2 : a_3$ zueinander stehen, wobei a_2 und a_3 beliebige positive Konstante sind.

Mit dieser Annahme kann nun die wichtige Frage beantwortet werden, mit welcher Durchschnittswahrscheinlichkeit ein Lernsubjekt im n-ten Lernversuch die Alternative A_1 wählen und damit die Lösung der gestellten Aufgabe erlernt haben wird. Diese Frage kann auch wie folgt formuliert werden: Angenommen bei einer großen Anzahl N von Lernsubjekten mit identischen Ausgangswahrscheinlichkeiten p_1, p_2 und p_3 werde das Lernexperiment gleichzeitig durchgeführt. Wieviel Lernsubjekte werden dann im n-ten Versuch die Alternative A_1 wählen? Wird diese Alternative von $N_n \in \mathbb{N}$ Lernsubjekten gewählt, dann ist die Durchschnittswahrscheinlichkeit, mit der ein einzelnes Lernsubjekt im n-ten Versuch lernt, gleich N_n/N.

Im ersten Lernversuch lernen voraussetzungsgemäß $N_1 = p_1 N$ Schüler, und es ist $p_{1,1} = p_1$. Von diesen N_1 Schülern lernen im zweiten Versuch

$$(L_1 p_1) N_1 \quad = \quad (L_1 p_1) p_1 N$$

Schüler. Im ersten Versuch haben $p_2 N$ Schüler die Aufgabe nicht vollständig gelöst, d.h. die Alternative A_2 gewählt. Von Ihnen wird im zweiten Versuch die Anzahl

$$(L_2 p_1) p_2 N$$

lernen. p_3N Schüler haben im ersten Versuch die Alternative A_3 gewählt. Von ihnen werden im zweiten Lernversuch

$$(L_3p_1)p_3N$$

lernen.

Insgesamt lernen dann im zweiten Versuch

$$(L_1p_1)p_1N + (L_2p_1)p_2N + (L_3p_1)p_3N$$

die Aufgabe richtig zu lösen. Die Durchschnittswahrscheinlichkeit, daß ein bestimmtes Lernsubjekt im zweiten Lernversuch die Aufgabe richtig löst, ist also

(1. 8) $$p_{1,2} = (L_1p_1)p_1 + (L_2p_1)p_2 + (L_3p_1)p_3.$$

Dabei läßt sich zeigen, daß $p_{1,2} > p_{1,1}$, wenn p_1 kleiner ist als die Fixpunkte der Operatoren. Formt man die Ungleichung

$$(L_1p_1)p_1 + (L_2p_1)p_2 + (L_3p_1)p_3 > p_{1,1}$$

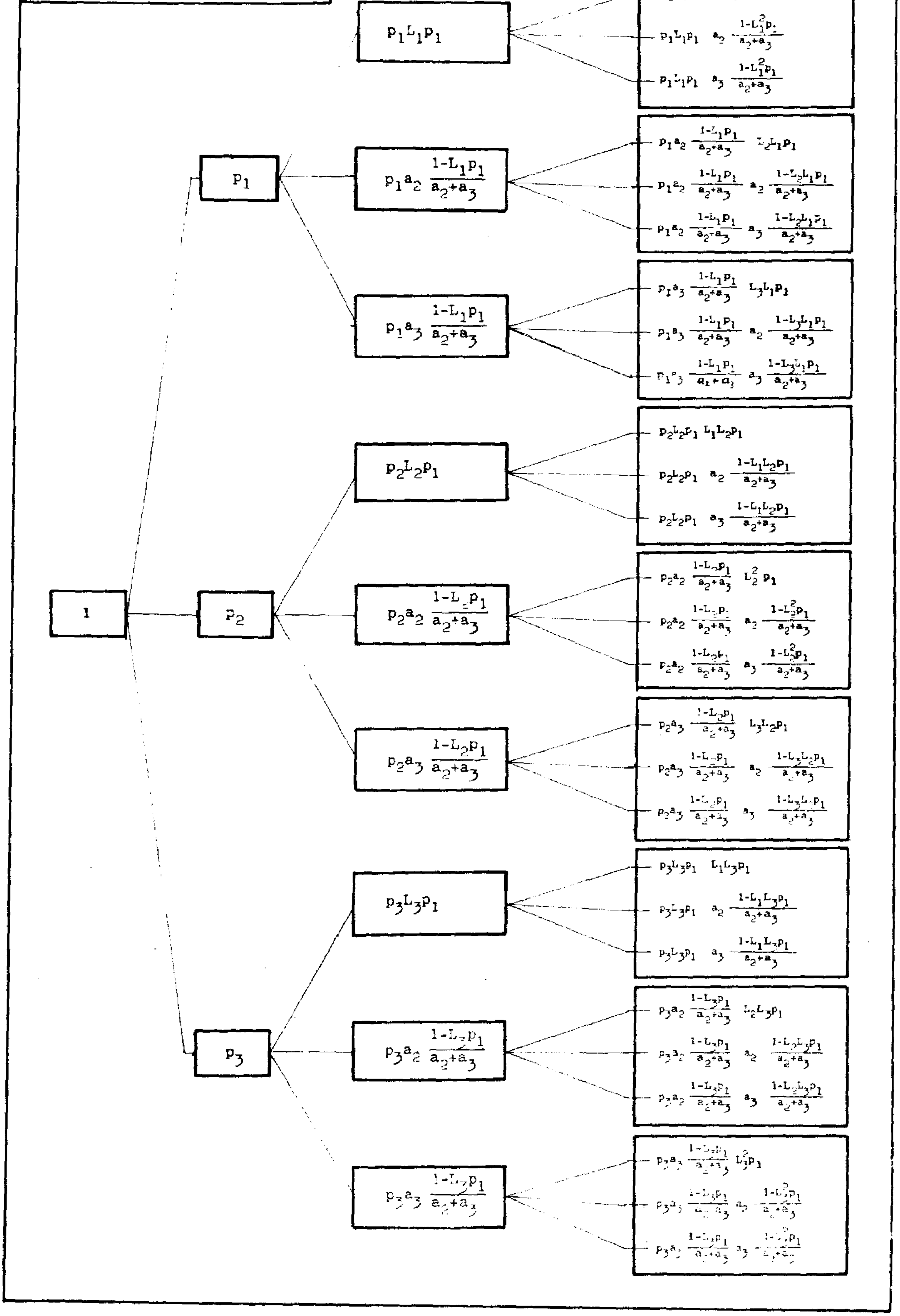
in den Ausdruck

$$L_1p_1 + \frac{(L_2p_1)}{p_{1,1}}p_2 + \frac{(L_3p_1)}{p_{1,1}}p_3 > 1$$

um, so stehen vor p_2 und p_3 Koeffizienten, die beide größer als 1 sind. Da definitionsgemäß $p_1 + p_2 + p_3 = 1$, muß die linke Seite größer als 1 sein.

Die Wahrscheinlichkeit, mit der ein bestimmtes Lernsubjekt während des Lernprozesses einen bestimmten Verzweigungspunkt im Übersichtsschema 1 durchläuft, kann für jeden dieser Punkte berechnet werden. Diese Wahrscheinlichkeiten sind im Übersichtsschema 2 für die Verzweigungspunkte angegeben, die nach drei Lernver-

Übersichtsschema 2



suchen möglich sind. Anhand dieses Schemas kann jeder beliebige Wert $p_{1,n}$ berechnet werden, indem man in der n-ten Spalte alle Ausdrücke summiert, die bei den Positionen der Alternative A_1 stehen. Der Versuch, den Wert $p_{1,n}$ in einer kurzen Formel auszudrücken, würde seinen Zweck verfehlen: schon in diesem recht einfachen Modell wäre eine derartige Formel vermutlich so unhandlich, daß sich der Schreibaufwand nicht lohnen würde. Bush und Mosteller, die diesen Versuch unternommen haben, scheiterten bereits an einem noch einfacheren Modell mit nur zwei Alternativen[1]. Der von ihnen verwendete Begriff "durchschnittliche Wahrscheinlichkeit der Alternative A_1" entspricht der hier gebrauchten Definition für $p_{1,n}$. Der Begriff durchschnittliche Wahrscheinlichkeit ist sprachlich gesehen keine sehr glückliche Konstruktion. Da dieser Begriff jedoch in der Literatur eingeführt ist, soll er auch in dieser Arbeit verwendet werden.

Zur Berechnung des Wertes $p_{1,n}$ müssen 3^{n-1} Einzelausdrücke summiert werden. Die Ausdrücke ergeben sich aus der paarweisen Multiplikation aller Glieder, die in der (n-1)-ten Spalte von Übersichtsschema 1 und 2 vorkommen.

---

1) Vgl. Bush, R.R. u. Mosteller, F., "Stochastic Models for Learning", op. cit., S. 78-81

Besonders wichtig für die Zwecke dieser Untersuchung ist es, zu zeigen, daß die Wahrscheinlichkeit, mit der die Alternative A_1 auftritt, von Lernversuch zu Lernversuch beständig bis zum niedrigsten Fixpunkt λ steigt:[1)]

(1.9) $p_{1,n} > p_{1,n-1}, \quad n = 2,3,\ldots \quad \text{für } p_{1,n-1} < \lambda_1, \lambda_2, \lambda_3$

Der Beweis, der im folgenden an Hand des Beispiels durchgeführt wird, für das oben angenommen wurde, daß sich die Wahrscheinlichkeiten der Alternativen A_2 und A_3 nach einer Erhöhung der Wahrscheinlichkeit von A_1 so verringern, daß sie im Verhältnis $a_2 : a_3$ stehen, gilt auch allgemein. (Um dies zu zeigen, sind in der zur Beweisführung verwendeten Ungleichung (1.10), s.u., die speziellen Funktionen

$$a_2 \frac{1-x}{a_2 + a_3} \quad , \quad a_3 \frac{1-x}{a_2 + a_3}$$

lediglich durch zwei allgemeine Funktionen $f(x)$ und $g(x)$ zu ersetzen.)

---

1) Wendet man n-mal hintereinander den linearen Operator L_i auf die Größe $p_{1,1}$ an, so ergibt sich

$$L_i^n = \alpha_i^n p_{1,1} + (1-\alpha_i^n)\lambda_i \ .$$

Im Grenzübergang $n \to \infty$ erhält man für diesen Ausdruck

$$\lim_{n \to \infty} L_i^n = \lambda_i ,$$

Vgl. Bush, R.R., u. Mosteller, F., "Stochastic Models for Learning", op. cit., S. 59-60. Die Prämisse, daß die Ausgangswahrscheinlichkeit $p_{1,1}$ kleiner ist als der Wert λ, der sich nach einer unendlichen Anzahl von linearen Transformationen der Größe $p_{1,1}$ ergibt, ist nicht unrealistisch.

Nach n Lernversuchen gibt es in Übersichtsschema 1 und 2 in der n-ten Spalte jeweils 3^n Verzweigungspunkte. Die ersten $3^2 = 9$ Verzweigungspunkte in der n-ten Spalte von Übersichtsschema 1 und die entsprechenden Punkte von Übersichtsschema 2 sind zusammen in Übersichtsschema 3 eingetragen. Im Schema 3 stehen deshalb an jedem Verzweigungspunkt zwei Werte übereinander: der obere Wert ist dem Schema 1, der untere dem Schema 2 entnommen. Dabei wurde davon ausgegangen, daß x der erste Wert in der (n-2)-ten Spalte von Schema 1 und y der entsprechende Wert von Schema 2 sei. Es kann nun gezeigt werden, daß für jedes beliebige Paar (x,y), $0 < x,\ \ y < 1$, folgende Ungleichung gilt:

$$(1.10) \qquad xy < xyL_1x + ya_2\frac{1-x}{a_2+a_3}L_2x + ya_3\frac{1-x}{a_2+a_3}L_3x.$$

Wenn die Ungleichung

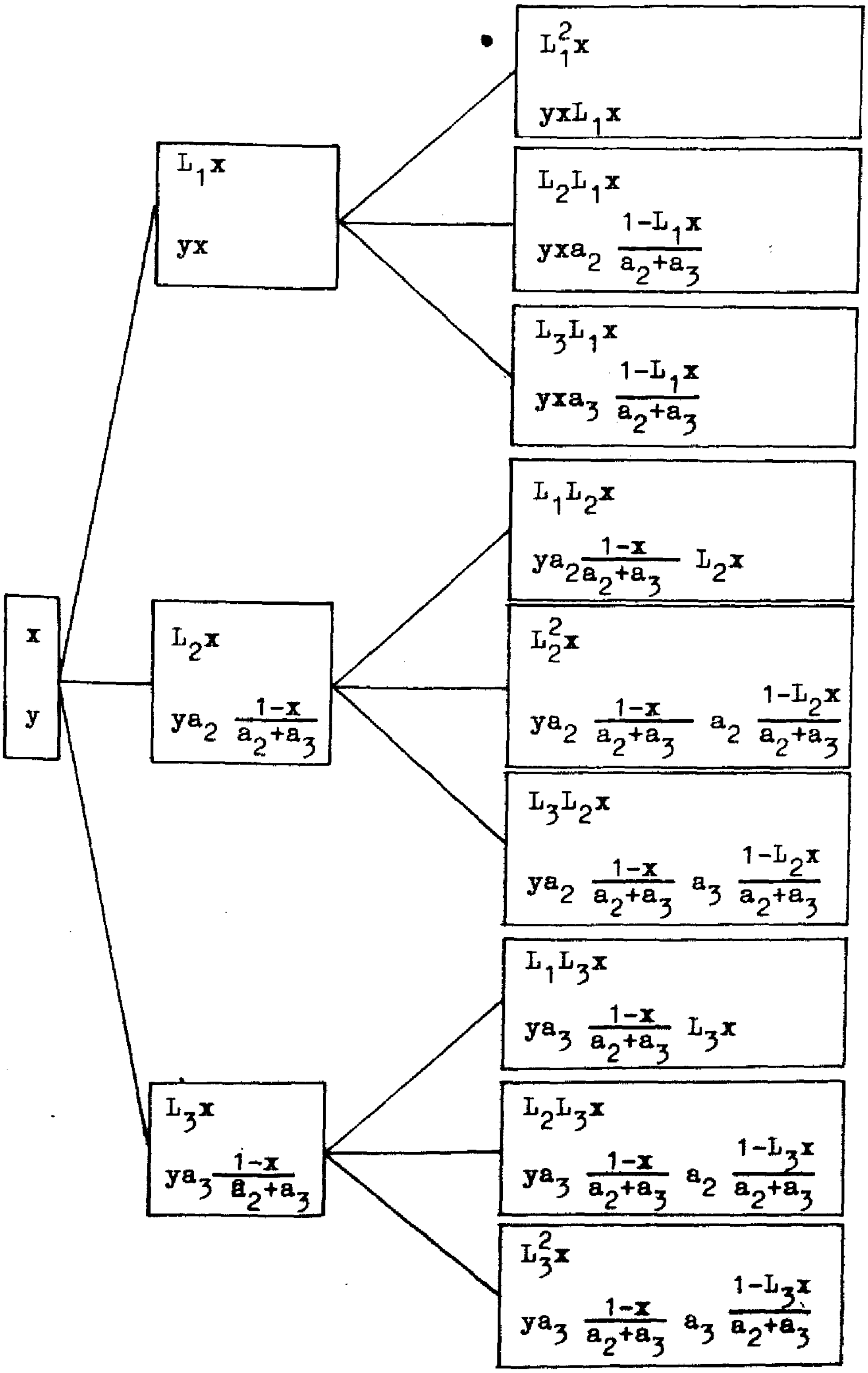
$$(1.11) \qquad xy < xyL_1x + ya_2\frac{1-x}{a_2+a_3}x + ya_3\frac{1-x}{a_2+a_3}x$$

gilt, dann gilt auch Ungleichung (1.10), denn es ist unter den obigen Voraussetzungen stets $L_2x > x$ und $L_3x > x$. Ungleichung (1.11) ist jedoch stets richtig, wie man aus folgender Umformung von (1.11) ersieht:

$$(1.11.1) \qquad 1 < L_1x + (1-x),$$

denn L_1x ist voraussetzungsgemäß größer x. Ungleichung (1.10) besagt nun nichts anderes, als daß die Wahrscheinlichkeit, mit der ein Lernprozeß einen der ersten drei Verzweigungspunkte, die im n-ten Versuch der Alternative A_1 entsprechen, durchläuft, größer ist als die Wahrscheinlichkeit, mit der dieser Prozeß im

Übersichtsschema 3



(n-1)-ten Versuch den ersten Punkt durchläuft, der der Alternative A_1 entspricht. Dieselbe Relation gilt bezüglich des zweiten A_1-Punktes im (n-1)-ten Versuch und den 4., 5. und 6. A_1-Punkten im n-ten Versuch usf. Da es zu jedem A_1-Punkt im (n-1)-ten Versuch drei A_1-Punkte im n-ten Versuch gibt, zwischen denen die Relation (1.10) gilt, muß auch bezüglich der Summe aller A_1-Punkte im (n-1)-ten Versuch und der Summe aller A_1-Punkte im n-ten Versuch die Relation (1.10) gelten. Da die Summe der Wahrscheinlichkeiten aller A_1-Punkte im (n-1)-ten Versuch und die Summe der Wahrscheinlichkeiten aller A_1-Punkte im n-ten Versuch identisch sind mit $p_{1,n-1}$ bzw. $p_{1,n}$, gilt stets $p_{1,n+1} > p_{1,n}$.

Entsprechend zu dem dargestellten Fall läßt sich dieser Beweis für das erweiterte Lernmodell mit r Alternativen $A_1, A_2, \ldots, A_r$, r Ereignissen $E_1, E_2, \ldots, E_r$, bzw. r Operatoren $L_1, L_2, \ldots, L_r$ verallgemeinern. An Stelle von (1.10) erhält man dann

$$
\begin{aligned}
(1.10.1) \quad xy < xyL_1x &+ ya_2 \frac{1-x}{a_2+a_3+..+a_r} L_2x + \\
&+ ya_3 \frac{1-x}{a_2+a_3+..+a_r} L_3x + \ldots + \\
&+ ya_{r-1} \frac{1-x}{a_1+a_2+..+a_r} L_{r-1}x + \\
&+ ya_r \frac{1-x}{a_1+a_2+..+a_r} L_rx.
\end{aligned}
$$

Auch in diesem Fall ist unter den obigen Annahmen stets $p_{1,n} > p_{1,n-1}$ für n = 2,3,... und für $p_{1,n-1} < \lambda_1, \lambda_2, \ldots, \lambda_r$.

Kapitel 2: Die wichtigsten Prämissen - Zielfunktionen und Organisationsprinzipien

2.1 Lerntheoretische Prämissen

(1) In Abschnitt 1.3 wurde ein Lernmodell diskutiert, für das sich die grundlegende Beziehung

$$(2.1) \qquad p_{1,n+1} > p_{1,n}, \qquad n = 1,2,\ldots \text{ für } p_{1,n} < \lambda_i$$

ableiten ließ. Diese Beziehung galt auch für das erweiterte Modell mit r Alternativen $A_1, A_2, \ldots, A_r$ und r Ereignissen $E_1, E_2, \ldots, E_r$. Der Beweis wurde für den Fall durchgeführt, daß die linearen Operatoren λ_1, λ_2 und λ_3 ungleiche Fixpunkte besitzen. Diesem Beweis lag die realistische Prämisse zugrunde, daß die Ausgangswahrscheinlichkeit $p_{1,1}$, mit der im ersten Lernversuch die Alternative A_1 gewählt bzw. die gestellte Lernaufgabe gelöst (oder kurz: "gelernt") wird, kleiner ist als der kleinste Fixpunkt.

Das Modell, für das die Beziehung (2.1) abgeleitet wurde, gehörte der Klasse von Lernmodellen an, bei denen die Ereignisse vom Lernsubjekt kontrolliert werden. Im folgenden wird diese Beschränkung aufgehoben: Die Untersuchung der optimalen Organisation von Lernprozessen in den Kapiteln 3 bis 7 stützt sich nicht mehr auf ein derartig spezielles Lernmodell. Alle folgenden Ableitungen gehen vielmehr lediglich davon aus, daß die konkreten Bedingungen des Lernens so geartet sind, daß sich aus ihnen auf die Beziehung (2.1) schließen läßt. Lernmodelle mit dieser Eigenschaft gibt es in allen drei Klassen von Lernmodellen (vom Lernsubjekt kontrollierte Ereignisse, vom Experimentator kontrollierte Ereignisse, vom Lernsubjekt und vom Experimentator gemeinsam kontrollierte Ereignisse). Da mit dem in Schulen systematisierten Lernen, das in dieser Untersuchung besonders interessiert, der Zweck verfolgt wird, die Wahrscheinlichkeit bestimmter erwünschter Verhaltensweisen so weit wie möglich zu erhöhen, steht die Prämisse (2.1) nicht im Widerspruch zum Gegenstand dieser Untersuchung; der Rahmen, der durch diese Prämisse gesetzt wird, ist vielmehr so weit, daß er die Mehrzahl aller in Schulen relevanten Lernbedingungen und Lernphänomene umfaßt.

Bezüglich der Anzahl, Abgrenzung und Definition von Alternativen werden lediglich folgende Bedingungen gesetzt:

(a) Die Menge aller möglichen Reaktionsweisen des Lernsubjekts wird in r Alternativen $A_1, A_2, \ldots, A_r$ eingeteilt, wobei $r \geqq 2$.

(b) Von diesen Alternativen wird nur die erste inhaltlich definiert: mit A_1 ist diejenige Verhaltensweise gemeint, die sich als "Lösung des Lernproblems" oder kurz: "Lernen" beschreiben läßt. Demnach entspricht die Wahl einer Alternative A_j, $j \neq 1$, im n-ten Lernversuch nicht dem Sachverhalt, den der Experimentator oder Lehrer mit dem Satz beschreiben würde, "das Lernsubjekt S hat im n-ten Lernversuch das Lernproblem gelöst, d.h. eine bestimmte Verhaltensweise erlernt". (Würde das Lernsubjekt schon im ersten Lernversuch die Alternative A_1 wählen, so würde ein Lehrer in der Regel nicht geneigt sein, davon zu sprechen, daß das Lernsubjekt vollständig gelernt habe, weil er vermutlich annähme, daß die richtige Lösung der Lernaufgabe dem Zufall zugeschrieben werden muß. Von "Lernen" im gewöhnlichen Sinn des Wortes würde ein Lehrer erst dann sprechen, wenn er genügend Gründe für die Annahme hätte, daß die Alternative A_1 auch in allen folgenden Versuchen mit einer bestimmten Mindestwahrscheinlichkeit gewählt würde. Der Begriff Lernen, wie er hier gebraucht wird, deckt sich also nicht immer mit dem umgangssprachlichen Ausdruck. Durch die Vorgabe einer bestimmten Mindestwahrscheinlichkeit für die Alternative A_1 in den folgenden Ableitungen wird jedoch gewährleistet, daß diese terminologische Inkongruenz nicht zu sachlichen Fehlschlüssen führt.)

