Algorithmen und zufällige Folgen

Vier Vorträge von

P e r M a r t i n - L ö f
(Stockholm)

gehalten am Mathematischen Institut der
Universität Erlangen-Nürnberg

Als Manuskript vervielfältigt

Erlangen, 1966

Errata

S. 2.3	Z. 4	statt: x	lies:	a
S. 2.11,	Z. 1	statt: $Z_{x\alpha}$ / Z_x	lies:	Z_x / $Z_{x\alpha}$
S. 2.12,	Z. 14	statt: ... und P eine Produktion...	lies:	... und y eine Produktion...
S. 2.13,	Z.2v.u.	statt: Ez	lies:	Bz
S. 2.14,	Z. 2	statt: Ez	lies:	Bz
S. 4.5	Z.5v.u.	statt: $0 < N \leqq N_1$	lies:	$0 < n \leqq N_1$

Vorbemerkung

Am 5., 6., 14. und 15. April 1966 hielt Herr Per M a r t i n - L ö f (Stockholm) 4 Vorträge am Mathematischen Institut der Universität Erlangen, Gegenstand der Vorträge waren neuere, z.T. noch unveröffentlichte Untersuchungen von Kolmogoroff und Martin-Löf, die u.a. eine Rechtfertigung der Ideen von von Mises zur Grundlegung der Wahrscheinlichkeitstheorie enthalten. Herr Martin-Löf berichtete eingehend über die Geschichte des gesamten Fragenkreises und gab auch eine kurze Einführung in den gegenwärtigen Untersuchungen zugrundeliegenden Zweig der mathematischen Logik. Um einem größeren Kreis von Kollegen und Kommilitonen den Zugang zu diesen Dingen zu erleichtern, haben wir die Vorträge ausgearbeitet. Daß der hiermit vorgelegte Text nicht annähernd die Lebendigkeit der Vorträge wiedergibt, liegt natürlich an uns. Wir sind auch durch leidlich konsequente Verwendung der mengentheoretischen Schreibweise von Herrn Martin-Löf's Darstellung abgewichen. Es sollte also klar sein, daß ein von Herrn Martin-Löf selbst geschriebener Text anders aussähe als dieser. Ebenso gehen alle Fehler zu unseren Lasten. Herr Martin-Löf hatte die große Freundlichkeit, unsere Manuskripte mit uns durchzusprechen. Wir möchten ihm nochmals wärmstens danken.

Erlangen, den 16. April 1966

K. Jacobs
W. Müller

Inhalt

§1. Die ursprünglichen Ideen von von Mises und Kolmogoroff.

Wir wollen versuchen, Wahrscheinlichkeitstheorie im Raum

$$\Omega = \{\omega=(x_1, x_2,\ldots) \mid x_t = 0 \text{ oder } 1,\ t = 1,2,\ldots\}$$

(auf den wir uns hier der Einfachheit halber beschränken) zu treiben.

Es gibt zwei verschiedene Vorschläge, diesen Versuch auf eine exakte Grundlage zu stellen.

1. Der von Mises'sche Vorschlag

legt den Begriff der zufälligen Folge $\omega \in \Omega$ (Kollektiv) zugrunde. Eine Folge $\omega = (x_1, x_2,\ldots) \in \Omega$ heißt zufällig (ein Kollektiv), wenn sie folgende beide Axiome erfüllt:

Axiom I: Die relativen Häufigkeiten des Auftretens von 0 bzw. 1 in ω existieren.

Axiom II: Diese relativen Häufigkeiten ändern sich nicht, wenn man vermöge einer Auswahlregel zu einer Teilfolge von ω übergeht.

Die Aufgabe der Wahrscheinlichkeitstheorie besteht nach von Mises darin, aus gegebenen Kollektivs neue Kollektivs herzuleiten, und die zugehörigen relativen (Limes-) Häufigkeiten, die dann Wahrscheinlichkeiten genannt werden, zu bestimmen.

In Axiom II steckt der Begriff 'Auswahlregel', den von Mises bei der in den 30er Jahren geführten Diskussion nicht völlig mathematisch präzisieren konnte. Er konnte es im Wesentlichen deshalb nicht, weil die formale Logik damals noch nicht weit genug entwickelt bzw. bekannt war.

Immerhin konnte A. Wald [4] , [13] eine Präzisierung in folgendem Sinne geben: Er definierte exakt so etwas wie 'sequentielle Auswahlregeln' und den Begriff des 'Kollektivs bezüglich eines gegebenen Systems W solcher Auswahlregeln' und konnte einen Satz beweisen, der besagt, daß es zu einer vorgeschriebenen Wahrscheinlichkeit p für das Auftreten von '1' und einem abzählbaren System W von Auswahlregeln stets kontinuierlich-viele Kollektivs bezüglich W gibt, für die die relative Häufigkeit der Einsen gleich p ist.

Die Diskussion von Axiom II wurde 1939 von Ville [12] durch ein Beispiel vorläufig beendet, das zeigte: Wählt man (z.B.) $p = \frac{1}{2}$, so gibt es zu jedem abzählbaren System W von sequentiellen Auswahlregeln ein Kollektiv $\omega = (x_1, x_2, \ldots)$ bezüglich W mit $\frac{1}{2}$ als relative Häufigkeit von '1', derart, daß die finiten relativen Häufigkeiten der 1, d.h. die Ausdrücke

$$\frac{1}{t} \sum_{u=1}^{t} x_u$$

stets $\geq \frac{1}{2}$ sind (sie haben natürlich für $t \longrightarrow \infty$ gegen den Wert $\frac{1}{2}$ zu konvergieren).

Eine solche Folge entspricht nicht der intuitiven Vorstellung von einer 'zufälligen Folge', sie hat eine 'Vorliebe für 1'. Exakt kann man z.B. sagen: sie genügt nicht dem Gesetz vom iterierten Logarithmus.

Dies Beispiel bedeutet, daß man die von Mises'schen Ideen noch beträchtlich weiterentwickeln muß, um zu einer befriedigenden Theorie zu gelangen. Als vernünftiges Ziel erscheint die mathematische Präzisierung folgender intuitiven Defintion: Eine Folge $\omega \in \Omega$ heißt zufällig, wenn sie allen Fastüberallgesetzen der Wahrscheinlichkeitstheorie genügt. - Offenbar hatte von Mises nur das starke Gesetz der großen Zahl, sowie Sätze über das Nichtvorhandensein von Spielsystemen ins Auge gefaßt, dagegen Sätze wie den vom iterierten Logarithmus außer Betracht gelassen.

2. Der Kolmogoroff'sche Vorschlag von 1933

legt den Begriff der Wahrscheinlichkeit eines Ereignisses zugrunde. Dies hat den Vorteil, daß sich die Wahrscheinlichkeitstheorie in eine exakte mathematische Theorie - die Theorie der Mengenfunktionen, mit der Maßtheorie als hochausgebildetem Hauptzweig - einbauen läßt. Unklare mathematische Begriffe treten nicht auf. Dagegen hat man eine (bewußte) Unschärfe bei der Interpretation von Wahrscheinlichkeiten. Auf Seite 4 von Kolmogoroff [5] findet sich der Satz:

'Unter gewissen Bedingungen ... kann man voraussetzen, daß einem Ereignis A, welches infolge der Bedingungen σ auftritt oder nicht, eine gewisse reelle Zahl P(A) zugeordnet ist, welche folgende Eigenschaften besitzt:

A. Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen σ eine große Anzahl von n Malen wiederholt und dabei durch m die Anzahl der Fälle bezeichnet, bei denen das Ereignis A stattgefunden hat, das Verhältnis m/n sich von P(A) nur wenig unterscheidet.

B. Ist P(A) sehr klein, so kann man praktisch sicher sein, daß bei einer einmaligen Realisation der Bedingungen σ das Ereignis A nicht stattfindet.'

Der Kolmogoroff'sche Ansatz von 1933 liefert insbesondere keine Präzisierung der obigen intuitiven Definition des Begriffs 'zufällige Folge': Bei jeder Anwendung eines Fastüberall-Gesetzes der Wahrscheinlichkeitstheorie tritt eine Nullmenge $\subseteq \Omega$ als Ausnahmemenge auf; die Vereinigung aller derartigen Nullmengen ist aber Ω, wenn die zugrundegelegte Wahrscheinlichkeitsverteilung punktmassenfrei ist.

3. Ältere Literatur

von Mises hat einen Teil seiner Ideen bereits in [7] (1919) geäußert. Eine zusammenfassende Publikation ist v. Mises' Buch [8] (1936). Zu beachten sind die Beiträge der Genfer Tagung von 1937 ([4]) mit Arbeiten

von Feller, Fréchet, v. Mises u.a., sowie die Arbeiten von Wald [4], [13] und Ville [12].

Als Vorläufer kann man die Arbeit von Borel [1] (1909), als ersten Beitrag von logischer Seite die Arbeit von Church [2] (1940) ansehen.
Vgl. ferner Tornier [11], Copeland [3], Popper [9], Reichenbach [10].

4. Ziel dieser Vorlesung ist ein Bericht über eine neue Definition des Begriffs 'zufällige Folge', die Kolmogoroff neuerdings vorgeschlagen hat. Sie beruht auf der von verschiedenen Logikern in verschiedenen äquivalenten Fassungen entwickelten Theorie der Algorithmen. In § 2 wird daher ein Abriß einer dieser Fassungen gegeben.