(2) Bei dem in Abschnitt 1.3 diskutierten Modell wurde unterstellt, daß jede Variable $p_{2,n}$, $p_{3,n}, \ldots, p_{r,n}$ eine bestimmte Funktion

$$(2.2) \qquad p_{i,n} = p_{i,n}(p_{1,n}), \qquad i = 2,3,\ldots r; \quad n = 1,2,\ldots$$

der Variablen $p_{1,n}$ ist. Die Form dieser Funktionen wurde so gewählt, daß die Bedingung

$$(2.3) \qquad \sum_{i=1}^{i=r} p_{i,n} = 1 \qquad n = 1,2,\ldots$$

erfüllt blieb. Im folgenden wird ebenfalls angenommen, daß diese Bedingung stets erfüllt ist, ohne daß über die Funktionen (2.2), die dieser Annahme entsprechen, weitergehende Voraussetzungen getroffen werden. Auch hinsichtlich der Spezifikation und Abgrenzung der einzelnen Ereignisse, sowie hinsichtlich der Form der linearen oder nicht linearen Operatoren, mit denen die Wirkung der Ereignisse beschrieben werden kann, sollen keine weiteren Annahmen getroffen werden, außer daß (2.3) stets erfüllt ist.

Unter diesen Bedingungen läßt sich die Wahrscheinlichkeit $p_{1,n}$, mit der A_1 im n-ten Versuch auftritt, als eine Funktion beschreiben, in der außer den Parametern der Funktionen (2.2) und den Parametern der Operatoren nur noch die Größe $p_{1,1}$ enthalten ist. So erhält man beispielsweise für das in Kapitel 1, Abschnitt 3, dargestellte Beispiel im zweiten Lernversuch:

$$(2.4) \qquad \begin{aligned} p_{1,2} = {} & (\alpha_1 p_{1,1} + \beta_1) p_{1,1} + \\ & + (\alpha_2 p_{1,1} + \beta_2) a_2 \frac{1 - \alpha_1 p_{1,1} + \beta_1}{a_2 + a_3} + \\ & + (\alpha_3 p_{1,1} + \beta_3) a_3 \frac{1 - \alpha_1 p_{1,1} + \beta_1}{a_2 + a_3} \end{aligned}$$

Unter der Voraussetzung (2.1) ergibt sich für die Folge der Ausdrücke $p_{1,1}$, $p_{1,2}$, $p_{1,3}, \ldots, p_{1,n}$ der in

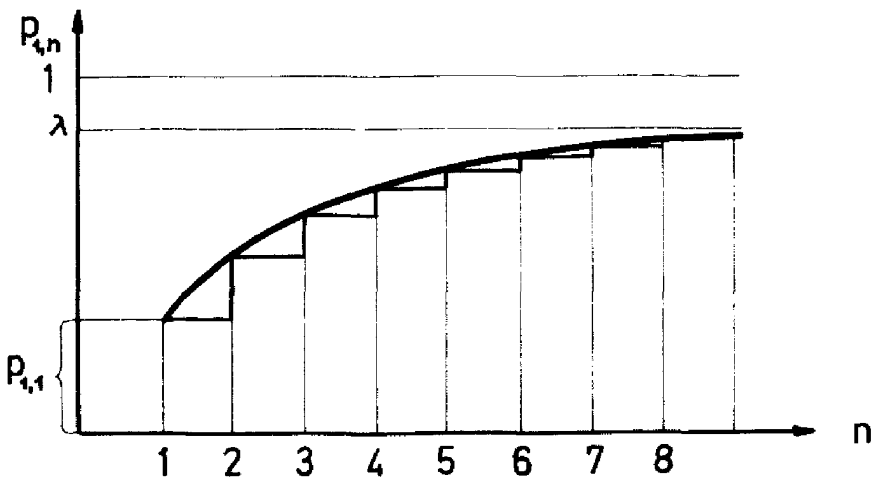
Fig. 2 dargestellte Verlauf. In diesem Schaubild wurde die Annahme berücksichtigt, daß $p_{1,n}$ einen oberen Grenzwert λ besitzt, der kleiner als 1 ist. In den folgenden Überlegungen ist indessen auch der Fall

$$\lambda = \lim_{n \to \infty} p_{1,n} = 1$$

mit eingeschlossen.

(3) Die Variable Zeit spielte in den bisherigen Überlegungen keine Rolle. Diese Größe soll nun eingeführt werden, indem angenommen wird, daß für jeden Lernversuch die gleiche Anzahl von Zeiteinheiten zur Verfügung gestellt wird.

Fig. 2



Ferner sei eine approximierende stetige und differenzierbare Funktion

(2.5) $p = f(t), \qquad t = 0$

eingeführt, auf der sämtliche Punkte $p_{1,n}$, $n = 1,2,...$ liegen. Die Funktion (2.5) hat dann unter Berücksichtigung der bisherigen Prämissen die Eigenschaften

(2.5.1) $\frac{df(t)}{dt} > 0, \qquad t \geqq 0$

(2.5.2) $\lim_{t \to \infty} \frac{df(t)}{dt} = 0$

(2.5.3) $f(0) = p_{1,1}$.

Die Einführung der Funktion (2.5) beruht auf folgender Hypothese: Das Ausmaß, in dem sich die Wahrscheinlichkeit der Alternative A_1 erhöht, hängt ausschließlich von der aufgewandten Lernzeit ab und ist unabhängig von der Anzahl der Lernversuche, die in einem Zeitintervall durchgeführt werden. Oder anders ausgedrückt: die Erhöhung der Wahrscheinlichkeit der Alternative A_1 wird ausschließlich von dem Produkt aus der Anzahl der Lernverscuhe und der Zeitdauer eines Lernversuchs bestimmt. Bugelski, der sich mit den Problemen befaßte, die bei der Anwendung der stochastischen Lerntheorie auf schulische Lernprobleme entstehen, zitiert eine Reihe von Experimenten,[1] die diese plau-

1) Bugelski, R.R., "The Psychology of Learning Applied to Teaching", New York 1964, S. 26

sible Hypothese fundieren: Er hält die Hypothese für genügend gesichert, um aus ihr praktische Folgerungen für die Lernpraxis zu fordern: "Some subjects learn slowly; some rapidly. Those who learn one kind of material slowly might not be so slow with other materials. It would, however, still take the subjects a certain amount of time to learn anything. If this is so, we might have a word of practical advice for curriculum planners. They should remember that there is only so much time in a school day (or school year)".[1] Bugelski begnügt sich mit dem Hinweis darauf, daß praktische Folgerungen aus dieser Hypothese möglich sind. Eine detaillierte Ausarbeitung der praktischen Konsequenzen wird von ihm nicht geliefert.

F. Restle war der erste, der die Analyse in dieser Richtung vorangetrieben hat.[2] Er ging von folgendem Sachverhalt aus: "Even if the students in an experiment (or classroom) are about of the same ability, we know that they will not learn at exactly the same time...".[3] Daraus ergibt sich, "that if the whole class... is taught at once, a great many students will waste a lot of time waiting until the last student learns. If the classes are smaller, less student time is wasted. However, with smaller classes more instruc-

---

1) Bugelski, R.R., op. cit., S. 26-27
2) Restle, F., "The Relevance of Mathematical Models for Education", in: "Theory of Learning and Instruction", The Sixty-third Yearbook of the National Society for the Study of Education, Chicago 1964, S. 111-131
3) op. cit., S. 114

tor-time is needed, since the instructor now must wait, separately, for the completion of learning in each class".[1] Anhand eines einfachen stochastischen Lernmodells leitet Restle die Bedingungen für die optimale Größe einer Klasse ab.

Außer Restle ist G. Feichtinger der einzige Autor, der die stochastische Lerntheorie in dieser Weise auf die Erziehungsplanung angewandt hat.[2] Feichtinger stellte folgendes Problem. Eine Grundgesamtheit von gleich begabten Schülern soll N verschiedene Probleme (items) von gleichem Schwierigkeitsgrad erlernen. Es wird angenommen, daß jeder Schüler eine gewisse Lernzeit L benötigt, um jedes dieser Probleme zu erlernen, wobei L eine exponential verteilte Zufallsgröße ist. Nun wird gefragt, wieviel Zeit zum Erlernen jedes Problems zur Verfügung gestellt werden soll, wenn jede von den Schülern zum Lernen aufgewandte Zeitperiode einen bestimmten Kostensatz verursacht und der volkswirtschaftliche Ertrag eines jeden Schülers, der sämtliche Probleme erlernt hat, gegeben ist. Die N Interventionspunkte, bei denen die Unterrichtung eines Problems zugunsten der Unterrichtung des nächstfolgenden aufgegeben werden muß, damit die Differenz zwischen Ertrag und Kosten ein Maximum wird, lassen sich unter Anwendung der dynamischen Programmierung

---

1) Restle, F., op. cit., S. 121
2) Feichtinger, G., "Über ein Entscheidungsproblem in der Erziehungsplanung", in: Unternehmensforschung, Bd. 14, Heft 1, 1970, S. 140-147

ermitteln.

Sowohl Restle als auch Feichtinger gehen in ihrer Analyse davon aus, daß alle Schüler gleich begabt sind, d.h. daß für jeden Schüler dieselbe Lernfunktion gilt - eine Annahme, die im folgendem durch die realistischere Prämisse zufallsverteilter Begabungen ersetzt wird. Im Gegensatz zu den Ansätzen von Restle und Feichtinger, die die Lernzeit eines Schülers als eine stochastische Variable behandeln, wird im folgenden von zwar unterschiedlichen, aber nicht stochastischen Lernzeiten ausgegangen.

(4) Fragt man danach, wie lange es dauert, bis eine bestimmte Ausgangswahrscheinlichkeit $p_{1,1}$ auf den Wert $p_{1,n}$ gestiegen ist, so läßt sich der dafür benötigte Zeitaufwand aus der Gleichung (2.5) durch Einsetzen von $p_{1,n}$ und Auflösen nach t ermitteln. Diese Fragestellung entspricht inhaltlich dem folgendem Problem: bei einer großen Anzahl N von Schülern mit identischen Ausgangswahrscheinlichkeiten ist zu bestimmen, wie lange es dauert, bis eine bestimmte Teilmenge dieser Schüler die Alternative A_1 wählt, d.h. lernt. Ist die Anzahl der Schüler in dieser Teilmenge in Höhe von N_n vorgegeben, so ist $N_n/N = p_{1,n}$ derjenige Anteil an der Gesamtheit der Schüler, der nach dem n-ten Versuch die Alternative A_1 gewählt hat, und die benötigte Zeit ergibt sich dabei aus Gleichung (2.5). Die Größe $p_{1,n}$ ist dabei der Erwartungswert, der in Über-

sichtsschema 1 in der n-ten Spalte aufgeführten Einzelgrößen. Er läßt sich dadurch ermitteln, daß die in der n-ten Spalte von Übersichtsschema 1 und 2 vorkommenden Einzelausdrücke paarweise multipliziert und die Produkte anschließend addiert werden. Wird dieser Erwartungswert vorgegeben, so resultiert daraus eine ganz bestimmte Anzahl von Perioden.

Geht man bei der Interpretation der Größe $p_{1,n}$ nicht von einer großen Anzahl von Schülern aus, die identische Ausgangswahrscheinlichkeiten $p_{1,1}$ besitzen, sondern von einem bestimmten individuellen Lernsubjekt mit der Ausgangswahrscheinlichkeit $p_{1,1}$, so kann $p_{1,n}$ nach wie vor als Erwartungswert interpretiert werden: $p_{1,n}$ ist dann der Erwartungswert derjenigen Wahrscheinlichkeit, mit der ein individuelles Lernsubjekt das Lernproblem im n-ten Lernversuch lernen wird. Die Wahrscheinlichkeit selbst, mit der dieses Lernsubjekt das Lernproblem im n-ten Lernversuch löst, ist dabei eine Zufallsvariable. Daraus ergibt sich, daß die zum Erreichen einer bestimmten Lernwahrscheinlichkeit erforderliche Zeitspanne ebenfalls eine Zufallsvariable ist. Stellt man nun wieder die Frage, wie lange es dauert, bis ein bestimmtes individuelles Lernsubjekt die Lernaufgabe mit einer bestimmten vorgegebenen Mindestwahrscheinlichkeit p_e von beispielsweise 0,9 löst, dann ist Gleichung (2.5) zur Beantwortung dieser Frage nicht mehr anwendbar, weil es sich nicht mehr darum handelt, die Zeitspanne zu ermitteln, die zum Erreichen eines bestimmten Erwartungswertes nötig ist. Die gesuchte Zeitspanne ist jetzt vielmehr eine Zufallsvariable, die einer bestimmten Wahrscheinlichkeitsverteilung unterliegt, und sie kann größer oder kleiner sein als die Zeitspanne, die zum Erreichen einer bestimmten Durchschnittswahrscheinlichkeit benötigt wird, selbst wenn diese Durchschnittswahrscheinlichkeit in der gleichen Höhe vorgegeben wird wie p_e.

Um diese Zeitspanne zu finden, die nötig ist, bis der Wert p_e gerade erreicht wird, kann man in Übersichtsschema 1 diejenigen Spalten aufsuchen, in denen mindestens eine der dort angegebenen Wahrscheinlichkeiten den Wert p_e gerade erreicht. Je weiter rechts die Spalten liegen, d.h. je größer die Lernzeit ist, desto häufiger werden in einer Spalte Wahrscheinlichkeiten auftreten, die dem Wert p_e gleich sind oder ihn übersteigen und umgekehrt. Da die Anzahl der Perioden, die aufgewandt werden müssen, bis ein individuelles Lernsubjekt den vorgegebenen Wert p_e erreicht, davon abhängt, auf welcher Route sich der individuelle Lernprozeß im Übersichtsschema 1 bewegt, kann die Frage gestellt werden, wie groß die Wahrscheinlichkeit für eine bestimmte Route bzw. für eine bestimmte Lernzeit ist. Um diese Frage zu beantworten, sucht man zunächst diejenige Route, die zu der längsten Lernzeit führt. Es ist dies für das dargestellte Modell der unterste Pfad in Übersichtsschema 1 und 2. Die Lernzeit, die aufzuwenden ist, wenn der individuelle Lernprozeß dieser Route folgt, läßt sich dadurch bestimmen, daß man denjenigen Wert in Schema 1 aufsucht, der auf dieser Route zum ersten mal den Wert p_e gerade erreicht. Die dazugehörige Nummer der Spalte, in der dieser Wert steht, entspricht der längsten Lernzeit zum Erreichen des Wertes p_e. Die Wahrscheinlichkeit dieser Lernzeit kann aus Übersichtsschema 2 abgelesen werden: sie ist derjenige Wert, der in der dem Schema 1 entsprechenden Spalte an unterster Stelle steht. Die Wahrscheinlichkeit jeder anderen Lernzeit läßt sich dann auf folgende Weise ermitteln. In Schema 1 wird zunächst die Position ermittelt, die einen Wert ergibt, der gerade p_e erreicht. Die Nummer der dazugehörigen Spalte gibt dann die Anzahl der Perioden bzw. die Lernzeit an. Die Wahrscheinlichkeit dieser Lernzeit läßt sich aus Schema 2 ermitteln, indem man sämtliche Routen, in die sich die gefundene Position in Schema 1 verzweigt, so lange weiterverfolgt, bis sie in derjenigen Spalte enden, der im Übersichtsschema 2 die bereits festgestellte längste Lernzeit entspricht. Jede dieser Routen endet in dieser Spalte in einem bestimmten Wert, der ihrer Wahrscheinlichkeit entspricht. Die Summe dieser Wahrscheinlichkeiten ist dann die gesuchte Wahrscheinlichkeit für die untersuchte Lernzeit. Allen übrigen Lernzeiten lassen sich auf diese Weise die dazugehörigen Wahrscheinlichkeiten ihrer Realisation ebenfalls zuordnen, wobei durch das beschriebene Verfahren garantiert ist, daß die Summe der Wahrscheinlichkeiten der einzelnen Lernzeiten 1 ergibt.

Diese Überlegungen gelten auch für Modelle, die nicht so einfach aufgebaut sind, wie das den Übersichtsschemata 1 und 2 zugrunde liegende Modell. Deshalb scheint es gerechtfertigt, folgende Feststellungen zu treffen:

(a) Die zum Erreichen einer vorgegebenen Mindesterfolgswahrscheinlichkeit p_e benötigte Zeit ist für jedes Lernsubjekt eine Zufallsvariable, für die es eine bestimmte Häufigkeitsverteilung gibt.

(b) Zu jedem Lernsubjekt gehört eine andere Häufigkeitsverteilung, deren Form davon abhängt, wie groß die individuelle Ausgangswahrscheinlichkeit $p_{1,1}$ ist.

Aus diesen Feststellungen resultiert eine erhebliche Schwierigkeit. Will man zum Beispiel drei Schüler, die unterschiedliche Häufigkeitsverteilungen der zum Erreichen eines bestimmten p_e-Wertes nötigen Lernzeit besitzen, so in zwei Klassen einteilen, daß die Lernzeit, die die erste Klasse benötigt, bis jeder Schüler den Wert p_e erreicht, zuzüglich der Lernzeit, die die zweite Klasse benötigt, bis jeder ihrer Schüler ebenfalls den Wert p_e erreicht, ein Minimum ist, so läßt sich daraus nicht die an sich naheliegende Regel ableiten, daß die beiden Schüler mit den größten Lernzeit-Mittelwerten zu einer Klasse zusammengefaßt werden müssen, und daß der Schüler mit dem kleinsten Lernzeit-Mittelwert die zweite Klasse bildet. Diese zunächst vermutete Regel würde nur dann zu dem gesuchten Minimum führen, wenn aus der bekannten Rangfolge der Mittelwerte der individuellen Häufigkeitsverteilungen darauf geschlossen werden könnte, daß die tatsächlich realisierten Lernzeiten der Schüler die gleiche Rangfolge ergeben würden wie die ihrer Mittelwerte. Dieser Schluß ist jedoch nicht möglich, ja es ist z.B. denkbar, daß derjenige Schüler mit dem größten Mittelwert die kleinste Zeitspanne benötigt, und daß der Schüler mit dem kleinsten Mittelwert die meiste Lernzeit aufwenden muß. Das Problem kann auch nicht dadurch gelöst werden, daß man die wahrscheinlichste Kombination der Lernzeiten ermittelt und die aus dieser Kombination resultierende Rangfolge der Lernzeiten für eine Einteilung der Schüler nach der vermuteten Regel zugrunde legt. Denn es handelt sich darum, den Erwartungswert der Summe der beiden Lernzeiten zu minimieren, weil diese Summe eine Zufallsvariable ist. In dem Falle, daß der Schüler mit dem kleinsten Lernzeit-Mittelwert allein die eine Klasse bildet und die beiden anderen Schüler mit höheren Mittelwerten die andere Klasse, besteht der zu minimierende Erwartungswert der Summe aus dem Mittelwert des Schülers, der allein in einer Klasse ist, zuzüglich dem Erwartungswert der aufzuwendenden Zeit, die benötigt wird, bis jeder Schüler in der anderen Klasse den Wert p_e erreicht. Die Größe des zweiten Erwartungswertes läßt sich nicht aus den Lernzeit-Mittelwerten dieser beiden Schüler allein berechnen, sondern erfordert die Kenntnis der beiden Lernzeitverteilungen. Gibt man diese Verteilungen für alle drei Schüler vor und berechnet für jede mögliche Klasseneinteilung den Erwartungswert der Zielfunktion, so kann sich dabei ergeben, daß die Klasseneinteilung, die die Zielfunktion am besten erfüllt, von derjenigen Klasseneinteilung abweicht, die sich unter Anwendung der vermuteten Regel ergibt, wenn man von der (nicht begründeten) Annahme ausgeht, daß die Schüler in jedem Fall genau jene Lernzeiten realisieren würden, die durch ihre Lernzeit-Mittelwerte gegeben sind.

Wollte man versuchen, für eine größere Anzahl von Schülern und für allgemeinere Zielfunktionen die optimale Klasseneinteilung nach dem geschilderten

Verfahren zu bestimmen, indem man jede Klasseneinteilung daraufhin untersucht, wie hoch der resultierende Erwartungswert der Zielfunktion ist, so würde sich daraus ein vermutlich untragbar hoher Rechenaufwand ergeben. Es ist deshalb interessant, zu überprüfen, ob sich einfachere Verfahren entwickeln lassen, sodaß vermieden werden kann, für jede denkbare Klasseneinteilung den dazugehörigen Erwartungswert der Zielfunktion numerisch zu ermitteln, um anschließend die beste Klasseneinteilung auswählen zu können. Die gewünschten Vereinfachungen sind zu erreichen, wenn man bezüglich der Häufigkeitsverteilungen bestimmte Annahmen setzt.
Eine Annahme, die den gewünschten Zweck erfüllt, besteht darin, daß man voraussetzt, daß sich die einzelnen Häufigkeitsverteilungen nicht überschneiden. Unter dieser Prämisse gibt es für einen bestimmten p_e-Wert nur eine einzige Rangfolge der Lernzeiten, obwohl dabei jeder individuelle Lernzeitwert eine Zufallsvariable bleibt. Die unter dieser Voraussetzung ableitbaren Optimierungsregeln unterscheiden sich nicht von denjenigen Regeln, die sich gewinnen lassen, wenn man annimmt, daß die Lernzeiten individuell verschiedene aber feste Größen sind und keinen stochastischen Verteilungen unterliegen. Obwohl in beiden Fällen die ableitbaren Regeln identisch sind, ist es formal einfacher, die Regeln unter der Annahme fester Lernzeiten abzuleiten. Dies ist einer der Grün-

de dafür, daß in den Kapitel 3 bis 7 bei der Ableitung der Optimierungsregeln von individuell verschiedenen aber festen Lernzeiten ausgegangen wird.

Außer diesen pragmatischen Gründen sind vermutlich folgende Überlegungen geeignet, die Annahme fester Lernzeiten zu rechtfertigen. Das Problem der optimalen Klasseneinteilung wird praktisch wohl erst dann relevant, wenn die Schüler längere Zeit, z.B. während eines ganzen Schuljahres, in der ihnen einmal zugewiesenen Klasse verbleiben. Die Lernprozesse, die die Schüler in diesen Fällen durchlaufen, lassen sich gedanklich in eine sehr große Anzahl von Lernexperimenten zerlegen. Man wird dabei davon ausgehen können, daß die Streuung der individuellen Lernwahrscheinlichkeiten währen der ersten Lernversuche noch sehr groß ist und sich mit einer steigenden Anzahl von Lernversuchen verringert, was dazu führt, daß die Wahrscheinlichkeit abnimmt, daß ein Schüler mit dem kleineren Lernzeit-Mittelwert am Ende eine Lernzeit realisiert, die größer ist als die realisierte Lernzeit eines Schülers mit größerem Lernzeit-Mittelwert. Ist die Anzahl der Lernversuche sehr groß, so werden die stochastischen Einflüsse zurücktreten gegenüber jenen Faktoren, die die Lernfähigkeit der Schüler kausal bestimmen und auf denen die Unterschiede zwischen den Lernfähigkeiten und damit den Lernzeiten der Schüler nicht zuletzt beruhen.

Die Frage, ob nicht auch bei einer stochastischen Behandlung der Optimierungsprobleme wenn nicht alle, so doch aber die wesentlichen Sätze abgeleitet werden können, die sich im Falle eines deterministischen Ansatzes ergeben, kann in dieser Arbeit nicht überprüft werden, obwohl die Vermutung sehr nahe liegt, daß dies der Fall sein würde. Es war nicht zuletzt diese Vermutung, die es lohnend erscheinen ließ, das Problem zunächst mit dem deterministischen Ansatz zu durchleuchten.

(5) Im folgenden wird davon ausgegangen, daß es eine Anzahl s von Schülern $S_1, S_2, \ldots, S_s$ gibt, die in eine Anzahl k von Klassen eingeteilt werden soll. Entsprechend dem deterministischen Ansatz wird jedem Schüler S_j, $j \in \{1,2,\ldots,s\}$, eine positive Zahl t(j) zugeordnet, diedie zum Erreichen der Mindest-Erfolgswahrscheinlichkeit p_e benötigte Zeitspanne angibt. Die Schüler sollen so numeriert werden, daß

$$t(j) > t(j') \text{ für } j > j' \qquad j, j' \in \{1,2,\ldots,s\} .$$

Dabei ist impliziert, daß die Lernzeiten aller Schüler unterschiedlich sind.

Zur Definition einer Klasse K wird folgende Symbolik eingeführt. Besteht z.B. die i-te Klasse K_i aus den Schülern $S_u, S_v, S_w, \ldots$, dann ist unter K_i die Menge

$$K_i = \{u, v, w, \ldots\}$$

zu verstehen, deren Elemente aus den Nummern derjenigen Schüler bestehen, die in der Klasse K_1 zusammengefaßt sind.

(6) Wird aus der Menge $S = \{1,2,...,s\}$ aller Schüler eine beliebige Teilmenge $K \subset S$, $K \neq \emptyset$, ausgewählt, und wird diese Teilmenge von Schülern gemeinsam in einer Klasse unterrichtet, dann soll angenommen werden, daß die zum Erreichen des Wertes p_e aufzuwendende Lernzeit jedes Schülers aus K nicht davon abhängt, (a) mit wievielen Schülern er sich gemeinsam in der Klasse befindet und (b) mit welchen Schülern er die Klasse gemeinsam besucht. Diese Prämisse ist bezüglich der Annahme (a) nur bei solchen Unterrichtsformen annähernd erfüllt, in denen es nicht darauf ankommt, daß während einer bestimmten Unterrichtszeit eine möglichst große Zahl von Schülern Gelegenheit erhält, mit dem Lehrer einen Dialog zu führen (Großvorlesungen, Vorträge etc.). Die Annahme (a) wird deshalb auch an entsprechender Stelle modifiziert. Auch die Annahme (b) ist in vielen Fällen nicht erfüllt. Psychologen und Soziologen werden immer Beispiele dafür finden können, die beweisen, daß die Lernfähigkeit des einzelnen Schülers in starkem Maße davon abhängt, mit welchen anderen Schülerpersönlichkeiten er in einer Klasse gemeinsam unterrichtet wird. Bei einer Anzahl s von Schülern kann man jeden Schüler auf $2^{s-1} - 1$ Möglichkeiten mit an-

deren Schülern in einer Klasse zusammenfassen. Für jeden Schüler hätten also $2^{s-1} - 1$ verschiedene Lernzeiten unterstellt werden müssen, um diese Effekte zu berücksichtigen. Abgesehen von dem allzu großen formalen Aufwand, der sich hieraus ergeben hätte, schien es schon deshalb angebracht, auf diese Verfeinerung zu verzichten, weil über diese Tatbestände noch zu wenig brauchbare Unterlagen existieren: "We have practically no useful information about how this isolated learning exercise differs from learning the same assigment in group situations, which is not surprising because psychologists are in rather general agreement that they would prefer to teach or observe one subjekt at a time in a learning situation".[1)]

2.2 Zielfunktionen und Organisationsprinzipien

Unter dem Begriff Organisationsprinzip soll im folgenden die Gesamtheit der Regeln verstanden werden, die etwas darüber aussagen,

(a) in wieviele Klassen die Grundgesamtheit der Schüler eingeteilt werden soll,

(b) wie groß jede dieser Klassen sein soll und

(c) mit welchen Schülern aus der Grundgesamtheit die

1) Bugelski, R.R., "The Psychology of Learning Applied to Teaching", op. cit., S. 23-24

Klassen besetzt werden müssen, wenn es gilt, ein bestimmtes vorgegebenes Ziel bestmöglich zu erfüllen. Das Ziel kann z.B. darin bestehen, die Zeit, die die Lehrer zur Unterrichtung der Schüler aufwenden müssen, zu minimieren. Außer diesem Ziel werden folgende Zielfunktionen untersucht und die aus ihnen resultierenden Organisationsprinzipien abgeleitet: Minimierung der von den Schülern aufzuwendenden Zeit, Minimierung der Ausbildungskosten, Maximierung der Ausbildungserträge, Maximierung der Nettoerträge der Ausbildung.

Bei allen diesen Zielfunktionen wird angenommen, daß folgendes übergeordnetes Ziel, das man als eine Gerechtigkeitsmaxime interpretieren könnte, erfüllt sein muß: Die Wahrscheinlichkeit, mit der ein Schüler nach Beendigung der Ausbildungszeit dazu in der Lage ist, das gestellte Problem zu lösen, darf eine für alle Schüler geltende gleich große Mindestwahrscheinlichkeit p_e von beispielsweise 0,9 nicht unterschreiten.
Diejenige Art der Zusammenfassung von Schülern zu Klassen, die eine vorgegebene Zielfunktion am besten erfüllt, wird optimale Klassenorganisation (optimales Klassensystem) genannt. In den folgenden Kapiteln gilt es, die Regeln abzuleiten, nach denen aus der Grundgesamtheit der Schüler das optimale Klassensystem gebildet werden kann. Dabei wird sich zeigen, daß ein

Klassensystem, das eine bestimmte Zielfunktion optimiert, nicht notwendig auch in bezug auf eine andere Zielfunktion optimal zu sein braucht.

Kapitel 3: Minimierung der Lehrerzeit

3.1 Vorgegebene relative Klassengrößen, vorgegebene Anzahl von Klassen

In diesem Abschnitt wird von der Frage ausgegangen, auf welche Weise eine Menge von Schülern in eine vorgegebene Anzahl k von Klassen eingeteilt werden soll, wenn (a) die relative Größe jeder Klasse gegeben ist, (b) jeder Schüler am Ende der Unterrichtszeit, die er in immer derselben Klasse verbringt, mindestens mit der Wahrscheinlichkeit p_e das Lernproblem lösen soll und (c) die Summe der Zeit, die alle Lehrer zur Unterrichtung der Schüler aufwenden müssen, ein Minimum sein soll. Dabei wird angenommen, daß jeder Klasse ein Lehrer zur Verfügung steht.

Die Lösung dieses Problems ist evident: Die Klasse mit der größten Anzahl von Schülern muß sich aus den Schülern zusammensetzen, deren Lernzeiten in der geordneten Menge $\{t(1), t(2),\dots,t(s)\}$ die größten Elemente bilden. Die Klasse mit der zweitgrößten Anzahl von Schülern muß alle diejenigen Schüler enthalten, deren Lernzeiten eine lückenlos geordnete Menge[1] bilden, wobei sich diese Werte an die Lernzeiten der größten Klasse anschließen. Entsprechend muß die drittgrößte Klasse gebildet werden usf. bis zur k-ten Klasse. Obwohl sich die Lösung dieses Prob-

1) Die Menge enthält alle Elemente, die zwischen ihrem kleinsten und größten Element vorkommen.

lems intuitiv angeben läßt, soll die Lösung bewiesen werden, weil bei der Behandlung der komplizierteren Probleme, die sich nicht intuitiv lösen lassen, dieser Beweis vorausgesetzt werden muß.

Das Problem läßt sich wie folgt formulieren. Aus den Elementen der Menge $S = \{1,2,\ldots,s\}$ sollen k Teilmengen $K_1, K_2, \ldots, K_k$ so gebildet werden, daß

(3.1) $\bigcup_{i=1}^{i=k} K_i = S$,

(3.2) $K_j \cap K_i = \emptyset$ für $j \neq i$.

Bezeichnet man mit a_i die Anzahl der Schüler, aus denen K_i gebildet wird, dann soll gelten

(3.3) $\sum_{i=1}^{i=k} a_i = s$, wobei

(3.4) $a_i \in \mathbb{N}$ für $i = 1,2,\ldots,k$,

(3.5) $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_{k-1} \leqslant a_k$.

Die Werte a_i/s entsprechen dabei den relativen Klassengrößen. Ferner soll die Zielfunktion, die die von den Lehrern aufzuwendende Zeit angibt, ein Minimum annehmen:

(3.6) $z = \sum_{i=1}^{i=k} t_i$,

wobei die Gleichung (3.6) folgende Definition enthält

(3.7) $t_i := \max_{j \in K_i} t(j)$, $i = 1,2,\ldots,k.$

Durch (3.7) wird gewährleistet, daß der Unterricht in jeder Klasse so lange fortgesetzt wird, bis derjenige Schüler, dem in jeder Klasse der größte t-Wert entspricht, die vorgegebene Mindest-Erfolgswahrscheinlichkeit p_e erreicht hat.

Die Zielfunktion (3.6) wird unter Einhaltung der gesetzten Nebenbedingungen durch folgende Klassenorganisation minimiert:

$$\begin{aligned} &K_1 = \{1,2,\ldots,a_1\} \\ (3.8)\qquad &K_2 = \{a_1+1, a_1+2,\ldots,a_1+a_2\} \\ &\vdots \\ &K_k = \{a_1+a_2+\ldots+a_{k-1}+1,\quad a_1+a_2+\ldots+a_{k-1}+2,\ldots \\ &\qquad \ldots, a_1+a_2+\ldots+a_{k-1}+a_k\} \end{aligned}$$

Zum Beweis dieses Ergebnisses werden die folgenden beiden Sätze formuliert und bewiesen.

Satz 1

Sei $K = K_1 \cup K_2 \cup \ldots \cup K_k$ ein Klassensystem, das die Bedingungen (3.1) bis (3.4) erfüllt, und dessen Klassen K_i nicht alle lückenlos geordnete Teilmengen der Menge

$$S = \{1,2,\ldots,s\}$$

sind. Sei ferner $f(x)$ eine positive strikt monoton steigende und $g(x)$ eine positive strikt monoton sinkende Funktion im Wertbereich $x > 0$. Dann gibt es stets ein

System $K^o = K_1^o \cup K_2^o \cup \ldots \cup K_k^o$, dessen Klassen K_i^o alle lückenlos geordnete Teilmengen der Menge S sind, und das die Bedingungen (3.1) bis (3.4) ebenfalls erfüllt. Dabei gilt immer

$$(3.9) \qquad \sum_{i=1}^{i=k} f(t_i^o) \quad < \quad \sum_{i=1}^{i=k} f(t_i) \ ,$$

$$(3.10) \qquad \sum_{i=1}^{i=k} g(t_i^o) \quad > \quad \sum_{i=1}^{i=k} g(t_i) \ ,$$

wobei $t_i^o = \max\limits_{j \in K_i^o} t(j)$, $i = 1,2,\ldots,k$,

$t_i = \max\limits_{j \in K_i} t(j)$, $i = 1,2,\ldots,k$

<u>Beweis des Satzes 1</u>

Im Klassensystem $K = K_1 \cup K_2 \cup \ldots \cup K_k$, das die Bedingungen (3.1) - (3.4) erfüllt, sei die Klasse K_x, $x \in \{1,2,\ldots,k\}$, eine nicht lückenlos geordnete Teilmenge der Menge S. Die i-te Klasse K_i, $i = 1,2,\ldots,k$, enthalte die Anzahl a_i von Elementen der Menge S. Dann enthält die Klasse K_x die Anzahl a_x von diesen Elementen.

Sei u das kleinste und v das größte Element von allen Elementen in K_x. Da K_x voraussetzungsgemäß eine nicht lückenlos geordnete Teilmenge der Menge S ist, gibt es in der Menge

$$(3.11) \qquad Q_x = \{u, u+1, \ldots, v\}$$

mindestens ein Element w mit

$$w \in K_x \quad .$$

Es sei w_1 das größte Element dieser Art und es sei K_z diejenige Klasse, die w_1 enthält. Ferner seien w_2, w_3,... die anderen etwa noch in K_z enthaltenen Elemente aus Q_x. Die Numerierung dieser Elemente sei so vorgenommen, daß

$$w_1 > w_2 > w_3 > \dots$$

gilt.

Es wird nun folgende Fallunterscheidung getroffen.

Es gilt für w_1 entweder

$$(3.12) \qquad w_1 = \max_{j \in K_z} j$$

oder

$$(3.13) \qquad w_1 \neq \max_{j \in K_z} j$$

(a) Angenommen es gilt (3.12). Dann läßt sich der Ausdruck

$$(3.14) \qquad \sum_{i=1}^{i=k} f(t_i)$$

$$t_i = \max_{j \in K_i} t(j) \quad , \qquad i = 1,2,\dots,k$$

verringern und der Ausdruck

$$(3.15) \qquad \sum_{i=1}^{i=k} g(t_i)$$

$$t_i = \max_{j \in K_i} t(j) \quad , \qquad i = 1,2,\dots,k$$

erhöhen, indem man folgenden Elemententausch durchführt:

(3.16) $\quad K'_x = (K_x \cup \{w_1\}) - \{u\}$

(3.17) $\quad K'_x = (K_z - \{w_1\}) \cup \{u\}$

(3.18) $\quad K'_i = K_i \,, \quad i = 1,2,\ldots,k; \quad i \neq x; \quad i \neq z.$

Da aufgrund dieses Elemententausches

$$\max_{j \in K'_z} j < \max_{j \in K_z} j$$

und

$$\max_{j \in K'_i} j = \max_{j \in K_i} j \,, \qquad \begin{matrix} i = 1,2,\ldots,k \,, \\ i \neq z \end{matrix}$$

ist der Ausdruck (3.14) gesunken und der Ausdruck (3.15) gestiegen.

(b) Angenommen es gilt (3.13). Es sei v' das größte Element von K_z. Da w_1 das größte nicht zu K_x gehörende Element in Q_x ist, muß v' größer sein als v. Es wird nun folgender Elemententausch durchgeführt:

(3.19) $\quad K'_x = (K_x \cup \{w_1\}) - \{v\}$

(3.20) $\quad K'_z = (K_z \cup \{v\}) - \{w_1\}$

(3.21) $\quad K'_i = K_i \,; \; i = 1,2,\ldots,k \,; \quad i \neq x \,; \quad i \neq z \,.$

Es ergibt sich:

(3.22) $$\max_{j \in K'_x} j < \max_{j \in K_x} j$$

$$(3.23) \qquad \max_{j \in K'_i} j = \max_{j \in K_i} j \quad , \qquad \begin{array}{l} i = 1,2,\ldots,k \,. \\ i \neq x \end{array}$$

(3.23) gilt wegen $v' > v > w_1$ auch für $i = z$.
Aufgrund von (3.22) und (3.23) ist wieder der Ausdruck (3.14) gesunken und der Ausdruck (3.15) gestiegen.

(c) Für jedes weitere Element w_2, w_3, ... läßt sich ebenfalls ein Elemententausch entweder nach Schema (a) oder (b) durchführen, wodurch jedes mal der Ausdruck (3.14) vermindert und der Ausdruck (3.15) erhöht wird, bis K_x eine lückenlos geordnete Menge geworden ist. Dasselbe gilt in bezug auf jede andere nicht lückenlos geordnete Menge.

<u>Satz 2</u>

Sei $K = K_1 \cup K_2 \cup \ldots \cup K_k$ ein Klassensystem, für das die Bedingungen (3.1) - (3.4) erfüllt sind, und für das alle Klassen K_i, $i = 1,2,\ldots,k$, lückenlos geordnete Teilmenge der Menge S sind.
Dann gilt folgender Satz:
Der Ausdruck

$$(3.14) \qquad \sum_{i=1}^{i=k} f(t_i)$$

$$t_i = \max_{j \in K_i} t(j) \quad , \qquad i = 1,2,\ldots,k$$

läßt sich unter Einhaltung der Bedingungen (3.1) - (3.4) verringern und der Ausdruck

$$(3.15)\qquad \sum_{i=1}^{i=k} g(t_i)$$

$$t_i = \max_{j \in K_i} t(j) \quad , \qquad i = 1,2,\ldots,k$$

unter Einhaltung der Bedingungen (3.1) - (3.4) vergrößern, so lange es in K zwei Klassen K_x und K_y , $x,y \in \{1,2,\ldots,k\}$, $x \neq y$, gibt, für die gilt:

$$(3.24)\qquad K_y = \{ u,u+1,u+2,\ldots,u+a_y-1 \}$$

$$(3.25)\qquad K_x = \{ u+a_y,u+a_y+1,u+a_y+2,\ldots,u+a_y+a_x-1 \}$$

$$(3.26)\qquad a_y > a_x \quad .$$

<u>Beweis des Satzes 2</u>

Vertauscht man die Elemente der beiden Klassen K_x und K_y so, daß aus K_x und K_y vor dem Tausch die beiden Klassen K'_x und K'_y nach dem Tausch entstehen:

$$(3.27)\qquad K'_x = \{ u,u+1,u+2,\ldots,u+a_x-1 \}$$

$$(3.28)\qquad K'_y = \{ u+a_x,u+a_x+1,u+a_x+2,\ldots,u+a_x+a_y-1 \} \quad ,$$

dann ist

$$(3.29)\qquad \max_{j \in K_x} t(j) = \max_{j \in K'_y} t(j)$$

$$(3.30)\qquad \max_{j \in K_y} t(j) > \max_{j \in K'_x} t(j) \quad ,$$

während alle übrigen t_i unverändert geblieben sind. Folglich ist durch diesen Elemententausch der Ausdruck (3.14) gesunken und der Ausdruck (3.15) gestiegen.

Die Behauptung, daß (3.8) das für die Minimierung der Lehrerzeit optimale Klassensystem darstellt, kann nun mit Hilfe der Sätze 1 und 2 bewiesen werden.
Die zu minimierende Zielfunktion lautet

$$(3.31) \qquad z = \sum_{i=1}^{i=k} t_i$$

$$t_i = \max_{j \in K_i} t(j) \quad , \quad i = 1,2,\dots,k \; .$$

Die Gleichung (3.31) ist ein Spezialfall der Funktion (3.14). Folglich kann Satz 1 direkt angewandt werden. Diese Anwendung ergibt, daß sich (3.31) so lange verringern läßt, bis alle Klassen K_i lückenlos geordnete Teilmengen der Menge S sind.
Bildet man entsprechend Satz 1 ein Klassensystem K mit lauter lückenlos geordneten Teilmengen von S, dann läßt sich die Zielfunktion aufgrund von Satz 2 weiter verringern, falls es in K zwei Klassen K_y und K_x gibt, für die die in (3.24) und (3.25) formulierten Eigenschaften gelten.

Unter der Nebenbedingung (3.5) gibt es nur ein einziges Klassensystem, für das in bezug auf jedes benachbarte Paar von Klassen K_x und K_y die in (3.24) und (3.25) formulierten Eigenschaften nicht gelten. Es ist dies das in (3.8) formulierte System.

Besitzt eine Gruppe von Schülern die gleiche Lernzeit, so kann es sein, daß die Anwendung der Sätze 1 und 2 nicht zu einem einzigen optimalen Klassensystem führt. Dieser Fall tritt dann ein, wenn nach Anwendung der Sätze 1 und 2 ein Teil der Schülergruppe mit gleichen Lernzeiten einer Klasse K_x und der restliche Teil einer Klasse K_{x+1} zugeordnet werden muß. Tauscht man in diesem Fall je zwei Schüler mit gleichen Lernzeiten zwischen den Klassen K_x und K_{x+1} aus, so bleibt der Wert der Zielfunktion unverändert. Dieser Wert bleibt auch dann unverändert, wenn alle Schüler dieser Gruppe in die Klasse K_x aufgenommen werden.

Muß indessen nach Anwendung der Sätze 1 und 2 die Schülergruppe mit gleichen Lernzeiten einer bestimmten Klasse ganz zugeordnet werden, weil sich nur so die Nebenbedingungen (3.4) und (3.5) einhalten lassen, dann gibt es nur ein einziges optimales Klassensystem.

3.2 Variable Anzahl von Klassen und variable relative Klassengrößen

Betrachtet man zunächst den Fall, daß die Anzahl der Klassen gegeben ist, während die relative Größe jeder Klasse variabel bleibt, so ergibt sich unmittelbar, daß dasjenige Klassensystem den Wert der Zielfunktion minimiert, für das die ersten k-1 Klassen aus je einem Schüler der k-1 Schüler mit den kleinsten Lernzeiten bestehen, während die k-te Klasse alle übrigen Schüler enthält.

Für den Fall, daß nicht nur die relative Größe der Klassen sondern auch die Anzahl k variabel ist, besitzt die Zielfunktion ihr absolutes Minimum, wenn alle Schüler in einer einzigen Klasse zusammengefaßt werden. Dieses Resultat ist indessen unrealistisch, weil es voraussetzt, daß die Lernzeit der Schüler unabhängig von der Klassengröße ist. Es soll deshalb auch der Fall untersucht werden, daß die Lernzeiten der Schüler eine Funktion der Klassengröße sind.