Literatur zu § 1

[1] B o r e l , E.
Les probabilités dénombrables et leurs applications arithmétiques
Rend. Circ. Mat. Palermo 27 (1909), 247 - 271
abgedruckt als Note V zu:
Leçons sur la Théorie des Fonctions
(Paris 1928 bei Gauthier-Villars)

[2] C h u r c h , A.
On the concept of a random sequence
Bull. Amer. Math. Soc. 46 (1940), 130 - 135

[3] C o p e l a n d , A. H.
Admissable numbers in the theory of probability
Amer. J. Math. 50 (1928), 535 - 552

[4] Actualités Scientifiques et Industrielles 735
(Colloque consacré à la Théorie des Probabilités et présidé par M.M. F r é c h e t , organisé à l'université de Genève, 1938, 2ième partie)

[5] K o l m o g o r o f f , A.N.
Grundbegriffe der Wahrscheinlichkeitsrechnung
Springer, Berlin 1933

[6] ----, On the tables of random numbers
Sankhya 25 (1963), 369 - 376

[7] von M i s e s , R.
Grundlagen der Wahrscheinlichkeitsrechnung
Math. Z. 5 (1919), 52 - 99

[8] ----, Wahrscheinlichkeit, Statistik und Wahrheit,
Springer, Wien 1936

[9] P o p p e r , K.
Logik der Forschung
Springer, Wien 1935

10. R e i c h e n b a c h , H.
Wahrscheinlichkeitslehre
Sijthoff, Leiden 1935

11. T o r n i e r , E.
Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie
Teubner, Leipzig 1936

12. V i l l e , J.
Etude critique de la notion de collectif
Gauthier-Villars, Paris 1939

13. W a l d , A.
Die Widerspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung
Ergebnisse eines mathematischen Kolloquiums 8 (1937), 38 - 72

§2. Rekursiv aufzählbare Mengen

Dieser Abschnitt enthält Definitionen und Sätze aus der Theorie der rekursiv aufzählbaren Mengen, auf die sich die späteren Ausführungen stützen werden.

Es gibt mehrere im wesentlichen äquivalente Theorien der rekursiven Aufzählbarkeit, die auf Herbrand, Gödel, Church, Kleene, Turing, Post und Markow zurückgehen. Die hier gegebene Darstellung folgt im wesentlichen den Ideen von Post; eine zentrale Rolle spielen die Begriffe 'rekursiv aufzählbare Menge', 'kanonischer Kalkül' (vgl. [1],[2]).

1. Beispiele rekursiv aufzählbarer Mengen

Intuitiv stellt man sich unter einer rekursiv aufzählbaren Menge eine Menge vor, deren Elemente sich in irgendeiner Reihenfolge (Wiederholung ist erlaubt) nach einem festgelegten Verfahren (Algorithmus) aus endlichem Grundmaterial erzeugen lassen; dabei dürfen schon erzeugte Elemente bei der Erzeugung weiterer Elemente mitbenützt werden - daher das Wort 'rekursiv'. Wir verdeutlichen diese Vorstellung durch einige Beispiele:

Beispiel 1 (alternierende Wörter).

Seien α, ß zwei verschiedene Zeichen. Endliche Folgen, die mit Hilfe dieser Zeichen gebildet sind, heißen Wörter über α, ß, und werden auch mit a,b,... bezeichnet. Das von α, ß verschiedene Verlegenheitssymbol $\square$ bezeichnet das leere Wort, bzw. eine Leerstelle und kann, wenn keine Verlegenheit besteht, auch weggelassen werden; $a\alpha$ bezeichnet das aus dem Wort a durch Anhängen von α entstehende verlängerte Wort etc.; insbesondere ist stets $a\square = a = \square a$. Die Menge

$$R = \{\square, \alpha, ß, \alpha ß, ß\alpha, \alpha ß \alpha, ß\alpha ß, \ldots\}$$

aller alternierenden Wörter über $\{\alpha, ß\}$ läßt sich nach dem durch folgende drei Regeln festgelegten Verfahren aus dem Grundmaterial $\{\alpha, ß\}$ erzeugen:

A1) $\square$, α und ß sind alternierende Wörter.

A2) Ist $x\alpha$ ein alternierendes Wort, so ist $x\alpha ß$ ein alternierendes Wort.

A3) Ist xß ein alternierendes Wort, so ist $xß\alpha$ ein alternierendes Wort.

Beispiel 2 (die natürlichen Zahlen).

Sei α ein Symbol. Die Menge

$$N = \{\square, \alpha, \alpha\alpha, \alpha\alpha\alpha, \ldots\},$$

die man als die Menge der natürlichen Zahlen ansehen kann, läßt sich aus dem endlichen Grundmaterial $\{\alpha\}$ nach den folgenden beiden Regeln erzeugen:

N1) □ ist eine natürliche Zahl.

N2) Ist x eine natürliche Zahl, so ist xα eine natürliche Zahl.

Beispiel 3 (die Vielfachen einer natürlichen Zahl x).

Sei a = $\alpha \ldots \alpha \in N$ (auch a = □ ist zugelassen) eine natürliche Zahl.

Die Menge

$$aN = \{ \square , a, aa, aaa, \ldots \}$$

der natürlichen Vielfachen von a läßt sich aus dem endlichen Grundmaterial {a} nach folgenden Regeln erzeugen:

aN1) □ ist ein Vielfaches von a

aN2) Ist y ein Vielfaches von a, so ist ya ein Vielfaches von a.

2. Formalisierung des Begriffs der rekursiv aufzählbaren Menge.

Geht man die in den Beispielen aus Nr. 1 angegebenen Bildungsregeln durch, so stellt man fest, daß sie aus

Axiomen, wie z.B. '□ ist eine natürliche Zahl'

und

Ableitungsregeln (Produktionen) wie z.B.

'Ist xα eine alternierende Folge, so ist auch x$\alpha\beta$ eine alternierende Folge'

bestehen.

Wir formalisieren jetzt diesen Tatbestand.

Sei A eine endliche nichtleere Menge, die wir als Alphabet bezeichnen, und

$$A^0 = \{\square\}$$

$$A^1 = \{a_1 \ldots a_1 \mid a_1, \ldots, a_1 \in A\}$$

die Menge der Wörter der Länge 1 über A (also das Kartesische Produkt von 1 Exemplaren von A) und

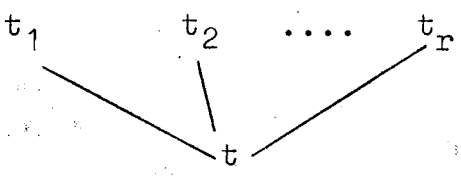
$$\bar{A} = \bigcup_{1=0}^{\infty} A^1$$

die Menge aller endlichen Wörter über A.

Wir wollen den Begriff 'rekursiv aufzählbare Teilmenge von $\bar{A}$' exakt definieren. Hierzu werden zunächst einige einfachere Begriffe erklärt.

Definition 2.1

Sei $V = \{x, y, \ldots, z\}$ eine endliche, zu A disjunkte Menge, deren Elemente wir als Variable bezeichnen; wir werden sie für das später anzuwendene Einsetzverfahren als 'Leerstellen' benötigen; Wörter über A + V (disjunkte Vereinigungen werden auch durch + bezeichnet), nennen wir auch Terme. Eine endliche evtl. leere Serie $t_1, \ldots, t_r$ $(r \geq 0)$ von Termen und ein einziger Term t aus A + V bilden zusammen eine Produktion (über A, mit Variablen aus V). Wir schreiben sie



Man nennt $t_1, \ldots, t_r$ auch die Prämissen, t die Conclusio dieser Produktion. Ist keine Prämisse vorhanden, so wird die Produktion einfach

t

geschrieben und auch als Axiom bezeichnet.

Speziell kann $\square$ ein Axiom sein.

Definition 2.2

Eine endliche Menge von Produktionen über A wird als ein kanonischer Kalkül über A (mit Variablen in der Vereinigung der bei den Produktionen auftretenden Variablenmengen) bezeichnet. Ein Wort a aus $\bar{A}$ (d.h. über A) heißt in diesem Kalkül ableitbar, wenn es sich mittels dessen Produktionen nach dem Einsetzungsverfahren ableiten läßt.

Statt den jedermann geläufigen Begriff 'Einsetzungsverfahren' hier formal komplett zu definieren, geben wir ein Beispiel an:

Beispiel 1. (Fortsetzung)

Das Wort $\alpha\beta\alpha\beta\alpha$ über $A = \{\alpha, \beta\}$ ist in dem durch die Axiome $\square, \alpha$, β und die weiteren Produktionen

$$\begin{array}{c} x\alpha \\ | \\ x\alpha\beta \end{array} \qquad\qquad \begin{array}{c} x\beta \\ | \\ x\beta\alpha \end{array}$$

gegebenen kanonischen Kalkül folgendermaßen ableitbar

Substitution		$x = \square$	$x = \alpha$	$x = \alpha\beta$	$x = \alpha\beta\alpha$
Produktion	α	$\begin{array}{c} x\alpha \\ \vert \\ x\alpha\beta \end{array}$	$\begin{array}{c} x\beta \\ \vert \\ x\beta\alpha \end{array}$	$\begin{array}{c} x\alpha \\ \vert \\ x\alpha\beta \end{array}$	$\begin{array}{c} x\beta \\ \vert \\ x\beta\alpha \end{array}$
Abgeleitetes Wort	α	$\alpha\beta$	$\alpha\beta\alpha$	$\alpha\beta\alpha\beta$	$\alpha\beta\alpha\beta\alpha$

Die Menge der Variablen ist $V = \{x\}$.

Definition 2.3

Eine Teilmenge S von $\bar{A}$ heißt rekursiv aufzählbar (= effektiv aufzählbar) über A, wenn es einen kanonischen Kalkül über einem Alphabet $B \supseteq A$ (mit endlicher Variablenmenge V) gibt, derart, daß S genau die Menge aller in diesem Kalkül ableitbaren Wörter über A ist. Man sagt dann, der Kalkül erzeugt S.

Theorem 2.1 Sind S, T rekursiv aufzählbare Mengen über A, so sind auch $S \cap T$ und $S \cup T$ rekursiv aufzählbare Mengen über A.

Beweis-Skizze. Man schreibe die kanonischen Kalküle für S und T hin und sorge durch Verwenden von Indices (z.B. 'S' und 'T') dafür, daß sie disjunkt sind: Alle im Kalkül für S vorkommenden Wörter erhalten die Variable S vorangestellt, so daß man jetzt mit ihm nur mehr Wörter der Gestalt Sx ableiten kann. Ebenso verfährt man mit T. Den aus diesen beiden disjunkten Kalkülen bestehenden Kalkül über A mit um {S, T} erweiterten Variablenmengen hat man jetzt nur noch um die

1) Produktion $\begin{array}{ccc} Sx & & Tx \\ & \diagdown\ \diagup & \\ & x & \end{array}$ zu ergänzen,

um den Kalkül für $S \cap T$ zu erhalten.

2) Produktionen $\begin{array}{c} Sx \\ | \\ x \end{array}$ und $\begin{array}{c} Tx \\ | \\ x \end{array}$

zu ergänzen, um den Kalkül für $S \cup T$ zu erhalten.

Die neuen Symbole S, T gehören in die Menge $B \supseteq A$ für den neuen Kalkül (Definition 2.1).

Wir werden später eine rekursiv aufzählbare Menge S (über einem geeigneten Alphabet A) angeben, deren Komplement $S^c = \bar{A} - S$ nicht rekursiv aufzählbar ist, und dabei einen Satz gewinnen, der zum Gödel'schen Unentscheidbarkeitssatz analog ist. Dies bedeutet, daß die folgende Definition nichttrivial ist.

Definition 2.4

Eine Teilmenge S von $\bar{A}$ heißt rekursiv (= entscheidbar), wenn sowohl S als auch $S^c = \bar{A} - S$ rekursiv aufzählbar sind.

Wenn $S \subseteq \bar{A}$ rekursiv ist, so kommt jedes Element $a \in \bar{A}$ entweder beim Ableiten von S oder beim Ableiten von S^c einmal vor; es wird also einmal entschieden, ob a zu A gehört oder nicht. Allerdings läßt sich bei gegebenem a i.a. nicht voraussagen, bis wann die Entscheidung sicher gefallen ist.

3. Der Hauptsatz der Theorie der rekursiv aufzählbaren Mengen

befaßt sich mit den rekursiv aufzählbaren Teilmengen der aus dem ein-elementigen Alphabet $A = \{\alpha\}$ gebildeten Menge

$$\bar{A} = N = \{\square, \alpha, \alpha\alpha, \alpha\alpha\alpha, \ldots\}$$

aller natürlichen Zahlen, die wir auch mit m, n,... bezeichnen werden.

Das kartesische Produkt N×N kann man als Teilmenge von $\bar{B}$ auffassen, wobei $B = \{\alpha, \beta\}$ ist:

N × N ist: Man identifiziere das Paar $(x, y) \in N \times N$ mit dem Wort $x\beta y \in \bar{B}$. Somit hat es einen Sinn, von rekursiv aufzählbaren Teilmengen von N × N zu sprechen. Man kann N × N auch mit N identifizieren (diagonale Durchzählung).

Der uns interessierende Hauptsatz lautet nun:

Theorem 2.2 Es gibt eine rekursiv aufzählbare Teilmenge U von N × N, derart, daß die Menge der 'Schnitte'

$$U_m = \{n \mid n \in N, (m, n) \in U\} \qquad (m \in N)$$

mit der Menge der rekursiv aufzählbaren Teilmengen von N übereinstimmt. Das bedeutet: Läuft m durch N, so durchläuft U_m genau das System aller rekursiv aufzählbaren Teilmengen von N: jede dieser Teilmengen kommt mindestens einmal als ein U_m vor, und jedes U_m ist so eine Teilmenge.

Anmerkung:
Ist U irgendeine rekursiv aufzählbare Teilmenge von N × N, so ist jeder ihrer Schnitte U_m eine rekursiv aufzählbare Teilmenge von N; man gewinnt einen U_m erzeugenden Kalkül, indem man dem U erzeugenden Kalkül die Produktion

$$\begin{array}{l} m\ \beta\ x \\ | \\ x \end{array}$$

hinzufügt.

Beweis. Der exakte Beweis dieses Theorems besteht in der Formalisierung folgender Idee:

Man schreibe sämtliche Kalküle über Alphabeten $\supseteq \{\alpha\}$ hin und zeige, daß man sie in 'rekursiv aufzählbarer Weise' numerieren kann; die Nummer eines Kalküls nennt man dann auch seine 'Gödel-Nummer'. Dann baue man die so numerierten Einzelkalküle zu einem Gesamtkalkül zusammen, der eine Teilmenge U von N × N erzeugt, wobei U_m gerade die vom Kalkül mit der Gödel-Nummer m erzeugte Teilmenge von N ist.
Bei der exakten Durchführung dieses Programms ergibt sich zunächst die Aufgabe, dafür zu sorgen, daß der Gesamtkalkül mit endlichvielen Variablen auskommt, obwohl er aus abzählbarvielen Einzelkalkülen zusammgenbaut ist.
Dies macht man, indem man die in den Einzelkalkülen auftretenden, evtl. insgesamt unendlichvielen Variablen nicht mit x, y,..., sondern mit ξ_0, ξ_1, ξ_2,... bezeichnet und die Indexfolge 0, 1, 2,... mit einem Symbol ß aufbaut:

$$\xi,\ \xi_{ß},\ \xi_{ßß},\ \xi_{ßßß},\ldots .$$

Ferner schreiben wir die Indices auf die Zeile:

$$\xi,\ \xi ß,\ \xi ßß,\ldots .$$

Ebenso verfahren wir mit den zur Formalisierung der Einzelkalküle benötigten Buchstaben, die dort zu den in Definition 2.1 - als A bezeichneten Mengen gehören.
Wir bauen sie in der Form α, $\alpha ß$, $\alpha ßß$, $\alpha ßßß$,... auf.
Aus Formalisierungsgründen wollen wir den Pfeil ----> als neues Symbol ins Alphabet aufnehmen und Produktionen

$$t_1 \quad t_2, \quad \ldots \quad t_r \longrightarrow t$$

aus den Einzelkalkülen in der neuen Gestalt

$t_1 \dashrightarrow t_2 \dashrightarrow \ldots \dashrightarrow t_r \dashrightarrow t$

schreiben. Axiome behalten also ihre alte Gestalt. Um nebeneinanderstehende Produktionen gegeneinander abzusetzen, benützen wir das - ebenfalls ins Alphabet aufzunehmende - Symbol $*$. Ein kanonischer (Einzel-)Kalkül wird nunmehr einfach als ein Wort über dem Alphabet $\{\alpha, \xi, ß, \dashrightarrow, *\}$ aufgefaßt. Diese Wörter kann man lexikographisch anordnen und dadurch mit den natürlichen Zahlen identifizieren. Wir haben damit die kanonischen Einzelkalküle mittels eines recht kleinen Alphabets einheitlich hingeschrieben. Natürlich gibt es unter ihnen viele, die nichts produzieren und somit die leere Menge $\phi \subseteq N$ erzeugen.

Für den zu konstruierenden Gesamtkalkül benötigen wir nun ein größeres Alphabet, nämlich

$\{\alpha, \xi, ß, \dashrightarrow, *, ?, Z, E, C, F, G, L, B, P, X, T, V, W\}$,

sowie die Variablenmenge

$\{t, a, v, b, x, y, z\}$.

Nun sind einige 30 Produktionen nötig. Wir schreiben sie mitsamt ihren Interpretationen hin, u.z. in der alten vertikalen Schreibweise (die horizontale wird nur für die Einzelkalküle verwendet).

1) Die Einzelkalküle sollen sich im Gesamtkalkül wiederfinden.

a) Wir verwenden das Symbol Z für 'ein Wort über $A = \{\alpha\}$', lesen Z x also: 'x ist ein Wort über A'.
Unser Gesamtkalkül enthält nun u.a. die Produktionen

$$Z\square, \qquad \begin{array}{l} Zx\alpha \\ | \\ Zx \end{array}$$

mit deren Hilfe man $\square, \alpha, \alpha\alpha, \ldots$ ableitet.

b) Wir verwenden das Symbol V für 'Variable'. Aus den Produktionen

$V\xi$ ('ξ ist eine Variable')

Vx

$Vx\beta$

des Gesamtkalküls leiten wir alle in den Einzelkalkülen benötigten Variablen $\xi, \xi\beta, \xi\beta\beta, \ldots$ ab.

c) Wir verwenden das Symbol L für 'Buchstabe' (letter), d.h. 'Element von $\{\alpha, \alpha\beta, \alpha\beta\beta, \ldots\}$'. Diese Menge wird mittels der Produktionen

$L\alpha$ ('α ist ein Buchstabe')

Lx

$Lx\beta$

erzeugt.

d) Um die aus Buchstaben zusammengesetzten Wörter zu erhalten, führen wir die Produktionen (W wird als 'ist ein Wort' gelesen)

$W\square$ ('$\square$ ist ein Wort')

$$\begin{array}{ll} Lx & Wy \\ & \diagdown\ \diagup \\ & Wxy \end{array}$$

ein.

e) Um die aus Buchstaben und Variablen zusammengesetzten sog. 'Terme' zu bekommen, benützen wir das Symbol T ('ist ein Term') und die Produktionen

T □ ('□ ist ein Term')

Lx Ty
Txy

Vx Ty
Txy

f) Um die Produktionen der Einzelkalküle zu bekommen, benützen wir das Symbol P ('ist eine (Einzel-)Produktion') und die Produktionen

Tx
|
Px ('Jeder Term ist eine Produktion')

Tx Py
Px ----> y ('Ist x ein Term und P eine Produktion, so ist Px ---> y eine Produktion')

g) Um die Einzelkalküle zu bekommen, benützen wir das Symbol E ('ist (Einzel-)Kalkül') und die Produktionen

Px
|
Ex ('Jede (Einzel-)Produktion ist ein (Einzel-)Kalkül)

Px Ex
Ex * y

2) Nun ist eine Formalisierung der Ableitungsvorgänge in den Einzelkalkülen anzugeben. Sie bestehen aus Substitution und der unter dem Namen 'modus ponens' bekannten

Schlußregel: 'F gilt, wenn E gilt und F aus E folgt'.