3.3 Berücksichtigung des Einflusses der Klassengröße auf die Lernzeit

(a) Hypothesen

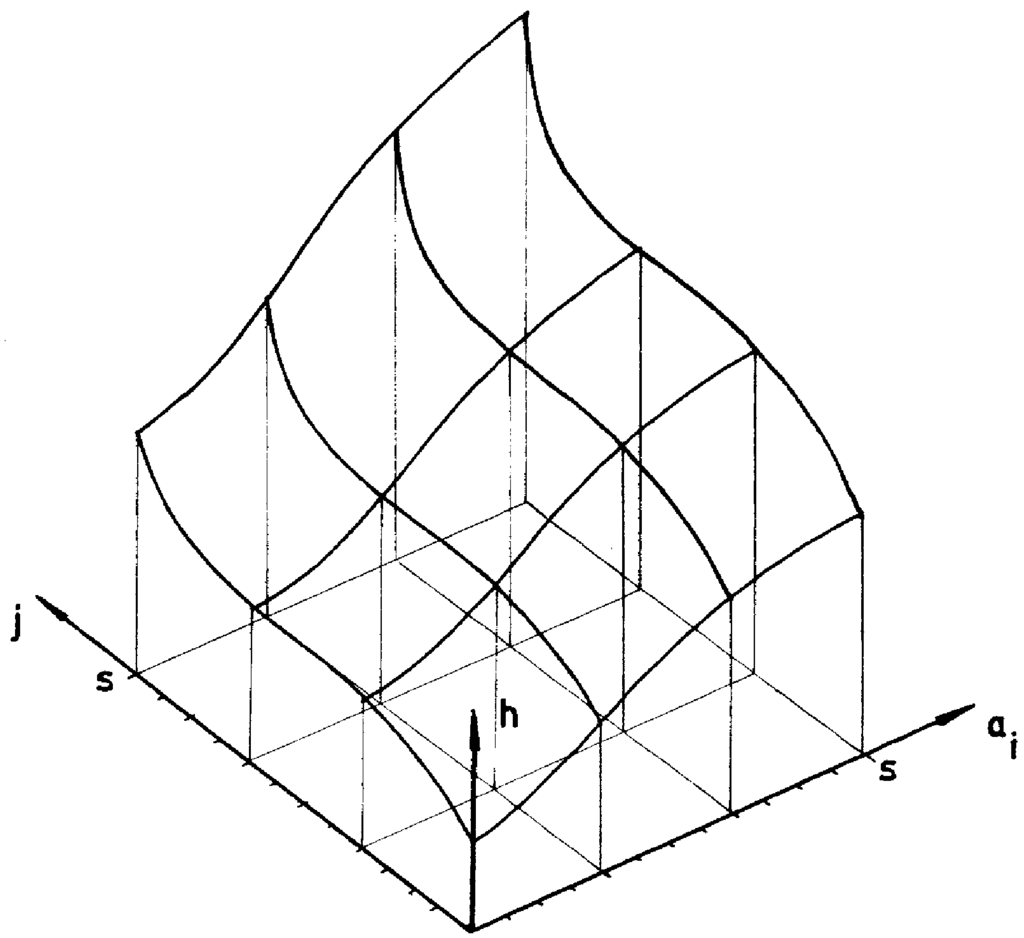
Da es bisher kaum experimentell gesicherte empirische Hypothesen über die Abhängigkeit der Lernzeit von der Klassengröße gibt, soll im folgenden von einer allgemeinen Funktion h

$$(3.32) \qquad h(j,a_i) \quad , \quad j \in \{1,2,\ldots,s\} \; , \quad a_i \in \mathbb{N} \; , \quad 1 \leqq a_i \leqq s \text{ für } i = 1,2,\ldots,k$$

ausgegangen werden, die für ein bestimmtes j und a_i diejenige Zeit angibt, die derjenige Schüler j der Klasse i mit a_i Schülern benötigt, der, falls er sich allein in der Klasse befände, eine Lernzeit von $t(j)$ benötigen würde. Es soll angenommen werden, daß die Rangfolge der Lernzeiten der Schüler unverändert bleibt, wenn sich die Größe der Klasse verändert, d.h. der langsamste Schüler bleibt stets der langsamste, ganz gleich, wie groß die Klasse ist. Es genügt für die folgenden Ableitungen anzunehmen, daß h eine strikt monoton steigende Funktion von a_i ist. Es gibt eine Reihe von Gründen, die die Annahme eines S-förmigen Verlaufs für diese Funktion nahelegen. Dies würde bedeuten, daß die Lernzeit eines Schülers zunächst überproportional und dann unterproportional mit der Schülerzahl steigt (vgl. Fig. 3). Ebenso plausibel erscheint indessen der Verlauf, für den der zusätzliche Zeitaufwand bei Vergrößerung der Klassengröße um einen Schüler von Anfang an abnimmt und schließlich verschwindend klein wird. Im folgenden soll keine spezielle Annahme getroffen werden, außer das h eine in j und a_i strikt monoton steigende Funktion ist. Zur Vereinfachung der Schreibweise wird als T_i diejenige Lernzeit definiert, die der langsamste Schüler von allen Schülern in Klasse K_i benötigt, wenn die Klasse K_i die Anzahl a_i von Schülern enthält:

$$\left.\begin{aligned} T_i &:= h(j^o, a_i) \\ j^o &= \max_{j \in K_i} j \end{aligned}\right\} \quad i = 1,2,\ldots,k \tag{3.33}$$

Fig. 3



(b) Vorgegebene Klassenzahl, vorgegebene relative Klassengrößen.

In diesem Fall besteht das Optimierungsproblem bei Minimierung der Lehrerzeit darin, ein Klassensystem zu finden, das die Funktion

$$(3.34) \qquad z = \sum_{i=1}^{i=k} h(j^o, a_i) = \sum_{i=1}^{i=k} T_i$$

unter den Nebenbedingungen (3.1) - (3.5) minimiert, wobei z die von allen Lehrern aufgewandte Zeit darstellt. Da die Größen a_i in diesem Fall vorgegeben sind, kann die Funktion h als ein Spezialfall derjenigen Funktion $f(t_i)$ in (3.14) aufgefaßt werden, für die Satz 1 bewiesen wurde. Wendet man Satz 1 auf dieses Optimierungsproblem an, so ergibt sich, daß das Klassensystem, das (3.34) minimiert, aus lückenlos geordneten Klassen bestehen muß. Satz 2 kann indessen nicht zur weiteren Bestimmung des optimalen Klassensystems verwendet werden, weil im Beweis für diesen Satz ein Elemententausch durchgeführt wurde, der nur das Argument der Funktion $f(t_i)$ in (3.14) veränderte, die Funktion selbst aber unverändert ließ. In dem vorliegenden Fall ändert sich jedoch durch diesen Elemententausch nicht nur das Argument j^o, sondern auch die als Parameter interpretierbare Größe a_i der Funktion $h(j^o, a_i)$. Es muß also aus der Menge aller möglichen Klassensysteme, deren Klassen lückenlos geordnete Mengen sind, das optimale Klassensystem auf andere Weise bestimmt werden. Bei einer vorgegebenen Anzahl k von Klassen und vorgegebenen relativen Klassengrößen gibt es eine Anzahl k! von Klassensystemen, deren Klassen lückenlos geordnete Mengen sind und die darüberhinaus die Bedingungen (3.1) - (3.5) nicht verletzen. Da die Anzahl dieser Systeme mit steigendem k stark zunimmt, ist eine direkte Auswahl des optimalen Systems durch Berechnung des Wertes der Zielfunktion für jedes der k! Systeme mit einem erheblichen Rechenaufwand verbunden. Dieser Rechenaufwand läßt sich bis zu einem gewissen Grade verringern, indem man das Problem als Zuordnungsproblem behandelt und mit Hilfe des ganzzahligen linearen Programmierens löst.

(c) Vorgegebene Klassenzahl, variable relative Klassengrößen.

Nimmt man an, daß die Funktion h, die bisher nur für ganzzahlige Werte j und a_i definiert wurde, durch eine für alle Punkte (j,a_i), $0 \leq j \leq s$, $1 \leq a_i \leq s$, definierte, stetige und differenzierbare Approximationsfunktion $h^*(j,a_i)$ ersetzt wird, so kann das Problem mit Hilfe eines Lagrange-Ansatzes approximativ gelöst werden: Die Funktion z

$$\begin{aligned} z = {} & h^*\left[x_1,x_1\right] + h^*\left[(x_1+x_2),x_2\right] + \ldots \\ & \ldots + h^*\left[(x_1+x_2+\ldots+x_k),x_k\right] - \\ & - \mu\left[\sum_{i=1}^{i=k} x_i - s\right] \end{aligned} \tag{3.35}$$

ist in bezug auf die Variablen x_i unter der Nebenbedingung zu minimieren, daß die Summe aller x_i die vorgegebene Anzahl s von Schülern ergibt. Dabei ist x_i die Anzahl der Schüler in Klasse i. Sind die Lösungen von (3.35), die durch Nullsetzen der partiellen Ableitungen und unter Beachtung der hinreichenden Bedingungen für ein Minimum gefunden wurden, nicht ganzzahlig, so ist es vermutlich am sinnvollsten, jedes nicht ganzzahlige x_i durch diejenige ganze Zahl zu ersetzen, zu der x_i den kleinsten Abstand besitzt. Sind alle nicht ganzzahligen x_i auf diese Weise durch ganzzahlige Werte x_i' ersetzt worden, so ist das optimale approximierte Klassensystem

$$
\begin{aligned}
K_1 &= \left\{ 1,2,\ldots,x_1' \right\} \\
K_2 &= \left\{ x_1'+1, x_1'+2,\ldots,x_1'+x_2' \right\} \\
&\vdots \\
K_k &= \left\{ x_1'+x_2'+\ldots+x_{k-1}'+1, x_1'+x_2'+\ldots+x_{k-1}'+2,\ldots \right. \\
&\qquad \left. \ldots, x_1'+x_2'+\ldots+x_{k-1}'+x_k' \right\}
\end{aligned}
\tag{3.36}
$$

(d) Variable Anzahl von Klassen.

Unter der Annahme, daß die Lernzeiten von der Klassengröße unabhängig sind, ist mit jeder Erhöung der Klassenzahl eine Vergrößerung der Lehrerzeit verbunden. Zu einem anderen Resultat kann man gelangen, wenn angenommen wird, daß die Lernzeiten eine strikt monoton steigende Funktion der Klassengröße sind. Ob eine Erhöhung der Klassenzahl den Wert der Zielfunktion vergrößert oder verkleinert, hängt von der Funktion h ab und außerdem davon, auf welche Weise die Anzahl der Klassen um eine weitere lückenlos geordnete Klasse bei Konstanz der gesamten Schülerzahl s erhöht wird. Die Erhöhung der Klassenzahl um eine Einheit kann dadurch erfolgen, daß eine bestimmte Klasse beispielsweise in zwei gleich große lückenlos geordnete Teilklassen zerlegt wird. Durch diese Zerlegung tritt in der Zielfunktion ein neuer Summand auf. Die daraus resultierende zusätzliche Lehrerzeit kann jedoch durch eine Abnahme der Lehrerzeit in der Restklasse, die nach Zweiteilung der ursprünglichen Klasse übrig geblieben ist, überkompensiert werden, weil die Lehrerzeit in dieser Restklasse infolge der nur noch halb so großen Schülerzahl gesunken ist. Die Frage, ob diese Überkompensation tatsächlich eintritt oder nicht, läßt sich ohne spezielle Annahmen über die Funktion h nicht beantworten.

3.4 Zusammenfassung

Unter der Annahme, daß die Lernzeit der Schüler nicht von der Klassengröße abhängt, läßt sich das Minimum der Lehrerzeit bei vorgegebenen relativen Klassengrößen und vorgegebener Klassenzahl allein unter Anwendung der Sätze 1 und 2 bestimmen. Das optimale Klassensystem besteht aus Klassen, die alle lückenlos geordnete Teilmengen der Menge $S = \{1, 2, .., s\}$

sind. Dabei muß die Rangfolge der relativen Klassengrößen der Rangfolge entsprechen, die von den Lernzeiten derjenigen Schüler gebildet wird, die in jeder Klasse die meiste Lernzeit benötigen. Diejenige Klasse, die den Schüler s enthält, muß die größte und diejenige Klasse, die den Schüler 1 enthält, die kleinste relative Klassengröße besitzen. Bei vorgegebener Klassenzahl und variablen relativen Klassengrößen erreicht die Zielfunktion ihr Minimum, wenn die k-1 schnellsten Schüler in je einer Klasse und die restlichen Schüler in der k-ten Klasse zusammengefaßt werden. Das absolute Minimum der Zielfunktion ist dann erreicht, wenn sich alle Schüler in einer Klasse befinden. Da in diesem Fall nicht mehr damit gerechnet werden kann, daß die Lernzeiten der Schüler unabhängig von der Klassengröße sind, wurde diese restriktive Annahme abgewandelt.

Die Einführung einer in der Variablen "Klassengröße" strikt monoton steigenden Lernzeitfunktion führte zu neuen Optimierungsproblemen. Da Satz 1 auch unter dieser neuen Annahme anwendbar blieb, ergab sich sowohl für die Fälle variabler wie konstanter relativer Klassengröße und Klassenzahl als Voraussetzung des optimalen Systems stets eine Klasseneinteilung, bei der alle Klassen lückenlos geordnete Teilmengen der Menge $S = \{1,2,\ldots,s\}$ waren. Da Satz 2 nicht mehr angewandt werden konnte, mußten zur weiteren Bestimmung des optimalen Klassensystems zusätzliche Verfahren gesucht werden.

Für den Fall vorgegebener relativer Klassengrößen und vorgegebener Klassenzahl gab es k! Klassensysteme, die die Bedingung lückenlos geordneter Klassen erfüllten. Eine Möglichkeit, aus diesen k! Systemen das optimale auszuwählen, besteht darin, daß man die Aufgabe als Zuordnungsproblem auffaßt und mit den Mitteln der ganzzahligen linearen Programmierung löst.

Bei variablen relativen Klassengrößen und vorgegebener Klassenzahl bietet sich als Lösungsverfahren die Formulierung der Zielfunktion als Lagrange-Ansatz an, doch muß hierfür die diskrete Funktion h durch eine geeignete Approximationsfunktion ersetzt werden.

Bei variabler Anzahl von Klassen ergab sich eine wichtige Abweichung zu dem entsprechenden Fall, bei dem die Lernzeit nicht als Funktion der Klassengröße behandelt wurde: eine Erhöhung der Klassenzahl hatte nun nicht notwendig ein Ansteigen der Lehrerzeit zur Folge, sondern konnte im Gegenteil auch dazu führen, daß die Lehrerzeit abnahm. Wann das eine und das andere der Fall ist, hängt von den Annahmen über den Einfluß der Klassengröße auf die Lernzeit, d.h. von den Eigenschaften der Funktion h ab.

Kapitel 4: Minimierung der Schülerzeit

4.1 Vorgegebene Anzahl von Klassen und vorgegebene relative Klassengrößen

Auch bei der Zielfunktion Minimierung der Schülerzeit soll gewährleistet werden, daß kein Schüler vom Ausbildungsprozeß ausgeschlossen wird. Außerdem sollen die Nebenbedingungen (3.1) - (3.5) bei der Klassenbildung erfüllt sein. Darüber hinaus gilt wieder die übergeordnete Maxime, daß jeder Schüler nach Beendigung der Ausbildung die Mindest-Erfolgswahrscheinlichkeit p_e erreicht.

Die Summe der von den Schülern aufzuwendenden Lernzeit ist durch

$$(4.1) \qquad z = t_1 a_1 + t_2 a_2 + \ldots + t_k a_k$$

gegeben, wobei a_i, $i = 1,2,\ldots,k$, die Anzahl der Schüler in Klasse K_i und

$$t_i = \max_{j \in K_i} t(j) \quad , \quad i = 1,2,\ldots,k$$

die Lernzeit des Schülers mit dem größten $t(j)$ von allen Schülern in Klasse K_i angibt. Definiert man

$$c_i := a_i/s \quad , \quad i = 1,2,\ldots,k$$

als die relative Klassengröße der Klasse K_i, dann kann Gleichung (4.1) auch als

$$(4.2) \qquad z = t_1 s c_1 + t_2 s c_2 + \ldots + t_k s c_k$$

geschrieben werden.

Bei der gegebenen Gesamtmenge S der Schüler erreicht die Funktion (4.2) für das gleiche Klassensystem ein Minimum wie folgende Funktion:

$$(4.3) \qquad z = t_1c_1 + t_2c_2 + \ldots + t_kc_k \quad .$$

Werden in (4.3) die Größen $c_1, c_2, \ldots, c_k$ als Konstante angesehen, dann kann Satz 1 unmittelbar auf dieses Problem angewandt werden: Notwendige Bedingung der optimalen Klassenorganisation im Hinblick auf die Minimierung der Schülerzeit ist wieder ein Klassensystem, dessen Klassen alle lückenlos geordnete Teilmengen der Menge S sind.

Der Satz 2 kann zur Ableitung der hinreichenden Bedingungen des Minimums der Zielfunktion (4.1) ebenfalls verwendet werden, weil ein nach Satz 2 vorgenommener Elemententausch die hier als gegeben betrachteten relativen Klassengrößen nicht verändert. Die Klassen müssen so gebildet werden, daß die Rangfolge der Klassengrößen der Rangfolge der Lernzeiten der jeweils langsamsten Schüler in jeder Klasse entspricht.

4.2 Vorgegebene Anzahl von Klassen und variable relative Klassengrößen

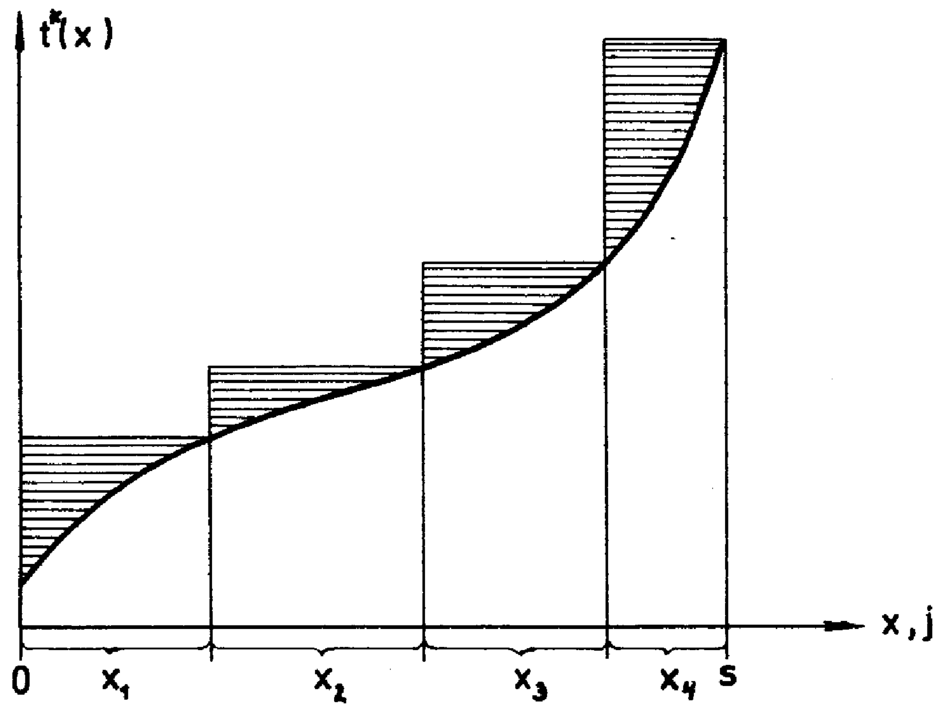
Faßt man die Funktion $t^*(x)$, $0 \leqq x \leqq s$, als eine Approximation der Funktion $t(j)$, $j \in \{1,2,\ldots,s\}$, auf, dann kann das Problem durch Anwendung der Differentialrechnung nährungsweise gelöst werden. Die Zielfunktion als Lagrange-Ansatz lautet in diesem Fall:

$$(4.4) \qquad z = x_1 t^*(x_1)+x_2 t^*(x_1+x_2)+\ldots+x_k t^*(x_1+x_2+\ldots+x_k) - \mu\left[\sum_{i=1}^{i=k} x_i - s\right] .$$

Hier ist wie in (3.35) die Zielfunktion unter der Nebenbedingung zu minimieren, daß die Summe der Schüler in allen Klassen gleich der Gesamtzahl s aller Schüler ist. Nicht ganzzahlige Lösungen x_i müssen wieder durch die ganzzahligen Werte mit dem kleinsten Abstand von x_i ersetzt werden. In Fig. 4 ist die Aufgabe für k = 4 graphisch veranschaulicht. Die Zielfunktion erreicht ihr Minimum, wenn die Summe der schraffierten Flächen ein Minimum wird. Der in Fig. 4 dargestellte S-förmige Verlauf spiegelt die impliziten Annahmen über die Besetzungszahlen einzelner Lernzeit-Intervalle mit alternativen Schülermengen wider. Bei dem dargestellten Verlauf ist die plausible Annahme impliziert, daß die relativen Häufigkeiten für die Lernzeit-Intervalle einer eingipfligen Häufigkeitsverteilung entsprechen.[1]

1) In der Psychologie geht man oft davon aus, daß das Merkmal Intelligenz normalverteilt ist oder zumindest einer eingipfligen glockenförmigen Verteilungsfunktion unterliegt. Dabei hat man festgestellt, daß die Fähigkeit "to learn ideas and symbols" eng mit dem Merkmal Intelligenz korreliert ist. Vgl. Thorndike, E.L., "Human Learning", Cambridge (Mass.), 1966, S. 184-185.

Fig. 4



Der Spezialfall, daß $t^*(x)$ eine Gerade darstellt, führt zu der Lösung $x_1 = x_2 = \dots = x_k$. In diesem Fall lautet die Zielfunktion

$$\begin{aligned} z = {} & x_1(a+bx_1) + x_2\left[a+b(x_1+x_2)\right] + \dots \\ (4.5) \qquad & \dots + x_k\left[a+b(x_1+x_2+\dots+x_k)\right] - \\ & - \mu\left[\sum_{i=1}^{i=k} x_i - s\right], \end{aligned}$$

wobei $t = a+bx$ eine Gerade mit positiver Steigung ist (vgl. Fig. 5). Differenziert man diese Zielfunktion partiell nach der Variablen x_i und x_{i+1}, $i \in \{1,2,\dots,k-1\}$, und setzt diese Ableitungen gleich Null, so erhält man

$$\frac{\partial z}{\partial x_i} = a+b \sum_{j=1}^{j=i} x_j + b \sum_{j=i}^{j=k} x_j - \mu = 0$$

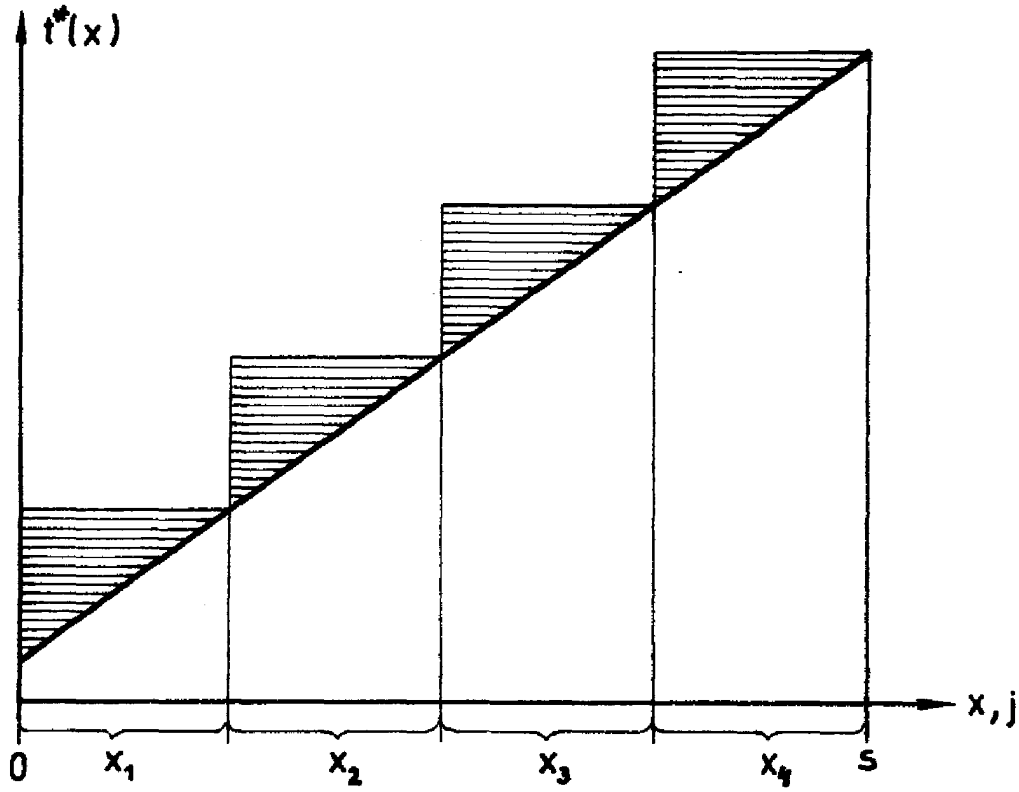
$$\frac{\partial z}{\partial x_{i+1}} = a+b \sum_{j=1}^{j=i+1} x_j + b \sum_{j=i+1}^{j=k} x_j - \mu = 0 \quad .$$

Setzt man diese beiden Ausdrücke einander gleich, so ergibt sich

$$x_i = x_{i+1} \quad .$$

Das Minimum der Zielfunktion ist also dann erreicht, wenn alle Klassen mit der gleichen Anzahl von Schülern besetzt werden. Die dabei vorausgesetzte Annahme, daß t(x) eine Gerade ist, impliziert eine spezielle Häufigkeitsverteilung der einzelnen Lernzeitwerte: zu jedem Lernzeitwert läßt sich genau ein Schüler zuordnen, wobei die Differenz zwischen zwei benachbarten Lernzeitwerten stets gleich ist. Da in diesem Fall $x_1 = x_2 = \ldots = x_k$, sind auch die schraffierten Flächen in Fig. 5 einander gleich.