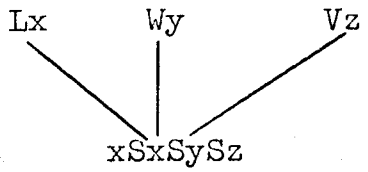
Um eine Substitution zu bezeichnen, verwenden wir die Symbolfolge

t S u S v S w ;

sie ist 'die Produktion t entsteht aus der Produktion u, wenn man das Wort v für die Variable w einsetzt' zu lesen.

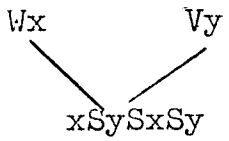
Wie das Substituieren vor sich geht, legen folgende Produktionen fest:

a)

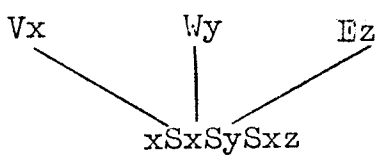


('Wenn man eine Variable z durch ein Wort y ersetzt, bleibt ein schon dastehender Buchstabe x unberührt')

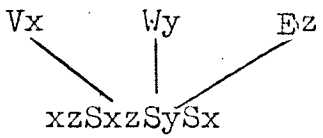
b) Wir verwenden das Symbol B ('ist eine Folge von mindestens einem Buchstaben ß (ß-Folge)') und fassen zunächst den puren Substitutionsvorgang durch die Produktion



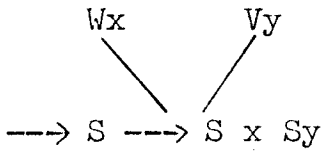
c) Wenn man die von der Variablen x durch eine höhere Nummer unterschiedene Variable xz durch y ersetzt, bleibt x unberührt ('Unterscheidung der Variablen')



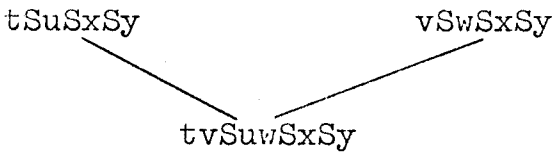
Analog



d) Bei Substitutionen bleiben Pfeile unberührt:



e) Die Hintereinanderausführung von Substitutionen regelt



Nun ist über die Ableitbarkeit aus den Einzelkalkülen einiges festzulegen. Wir benützen die Symbolfolge

x C y ,

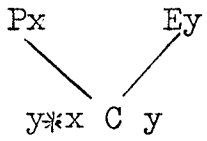
die 'y ist aus x ableitbar' zu lesen ist.

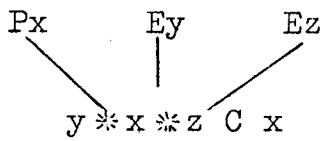
Die hier benötigten Produktionen sind:

f) Px / xCx ('Die Produktion x ist aus der Produktion x ableitbar')

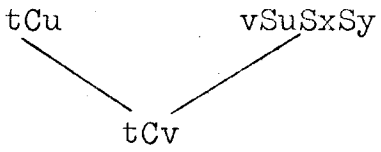
g) Px Ey / x*y C x (Übergang von einem Einzelkalkül zu einer in ihm enthaltenen Produktion)

Analog

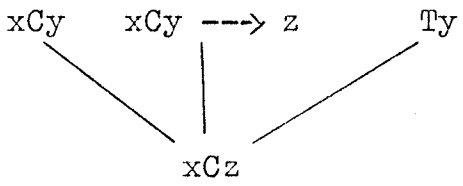




h) Das Verhalten der Ableitbarkeit bei Substitutionen regelt



i) Den 'modus ponens' enthält die Produktion



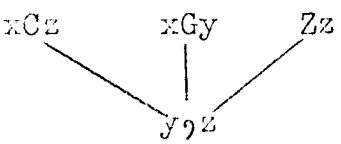
3) Um die lexikographische Anordnung der Einzelkalküle zu beschreiben, benützen wir das Symbol F ('ist Nachfolger von') und die Produktionen

□ F α	('α ist Nachfolger von □')
x α F x ξ	('x ξ ist Nachfolger von x α')
x ξ F x ß	('x ß ist Nachfolger von x ξ')
x ß F x --->	
x ---> F x ⁎	
xFy	
x ⁎Fyα	

4) Um die Numerierung der Einzelkalküle durch gewisse natürliche Zahlen zu beschreiben, benützen wir das Symbol G ('ist Gödel Nummer von') und die Produktionen

□ G □	('□ erhält die Gödel-Nummer □')
xFy xGz / yGzα	('der Nachfolger y des x mit der Gödel-Nummer z erhält die auf z folgende Gödel-Nummer zα')

5) Die Projektion der mittels des Gesamtkalküls abgeleiteten Wörter über $A = \{\alpha\}$ in die Ebene $N \times N$ leistet die Produktion



('ist y die Gödel-Nummer von x und ist z ein aus x ableitbares Wort über $\{\alpha\}$, so ist (y,z) ein Element von $U \subseteq N \times N$').

Damit ist der Beweis von Theorem 2.2 vollständig angegeben.

4. Nicht-entscheidbare Mengen

Mit Hilfe des Hauptsatzes (Theorem 2.2) gelingt jetzt der Beweis von

Theorem 2.3 Es gibt eine rekursiv aufzählbare Teilmenge S von N, derart, daß $S^c = N - S$ nicht rekursiv aufzählbar ist.

Beweis: Sei $U \subseteq N \times N$ wie in Theorem 2.2. Wir definieren S als den Schnitt von U mit der Diagonale von $N \times N$ genauer:

$$S = \{m \mid m \in N, \quad (m, m) \in U\}$$

1) S ist rekursiv aufzählbar: Man füge die Produktion

zu dem U erzeugenden Kalkül hinzu.

2) $S^c = N - S$ ist nicht rekursiv aufzählbar. Denn sei $F \subseteq N$ rekursiv aufzählbar, und $m \in N$ derart, daß $U_m = F$, d.h.

$$F = \{u \mid u \in N, (m, u) \in U\}$$

Fall I: $m \in F$, dann ist $(m, m) \in U$, also $m \in S$, d.h. $m \notin S^c$.

Fall II: $m \notin F$, dann ist $(m, m) \notin U$, also $m \notin S$, d.h. $m \in S^c$.

Die Mengen S^c und F unterscheiden sich also mindestens in der Zugehörigkeit von m, d.h. $S^c \neq F$. Da F eine beliebige rekursiv aufzählbare Teilmenge von N war, folgt, daß S^c nicht rekursiv aufzählbar sein kann.

Wir wollen die in diesem Beweis auftretende Menge S noch etwas genauer betrachten. $m \in S$ bedeutet soviel wie $(m, m) \in U$, d.h. m ist im Einzelkalkül Nr. m ableitbar. Daß S^c nicht rekursiv aufzählbar, S also nicht entscheidbar ist, bedeutet also: Die Aussage 'm ist im Einzelkalkül Nr. m ableitbar' ist zwar entweder wahr oder falsch, aber es gibt keinen Kalkül, der für ein beliebig vorgelegtes m in endlicher (wenn auch nicht beschränkter) Zeit entscheiden kann, ob diese Aussage für dies m wahr oder falsch ist. Dies ist ein genaues Analogon zu dem berühmten Unentscheidbarkeitssatz von Gödel.

Die Beziehung zu bekannten Paradoxien der Mengenlehre tritt zutage, wenn man die Aussage '$m \in S$' = 'm ist im Einzelkalkül Nr. m ableitbar' durch die Aussage 'm ist in sich selbst enthalten' ersetzt; wenn man m mit dem Schnitt U_m von U identifiziert, ist das eine legitime mengentheoretische Aussage. Dasselbe Argument, das bei Zulassung aller Mengen zu einer Paradoxie führt, liefert bei Beschränkung auf rekursiv aufzählbare Mengen eine Unvollständigkeitsaussage, eben Theorem 2.3.

5. Der Hauptsatz der Theorie der berechenbaren Funktionen

wird als Folgerung aus Theorem 2.2 erhalten, wenn man Funktionen mit ihren Graphen identifiziert.