Fig. 5



4.3. <u>Die optimale Anzahl von Klassen</u>

Zerlegt man eine beliebige lückenlos geordnete Klasse in zwei oder mehr lückenlos geordnete Teilklassen, so ist aus Fig. 4 und 5 sofort ersichtlich, daß dadurch die Summe der schraffierten Flächen und damit auch der Wert der Zielfunktion kleiner wird. Ist $t(j)$, $j \in \{1,2,\ldots,s\}$, eine strikt monoton steigende Funktion, so ist das absolute Minimum der Zielfunktion dann erreicht, wenn alle Klassen so weit zerlegt werden, bis ein Klassensystem entsteht, dessen Klassen alle nur mit einem Schüler besetzt sind. Besitzen zwei oder mehr Schüler den gleichen t-Wert, so daß $t(j)$ zwar monoton steigend, aber nicht strikt monoton steigend ist, dann erreicht die Zielfunktion ihr absolutes Minimum auch schon bei einer Klasseneinteilung, in der alle Schüler mit gleichem t-Wert in einer Klasse zusammengefaßt werden, während die anderen Schüler sich in je einer Klasse befinden. Bei einer monoton aber nicht strikt monoton steigenden Funktion $t(j)$ gilt jedoch ebenso wie bei einer strikt monoton steigenden Funktion, daß die Zielfunktion ihr absolutes Minimum für $k = s$ erreicht.

4.4 Berücksichtigung des Einflusses der Klassengröße auf die Lernzeit

(a) Vorgegebene Anzahl von Klassen und vorgegebene relative Klassengrößen.
Die zu minimierende Zielfunktion lautet in diesem Fall:

$$(4.6) \qquad z = \sum_{i=1}^{i=k} T_i a_i = \sum_{i=1}^{i=k} h(j^o, a_i) a_i \quad , \qquad j^o = \max_{j \in K_i} j$$

Da die a_i fest vorgegebene Größen sind, fällt die Funktion h unter diejenige Klasse von Funktionen $f(t_i)$, für die Satz 1 bewiesen wurde. Das optimale Klassensystem muß daher aus lückenlos geordneten Klassen bestehen. Satz 2 kann zur weiteren Bestimmung des optimalen Systems nicht verwendet werden.

Eine einfache Lösung erhält man nur, wenn $a_1 = a_2 = \ldots = a_k = s/k$ vorausgesetzt wird, weil dann das optimale System allein unter Anwendung des Satzes 1 bestimmt werden kann. In allen anderen Fällen kann die Aufgabe als ein Zuordnungsproblem behandelt und beispielsweise mit Hilfe des linearen ganzzahligen Programmierens gelöst werden. Die Figuren 6 und 7 veranschaulichen das Problem graphisch: Aus der Menge $S = \{1,2,\ldots,12\}$ wurden einmal die beiden Klassen $\{1,2,3,4\}$ und $\{5,6,\ldots,11,12\}$ gebildet und (Fig.7) die beiden Klassen $\{1,2,\ldots,7,8\}$ und $\{9,10,11,12\}$, wobei als vorgegeben angenommen wurde, daß $a_1 = 4$ und $a_2 = 8$. Die schraffierten Flächen geben die Lernzeit in einer bestimmten Klasse an. Optimal ist dasjenige System, für das die Summe dieser Flächen ein Minimum ist.

(b) Vorgegebene Anzahl von Klassen und variable relative Klassengrößen.

Die der Gleichung (4.4) entsprechende Zielfunktion lautet jetzt:

$$
\begin{aligned}
z = {} & x_1 h^*\left[x_1, x_1\right] + x_2 h^*\left[(x_1+x_2), x_2\right] + \ldots \\
& \ldots + x_k h^*\left[(x_1+x_2+\ldots+x_k), x_k\right] - \\
& - \mu\left[\sum_{i=1}^{i=k} x_i - s\right] .
\end{aligned}
\tag{4.7}
$$

Fig. 6

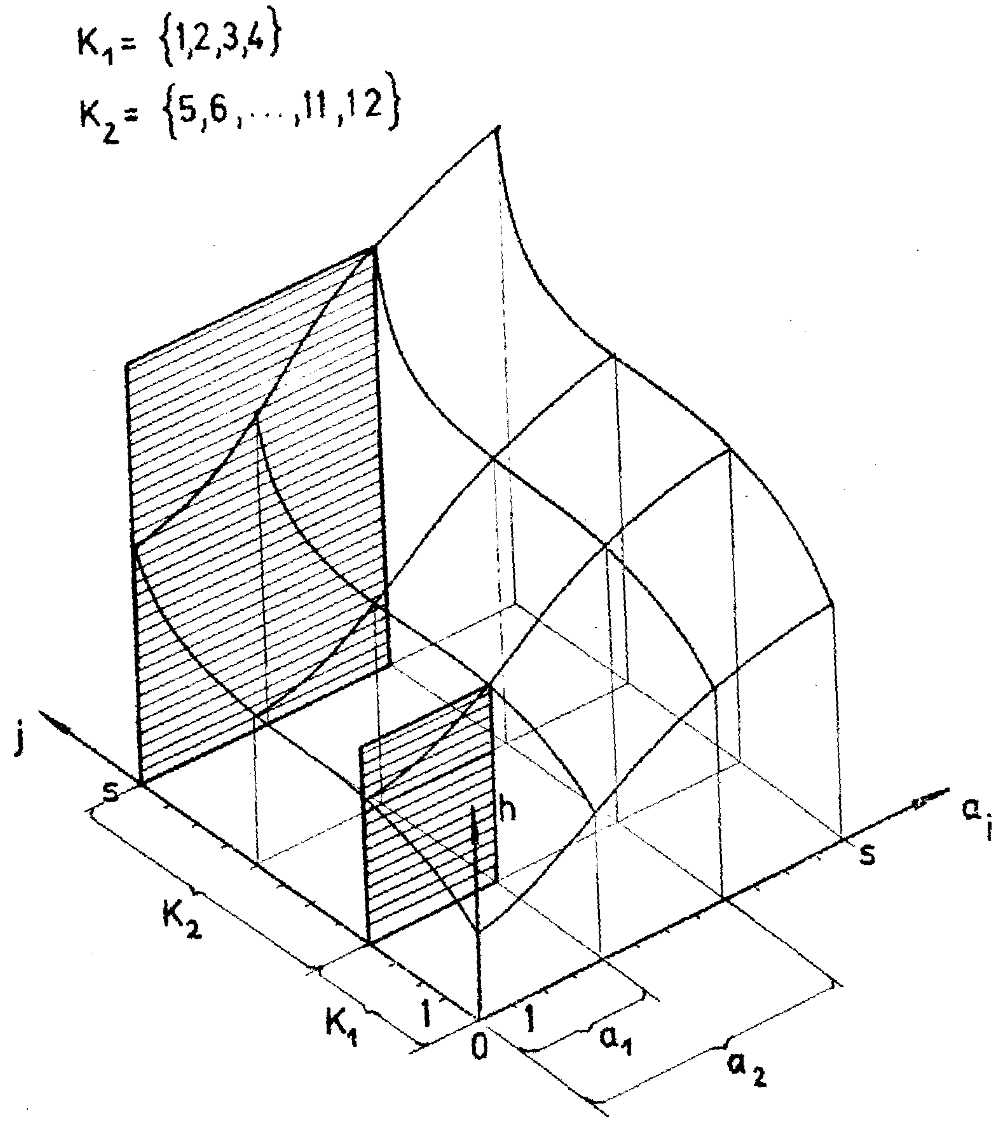
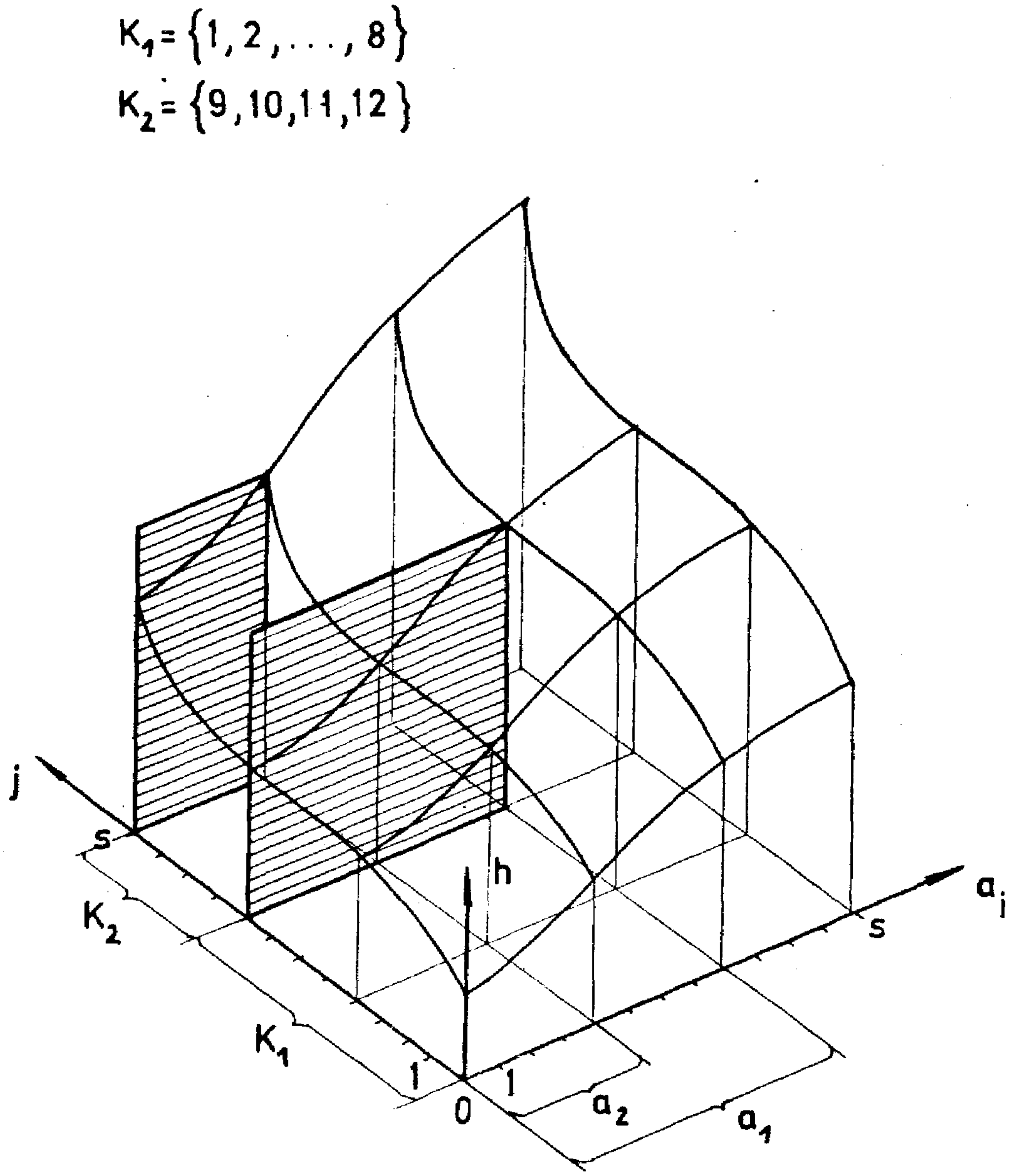


Fig. 7



Es ergibt sich also kein neues Problem bei der Ableitung des optimalen Klassensystems, wenn es gelingt, die Funktion h durch eine stetige und differenzierbare Approximationsfunktion h^* zu ersetzen. Fig. 6 kann hier gleichfalls zur Veranschaulichung des Problems herangezogen werden: Es sind k Flächen so zu bestimmen, daß die Summe der Flächeninhalte minimiert wird.

(c) Die optimale Anzahl von Klassen.
Wie in dem Fall, daß die Lernzeit nicht von der Klassengröße abhängt, ist auch hier das Minimum der Zielfunktion dann erreicht, wenn jede Klasse mit einem Schüler besetzt ist ($k = s$). Besitzen mehrere Schüler gleiche Lernzeiten, und werden diese Schüler in einer Klasse zusammengefaßt, während man jede andere Klasse nur mit einem Schüler besetzt, so erreicht die Zielfunktion nicht ihr absolutes Minimum. Hier ergibt sich also ein Unterschied zu dem Fall, daß die Lernzeit nicht von der Klassengröße abhängt.

4.5 Zusammenfassung

Durch Anwendung von Satz 1 erhält man für das optimale Klassensystem die wichtige Eigenschaft, daß alle Klassen aus lückenlos geordneten Teilmengen der Menge $S = \{1,2,\ldots s\}$ bestehen müssen, und zwar unabhängig davon, ob die Anzahl und die relative Größe der Klassen variabel oder vorherbestimmt ist. Dieses Ergebnis ändert sich nicht, wenn angenommen wird, daß die Lernzeit eine strikt monoton steigende Funktion der Klassengröße ist.

Satz 2 kann zur weiteren Bestimmung des optimalen Klassensystems nicht immer angewandt werden. Sind die Anzahl und die relative Größe der Klassen vorgegeben, so gibt es noch k! Möglichkeiten, ein Klassensystem zu bilden, dessen Klassen alle lückenlos geordnete Teilmengen der Menge S sind. Aus der Menge dieser Klassensysteme kann das optimale z.B. unter Anwendung der ganzzahligen Programmierung (Behandlung als Zuordnungsproblem) ausgewählt werden. Ist die relative Größe jeder Klasse gleich, so ergibt sich die vollständige Lösung schon aus der Anwendung von Satz 1.

Sind die relativen Klassengrößen variabel, während die Anzahl der Klassen vorgegeben wird, so kann die Zielfunktion als Lagrangesche Funktion unter der Nebenbedingung optimiert werden, daß die Anzahl der Schüler in allen Klassen gleich der Gesamtzahl der Schüler ist. Hierfür müssen allerdings die Funktionen t(j) und h durch stetige differenzierbare Funktionen approximiert werden. Im Spezialfall, daß sich t(j) durch eine Gerade approximieren läßt, ergibt sich unter der Voraussetzung einer von der Klassengröße unabhängigen Lernzeit ein optimales Klassensystem, bei dem alle Klassen gleich groß sind.

Ist die Anzahl der Klassen eine Variable, so erreicht die Zielfunktion ihr absolutes Minimum, wenn jede Klasse mit einem Schüler besetzt ist. Hängt die Lernzeit nicht von der Klassengröße ab, so erreicht die Zielfunktion das absolute Minimum auch schon bei einer Klasseneinteilung, bei der alle Schüler mit gleichen Lernzeiten in einer Klasse zusammengefaßt werden und bei der die übrigen Klassen nur mit einem Schüler besetzt sind.

Kapitel 5: Ertragsmaximierung

5.1 Erträge der Ausbildung

Als Ertrag der Ausbildung eines Schülers kann zunächst die Summe seiner Einkünfte definiert werden, die er aufgrund seiner Ausbildung durch den Verkauf seiner Arbeitskraft auf einem Arbeitsmarkt erzielt. Dieser Ertrag hängt cet. par. davon ab, in welcher Art von Fertigkeit er ausgebildet wurde, bis zu welchem Grade er diese Fertigkeit erlernt hat und wie lange er seine Arbeitskraft dem Käufer zur Verfügung stellt. Die Art der Fertigkeit, in der sich ein Schüler ausbilden läßt, und die von der Wahl der Schule und des Unterrichtsfaches bestimmt wird, soll als gegeben betrachtet werden. Ebenso ist der Ausbildungsgrad in dieser Untersuchung eine vorgegebene Konstante, die durch die Erfolgswahrscheinlichkeit p_e definiert wird, mit der ein Schüler nach Beendigung seiner Ausbildung ein der Ausbildung adäquates Problem zu lösen in der Lage ist. Die dritte Größe, von der der Ertrag der Ausbildung abhängt, ist die Länge des Zeitraums, die dem Schüler nach Beendigung seiner Ausbildung für den Einsatz seiner Arbeitskraft zur Verfügung steht. Bei vorgegebener Lebenserwartung des Schülers und vorgegebenem Ausbildungsbeginn ist diese Zeitspanne und damit der Ausbildungsertrag um so größer, je kürzer die Ausbildungszeit ist, die zum Erreichen der Mindest-Erfolgswahrscheinlichkeit

p_e nötig ist. Neben dem Ertrag, der dem Schüler als Einkommen zufließt, gibt es eine Reihe von Ertragsströmen, die indirekt auf die Ausbildung des Schülers zurückgeführt werden können. Ein derartiger Ertragsstrom ist beispielsweise die Einkommensdifferenz, um die das Einkommen eines Unternehmers, der die Arbeitskraft des Schülers im Produktionsprozeß einsetzt, dasjenige Einkommen übersteigt, das der Unternehmer bei Einsatz einer nicht ausgebildeten Arbeitskraft erzielt hätte. Weitere durch die Ausbildung des Schülers hervorgerufenen Ertragsströme können bei allen übrigen Wirtschaftssubjekten einer Volkswirtschaft entstehen, die als Glieder in der Kette von Güterkäufen und -verkäufen mit dem Verkäufer der Arbeitskraft verbunden sind. Die Summe dieser indirekt auf die Ausbildung eines Schülers zurückführbaren Ertragsströme ist cet. par. ebenfalls um so größer, je kürzer die Ausbildungszeit ist, d.h. je länger der Schüler seine Arbeitskraft zur Verfügung stellt.

Bezeichnet man mit t_{ℓ_j} die Lebenserwartung des Schülers j, mit t_{b_j} diejenige Zeitspanne, die verstreicht, ehe der Schüler seine Ausbildung beginnt und mit t_{u_j} die Zeitspanne, die zum Einsatz der Arbeitskraft zur Verfügung steht, dann ist definitionsgemäß

$$(5.1) \qquad t_{\ell_j} := t_{b_j} + t_i + t_{u_j} \; .$$

Im folgenden sei angenommen, daß der Ausbildungsbeginn und die Lebenserwartung bei allen Schülern gleich sind:

$$(5.2) \qquad \left. \begin{aligned} t_{b_j} &= t_b = \text{konst.} \\ t_{\ell_j} &= t_\ell = \text{konst.} \end{aligned} \right\} \quad j = 1,2,\ldots,s$$

Die Summe der Einkünfte, die die Gesamtheit der Absolventen der Klasse K_i aufgrund ihrer Ausbildung bezieht, einschließlich aller indirekt hervorgerufenen Einkommensströme, ist dann mit (5.2)

$$(5.3) \qquad c_i s \varphi(t_\ell - t_b - t_i) \quad .$$

In Gleichung (5.3) ist c_i der Anteil der Klasse K_i an der Gesamtmenge $S = \{1,2,\ldots,s\}$ aller Schüler. Die Funktion φ gibt die einem Schüler zurechenbaren Ertragsströme in Abhängigkeit von der Länge der Zeitspanne an, in der der Schüler seine Arbeitskraft zum Einsatz bringen kann. Es sei vereinfachend angenommen, daß diese Funktion linear ist:

$$(5.4) \qquad \varphi(t_\ell - t_b - t_i) := \gamma t \quad .$$

Die Gleichung (5.4) entspricht der Prämisse, daß der einem Schüler zurechenbare Ertrag in jeder Periode in Höhe eines bestimmten konstanten Einkommensbetrages γ anfällt.

Definiert man zur Vereinfachung der Schreibweise

$$(5.5) \qquad t_d := t_\ell - t_b \quad ,$$

dann erscheint der Ausdruck (5.3) unter Verwendung von (5.5) in der Form

(5.6) $$c_i s\gamma(t_d-t_i).$$

Die Summe der Ertragsströme, die allen Schülern in allen Klassen zurechenbar sind, beträgt dann mit Gleichung (5.6)

(5.7) $$s\gamma\sum_{i=1}^{i=k} c_i(t_d-t_i) .$$

5.2 Ertragsmaximierung bei Ausbildung von einer Kohorte, vorgegebener Anzahl von Klassen und vorgegebenen relativen Klassengrößen

Wenn in jeder Klasse nur eine einzige Kohorte [1] ausgebildet wird, so genügt es, die Organisation der Schüler in Klassen so vorzunehmen, daß der in Gleichung (5.7) definierte Gesamtertrag maximiert wird. Wenn die Schüler in verschiedenen Klassen zu unterschiedlichen Zeitpunkten ihre Ausbildung beenden, dann fallen auch die Ertragsströme, die den Schülern zuzurech-

1) In der bildungsökonomischen Literatur bezeichnet der Begriff Kohorte gewöhnlich eine Gruppe von Schülern, die dem gleichen Jahrgang angehören. In dieser Untersuchung sollen unter dem Begriff nicht nur die jahrgangsgleichen, sondern alle nach bestimmten Gesichtspunkten zusammengefaßten Schülergruppen verstanden werden, die nacheinander dieselbe Klasse durchlaufen.

nen sind, zu verschiedenen Zeitpunkten an. Um die Zeitpräferenz zu berücksichtigen, die gegenwärtigem Einkommen zugemessen wird, müssen die einzelnen Einkommens- bzw. Ertragsströme vor der Addition auf einen gemeinsamen Zeitpunkt diskontiert werden. Der auf den Beginn der Ausbildung diskontierte Ertrag y_i, der allen in der Klasse K_i ausgebildeten Schülern zurechenbar ist, beträgt bei einem Zinssatz von q-Prozent und dem diesem Zinssatz entsprechenden Zinsfaktor $r = 1 + q/100$

$$(5.8) \qquad y_i = s\, c_i \gamma \frac{r^{t_d - t_i} - 1}{r^{t_d}\,(r-1)} \quad .$$

Mit Gleichung (5.8) ist

$$(5.9) \qquad \gamma s c_i \frac{r^{t_d - t_i} - 1}{r^{t_d - t_i}\,(r-1)} = \gamma s c_i \frac{1 - r^{t_i - t_d}}{(r-1)}$$

der Ertragsstrom der Schüler der Klasse K_i, der sich durch Diskontierung auf den Zeitpunkt, in dem die Klasse K_i ihre Ausbildung beendet, ergibt.