Definition 2.5 Eine Funktion f, welche einen Teil der Wörter über dem endlichen Alphabet A in Wörter über dem endlichen Alphabet B abbildet, heißt berechenbar, wenn ihr Graph rekursiv aufzählbar ist.

Theorem 2.4: Es existiert eine berechenbare Funktion

$$u: N \times N \dashrightarrow N$$

derart, daß u als Funktion des zweiten Arguments die Menge aller berechenbaren Funktionen $f : N \dashrightarrow N$ durchläuft, wenn das erste Argument N durchläuft;
d.h. zu jeder berechenbaren Funktion $f : N \dashrightarrow N$ existiert ein $m \in N$, derart, daß

$$f(n) = u(m,n)$$

für alle $n \in N$ gilt (m heißt die Gödelnummer von f).

Literatur zu § 2

[1] P o s t , E. L.
Recursively enumerable sets of positive integers and their decision problems
Bull. Amer. Math. Soc. 50 (1944), 284 - 316

[2] S m u l l y a n , R. M.
Theory of Formal Systems
Annals of Mathematics Studies 47 (1961)

§ 3. Die Definition endlicher zufälliger Folgen.

1. Intuitives über endliche zufällige Folgen

Betrachten wir die folgenden vier Beispiele von Folgen der Länge 30 aus Nullen und Einsen:

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0

0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 0

0 1 0 0 1 1 1 0 1 0 0 1 1 1 1 0 1 0 0 0 0 0 1 1 0 0 1 0 1 1

Wenn man sich dieselben aus einem wiederholten Zufallsexperiment entstanden denkt, das mit Wahrscheinlichkeit 1/2 Nullen und Einsen produziert, so sind nach der maßtheoretischen Auffassung der Wahrscheinlichkeitstheorie alle diese Folgen gleich wahrscheinlich; die Wahrscheinlichkeit einer jeden ist dieselbe und gleich 2^{-30}. Doch verhalten sie sich hinsichtlich ihrer Regelmäßigkeit ganz verschieden; die erste Folge enthält sozusagen sehr wenig Information, die (bei gegebener Länge) schon durch eine kürzere Aussage wie etwa: "nur Nullen" vermittelt werden könnte. Die zweite Folge ist ebenso durch "Periode 1001" beschreibbar, eine Aussage, welche sich für große Wortlängen n auch mit viel weniger als n Zeichen hinschreiben ließe. Auch eine Folge vom Typ des 3.Beispiels, die sich durch eine geringe Häufigkeit der Einsen auszeichnet, ist bei größerer Länge n durch eine kürzere Aussage beschreibbar: man ordne nämlich etwa die binären Folgen der Länge n nach wachsender Häufigkeit der Einsen und innerhalb der Klassen gleicher Häufigkeit lexikographisch (0 vor 1). Um in dieser Durchzählung die Nummer der

Folge (i.Dualdarstellung) anzugeben, braucht man ungefähr

$$n\, H\left(\frac{m}{n}\right) \quad \text{bits}$$

(= binäre Einheiten 0 oder 1), wo n die Anzahl der Einsen ($m/n \leq 1/2$) und allgemein

$$H(p) = -p\ {}^2\log p - (1-p)\ {}^2\log (1-p) \text{ ist.}$$

Wenn man für die allgemeine Beschreibung dieser Durchzählung die konstante Anzahl von c bits benötigt, so kann man unsere Folge mit insgesamt ungefähr

$$n\, H\left(\frac{m}{n}\right) + c \quad \text{bits}$$

angeben, und diese Anzahl ist für kleine Werte von m/n kleiner als n.

Im Gegensatz zu diesen 3 Beispielen kann man jedoch die 4.Folge allem Anschein nach nicht mit einem kürzeren Ausdruck angeben. Sie ist die einzige der 4 Folgen , die wir als Ergebnis eines Zufallsexperiments, welches weder Nullen noch Einsen bevorzugt, ansehen würden. Es liegt daher nahe, die Zufälligkeit einer Folge durch die Kompliziertheit ihrer Beschreibung zu definieren. Diese Idee geht auf Kolmogroff [5] zurück (vgl. auch Chaitin [3]). Genauer gesagt, wird man die Komplexität

$$K(x_1 \ldots x_n | n)$$

einer binären Folge $x_1 \ldots x_n$ als die minimale Anzahl der bits definieren, die man braucht, um jene anzugeben (die Länge n der Folge als gegeben betrachtet). Die Komplexität

ist stets höchstens gleich n ; man wird eine Folge zufällig nennen, wenn ihre Komplexität diesem Maximum nahekommt.

2. Das Kolmogoroffsche Komplexitätsmaß

Seien zwei endliche Alphabete gegeben. Eine berechenbare Funktion A, die jedem Paar (p,x), bei dem p eine binäre Folge und x ein endliches Wort über dem einen Alphabet ist, ein Wort y über dem anderen Alphabet zuordnet, wird im folgenden der Kürze halber als Algorithmus bezeichnet.
A kann als eine Rechenmaschine, p als Programm, mit dem y aus x auf ihr berechnet wird, interpretiert werden.

Für ein beliebiges Wort p über einem Alphabet bezeichne l(p) die Länge von p.

Definition 3.1

Seien x,p,y wie oben. Dann wird

$$K_A(y|x) := \min_{A(p,x)=y} l(p) \quad (= +\infty, \text{ falls } A(p,x)\neq y \text{ für alle } p)$$

die bedingte Komplexität bezüglich A von y bei gegebenem x genannt.
Sie ist also per def. gleich der Länge des kürzesten Programms p, mit welchem die Maschine A y aus x berechnen kann.

Wir wollen den Begriff der Komplexität an den obigen Beispielen verdeutlichen:

Beispiel 1. Sei $y = \underbrace{0\,0\,\ldots\,0}_{n\text{-mal}}$ (wo n eine natürliche Zahl ist)

Zur Bestimmung der Komplexität von y bei gegebenem n legen wir den Algorithmus

$$A(\ ,n) = \underbrace{0\,0\,\ldots\,0}_{n\text{-mal}}$$

zugrunde. Das kürzeste Programm zur Berechnung von y ist dann das leere Programm, es hat die Länge 0; daher gilt

$$K_A(y|n) = 0.$$

Beispiel 2. Sei $y = \underbrace{1\,0\,0\,1\,1\,0\,0\,1\,\ldots}_{n\text{ Zeichen}}$ und $A(\ ,n) = y$.

Dann ist ebenfalls $K_A(y|n) = 0$.

Beispiel 3. Sei $A(p,n)$ das k-te Element in der oben eingeführten Durchzählung der binären Folgen $x_1 \ldots x_n$, wo k diejenige natürliche Zahl ist, deren Dualentwicklung durch p gegeben wird. Dann erhalten wir

$$K_A(x_1 \ldots x_n|n) \approx n\,H(\tfrac{m}{n}),$$

wo $m = x_1 + \ldots + x_n$ und $m/n \leq 1/2$.

Die eben definierte bedingte Komplexität bezüglich eines Algorithmus A ist noch nicht als Komplexitätsmaß verwendbar, da sie von dem zugrunde gelegten Algorithmus abhängt. Kolmogoroff [5] hat jedoch die Existenz eines "universellen" Algorithmus bewiesen, der die Eigenschaft hat, daß die zugehörige bedingte Komplexität asymptotisch im wesentlichen nicht größer als die bedingte Komplexität bezüglich eines beliebigen anderen Algorithmus ist. Genauer, es gilt folgendes

Theorem 3.1. (Kolmogoroff [5] , Solomonoff [8])

Es existiert ein Algorithmus A derart, daß für jeden Algorithmus B

$$K_A(y|x) \leq K_B(y|x) + c$$

gilt, wo c eine Konstante ist, die von A und B , aber nicht von x und y abhängt. Einen solchen Algorithmus nennen wir universell (nach Solomonoff) oder asymptotisch optimal (nach Kolmogoroff).

Definition 3.2

A sei ein für das folgende fest gewählter universeller Algorithmus. Dann heißt

$$K(y|x) = K_A(y|x)$$

die (bedingte) Komplexität von y bei gegebenem x.

Beweis von Theorem 3.1:

Wir benützen den Hauptsatz über berechenbare Funktionen (Theorem 2.4); angewandt auf Algorithmen, sagt dieser aus:

es existiert eine berechenbare Funktion U: $(m,p,x) \to y$ (wo m eine natürliche Zahl ist), derart, daß zu jedem Algorithmus B eine natürliche Zahl b existiert mit

$$U(b,p,x) = B(p,x) \quad \text{für alle p und x}$$

(b wird auch Gödelnummer von B genannt).

Wir definieren nun den universellen Algorithmus A wie folgt:

$$A(m_1\, m_1\, m_2\, m_2 \ldots m_r\, m_r\, 0\, 1\, p,\, x) = U(m,p,x),$$

wo $m_1\, m_2 \ldots m_r$ die Dualentwicklung von m ist (die Verdopplung und die Zeichen "0 1" dienen lediglich zur Unterscheidung von m und p). A ist ein Algorithmus; er erfüllt die Be-

hauptung des Satzes:

sei B irgendein Algorithmus mit Gödelnummer b (bzgl.U):

$$B\,(p,x) = U\,(b,p,x) \text{ für alle } p \text{ und } x.$$

y sei eine beliebige Binärfolge.

Wenn kein Programm p mit $B\,(p,x) = y$ existiert, so ist die behauptete Ungleichung mit jeder Konstanten erfüllt.

Sei also p ein Programm von minimaler Länge und $B\,(p,x) = y$:

$$K_B(y|x) = l(p).$$

Ist $b_1 \ldots b_r$ die Dualentwicklung von b, so haben wir durch

$$A(b_1\ b_1 \ldots b_r\ b_r\ 0\ 1\ p,x) = U(b,p,x) = B(p,x) = y$$

die Folge y mit einem Programm der Länge $l(p) + 2r + 2$ berechnet. D.h. es gilt

$$K_A(y|x) \leq l(p) + 2r + 2 = K_B(y|x) + 2r + 2$$
$$= K_B(y|x) + c$$

mit $c = 2r + 2$. Damit ist das Theorem bewiesen .

Es ist übrigens möglich, den Wert von c weiter herab zu drücken : man kann erreichen, daß c fast so klein wird wie die Anzahl der bits der Gödelnummer des Algorithmus B, also ungeführ gleich der Zahl der bits in der exakten Beschreibung von B.

3. Der Begriff einer endlichen zufälligen Folge

Da für den reproduzierenden Algorithmus $B(p,n) = p$ stets

$$K_B(x_1\ x_2\ \ldots\ x_n \mid n) = n$$

gilt, haben wir nach dem obigen Hauptsatz

$$K\,(x_1\ x_2 \ldots\ x_n \mid n) \leq n + c \quad (c=\text{const.}).$$

Wir werden eine Folge $x_1\ x_2\ \ldots\ x_n$ als <u>zufällig</u> betrachten, wenn $K\,(x_1\ x_2\ \ldots\ x_n|n) \approx n$ ist.

Diese Definition legt nicht genau fest, welche Folgen als zufällig zu gelten haben und welche nicht, d.h. wie weit die Komplexität einer Folge von n abweichen darf, ohne daß ihr der Charakter der Zufälligkeit abgesprochen werden muß. Wesentlich ist nur die Rolle von K als Maß für die "Zufälligkeit" einer Folge. Wir werden später bei der Betrachtung der Menge aller binären Folgen (verschiedener, beliebig großer Längen) die Definition durch die Forderung

$$\text{" } n - K(x_1 \ldots x_n | n) \text{ beschränkt "}$$

präzisieren.

Der folgende Satz kann als eine Präzisierung der Aussage

"Beinahe alle Folgen sind zufällig"

aufgefaßt werden.

Theorem 3.2. Sei $c \geq 0$ eine ganze Zahl. Es gibt mehr als $(1 - 2^{-c})\, 2^n$ Folgen $x_1 \ldots x_n$ der Länge n, für welche

$$K(x_1 \ldots x_n | n) \geq n-c$$

ist.

Beweis: Die Anzahl der Folgen $x_1 \ldots x_n$ mit $K(x_1 \ldots x_n | n) < c$ ist kleiner als 2^c; denn $K(x_1 \ldots x_n | n) < c$ ist genau dann der Fall, wenn es ein Programm p mit $l(p) < c$ gibt; aber es existieren genau $2^c - 1$ solche Programme.

Corollar: Es gibt stets Folgen der Komplexität $K(.|n) \geq n$.

Beweis: man setze $c = 0$.

Dieser Beweis läßt sich übrigens durch keinen konstruktiven Beweis ersetzen; das zeigt

Theorem 3.3. Jeder Algorithmus (im weiteren Sinne des Kanonischen Kalküls) produziert nur Folgen beschränkter Komplexität; d.h. hinreichend lange Folgen weisen Regelmäßigkeiten auf.
(Man kann diesenSatz als Präzisierung der Aussage:

"Zufällige Folgen sind nicht konstruierbar"

auffassen).

Beweis: Sei $F : n \longrightarrow x_1 \ldots x_n$ eine berechenbare Funktion, die jeder natürlichen Zahl eine binäre Folge zuordnet. Damit definieren wir den Algorithmus B:

$$B(p,n) = F(n).$$

Also gilt $K_B(x_1 \ldots x_n | n) = 0$, d.h. $K(x_1 \ldots x_n | n) \leq c$.

Anmerkung: dieser Satz zeigt, daß $K(y|x)$ als Funktion von x und y (die man o.B.d.A. als binäre Folgen ansehen kann) keine berechenbare Funktion ist; denn sonst wäre auch $K(x|l(x))$ berechenbar, und man könnte einen Algorithmus finden, der die x mit $K(x|l(x)) \geq l(x)$ produziert, im Widerspruch zu Theorem 3.3.

4. Zufälligkeitstests

Mit dem Begriff der Zufälligkeit einer binären Folge $x_1 \ldots x_n$ verbinden wir die Eigenschaft der Gleichhäufigkeit von Nullen und Einsen. Wir wollen diese Hypothese für die

oben definierten zufälligen Folgen t e s t e n. D.h. wir lehnen die Hypothese ab, wenn

$$\left| \frac{s_n}{n} - \frac{1}{2} \right|$$

groß ist ($s_n = x_1 + \ldots + x_n$ ist die Anzahl der Einsen). Wir können dazu also kritische Regionen der Form

$$|2s_n - n| > c(m,n)$$

wählen, wo 2^{-m} die Sicherheitswahrscheinlichkeit des Tests ist (m ist dabei eine natürliche Zahl). Dabei soll c(m,n) im Rahmen der Forderung, daß die kritische Region höchstens 2^{n-m} Folgen der Länge n enthalte, so klein wie möglich sein.

Nun muß jeder Test vor der Durchführung des Experiments effektiv angegeben werden können; daraus folgt unter Annahme der These von Church (vgl. z.B. [4]), welche besagt, daß allein rekursiv aufzählbare Mengen konstruierbar sind, daß seine kritische Region rekursiv aufzählbar ist, und zwar sowohl in Abhängigkeit von der Sicherheitswahrscheinlichkeit als auch vom Stichprobenumfang, also in unserem Fall in Abhängigkeit von m und n. Die kritischen Regionen U_m zur Sicherheitswahrscheinlichkeit 2^{-m} können o.B.d.A. als ineinandergeschachtelt angenommen werden:

$$X = U_0 \supseteq U_1 \supseteq U_2 \supseteq \ldots$$

(X ist dabei die Menge aller endlichen binären Folgen). Die rekursive Aufzählbarkeit der Familie U_m bedeutet: es existiert eine rekursiv aufzählbare Menge $U \subseteq N \times X$, derart daß

$$U_m = \{ x \mid (m,x) \in U \}.$$

U bezeichnen wir als Test und benutzen diesen Begriff

künftig stets in dem so präzisierten Sinne.

Jedem $x \in X$ ordnen wir die Sicherheitswahrscheinlichkeit (genauer: ihren Logarithmus) der kleinsten kritischen Region zu, in der es enthalten ist:

$$m_U(x) := \max_{x \in U_m} m .$$

Ist $m_U(x)$ groß, so heißt das, daß für sehr viele Sicherheitswahrscheinlichkeiten die Hypothese des Tests abgelehnt werden muß; $m_U(x)$ ist also ein Maß für die Abweichung von x von der Hypothese des Tests.

Nun hängt m_U aber von der speziellen Wahl des Tests ab. Doch kann man wie oben durch Anwendung des Theorems 2.2 die Existenz eines universellen Tests erhalten.

Theorem 3.4. Es existiert ein universeller Test U derart, daß für jeden Test V

$$V_{m+c} \subseteq U_m$$

für alle natürlichen Zahlen m gilt; dabei ist c eine Konstante, die von U und V, aber nicht von m abhängt. (Die "konstante Verschiebung" um c bedeutet eine Multiplikation aller Sicherheitskoeffizienten mit der Konstanten 2^c).

Beweis: Nach einem zu Theorem 2.2 analogen Satz ist die Familie aller Tests rekursiv aufzählbar: es gibt also eine rekursiv aufzählbare Menge $T \subseteq N \times N \times X$, so daß

$$T_i = \{ (m,x) \mid (i,m,x) \in T \}$$

für i=1,2,3,... die Familie aller Tests durchläuft.

U sei das Bild von T unter der Abbildung

$$(i,m+i,\ x) \longrightarrow (m,x).$$

Sei nun V ein beliebiger Test. Dann existiert ein $i \in N$ derart, daß

$$V=\{(m,x) \mid (i,\ m,\ x) \in T\}\ .$$

Also ist $$V_{m+i} = \{x \mid (i,m+i,x) \in T\}$$

$$\subseteq \{x \mid (m,x) \in U\} \quad = \quad U_m\ .$$

Damit ist Theorem 3.4 bewiesen.

Wir fixieren nun einen solchen universellen Test U und bezeichnen

$$m(x) = m_U(x)$$

als die zu x gehörige Sicherheitswahrscheinlichkeit. Es gilt

$$m_V(x) \leq m(x) + c$$

mit einer Konstanten c, die von V, aber nicht von x, abhängt.

Die beiden Maße K und m stehen bisher anscheinend beziehungslos nebeneinander. Der folgende Satz zeigt jedoch, daß sie sozusagen komplementär sind:

Theorem 3.5. Es gilt die asymptotische Gleichheit

$$|m(x_1 \ldots x_n) - n + K(x_1 \ldots x_n \mid n)| \leq c$$

(c = const.).

Beweis: Sei $V_m = \{x_1 \ldots x_n \mid K(x_1 \ldots x_n \mid n) < n - m\}$.

Dann ist leicht zu zeigen, daß

$$V := \{(m,x) \mid K(x|n) < n - m\}$$

$$= \{(m,x) \mid (\exists p) \ [\, A(p,n) = x \ \& \ l(p) < n - m]\}$$

rekursiv aufzählbar ist.