Die Summe der den Schülern aller Klassen zurechenbaren Erträge ist aufgrund von (5.8)

$$(5.10) \qquad \sum_{i=1}^{i=k} y_i = s\gamma \sum_{i=1}^{i=k} c_i \frac{r^{t_d - t_i} - 1}{r^{t_d}\,(r-1)} \quad .$$

Statt den Ausdruck (5.10) zu maximieren, genügt es, den Ausdruck (5.11)

$$(5.11)\quad \begin{aligned} &(c_1 r^{t_d - t_1} - c_1) + (c_2 r^{t_d - t_2} - c_2) + \ldots + \\ &\ldots + (c_k r^{t_d - t_k} - c_k) \quad , \end{aligned}$$

bzw. die Ausdrücke (5.12) oder (5.13)

$$(5.12)\quad c_1 r^{t_d - t_1} + c_2 r^{t_d - t_2} + \ldots + c_k r^{t_d - t_k}$$

$$(5.13)\quad z = c_1 \frac{1}{r^{t_1}} + c_2 \frac{1}{r^{t_2}} + \ldots + c_k \frac{1}{r^{t_k}}$$

zu maximieren, weil alle diese Ausdrücke für die gleiche Klassenorganisation ihren höchsten Wert erreichen. Im folgenden wird das Maximum der Zielfunktion (5.10) analysiert, indem die einfachste, äquivalente Form von (5.10), nämlich die Funktion (5.13), zugrunde gelegt wird.

Das Problem, ein Klassensystem zu finden, das die Summe aller Erträge unter der Voraussetzung maximiert, daß (a) eine Mindest-Erfolgswahrscheinlichkeit p_e für alle Schüler garantiert und (b) kein Schüler vom Lernprozeß ausgeschlossen wird, kann unter Anwendung des Satzes 1 zu einem wesentlichen Teil gelöst werden.

Die Ausdrücke

$$(5.14)\quad c_i r^{-t_i} \quad , \quad i = 1,2,\ldots,k$$

in Gleichung (5.13) gehören zu jener Klasse von Funktionen, für die Satz 1 gilt. Es ist nämlich jeder dieser Ausdrücke eine strikt monoton sinkende Funktion in der Variablen t_i.

Die relativen Klassengrößen können in (5.13) als Konstante angesehen werden, weil sie sich durch die im Beweis zu Satz 1 durchgeführten Vertauschungen der Elemente nicht verändern. Aus der Anwendung von Satz 1 folgt, daß das ertragsmaximierende Klassensystem aus lückenlos geordneten Teilmengen der Schülermenge S bestehen muß.

Eine Anwendung des Satzes 2 auf das Problem der Ertragsmaximierung ist ebenfalls möglich, denn durch den in Satz 2 beschriebenen Elemententausch, der zur Erhöhung der Zielfunktion führt, bleiben die relativen Klassengrößen konstant.

5.3 Ertragsmaximierung bei Ausbildung von einer Kohorte, vorgegebener Anzahl von Klassen und variablen relativen Klassengrößen

Im vorangegangenen Abschnitt wurde gezeigt, daß die Funktion

$$(5.15) \qquad z = \sum_{i=1}^{i=k} c_i r_i^{-t_i}$$

die einfachste äquivalente Form der Zielfunktion (5.10) ist, weil der Wert der Zielfunktion für das gleiche Klassensystem ein Maximum annimmt wie die Funktion (5.13).

Wenn wieder $t^*(x)$, $0 \leqq x \leqq s$, als eine stetige differenzierbare Approximation der diskreten Funktion $t(j)$, $j \in \{1,2,\ldots,s\}$, betrachtet wird und x_i die Anzahl der Schüler in Klasse K_i bezeichnet, dann kann das Optimierungsproblem bei variablen relativen Klassengrößen folgendermaßen formuliert werden. Die Funktion

$$(5.16) \qquad z = \sum_{i=1}^{i=k} x_i r^{-t_i/s}$$

ist in bezug auf die Variablen x_i (vgl. Fig. 4) zu maximieren, wobei die Nebenbedingung

$$\sum_{i=1}^{i=k} x_i = s$$

erfüllt sein muß. Es kann deshalb folgender Lagrange-Ansatz gebildet werden:

$$(5.17) \qquad z = 1/s\left[x_1 r^{-t^*(x_1)} + x_2 r^{-t^*(x_1+x_2)} + \ldots \right.$$
$$\left. \ldots + x_k r^{-t^*(x_1+x_2\ldots+x_k)}\right] - \mu\left[\sum_{i=1}^{i=k} x_i - s\right]$$

Die Lösungen für die Variablen x_i, die man aus dem Gleichungssystem der partiellen Ableitungen erhält, sind dabei gegebenenfalls wieder durch Auf- oder -Abrunden durch ganzzahlige Werte zu ersetzen.

Während sich bei der Schülerzeitminimierung aus dem Lagrange-Ansatz eine einfache Lösung ergab, wenn angenommen wurde, daß $t(x)$ eine Gerade ist (vgl. Abschnitt 4.2), lassen sich in dem vorliegenden Fall durch derartige Substitutionen keine nennenswerten Vereinfachungen erzielen.

5.4 Ertragsmaximierung bei Ausbildung einer Kohorte, variabler Anzahl von Klassen und variablen relativen Klassengrößen

Teilt man eine beliebige Klasse eines Klassensystems mit lückenlos geordneten Klassen in zwei lückenlos geordnete Teilklassen auf, so steigt dadurch der zu maximierende Wert der Ertragsfunktion (5.10), bzw. der Wert der Funktion (5.13). Wird allgemein die i-te Klasse aufgeteilt, so wird der Term $c_i r^{-t_i}$ in (5.13) zu

$$c_i^{(1)} r^{-t_{i,1}} + c_i^{(2)} r^{-t_i} \quad , \tag{5.18}$$

worin $t_{i,1}$ der t-Wert des Schülers mit der größten Lernzeit von allen Schülern in der neuen Teilklasse $K_i^{(1)}$ ist. Da die neuen relativen Klassengrößen $c_i^{(1)}$ und $c_i^{(2)}$ zusammen gleich c_i sind, ist der Ausdruck (5.18) größer als der ursprüngliche Term $c_i r^{-t_i}$. Das Maximum der Zielfunktion bei variabler Klassenzahl ist deshalb dann erreicht, wenn jedem Schüler eine Klasse zugeordnet wird.

Hieraus darf jedoch nicht der an sich naheliegende Schluß gezogen werden, daß der Ertrag notwendig steigen muß, wenn sich die Anzahl der Klassen erhöht. Denn es ist möglich, die Anzahl der Klassen so um eine Einheit zu vergrößern, daß die dabei auftretenden Einsparungen an Lernzeit in solchen Klassen auftreten, deren relative Klassengrößen sich stark vermindern, während die Klassen mit Lernzeitverlusten durch hohe relative Klassengrößen gekennzeichnet sind. Eine Verringerung des Ertrages kann auch dadurch zustande kommen, daß nach der Erhöhung der Klassenzahl die relativen Klassengrößen gleich sind. Hierfür ein numerisches Beispiel:

Schülermenge $S = \{1,2,3,4,5,6\}$

Lernzeiten der Schüler $t(j)$:

$t(1) = 8$, $t(2) = 9$, $t(3) = 10$, $t(4) = 40$, $t(5) = 45$, $t(6) = 50$,

Die Menge S sei bisher in die beiden Klassen $K_1 = \{1,2,3\}$ und $K_2 = \{4,5,6\}$ aufgeteilt gewesen. Bei $r = 1{,}03$ ist für diese Klasseneinteilung der Wert der Funktion (5.13)

$$z = 1/2\left[(1/1{,}3439) + (1/4{,}3839)\right] \approx 0{,}49 \quad .$$

Errechnet man den entsprechenden Wert für die Klasseneinteilung $K_1' = \{1,2\}$, $K_2' = \{3,4\}$ und $K_3' = \{5,6\}$, so erhält man $z' = 0{,}43$. Da $z > z'$, ist der Ertrag hierbei gesunken.

5.5 Optimale Klassensysteme im Hinblick auf die Ertragsmaximierung bei Ausbildung mehrerer Kohorten

5.5.1 Unendlicher Planungshorizont

Die bisher in den einzelnen Abschnitten des Kapitels 5 abgeleiteten Regeln für die Bildung eines optimalen Klassensystems werden z.T. verändert, wenn in jeder Klasse statt nur einer Kohorte eine beliebig große Anzahl von Kohorten ausgebildet wird. Um dies zu zeigen, soll angenommen werden, daß sofort eine neue Kohorte noch nicht ausgebildeter Schüler bereit steht, nachdem eine Klasse ihre Ausbildung absolviert hat, um in der gleichen Klasse die bereits ausgebildeten Schüler zu ersetzen. Dabei wird vorausgesetzt, daß die neu eintretenden Schüler so ausgewählt sind, daß sie hinsichtlich der Lernzeiten dieselben Werte besitzen wie die Schüler der vorangegangenen Klasse.

Unter diesen Voraussetzungen ist der auf den Ausbildungsbeginn der ersten k Klassen diskontierte Gesamtertrag einer unendlichen Anzahl von Kohorten, die die Klasse K_i besucht haben, durch folgenden Ausdruck gegeben:

$$\bar{y}_i := c_i s \gamma \left[\frac{r^{t_d - t_{i-1}}}{r^{t_d}(r-1)} + \frac{r^{t_d - t_{i-1}}}{r^{t_d}(r-1)r^{t_i}} + \right.$$

$$(5.19) \qquad \left. + \frac{r^{t_d - t_{i-1}}}{r^{t_d}(r-1)r^{2t_i}} + \ldots \right]$$

$$= c_i s \gamma \frac{r^{t_d} - r^{t_i}}{r^{t_d}(r-1)(r^{t_i}-1)} \quad .$$

Mit (5.19) ist dann der Gesamtertrag sämtlicher Schüler aller Kohorten in allen Klassen:

$$(5.20) \qquad \sum_{i=1}^{i=k} \bar{y}_i \quad .$$

Um (5.20) zu maximieren, genügt es, die Funktion

$$(5.21) \qquad z = \sum_{i=1}^{i=k} c_i \; \frac{r^{t_d} - r^{t_i}}{r^{t_i} - 1}$$

zu maximieren, weil (5.21) für dieselbe Einteilung der Schüler in Klassen ein Maximum annimmt wie die Zielfunktion (5.20). Die Funktionen

$$(5.22) \qquad \frac{r^{t_d} - r^{t_i}}{r^{t_i} - 1} \quad , \qquad i = 1,2,\ldots,k$$

stimmen in einer wichtigen Eigenschaft mit den im Satz 1 eingeführten Funktionen $g(t_i)$ überein, denn es ist

$$(5.23) \qquad \frac{r^{t_d}-r^{t_i}}{r^{t_i}-1} > \frac{r^{t_d}-r^{t_j}}{r^{t_j}-1} \qquad \text{für } t_i < t_j \quad .$$

Es können deshalb die in den Abschnitten (5.2) - (5.4) unter alternativen Bedingungen für k und den Vektor $(c_1,c_2,...,c_k)$ formulierten Lösungsansätze für das vorliegende Problem übernommen werden. Im Gegensatz zu dem Fall, daß nur eine Kohorte ausgebildet wird, ist in der Zielfunktion bei Ausbildung mehrerer Kohorten (5.21) die Größe t_d enthalten. Im Gegensatz zum vorangegangenen Fall hängt das optimale Klassensystem deshalb auch davon ab, zu welchem Zeitpunkt die Schüler ihre Ausbildung aufnehmen.

5.5.2 Endlicher Planungshorizont

Wird in jeder Klasse eine endliche Anzahl von Kohorten ausgebildet (endlicher Planungshorizont), so gelten für die optimale Klassenorganisation die gleichen Regeln wie in Abschnitt 5.5.1 für den Fall eines unendlichen Planungshorizontes. Um dies zu zeigen, wird zunächst die Ertragsfunktion abgeleitet, die sich ergibt, wenn in einer Klasse K_i die Anzahl m_i, $m_i \in \mathbb{N}$, $\quad i = 1,2,...,k$, von Kohorten ausgebildet wird. Es sei angenommen, daß der Planungszeitraum die Anzahl τ von Perioden umfaßt, und daß für jede Klasse die Relation

$$t_i m_i = \tau = \text{konstant für } i = 1,2,...,k \quad ,$$

genau erfüllt ist.

Der auf den Beginn der Ausbildung bezogene diskontierte Ertrag der ersten Kohorte, die die Klasse K_i besucht, beträgt:

$$(5.24) \qquad y^{(i1)} = c_i \; s \, \gamma \, \frac{r^{t_d - t_i} - 1}{r^{t_d}(r-1)} \quad .$$

Dabei bezeichnet der 1. Index von y die Nummer der Klasse und der zweite die Nummer der Kohorte, die diese Klasse besucht. Der Ertrag der zweiten Kohorte ist dann

$$y^{(i2)} = \frac{1}{r^{t_i}} \, y^{(i1)}$$

und der der n-ten Kohorte

$$(5.25) \qquad y^{(in)} = \frac{1}{r^{t_i}} \, y^{(i,n-1)} = \frac{1}{r^{t_i(n-1)}} \, y^{(i1)} \quad .$$

Die Summe der Erträge, die sämtlichen m_i Kohorten der Klasse K_i zugerechnet werden können, ist dann

$$(5.26) \qquad y^{(i)} = y^{(i1)} \sum_{j=1}^{j=m_i} \frac{1}{r^{t_i(j-1)}} \quad .$$

Formt man den Ausdruck (5.26) unter Verwendung der Formel für die Summe einer endlichen geometrischen Reihe um, so ergibt sich

$$(5.27) \qquad y^{(i)} = y^{(i1)} \, \frac{r^{-t_i m_i} - 1}{r^{-t_i} - 1} \quad .$$

Setzt man Gleichung (5.24) in Gleichung (5.27) ein, so erhält man nach einigen Umformungen

$$(5.28) \qquad y^{(i)} = c_i \; s \; \gamma \frac{(r^{t_d} - r^{t_i})(1 - r^{-t_i m_i})}{r^{t_d}(r-1)(r^{t_i}-1)} \quad .$$

Die Gesamtsumme der Erträge, die sämtlichen ausgebildeten Schülern zurechenbar ist, ergibt sich dann aus Gleichung (5.28) durch Summation über alle k Klassen zu:

$$(5.29) \qquad \sum_{i=1}^{i=k} y^{(i)} = s \, \gamma \sum_{i=1}^{i=k} c_i \; \frac{(r^{t_d} - r^{t_i})(1 - r^{-\tau})}{r^{t_d}(r-1)(r^{t_i}-1)} \quad ,$$

wobei $t_i m_i = \tau$ substituiert wurde.

Sind die Größen v, t_d, s, γ und τ konstant, so kann Gleichung (5.29) auf Gleichung (5.21) reduziert werden, weil (5.29) für das gleiche Klassensystem ein Maximum annimmt wie (5.21). Folglich ist das optimale Klassensystem bei endlichem Planungshorizont das gleiche wie bei unendlichem.

5.6 Abhängigkeit der Lernzeit von der Klassengröße

(a) Bei Ausbildung einer Kohorte.

Sind die relativen Klassengrößen und die Anzahl der Klassen vorgegeben, so kann als Zielfunktion nach wie vor die Gleichung (5.10) verwendet werden, wenn dort das Symbol t_i durch das entsprechende Symbol T_i ersetzt wird. Da Satz 1 auch in diesem Fall angewandt werden kann, muß das optimale Klassensystem wieder aus lückenlos geordneten Teilmengen von S bestehen. Satz 2 kann hingegen zur weiteren Bestimmung des optimalen Systems nicht herangezogen werden. Das Problem besteht deshalb auch hier darin, aus der Menge der k! lückenlos geordneten Klassensysteme, die alle die Nebenbedingungen (3.1) - (3.5) erfüllen, das ertragsmaximale auszuwählen.

Sind die relativen Klassengrößen variabel und ist nur die Anzahl der Klassen vorgegeben, so kann das optimale System wieder durch einen Lagrange-Ansatz approximativ bestimmt werden. Die zu Gleichung (5.17) analoge Funktion lautet hierbei:

$$z = 1/s\Big[x_1 r^{-h^*[x_1, x_1]} + x_2 r^{-h^*[(x_1+x_2), x_2]} + \ldots$$
$$\ldots + x_k r^{-h^*[(x_1+x_2+\ldots+x_k), x_k]}\Big] -$$
$$-\mu\left[\sum_{i=1}^{i=k} x_i - s\right].$$

Sind die Anzahl der Klassen und die relativen Klassengrößen variabel, so gilt auch hier wieder, daß der Ertrag steigt, wenn eine lückenlos geordnete Klasse in zwei lückenlos geordnete Teilklassen zerlegt wird. Daraus darf jedoch auch hier nicht der Schluß gezogen werden, daß auch bei jeder anderen Art von Erhöhung der Klassenzahl der Ertrag steigen muß. Wie in Abschnitt 5.4. lassen sich hier ebenfalls numerische Beispiele konstruieren, bei denen der Ertrag sinkt.

(b) Ausbildung mehrerer Kohorten.

In den Abschnitten 5.5.1 und 5.5.2 wurde gezeigt, daß sowohl bei unendlichem wie bei endlichem Planungshorizont die Gleichung (5.21) die einfachste äquivalente Form der Zielfunktion ist. Dieses Ergebnis trifft auch für den Fall zu, daß die Lernzeit von der Klassengröße abhängig ist: In (5.21) müssen die Größen t_i lediglich durch die Größen T_i ersetzt werden. Im übrigen führt die Maximierung der so gewonnenen Gleichung zu den gleichen Lösungsansätzen zurück, die oben im Fall (a) unter der Annahme gelten, daß nur eine Kohorte ausgebildet wird.

5.7 Zusammenfassung

Die bei der Ertragsmaximierung auftretenden Probleme unterscheiden sich hinsichtlich der Lösungsansätze nicht wesentlich von den bereits bei der Schüler- und Lehrerzeitminimierung diskutierten Fragen. Hier wie dort konnte Satz 1 zur Bestimmung des optimalen Klassensystems angewandt werden: Die einzelnen Klassen des optimalen Systems müssen lückenlos geordnete Teilmengen der Menge $S = \{1,2,...,s\}$ sein. Zur weiteren Bestimmung des optimalen Systems bot sich je nach den Nebenbedingungen der Lagrange-Ansatz an, oder es mußte die Auswahl des optimalen Systems dadurch getroffen werden, daß die Aufgabe als Zuordnungsproblem interpretiert wurde. Die intuitive Vorstellung, daß der Ertrag mit steigender Klassenzahl steigt, erwies sich nicht in jedem Falle als richtig. Ein weiteres wichtiges Ergebnis war wie in den vorangegangenen Kapiteln die Abhängigkeit des optimalen Systems von der Häufigkeitsverteilung der Lernzeiten, die in der Funktion $t(j)$ implizite enthalten ist. An diesen Ergebnissen änderte sich nichts durch die Annahme, daß die Lernzeiten eine Funktion der Klassengröße sind.

Kapitel 6: Kosten und Kostenminimierung

6.1 Bei Ausbildung einer Kohorte

6.1.1 Ohne Diskontierung der Kosten

6.1.1.1 Vorgegebene Anzahl von Klassen und vorgegebene relative Klassengrößen

Die Zusammenfassung der Schüler zu mehreren Klassen und deren Ausbildung führt zu Kosten, die danach unterschieden werden können, ob sie im Hinblick auf das Ergebnis der Lernprozesse (Produktion einer bestimmten Mindest-Erfolgswahrscheinlichkeit für alle Schüler) variabel sind oder in einer vom Ergebnis der Lernprozesse unabhängigen fixen Größe anfallen (variable und fixe Kosten).

Da die Produktion der vorgegebenen Mindest-Erfolgswahrscheinlichkeit p_e in dem zugrundegelegten Lernmodell bei gegebenen Lernparametern der Schüler von der für das Lernen aufgewandten Zeit abhängt, soll angenommen werden, daß die variablen Kosten, die aus dem Entgelt für die Lehrer bestehen, der von den Lehrern aufgewandten Unterrichtszeit direkt proportional sind. Die fixen Kosten (Gebäudekosten, Lernmittelkosten, etc.), die mit der Höhe der produzierten Mindest-Erfolgswahrscheinlichkeit p_e voraussetzungsgemäß nicht in Beziehung stehen (in dem zugrundegelegten Lernmodell kann beispielsweise keine Funktion zwischen dem p-Wert nach n Perioden und dem für eine beliebige Fixkostenart aufgewandten Betrag abgeleitet werden), sollen den einzelnen Lernperioden zu gleichen Anteilsbeträgen zugerechnet werden. Obwohl aufgrund dieser Annahmen sowohl die variablen als auch die fixen Kosten die gleiche Dimension (DM pro Zeiteinheit) besitzen, bestehen zwischen ihnen Unterschiede, die für die kostenminimale Klassenorganisation von Bedeutung sind.

Unter den Voraussetzungen, daß (a) die fixen Kosten so lange aufgewandt werden müssen, bis die am langsamsten lernende Klasse den Wert p_e erreicht, und daß (b) die variablen Kosten pro Zeiteinheit (Lohnsatz der Lehrer) in jeder Klasse gleich sind, läßt sich die zu minimierende Zielfunktion wie folgt formulieren:

$$(6.1) \qquad z = wt(s) + v \sum_{i=1}^{i=k} t_i \quad .$$

In Gleichung (6.1) stellt v den Aufwand pro Zeiteinheit an variablen Kosten dar und w den Fixkostenbetrag, der auf eine Zeiteinheit entfällt. Die Größen t_i und t(s) sind wie folgt definiert:

$$(6.2) \qquad \begin{aligned} t(s) &= \max_{j \in S} t(j) \\ t_i &= \max_{j \in K_i} t(j) \quad , \quad i = 1,2,\ldots,k \quad . \end{aligned}$$

Der Ausdruck t(s) gibt die Anzahl von Perioden an, die diejenige Klasse an Lernzeit aufwenden muß, die als letzte von allen k Klassen den vorgegebenen Wert p_e erreicht.