Für das so konstruierte V gilt

$$m_V(x) = \max_{x \in V_m} m = \max_{K(x|n) < n-m} m = n - 1 - K(x|n),$$

folglich

$$m(x) - n + K(x|n) \leq c$$

für eine Konstante c.

Die umgekehrte Ungleichung ergibt sich so:

Da V rekursiv aufzählbar ist, gibt es eine berechenbare Funktion

$$f: N \longrightarrow N \times X,$$

die V ohne Wiederholungen durchzählt. Mit Hilfe dieser Funktion definieren wir folgenden Algorithmus B :

für $f(1) = (m_1, x_1)$ setzen wir $B(\underbrace{0\ldots0}_{l(x_1)-m_1}, l(x_1)) = x_1$;

für $f(2) = (m_2, x_2)$ müssen wir die folgenden Fälle unterscheiden:

1.Fall: $(m_1, l(x_1)) = (m_2, l(x_2))$.

In diesem Fall sei $B(\underbrace{0\ldots01}_{l(x_2)-m_2}, l(x_2)) = x_2$;

2.Fall: $(m_1, l(x_1)) \neq (m_2, l(x_2))$.

Hier definieren wir $B(\underbrace{0\ldots0}_{l(x_2)-m_2}, l(x_2)) = x_2$.

Auf diese Weise fahren wir fort, indem wir jedes Mal nachprüfen, ob das Paar $(m_i, l(x_i))$ früher schon einmal aufgetreten ist("1.Fall") oder nicht("2.Fall").

Im 1.Fall definieren wir B für die lexikographisch erste binäre Folge, für die B (, $l(x_i)$) noch nicht definiert wurde. Da sich allgemein in V_m höchstens 2^{n-m} Folgen der Länge n befinden - denn V ist ein Test - kann der 1.Fall für jedes Paar $(m_i, l(x_i))$ höchstens $(l(x_i) - m_i)$ -mal auftreten, und folglich ist unser Definitionsverfahren nicht mehrdeutig.

Offensichtlich ist $K_B(x_i | l(x_i)) = l(x_i) - m(x_i)$.
Da wegen $V_o = X$ der Wertebereich von B ganz X ist, gilt

$$K_B(x|l(x)) = l(x) - m(x),$$

also

$$K(x|l(x)) \leq l(x) - m(x) + c$$

(Zunächst für eine andere Konstante c als oben, aber wir denken uns für c sofort das Maximum beider gewählt).
Beide Ungleichungen zusammen ergeben die Behauptung des Theorems.

Nun können wir - übrigens unter Verwendung von Hilfsmitteln aus der klassischen Wahrscheinlichkeitstheorie - den Ausdruck $|2\, s_n - n|$ abschätzen. Sei W der eingangs beschriebene Test mit den kritischen Regionen

$$W_m = [\, |2\, s_n - n| > c(m,n)\,] \,.$$

Dann gilt

$$|2\, s_n - n| \leq c(m_W(x) + 1,\, n),$$

denn c(.,.) war minimal gewählt, so daß die Addition von 1 zum ersten Argument die Ungleichung umkehrt.
Vergleich mit dem universellen Test liefert weiter

$$|2\,s_n - n| \leq c(m(x) + c, n)$$

$$= c(n - K(x_1 \ldots x_n|n) + c', n) \qquad (c'=\text{const.}).$$

Nun sei $n-K(x_1 \ldots x_n|n)$ beschränkt(dies ist eine mögliche Präzisierung des Begriffs "zufällige Folge").

Da wir bei der Konstruktion des Tests W die Folgen wie in einem Bernoulli-Experiment mit den Wahrscheinlichkeiten 1/2, 1/2, bewertet haben, ist der Satz von de Moivre-Laplace anwendbar; er liefert

$$\frac{c(c,n)}{\sqrt{n}} \xrightarrow{n\to\infty} \Phi^{-1}\,(1-2^{-c-1}),$$

wo Φ die Verteilungsfunktion der Normalverteilung mit Mittelwert 0 und Varianz 1 ist; c bedeutet eine andere Konstante als in den früheren Formeln.

Damit erhalten wir das Ergebnis

$$\left|\frac{s_n}{n} - \frac{1}{2}\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

Speziell existiert die relative Limeshäufigkeit der Einsen in der oben genau festgelegten Familie von zufälligen Folgen und ist gleich 1/2.

§ 4. Unendliche Folgen.

Wir betrachten nun wieder den Raum

$$\Omega = \{\omega = (x_1, x_2, \ldots) \mid x_t = 0 \text{ oder } 1 \ (t = 1,2,\ldots)\}$$

aller unendlichen Folgen von Nullen und Einsen. Unser Ziel ist es, diese Folgen in 'zufällige' und 'nichtzufällige' einzuteilen.
Dabei werden wir verlangen, daß der zu schaffende Begriff 'zufällige Folge' allen vernünftigen intuitiven Forderungen genügt und damit die Ideen von von Mises verwirklicht.
Wie wir uns in § 1 überlegt haben, läuft das z.B. darauf hinaus, daß eine zufällige Folge die Fastüberall-Gesetze der Wahrscheinlichkeitstheorie (Gesetz der großen Zahl, Satz vom iterierten Logarithmus,...) erfüllt.

Eine naheliegende Definition wäre so etwas wie: eine Folge $(x_1, x_2, \ldots)$ von Nullen und Einsen heißt zufällig, wenn ihre endlichen Abschnitte $(x_1, \ldots, x_n)$ zufällig sind. Wir werden in Nr. 1 einen Satz beweisen (Theorem 4.1), aus dem folgt, daß diese Definition nicht funktionieren kann: Zu jeder unendlichen Folge $(x_1, x_2, \ldots)$ gibt es unendlichviele n, für welche der Abschnitt $(x_1, \ldots, x_n)$ wesentliche Regelmäßigkeiten aufweist und somit im Sinne unserer Vereinbarung nicht als zufällig anzusehen ist.

In Nr. 2 wird daher eine andere Definition vorgeschlagen: Eine unendliche Folge $(x_1, x_2, \ldots)$ heißt zufällig, wenn sie einen 'universellen Sequentialtest' überlebt (Definition 4.4).

Anschließend zeigen wir, daß die in diesem Sinne zufälligen Folgen allen intuitiven Forderungen genügen. Daß es wirklich welche gibt, wird - der Natur der Sache entsprechend (Theorem 4.6) - nicht mit konstruktiven, sondern mit Mitteln der Maßtheorie bewiesen: Die zufälligen Folgen bilden eine meßbare Menge vom (Bernoulli-)Maß 1 (Theorem 4.4). Die von Mises'schen Ideen sind damit gerechtfertigt.

1. Komplexitätsschwankungen in unendlichen Folgen.

Wir denken uns einen universellen Algorithmus fest gewählt, auf den wir unsere Komplexitätsmaße $K(x_1,\ldots,x_n|n)$ beziehen.

Theorem 4.1 (Martin-Löf [7]).

Sei $f(n)$ eine für $n = 1,2,\ldots$ definierte berechenbare Funktion, die

$$\sum_{n=1}^{\infty} 2^{-f(n)} = \infty$$

erfüllt. Dann gibt es zu jeder unendlichen Folge $(x_1, x_2,\ldots)$ von Nullen und Einsen unendlichviele n, für welche

$$K(x_1 \ldots x_n|n) < n - f(n)$$

gilt.

Beweis.

1) Aus beweistechnischen Gründen ersetzen wir die Funktion $f(n)$ vorübergehend durch eine etwas größere Funktion $g(n)$. Wir zeigen: Es gibt eine berechenbare, für $n = 1,2,\ldots$ definierte Funktion $g(n)$, derart, daß

$$\sum_{n=1}^{\infty} 2^{-g(n)} = \infty$$

$$g(n) - f(n) \longrightarrow \infty$$

gilt. - Hierzu haben wir lediglich die ganzen Zahlen $o = n_o < n_1 < n_2 < \ldots$ so zu bestimmen, daß

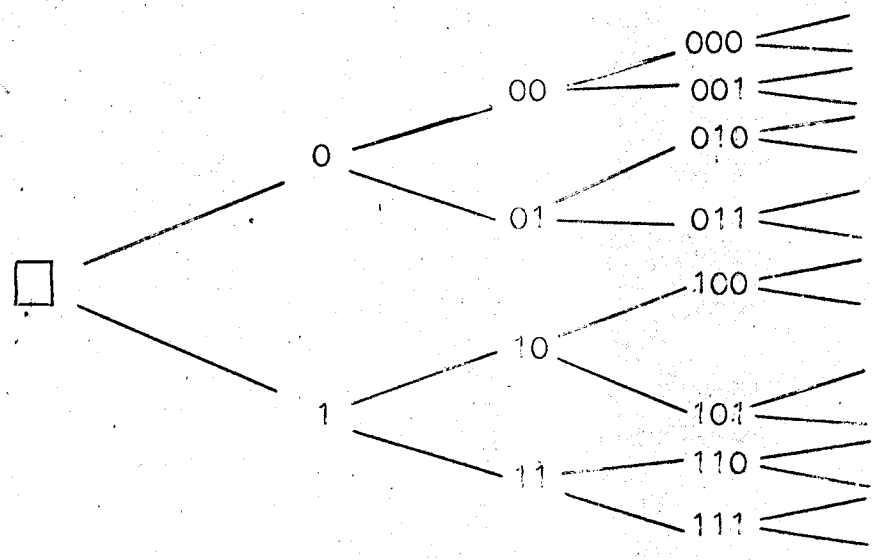
$$\sum_{n=1}^{n_1} 2^{-f(n)} > 2^o, \ldots, \quad \sum_{n=n_k+1}^{n_{k+1}} 2^{-f(n)} > 2^k, \ldots$$

gilt, und alsdann die Funktion g durch

$$g(n) = f(n) + k \qquad (n_k < n \leq n_{k+1})$$

zu definieren. Man überzeugt sich leicht, daß die Funktion $k \longrightarrow n_k$ berechenbar gewählt werden kann, wodurch auch g eine berechenbare Funktion wird.