Die Zielfunktion (6.1) ist wieder unter den Nebenbedingungen (3.1) - (3.5) zu minimieren. Da der Ausdruck wt(s) in Gleichung (6.1) eine Konstante ist, ergibt sich, daß jedes Klassensystem, das die innerhalb der Lehrerzeitminimierung bereits diskutierte Gleichung

$$(6.1.1) \qquad z = \sum_{i=1}^{i=k} t_i$$

minimiert, auch die Zielfunktion (6.1) minimiert. Das optimale Klassensystem in bezug auf die Minimierung von Gleichung(6.1.1) wurde bereits in Kapitel 3 abgeleitet. Das dort gefundene optimale System, das auch die Gesamtkosten minimiert, lautet:

$$\begin{aligned}
K_1 &= \{1,2,\ldots,a_1\} \\
K_2 &= \{a_1+1, a_1+2, \ldots, a_1+a_2\} \\
&\vdots \\
K_k &= \{a_1+a_2+\ldots+a_{k-1}+1, a_1+a_2+\ldots+a_{k-1}+2, \ldots \\
&\qquad \ldots, a_1+a_2+\ldots+a_{k-1}+a_k\} \quad .
\end{aligned}$$

Diese optimale Klasseneinteilung ist unabhängig von der absoluten und der relativen Höhe der beiden Kostensätze v und w.

6.1.1.2 Variable Anzahl von Klassen und variable relative Klassengrößen

Sollen die Gesamtkosten bei frei wählbaren relativen Klassengrößen $c_i = a_i/s$, $i = 1,2,\ldots,k$, sowie bei frei wählbarem k minimiert werden, so entfällt bei der Optimierung die Nebenbedingung (3.5). Da auch in diesem Fall die Zielfunktion (6.1) für die gleiche Klassenorganisation ein Minimum besitzt wie die Funktion

$$z = \sum_{i=1}^{i=k} t_i \quad ,$$

ergibt sich in diesem Fall dieselbe Lösung wie in Kapitel 3, Abschnitt 2: Die Gesamtkosten erreichen ein Minimum, wenn alle Schüler in einer einzigen Klasse zusammengefaßt werden. Allerdings muß auch hier die lerntheoretische Prämisse überprüft werden, derzufolge die Lernfunktion eines Schülers unabhängig von der Größe und der Zusammensetzung einer Klasse ist.

6.1.2 Mit Diskontierung der Kosten

6.1.2.1 Vorgegebene Anzahl von Klassen und vorgegebene relative Klassengrößen

Um hinsichtlich der Ausgaben, die zu verschiedenen Zeitpunkten anfallen, die Zeitpräferenz zu berücksichtigen, sollen die Folgerungen aus der Zielfunktion Kostenminimierung erneut abgeleitet werden. Hierbei ergeben sich keine neuen Organisationsprinzipien.

Die Formulierung der Kostenfunktion unter der Berücksichtigung der Zeitpräferenz wird jedoch für die Diskussion der Zielfunktion Nettoertragsmaximierung benötigt. Sie soll deshalb ebenfalls diskutiert werden.

Bei einem Zinssatz von q-Prozent und einem Zinsfaktor von $r = 1 + q/100$ betragen die auf den Beginn der Ausbildung diskontierten Gesamtkosten aller Klassen:

$$(6.3) \qquad z = \frac{w(r^{t(s)}-1)}{r^{t(s)}(r-1)} + \sum_{i=1}^{i=k} \frac{v(r^{t_i}-1)}{r^{t_i}(r-1)} \quad .$$

Dasjenige Klassensystem, das den Ausdruck (6.4) minimiert, minimiert auch die Zielfunktion (6.3):

$$(6.4) \qquad \sum_{i=1}^{i=k} \frac{r^{t_i}-1}{r^{t_i}} \quad .$$

Statt den Ausdruck (6.4) zu minimieren, kann auch der Term

$$(6.5) \qquad \sum_{i=1}^{i=k} \frac{-1}{r^{t_i}}$$

minimiert, bzw. die Funktion

$$(6.6) \qquad z = \frac{1}{r^{t_1}} + \frac{1}{r^{t_2}} + \ldots + \frac{1}{r^{t_k}}$$

maximiert werden.

Die Funktionen

$$(6.7) \qquad \frac{1}{r^{t_i}} \quad , \qquad i = 1,2,\ldots,k$$

gehören zu der gleichen Klasse von Funktionen, für die die Sätze 1 und 2 bewiesen worden sind: Diese Funktionen sind strikt monoton fallend für positive Werte t_i. Aus der Anwendung des Satzes 1 folgt daher, daß das kostenminimale Klassensystem aus Klassen bestehen muß, die lückenlos geordnete Teilmengen der Menge $S = \{1,2,\ldots,s\}$ sind. Aus der Anwendung von Satz 2 folgt ferner, daß beim kostenminimalen System die Rangfolge der Schülerzahlen in den einzelnen Klassen derjenigen Rangfolge entsprechen muß, die durch die Lernzeiten der jeweils langsamsten Schüler in jeder Klasse gebildet wird. Das System, das diese beiden Forderungen erfüllt, ist identisch mit dem in Abschnitt 6.1.1.1 abgeleiteten System und ist ebenfalls unabhängig von den beiden Kostensätzen v und w.

6.1.2.2 Vorgegebene Anzahl von Klassen und variable relative Klassengrößen

Bezeichnet man zwei Klassen des im vorangegangenen Abschnitt abgeleiteten kostenminimalen Klassensystems als benachbart, wenn ihre Vereinigung zu einer lückenlos geordneten Teilmenge von $S = \{1,2,\ldots,s\}$ führt, dann ergibt sich unmittelbar folgender Satz: Der Wert der Zielfunktion (6.3) verringert sich, wenn zwei benachbarte Klassen sich so verändern lassen, daß nach der Veränderung die eine der beiden Klassen nur aus dem Schüler mit der geringsten Lernzeit von allen Schülern der beiden Klassen besteht und die zweite Klasse alle übrigen Schüler umfaßt. Hieraus ergibt sich, daß das Minimum der Gesamtkosten dann erreicht wird, wenn die k-1 Schüler mit den k-1 kleinsten Lernzeiten in je einer Klasse und alle übrigen Schüler in einer weiteren Klasse zusammengefaßt werden. Dieses Ergebnis kann sich ändern, wenn angenommen wird, daß die Lernzeiten von der Klassengröße abhängig sind (vgl. Abschnitt 6.3).

6.1.2.3 Variable Anzahl von Klassen und variable relative Klassengrößen

Faßt man zwei beliebige Klassen K_i und K_{i+1}, $i = 1,2,\ldots,k-1$, des optimalen Klassensystems, das nach den Regeln des vorausgegangenen Abschnitts gebildet wurde, in einer Klasse K_i'

$$K_i' = K_i \cup K_{i+1}$$

zusammen, so entfällt dadurch im Term (6.4) ein Summand und der Wert der Zielfunktion (6.3) verringert sich. Das Minimum der Zielfunktion ist deshalb dann erreicht, wenn alle Schüler in einer Klasse zusammengefaßt werden. Allerdings ist die hierbei implizierte Annahme, daß die Lernzeiten der Schüler von der Klassengröße nicht beeinflußt werden, kaum mehr erfüllt. Das Ergebnis wird deshalb weiter unten im Hinblick auf die alternative Prämisse hin überprüft, daß Lernzeiten und Klassengrößen voneinander abhängen.

6.2 Bei Ausbildung mehrerer Kohorten

6.2.1 Unendlicher Planungshorizont

6.2.1.1 Variable Kosten

Da jede ausscheidende Klasse sofort durch eine neue Klasse ersetzt wird, fallen für jede Klasse in jeder Periode Kosten in Höhe von v an, deren Barwert sich mit der Formel für den Barwert einer ewigen Rente errechnet als $v/(r-1)$. Entsprechend betragen die Kosten für alle k Klassen $kv/(r-1)$ Geldeinheiten.

6.2.1.2 Variable und fixe Kosten

Der Fixkostenbetrag w wurde als derjenige Geldbetrag definiert, der pro Zeiteinheit aufgewandt werden muß, solange sich noch eine Klasse in der Schule in Ausbildung befindet. Bei Ausbildung einer unendlichen Anzahl von Kohorten in jeder Klasse betragen die auf den Ausbildungabeginn der ersten Kohorten diskontierten fixen Kosten $w/(r-1)$ Geldeinheiten.
Diese Kosten sind unabhängig davon, wie die Schüler in eine gegebene Anzahl von Klassen zusammengefaßt werden und unabhängig von der Anzahl der Klassen.
Für die Gesamtkosten, die alle Kohorten in allen Klassen verursachen, ergibt sich der Gesamtbetrag

(6.17) $(kv+w)/(r-1)$.

Bei gegebener Anzahl von Klassen läßt sich deshalb aus der Zielfunktion minimaler Gesamtkosten kein optimales Klassensystem ableiten: alle Arten von Klassensystemen sind in bezug auf die Zielfunktion äquivalent.

6.2.2 Endlicher Planungshorizont

Die Ergebnisse des vorangegangenen Abschnitts gelten auch für den Fall eines endlichen Planungshorizontes, d.h. für die Ausbildung einer endlichen Anzahl von Kohorten in jeder Klasse. Denn in diesem Falle entstehen für die Dauer des Planungszeitraumes τ pro Periode feste Ausgabenströme. Deren diskontierte Summe beträgt:

(6.8) $$\frac{1-r^{\tau}}{r^{\tau}(1-r)}\,(kv+w)$$

6.3 Berücksichtigung des Einflusses der Klassengröße auf die Lernzeit

Der Einfluß der Klassengröße wird in den einzelnen Ansätzen zur Kostenminimierung formal dadurch berücksichtigt, daß die Größen t_i durch den allgemeineren Term $T_i = h(j^o, a_i)$ ersetzt werden. Satz 1 kann nach dieser Substitution weiterhin auf die Kostenfunktion angewandt werden, weil sich die Größen a_i im Term T_i durch den im Beweis von Satz 1 durchgeführten Elemententausch nicht verändern. Aus der Anwendung dieses Satzes ergibt sich auch in diesem

Fall als notwendige Bedingung des kostenminimalen Klassensystems, daß es aus lückenlos geordneten Teilmengen der Menge S bestehen muß. Satz 2 ist hingegen nicht länger anwendbar, denn ein Elemententausch, wie er im Beweis von Satz2 durchgeführt wurde, hätte bei der Funktion $h(j^o,a_i)$ zur Folge, daß sich außer der Variablen j^o auch die Größe a_i ändert. Somit wäre die Voraussetzung von Satz 2, daß die Lernzeit mit steigendem j steigt, nicht mehr in jedem Falle erfüllt. Die weiteren Eigenschaften des kostenminimalen Klassensystems müssen deshalb für die einzelnen Fälle gesondert abgeleitet werden.

(a) Die verschiedenen Fälle bei Ausbildung einer Kohorte.

Bei vorgegebener Klassenzahl und vorgegebenen relativen Klassengrößen muß das kostenminimale System nun nicht mehr notwendigerweise die Eigenschaft besitzen, daß die Rangfolge der Klassengrößen mit derjenigen Rangfolge übereinstimmt, die von den Lernzeiten der jeweils langsamsten Schüler der einzelnen Klassen gebildet wird (vgl. Abschnitte 6.1.1.1 und 6.1.2.1). Bei k Klassen gibt es wieder die Anzahl k! von Klassensystemen, die alle die Bedingung lückenlos geordneter Klassen und die Nebenbedingungen (3.1) bis (3.5) erfüllen. Die Optimierung besteht hier also wieder in der Lösung eines Zuordnungsproblems.

Bei vorgegebener Klassenzahl und variablen Klassengrößen kann wieder ein Lagrange-Ansatz zur approximativen Bestimmung der Klassenbesetzungszahlen verwendet werden. Für den Fall, daß die Ausbildungskosten nicht diskontiert werden, besteht der Lagrange-Ansatz in der Minimierung der Funktion (3.35). Bei Diskontierung der Kosten kann die Funktion (6.3) als Ausgangspunkt verwendet werden. Dividiert man diese Funktion durch das konstante Glied r-1 und löst die Brüche auf, so läßt sich (6.3) auf den Ausdruck

$$
\begin{aligned}
z = & -v\Big[r^{-h^{*}[x_1,x_1]} + r^{-h^{*}[(x_1+x_2),x_2]} + \ldots \\
& \ldots + r^{-h^{*}[(x_1+x_2+\ldots+x_k),x_k]}\Big] - \\
& - wr^{-h^{*}[(x_1+x_2+\ldots+x_k),x_k]} - \\
& - \mu\left[\sum_{i=1}^{i=k} x_i - s\right]
\end{aligned}
\tag{6.6.1}
$$

reduzieren. (6.6.1) basiert auf der Hypothese, daß die k-te Klasse die langsamste ist. Da diese Hypothese nicht zutreffen muß, ist der Ansatz für die k-1 alternativen Hypothesen ebenfalls zu lösen. Aus den resultierenden k Systemen ist dasjenige auszuwählen, das mit der verwendeten Hypothese übereinstimmt.

Bei variabler Klassenzahl und variablen relativen Klassengrößen muß das bisherige Ergebnis relativiert werden. Das kostenminimale Klassensystem besteht nun nicht mehr notwendigerweise aus einer einzigen Klasse, die alle Schüler umfaßt, sondern es kann der entgegengesetzte Extremfall eintreten, bei dem das Kostenminimum dann erreicht wird, wenn für jeden Schüler eine eigene Klasse gebildet wird. Das tatsächliche Ergebnis hängt in starkem Maße von der Funktion h ab. Um dies zu zeigen, sei angenommen, ein Klassensystem bestehe für $S = \{1,2,...,12\}$ aus den beiden Klassen $K_1 = \{1,2,3,4\}$, $K_2 = \{5,6,...,12\}$. Die Größen T_1 und T_2 dieser beiden Klassen sind in Fig. 8 eingetragen. Der Wert der Funktion (6.3) beträgt für dieses System

$$\frac{w}{r-1}\left[1-(1/r^{T_2})\right] + \frac{v}{r-1}\left[1-(1/r^{T_1}) + 1-(1/r^{T_2})\right]$$

Wird aus dem Klassensystem K_1, K_2 ein neues System gebildet, indem die beiden Klassen zu einer Klasse $K_3 = \{1,2,...,12\}$ zusammengefaßt werden, dann ist der Wert der Funktion (6.3) für dieses neue System

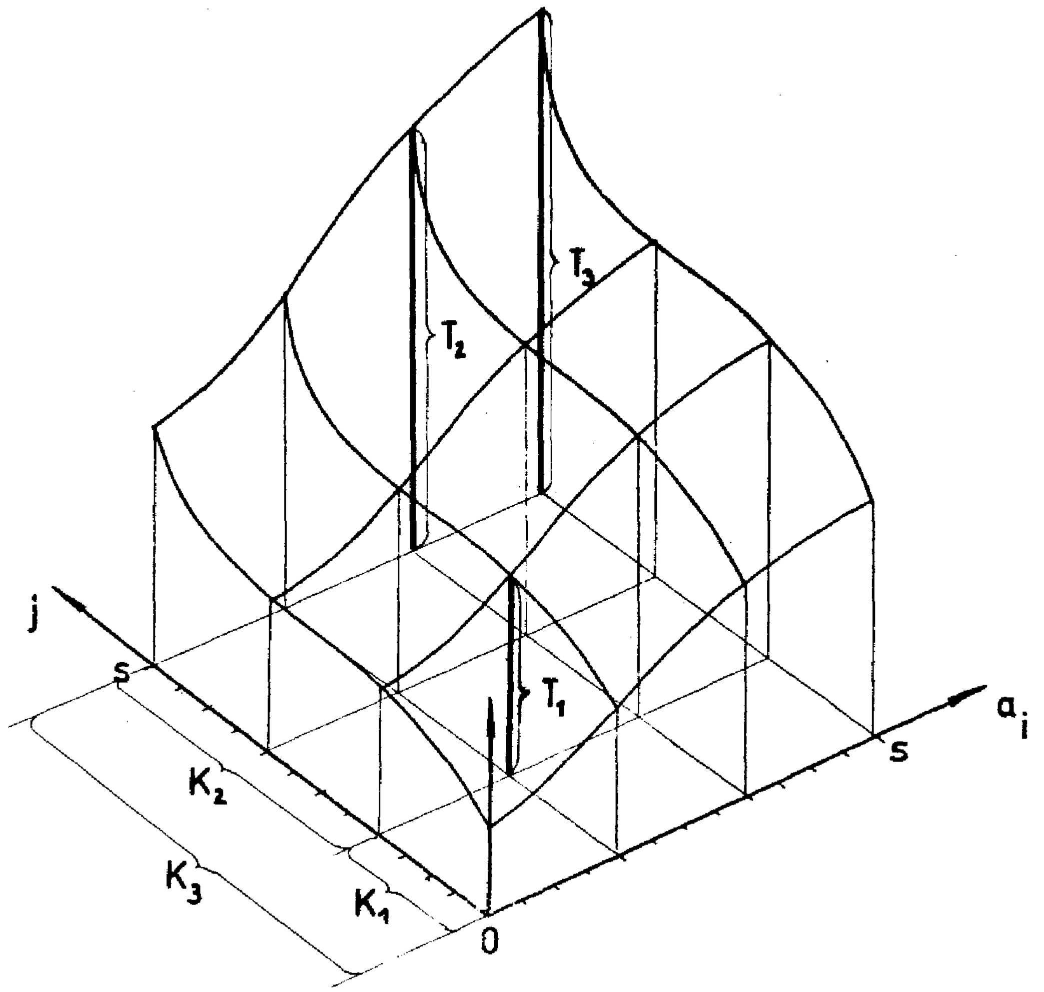
$$\frac{w}{r-1}\left[1-(1/r^{T_3})\right] + \frac{v}{r-1}\left[1-(1/r^{T_3})\right] .$$

Die Kosten können für dieses neue System größer als für das System K_1, K_2 sein, wenn die Größen T_1 und T_2 verglichen mit T_3 klein sind.

(b) Bei Ausbildung mehrerer Kohorten.

Unter der Annahme, daß mehrere Kohorten ausgebildet werden, ergab sich, daß die Klasseneinteilung für die Höhe der Kosten ohne Bedeutung war. Dieses Ergebnis bleibt unverändert, wenn angenommen wird,

Fig. 8



daß die Lernzeit von der Klassengröße abhängt.

6.4 Zusammenfassung

Unter der Annahme, daß die Lernzeit nicht von der Klassengröße abhängt, und daß nur eine Kohorte ausgebildet wird, konnten die kostenminimalen Systeme durch Anwendung der Sätze 1 und 2 bestimmt werden: Bei vorgegebener Klassenzahl und vorgegebenen relativen Klassengrößen mußten die Klassen so gebildet werden, daß die Rangfolge der Klassengrößen mit der Rangfolge übereinstimmte, die sich durch die Lernzeiten der jeweils langsamsten Schüler in einer Klasse ergab. Das Kostenminimum bei variabler Klassenzahl und variablen relativen Klassengrößen wurde bei einem Klassensystem erreicht, bei dem sich alle Schüler in einer Klasse befinden. Bei Abhängigkeit der Lernzeit von der Klassengröße ließ sich nur noch Satz 1 anwenden. Die Ableitung des kostenminimalen Klassensystems, das aus lückenlos geordneten Klassen bestehen muß, bestand bei vorgegebenen relativen Klassengrößen und vorgegebener Klassenzahl in der Lösung eines Zuordnungsproblems. Bei variablen relativen Klassengrößen und konstanter Klassenzahl konnte zur Bestimmung der Klassenbesetzungszahlen wieder ein Lagrange-Ansatz formuliert werden. Dabei zeigte sich, daß das Kostenminimum im Gegensatz zu den bisherigen Fällen auch von den beiden Kostensätzen v und w abhängt. Ein weiterer Unterschied bestand darin, daß die Kosten mit steigender Klassenzahl nicht mehr notwendigerweise steigen mußten, sondern auch fallen konnten, je nach dem, welche Form die Lernzeitfunktion h besitzt.

Bei Ausbildung mehrerer Kohorten und gegebener Klassenzahl hatte die Klasseneinteilung keinen Einfluß auf die Höhe der Kosten. An diesem Ergebnis änderte sich nichts, wenn angenommen wurde, daß die Lernzeit eine Funktion der Klassengröße ist.

Kapitel 7: Maximierung des Nettoertrages

7.1 Nettoertragsmaximierung bei Ausbildung einer Kohorte

Als Nettoertrag soll im folgenden die Differenz zwischen den einer Schülermenge zurechenbaren Erträgen und den für die Ausbildung dieser Schüler aufgewandten Kosten definiert werden. Unter Verwendung der Gleichung (5.10) für den Ertrag und der Gleichung (6.3) für die Kosten ergibt sich als Nettoertragsfunktion bei Ausbildung einer Kohorte die folgende Funktion

$$z = s\gamma\left[\sum_{i=1}^{i=k} c_i \frac{r^{t_d - t_i} - 1}{r^{t_d}(r-1)}\right] - v\left[\sum_{i=1}^{i=k} \frac{r^{t_i} - 1}{r^{t_i}(r-1)}\right] - \frac{w(r^{t(s)} - 1)}{r^{t(s)}(r-1)} \quad . \tag{7.1}$$

Multipliziert man (7.1) mit dem konstanten Faktor r-1, so erreicht die daraus resultierende Gleichung bei der gleichen Klasseneinteilung ein Maximum wie die Gleichung (7.1) selbst. Dasselbe gilt, wenn man von (7.1) den letzten Bruch, der konstant ist und die fixen Kosten darstellt, subtrahiert. Gleichung (7.1) läßt sich ferner dadurch vereinfachen, daß man den Bruch im ersten Summenausdruck auflöst und anschließend den ebenfalls konstanten Ausdruck $s\gamma r^{-t_d}$ dazu addiert. Aus diesen Umformungen erhält man als zu maximierenden Ausdruck die Gleichung

$$(7.2) \qquad z = s\gamma \sum_{i=1}^{i=k} c_i \frac{1}{r^{t_i}} - v \sum_{i=1}^{i=k} \frac{r^{t_i}-1}{r^{t_i}} ,$$

die sich auch in der Form

$$(7.3) \qquad z = \sum_{i=1}^{i=k} \frac{s\gamma c_i - v(r^{t_i}-1)}{r^{t_i}}$$

schreiben läßt. In Gleichung (7.3) sind die Größen s, γ und v enthalten. Im Gegensatz zu der reinen Ertragsmaximierung bzw. Kostenminimierung ist deshalb das optimale Klassensystem von der Anzahl der Schüler, der Ertragsrate γ und der Kostenrate v abhängig. Da jeder der einzelnen Summanden in Gleichung (7.3) eine strikt monoton sinkende Funktion in der Größe t_i ist, kann Satz 1 zur Bestimmung des Maximums angewandt werden: Das ertragsmaximale System muß aus lückenlos geordneten Klassen bestehen. Zur weiteren Bestimmung des optimalen Systems empfiehlt es sich wieder, einige Fälle zu unterscheiden.

(a)Vorgegebene Klassenzahl und vorgegebene relative Klassengrößen.

Wie bei der reinen Ertragsmaximierung bzw. Kostenminimierung kann Satz 2 auch in diesem Fall zur weiteren Bestimmung des optimalen Klassensystems herangezogen werden, weil im Beweis zu Satz 2 wie auch in Gleichung (7.3) die relativen Klassengrößen konstant gehalten werden. Wendet man beide Sätze an, so ist das optimale System vollkommen bestimmt.

(b) Vorgegebene Klassenzahl und variable relative Klassengrößen.

Verwendet man wieder $t^*(x)$, $o \leqq x \leqq s$, als eine stetige und differenzierbare Approximation der diskreten Funktion $t(j)$, so kann das Maximum von (7.3) durch Differentiation des Lagrange-Ansatzes

$$z = \frac{x_1\gamma - v(r^{t^*(x_1)}-1)}{r^{t^*(x_1)}} + \frac{x_2\gamma - v(r^{t^*(x_1+x_2)}-1)}{r^{t^*(x_1+x_2)}} + \ldots$$

$$+ \ldots \frac{x_k\gamma - v(r^{t^*(x_1+x_2+\ldots+x_k)}-1)}{r^{t^*(x_1+x_2+\ldots+x_k)}} - \tag{7.4}$$

$$- \mu \left[\sum_{i=1}^{i=k} x_i - s \right]$$

nach den unbekannten Klassenbesetzungszahlen x_i, $o \leqq x_i \leqq s$, und μ nährungsweise bestimmt werden (zur Definition der x_i vgl. Fig. 4).

(c) Variable Klassenzahl und variable relative Klassengrößen.

Aus Abschnitt 5.4 (Ertragsmaximierung) ergab sich, daß das Maximum des Ertrages für eine Klassenorganisation erreicht wird, bei der jede Klasse nur mit einem Schüler besetzt ist (k=s). Die Kosten erreichten dagegen ihr Minimum für ein Klassensystem, bei dem alle Schüler in einer einzigen Klasse zusammengefaßt sind (vgl. Abschnitt 6.1.2.3). Über die optimale Klassenzahl bei Maximierung des Nettoertrages lassen sich daher keine allgemeinen Aussagen treffen. Um das optimale k zu ermitteln, muß für jedes k der maximale Wert der Funktion (7.3) errechnet werden, indem für jedes k, $k \in \{2,3,\ldots,s-1\}$, ein Lagrange-Ansatz entsprechend Gleichung (7.4) im Hinblick auf die optimalen Klassenbesetzungszahlen maximiert und der Wert der Funktion (7.3) für die resultierenden x_i bzw. c_i errechnet wird. Für k=1 und k=s kann die Funktion (7.3) jeweils nur einen Wert annehmen, so daß sich eine Suboptimierung über den Ansatz (7.4) für k=1 und k=s erübrigt.

7.2 Nettoertragsmaximierung bei Ausbildung mehrerer Kohorten

Die Nettoertragsfunktion lautet in diesem Fall unter Verwendung der Gleichung (5.29) für den Ertrag und der Gleichung (6.8) für die Kosten:

$$(7.5) \qquad z = s\,\gamma\left[\sum_{i=1}^{i=k} c_i\,\frac{(r^{t_d}-r^{t_i})(1-r^{-\tau})}{r^{t_d}(r-1)(r^{t_i}-1)}\right]-\frac{1-r^{\tau}}{r^{\tau}(1-r)}(kv+w).$$

Da sich Gleichung (7.5) bei konstantem r,τ,v,t_d und w für jedes beliebige k auf die in Abschnitt 5.5.1 diskutierte Gleichung (5.21) reduzieren läßt, kann zur Lösung des Problems bei k = konst. auf die Abschnitte 5.5.1 und 5.5.2 verwiesen werden.

Da jeder einzelne Summand in der Ertragsfunktion (5.29) eine strikt monoton sinkende Funktion in der

Größe t_i ist, läßt sich der Ertrag durch Aufteilung einer lückenlos geordneten Klasse in zwei lückenlos geordnete Teilklassen erhöhen. Das Maximum des Ertrages ergibt sich somit für k=s. Für die Kosten erhält man aus der Kostenfunktion (6.8) ein gegensätzliches Ergebnis: Das Kostenminimum wird für k=1 erreicht, d.h. für eine Klasseneinteilung, bei der alle Schüler sich in einer Klasse befinden. Zur Bestimmung der optimalen Klassenzahl muß daher wie im vorangegangenen Abschnitt für jedes k, $k \in \{2,3,\ldots,s-1\}$, der maximale Wert der Funktion (7.5) errechnet werden. Hierfür muß aus (7.5) ein Lagrange-Ansatz analog zu (7.4) gebildet und im Hinblick auf die x_i maximiert werden. Für jedes k, $k \in \{2,3,\ldots,s-1\}$, ergeben sich dadurch optimale x_i bzw. c_i, mit denen der Wert der Funktion (7.5) errechnet werden kann. Für k=1 und k=s läßt sich der Wert der Funktion (7.5) direkt bestimmen.

7.3 Berücksichtigung des Einflusses einer Klassengröße auf die Lernzeit

(a) Bei Ausbildung einer Kohorte.

Die Nettoertragsfunktion bei Abhängigkeit der Lernzeit von der Klassengröße kann aus Gleichung (7.1) dadurch ermittelt werden, daß die Größen t_i durch die Ausdrücke T_i ersetzt werden. Außerdem muß der Term t(s), der der Lernzeit des langsamsten Schülers entspricht, durch die Größe T'

$$T' := \max_{\mu \in \{T_1, T_2, \ldots, T_k\}} \mu$$

substituiert werden. T' gibt dabei ebenfalls die Lernzeit des langsamsten Schülers an, doch muß dies nun nicht mehr die Lernzeit des Schülers s sein, weil die Lernzeit jedes Schülers auch davon abhängt, wie groß die Klasse ist, in der er sich befindet. Durch diese Substitutionen entsteht eine neue Nettoertragsfunktion. Löst man aus dieser neuen Funktion alle Summanden für ein spezielles $i \in \{1,2,\ldots,k\}$ heraus, so bildet die Summe dieser Summanden eine strikt monoton sinkende Funktion in der diesem i entsprechenden Größe T_i. Satz 1 kann somit angewandt werden: die Funktion $g(t_i)$ in Satz 1 muß dafür durch den Term $g(T_i)$ ersetzt werden. Die dafür nötige Voraussetzung, daß T_i ebenso wie t_i eine strikt monoton steigende Funktion in j, $j \in \{1,2,\ldots,s\}$ ist, ist dabei erfüllt, denn durch den in Satz 1 durchgeführten Elemententausch bleibt die in T_i enthaltene Größe a_i konstant. Wendet man Satz 1 auf dieses Problem an, so ergibt sich, daß auch in diesem Fall das nettoertragsmaximale Klassensystem aus lückenlos geordneten Teilmengen der Schülermenge S bestehen muß.

Bei vorgegebenen relativen Klassengrößen und vorgegebener Klassenzahl gibt es wieder k! Klassensysteme, die die Bedingung lückenlos geordneter Klassen und die Nebenbedingungen (3.1) - (3.5) erfüllen. Es muß deshalb wie bisher die optimale Klassenorganisation aus der Menge dieser Systeme durch geeignete Verfahren ausgewählt werden.

Bei vorgegebener Klassenzahl und variablen relativen Klassengrößen kann Gleichung (7.1) als Ausgangspunkt für eine approximative Lösung mittels eines Lagrange-Ansatzes verwendet werden. Hierfür sind die Größen sc_i durch die unbekannten Klassenbesetzungszahlen x_i und die Größen t_i durch die Ausdrücke
$h^*[(x_1, x_2+\ldots+x_i), x_i]$, $i = 1,2,\ldots,k$, zu ersetzen. Bei der Ersetzung des Terms t(s) in (7.1) ergeben sich die gleichen Schwierigkeiten wie bei der Kostenminimierung mit Gleichung (6.6.1): um den Lagrange-Ansatz auch für den Teilausdruck der fixen Kosten formulieren zu können, müßte schon bei der Formulierung bekannt sein, welche Klasse die langsamste sein wird. Es müssen deshalb auch hier k verschiedene Lagrange-Ansätze formuliert und gelöst werden, wobei bei jedem dieser Ansätze hypothetisch eine andere Klasse als die langsamste vorausgesetzt wird: statt t(s) werden alternativ die Ausdrücke $h^*[x_1, x_1]$, $h^*[(x_1+x_2), x_2], \ldots, h^*[(x_1+x_2+\ldots+x_k), x_k]$ verwendet. Aus der Maximierung dieser Lagrange-Ansätze resultieren k Klassensysteme, von denen jedes einzelne darauf überprüft werden muß, ob die hypothetisch angenommene langsamste Klasse diejenige Klasse ist, die sich aus der Lösung des Lagrange-Ansatzes als lang-

samste ergibt. Für den Fall, daß bei mehr als einem der rechnerisch ermittelten Systeme die Hypothese erfüllt ist, ist dann dasjenige System auszuwählen, für das der Wert der Zielfunktion am größten ist.

Bei variabler Klassenzahl muß dieses für ein vorgegebenes k beschriebene Verfahren für jedes $k, k \in \{2,3,\ldots,s-1\}$, wiederholt werden, damit nach der Ermittlung des maximalen Wertes der Zielfunktion für jedes k das optimale System ausgewählt werden kann. Für k=1 und k=s bedarf es dabei keiner Suboptimierung in bezug auf die relativen Klassengrößen nach dem oben beschriebenen Verfahren, weil für k=1 und k=s die Zielfunktion jeweils nur einen Wert annehmen kann.

(b) Bei Ausbildung mehrerer Kohorten.

Ausgangspunkt der Überlegungen ist in diesem Fall die Nettoertragsfunktion (7.5), in der wieder die Größen t_i durch die Ausdrücke T_i zu ersetzen sind. Da der Kostenterm in dieser Gleichung für jedes vorgegebene k eine Konstante ist, kann das Problem für k = konst. auf den bereits diskutierten Fall der reinen Ertragsmaximierung reduziert werden.

Sind die relativen Klassengrößen und die Klassenzahl variabel, so muß zur Bestimmung des optimalen k der maximale Wert der Funktion (7.5) für jedes k errechnet werden. Für k=1 und k=s besitzt die Funktion (7.5) jeweils nur einen einzigen Wert. Für jedes andere $k, k \in \{2,3,\ldots,s-1\}$, muß aus (7.5) ein Lagrange-

Ansatz gebildet werden, der wieder in bezug auf die unbekannten Klassenbesetzungszahlen x_i zu maximieren ist. Optimal ist dasjenige Klassensystem, für das der so ermittelte Wert der Funktion (7.5) maximal ist.

7.4 Zusammenfassung

Als Nettoertragsfunktion wurde die Differenz zwischen den jeweiligen Ertrags- und Kostenfunktionen verwendet. Dabei zeigte sich, daß jedes nettoertragsmaximale System aus lückenlos geordneten Teilmengen der Schülermenge S bestehen muß, und zwar unabhängig davon, ob die Lernzeit von der Klassengröße abhängt oder nicht (Anwendung von Satz 1).

Satz 2 konnte nur noch angewandt werden, wenn die Lernzeit nicht von der Klassengröße abhing. Bei vorgegebener Klassenzahl und vorgegebenen relativen Klassengrößen mußte aus den k! Klassensystemen mit lückenlos geordneten Klassen, die alle die Nebenbedingungen der Klassenorganisation erfüllen, das optimale System aus diesen k! Systemen durch Berechnung des Wertes der Zielfunktion ausgewählt werden. Bei vorgegebener Klassenzahl und variablen relativen Klassengrößen bot sich als Lösungsverfahren die Optimierung der Zielfunktion in Form eines Lagrange-Ansatzes in bezug auf die unbekannten Klassenbesetzungszahlen an. Bei variabler Klassenzahl und variablen relativen Klassengrößen mußte für jedes k, $k \in \{1,2,\ldots,s\}$, der Wert der Zielfunktion isoliert ermittelt werden, um die Auswahl des optimalen Systems treffen zu können. Da für jedes k, $k \in \{2,3,\ldots,s-1\}$, die Zielfunktion je nach den relativen Klassengrößen unterschiedliche Werte annehmen kann, mußte zur Bestimmung des maximalen Wertes der Zielfunktion für jedes dieser k die Zielfunktion in bezug auf die relativen Klassengrößen maximiert werden, wobei der Maximierungsansatz jeweils in der Formulierung einer Lagrange-Funktion bestand.

Resümee

Ausgangspunkt dieser Untersuchung war die Frage, ob sich aus der Hypothese, daß die Lernfähigkeiten bzw. die Lernzeiten einer bestimmten Menge von Schülern statistisch in einer bestimmten Weise verteilt sind, bestimmte Regeln ableiten lassen, die etwas darüber aussagen, auf welche Weise die Schüler zu Klassen zusammengefaßt werden müssen, damit verschiedene ökonomische Zielsetzungen unter Beachtung eines übergeordneten pädagogischen Ziels bestmöglich erfüllt sind. Die optimale Organisation der Schüler in Klassen wurde im Hinblick auf folgende Zielfunktionen untersucht: Minimierung der Lehrerzeit, Minimierung der Schülerzeit, Maximierung der Ausbildungserträge, Minimierung der Ausbildungskosten, Maximierung des Nettoertrages der Ausbildung. Bei allen diesen Zielfunktionen wurden die Nebenbedingungen berücksichtigt, daß (a) jeder Schüler eine vorgegebene Mindest-Erfolgswahrscheinlichkeit nach Abschluß der Ausbildung erreichen müsse und (b) kein Schüler von der Ausbildung ausgeschlossen werden dürfe.

Zur Lösung der verschiedenen Optimierungsprobleme wurden 2 Sätze bewiesen. Aus Satz 1 ergab sich, daß die notwendige Bedingung des Optimums bei allen Zielfunktionen darin besteht, die Klassen so zu bilden, daß sie lückenlos geordnete Teilmengen derjenigen

Menge darstellen, deren Elemente durch die Lernzeiten sämtlicher Schüler gebildet werden. Der Satz 2 diente dazu, das optimale Klassensystem bei alternativen Zielfunktionen und vorgegebenen relativen Klassengrößen, sowie einer vorgegebenen Anzahl von Klassen zu bestimmen.

Die Ableitung des optimalen Klassensystems wurde bei den einzelnen Zielfunktionen daraufhin überprüft, ob andere Ergebnisse zu erwarten sind, wenn angenommen wird, daß die Lernzeiten der Schüler eine Funktion der Klassengröße sind. Diese Erwartung traf in vielen Fällen zu.

Die Methode der Ableitung bestand bei vorgegebener Klassenzahl und variablen relativen Klassengrößen darin, daß die Zielfunktion in der Form eines Lagrange-Ansatzes formuliert wurde, der als Nebenbedingung die Forderung enthielt, daß kein Schüler vom Unterrichtsprozeß ausgeschlossen werden dürfe. Die Optimierung dieses Ansatzes in bezug auf die unbekannten Klassenbesetzungszahlen mit den Mitteln der partiellen Differentiation lieferte eine approximative Lösung des Problems.

Bei vorgegebener Klassenzahl und vorgegebenen relativen Klassengrößen ergab sich die Lösung in den Fällen, in denen die Lernzeit nicht von der Klassengröße abhing, aus der Anwendung der Sätze 1 und 2. In den übrigen Fällen gab es bei k Klassen jeweils k! Klassensysteme, von denen jedes die Nebenbedingungen erfüllte. Das optimale System kann aus der Menge der k! Systeme beispielsweise dadurch ausgewählt werden, daß die Aufgabe als Zuordnungsproblem interpretiert und mit den Mitteln der ganzzahligen linearen Programmierung gelöst wird.

Bei variabler Klassenzahl und variablen relativen Klassengrößen ergaben sich u.a. die extremen Lösungen, daß alle Schüler in einer Klasse zusammengefaßt werden müssen (Kostenminimierung), oder daß für jeden Schüler eine Klasse gebildet werden muß (Ertragsmaximierung). Bei der Nettoertragsmaximierung mußte zur Bestimmung der optimalen Klassenzahl der Wert der Zielfunktion für alternative Klassenzahlen errechnet werden. Da dieser Wert bei gleicher Klassenzahl nicht unabhängig von den relativen Klassengrößen ist, mußte hierbei für jede Klassenzahl eine Suboptimierung in bezug auf alternative Klassenbesetzungszahlen durchgeführt werden.

Eine besondere Erwähnung verdient das Ergebnis, daß entgegen der intuitiven Vorstellung die Ausbildungskosten mit sinkender Klassenzahl nicht notwendig sinken und die Ausbildungserträge mit steigender Klassenzahl nicht notwendig steigen müssen, obwohl das Minimum der Kosten bei einem Klassensystem erreicht wird, bei dem alle Schüler sich in einer Klasse befinden und das Maximum der Erträge dann realisiert wird, wenn für jeden Schüer eine Klasse gebildet wird.

Literaturverzeichnis

Arrow, K.J.	"The Economic Implications of Learning by Doing", in: Review of Economic Studies, Vol. 29, 1961-62, S. 155-173
Asher, H.	"Cost Quantity Relationship in the Airframe Industry", Projekt RAND Paper R-291, RAND Corp., Santa Monica (Calif.), Juli 1956
Blaug, M.	"Approaches to Educational Planning", in: Economic Journal, Bd. 77, 1967, S. 262-287
Bruner, J.S.	"Some Theorems on Instruction with Reference to Mathematics", in: 'Theories of Learning and Instruction', The Sixty-third Yearbook of the National Society for the Study of Education, Part I, Chicago: University Press, 1964, S. 306-335
Bugelski, R.R.	"The Psychology of Learning Applied to Teaching", New York 1964
Bush, R.R. u. Mosteller, F.	"Stochastic Models for Learning", New York 1955
Bush, R.R u. Sternberg, S.H.	"A Single-Operator Model", in: 'Studies in Mathematcal learning Theory', Hrsg.: Bush, R.R. u. Estes, W.K., Stanford: University Press, 1959, S. 204-214
Champion, R.A.	"Learning and Activation", New York 1969
Correa, H.	"A Survey of Mathematical Models in Educational Planning", in: 'Mathematical Models in Educational Planning', OECD: Education and Development, Technical Reports, Paris 1967, S. 21-93
Dantzig, G.B.	"Lineare Programmierung und Erweiterungen", Übersetzung von Jaeger, A., Heidelberg 1966
Das, J.P.	"Verbal Conditioning and Behaviour", New York 1969

Deese, J.	"The Psychology of Learning", New York 1952
Deutscher Bildungsrat	"Einrichtung von Schulversuchen mit Gesamtschulen", Empfehlungen der Bildungs-Kommission, Bad Godesberg 1969
Estes, W.K.	"The Statistical Approach to Learning Theory", in: 'Psychology: A Study of a Science', Hrsg.: Koch, S., New York 1959
Estes, W.K. u. Suppes, P.	"Foundations of Linear Models", in: 'Studies in Mathematical Learning Theory', Hrsg.: Bush, R.R. u. Estes, W.K., Stanford (Calif.): University Press, 1959, S. 137-179
Estes, W.K., Hopkins, B.L. u. Crothers, E.J.	"Al-Or-None and Conservation Effects in the Learning and Retentation of Paired Associates", in: Journal of Experimental Psychology, Vol. 60, No. 6, Dez. 1960, S. 329-339
Feichtinger, G.	"Stochastische Automaten als Grundlage linearer Lernmodelle", in: Statistische Hefte, N.F., Heft 1, 10.Jg., 1969, S. 5-21
-ders.	"Über ein Entscheidungsproblem in der Erziehungsplanung", in: Unternehmensforschung, Nr. 14, Heft 1, 1970, S. 140-147
Gody, C.S.	"Productivity Changes in Selected Wartime Shipbuilding Programs", in: Monthly Labor Review, Vol. 61, Dez. 1945
Harman, H.H.	"Modern Factor Analysis", Chicago 1960
Hilgard, E.R.	"Theories of Learning", London 1958
Hilgard, E.R. u. Marquis	"Conditioning and Learning", New York 1961
Hill, W.F.	"Learning, A Survey of Psychological Interpretations", London 1965

Hirsch, W.Z. — "Firm Progress Ratios", in: Econometrica, Vol. 24, April 1956, S. 136-143

-ders. — "Manufacturing Progress Functions", in: Review of Economics and Statistics, Vol. 34, Mai 1952, S. 143-155

Holzinger, K.J. — "Factor Analysis", Chicago 1941

-ders. — "Interpretation of Second-order Factors", in: Psychometrica, Vol. 10, 1945, S. 21-25

Hüfner K. u. Naumann, J. — "Die Problematik von Kosten/Ertragsvergleichen alternativer Schulsysteme - Ein Diskusionsbeitrag", in: 'Theorie u. Praxis der Infrastrukturpolitik', Hrsg.: Jochimsen, R. u. Simonis, U.E., Berlin 1970, S. 89-105

Kimble, G.A. — "Foundations of Conditioning and Learning", New York 1967

La Berge, D.L. — "The Effect of Preliminary Trials on Rate of Conditioning in a Simple Prediction Situation", in: Journal of Experimentical Psychology, No. 57, 1959, S. 20-24

Lawson, R. — "Learning and Behavior", New York 1960

Miller, G.A. — "Finite Markov Processes in Psychology", in: Psychometrica, No. 17, 1952, S. 149-167

Menzel, W. — "Theorie der Lernsysteme", Berlin 1970

Mowrer, O.H. — "Learning Theory and Behavior", New York 1960

Parreren, v.C.F. — "Lernprozeß und Lernerfolg", Braunschweig 1966

Restle, F. — "A Theory of Discrimination Learning", in: Psychological Review, No. 62, 1955, S. 11-19

Lebenslauf

Am 4.1.1939 wurde ich in Heufeld, Banat(Jugoslawien), geboren. Wie alle deutschstämmigen Familien des Banats wurde auch unsere Familie nach Kriegsende mehrere Jahre in jugoslawischen Lagern interniert. Nach der Flucht aus einem Internierungslager kamen wir nach zweijährigem Aufenthalt im Flüchtlingslager Linz in Österreich 1949 nach Deutschland. Hier besuchte ich erstmals in Wasseralfingen, Württemberg, eine Grundschule. Im Aufbaugymnasium Künzelsau, einem Internat, legte ich 1959 die Reifeprüfung ab.
Vom Wintersemester 1959/60 bis zum Wintersemester 1961/62 studierte ich an der Technischen Hochschule Stuttgart Flugzeugbau und an der Hochschule für Gestaltung in Ulm Publizistik. Im Sommersemester 1962 begann ich das Studium der Volkswirtschaftslehre an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Freien Universität Berlin. An dieser Fakultät erwarb ich im Sommer 1967 das Diplom für Volkswirte. Nach der Diplomprüfung begleitete ich zwei Semester lang als Tutor die Vorlesungen von Prof. Dr. Kurt Elsner. Seit Oktober 1968 bin ich wissenschaftlicher Mitarbeiter am Deutschen Institut für Wirtschaftsforschung, Berlin. Als Gastdozent an der Fachhochschule für Wirtschft, Berlin, halte ich Vorlesungen über Allgemeine Volkswirtschaftslehre .