2) Nun denken wir uns sämtliche endlichen Folgen von Nullen und Einsen folgendermaßen angeschrieben:



In der Spalte Nr. n dieses Schemas stehen die Folgen der Länge n in lexikographischer Anordnung.

Von jeder Folge der Länge n gehen zwei Verbindungsstriche nach rechts zu den durch Anhängen von 0 bzw. 1 aus ihr entstehenden Folgen der Länge n + 1. Die Folgen, die man von einer gegebenen Folge aus erreichen kann, indem man längs Verbindungsstrichen nach rechts wandert (die also aus ihr durch Anhängen von Nullen und Einsen entstehen), heißen die Nachfolger der gegebenen Folge. Die Abschnitte $(x_1,\ldots,x_n)$ $(n = 0, 1,\ldots)$ unserer vorgegebenen unendlichen Folge $(x_1, x_2,\ldots)$ sind also auf einem 'Faden' von Verbindungsstrichen aufgefädelt, der das Schema von links nach rechts durchläuft, da sie Nachfolger voneinander sind.

Wir wählen nun im

1. Wahlgang

die Menge M_1 von Folger der Länge 1 folgendermaßen:

M_1 ist leer, falls $1 - g(1) < 0$ ist;

M_1 besteht aus den obersten $2^{1-g(1)}$ Folgen der Länge 1, falls $1 - g(1) \geq 0$ ist;

sodann

die Menge M_2 von Folgen der Länge 2 folgendermaßen:

M_2 ist leer, falls $2 - g(2) < 0$ ist;

M_2 besteht aus den obersten $2^{2-g(2)}$ unter denjenigen Folgen der Länge 2, die keine Nachfolger von Folgen aus M_1 sind; falls $2-g(2)\geq 0$ ist,

sodann

die Menge M_3 von Folgen der Länge 3 folgermaßen:

M_3 ist leer, falls $3 - g(3) < 0$ ist;
M_3 besteht aus den obersten $2^{3-g(3)}$
unter denjenigen Folgen der Länge 3, die nicht Nachfolger von Folgen aus $M_1 \cup M_2$ sind; falls $3 - g(3) \geq 0$ ist.

etc., und setzen dieses Verfahren solange fort, bis wir bei einem N-ten Schritt zum erstenmal $N - g(N) \geq 0$ haben und weniger als $2^{N-g(N)}$ Folgen der Länge N vorfinden, die keine Nachfolger von Folgen aus $M_1 \cup \ldots \cup M_{N-1}$ sind. Eine solche Zahl existiert, denn andernfalls wäre die Anzahl der Nachfolger der Länge N von Folgen aus $M_1 \cup \ldots \cup M_{N-1}$, nämlich die Zahl

$$\sum_{\substack{n-g(n)\geq 0 \\ n\leq N-1}} 2^{N-n+(n-g(n))} =$$

$$= 2^N \sum_{\substack{n-g(n)\geq 0 \\ n\leq N-1}} 2^{-g(n)} ,$$

wegen $\sum_{n=1}^{\infty} 2^{-g(n)} = \infty$ gewiß einmal $> 2^N$, was nicht möglich ist.

Wir setzen nun $N_1 = N$, wählen M_{N_1} als die Menge <u>aller</u> Folgen der Länge N_1, die nicht als Nachfolger von Folgen aus $M_1 \cup \ldots \cup M_{N_1-1}$ auftreten, und beenden den 1. Wahlgang mit diesem Schritt Nr. N_1 in der Gewißheit, daß es ein n mit $0 < N \leq N_1$ und

$$(x_1, \ldots, x_n) \in M_n$$

gibt. Im

<u>2. Wahlgang</u>

wählen wir

die Menge M_{N_1+1} von Folgen der Länge $N_1 + 1$ folgendermaßen:

M_1 ist leer, falls $N_1 + 1 - g(N_1 + 1) < 0$ ist;

M_1 besteht aus den $2^{N_1+1-g(N_1+1)}$ obersten Folgen der Länge $N_1 + 1$; falls $N_1+1-g(N_1+1) \geq 0$ ist;

etc. Aus dem gleichen Grunde wie vorhin ($\sum 2^{-g(n)} = \infty$) endet der 2. Wahlgang beim Schritt $N_2 > N_1$, und man ist sicher, daß es ein n mit $N_1 < n \leq N_2$ und

$$(x_1,\ldots,x_n) \in M_n$$

gibt.

Insgesamt erhalten wir eine Folge $M_1, M_2,\ldots$, wobei M_n eine Menge von höchstens $2^{n-g(n)}$ Folgen der Länge n ist. Es gibt unendlichviele n mit

$$(x_1,\ldots,x_n) \in M_n$$

($(x_1,\ldots,x_n)$ ist der n-te Abschnitt von $(x_1, x_2,\ldots)$).

Nun betrachten wir den Algorithmus A, der für jedes $n = 1,2,\ldots$ dem aus der binären Folge für die natürliche Zahl k bestehenden Programm p die k-te Folge aus M_n (in der lexikographischen Anordnung von M_n) als Wert $A(p, n)$ zuordnet, solange, bis alle Elemente von M_n aufgetreten sind; dann werden noch alle übrigen Folgen der Länge n in lexikographischer Reihenfolge produziert. Daß A tatsächlich ein Algorithmus ist, beruht im wesentlichen auf der Rekursivität von g (wir verzichten auf die Details). Wenn der Abschnitt $(x_1,\ldots,x_n)$ von $(x_1, x_2,\ldots)$ zu M_n gehört - was für unendlichviele n eintritt - haben wir also

$$K_A(x_1 \ldots x_n \mid n) \leq n - g(n),$$

denn die Binärentwicklung einer Zahl $\leq 2^{n-g(n)}$, also das Programm zur Herstellung eines $(x_1,\ldots,x_n) \in M_n$, hat höchstens $n - g(n)$ Stellen.

Nach Theorem 3.1 gibt es eine Konstante C, derart, daß allgemein

$$K(x_1 \ldots x_n | n) \leq K_A(x_1 \ldots x_n\, n) + C \qquad (n = 0, 1, \ldots)$$

und somit

$$K(x_1 \ldots x_n | n) \leq n - g(n) + C$$

für die $(x_1,\ldots,x_n) \in M_n$ erfüllenden Abschnitte von $(x_1,x_2,\ldots)$ gilt. Für hinreichend großes n ist aber $g(n) > f(n) + C$, was

$$K(x_1 \ldots x_n | n) < n - f(n)$$

und somit die Behauptung des Theorems beweist.

Anmerkung: Die Konstruktion der M_n ist einer von Borel [1] bei der Approximation reeller Zahlen durch rationale benützten Idee analog.

Als Anwendung erhalten wir z.B. das

Corollar.

Zu jeder unendlichen Folge $(x_1, x_2,\ldots)$ von Nullen und Einsen gibt es unendlichviele n, für welche

$$K(x_1 \ldots x_n | n) < n - \log n$$

gilt.

Beweis. $f(n) = [{}^2\log n]$ ist eine berechenbare Funktion, die

$$\sum_{n=1}^{\infty} 2^{-f(n)} \geq \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

2. Sequentialtests und die Definition unendlicher zufälliger Folgen.

Wir wollen jetzt jeder endlichen Folge

$$(x_1,\ldots,x_n)$$

von Nullen und Einsen die Zylindermenge

$$[x_1,\ldots,x_n] = \{\omega = (y_1, y_2,\ldots) \mid y_1 = x_1,\ldots,y_n = x_n\}$$

zuordnen. Der leeren Folge entspricht dabei ganz Ω. Der Verlängerung einer endlichen Folge entspricht der Übergang zu einer kleineren Zylindermenge.

Jeder Teilmenge U' der Menge

$$X = \overline{\{0,1\}} = \{(x_1,\ldots,x_n) \mid n \geq 0,\ x_1,\ldots,x_n = 0 \text{ oder } 1\}$$

entspricht die offene Menge

$$\mathcal{U} = \mathcal{U}(U') = \bigcup_{(x_1,\ldots,x_n)\in U'} [x_1,\ldots,x_n] .$$

Umgekehrt kann man aus jeder (bezüglich der üblichen Produkttopologie in Ω) offenen Menge $\mathcal{U} \subseteq \Omega$ eine Menge

$$U = U(\mathcal{U}) = \{(x_1,\ldots,x_n) \mid [x_1,\ldots,x_n] \subseteq \mathcal{U}\} \subseteq X$$

gewinnen. Eine so gewonnene Menge $U = U(\mathcal{U})$ ist stets sequentiell in dem Sinne, daß

$$(x_1,\ldots,x_n, x_{n+1},\ldots,x_m) \in U ,$$

falls $(x_1,\ldots,x_n) \in U$. Es gilt

$$\mathcal{U}(U(\mathcal{U})) = \mathcal{U} ,$$

aber i.a. nicht $U(\mathcal{U}(U')) = U'$.

Ein $\omega = (x_1, x_2, \ldots)$ gehört genau dann zu der nichtleeren offenen Menge $\mathcal{U} \subseteq \Omega$, wenn es ein n mit $(x_1, \ldots, x_n) \in U(\mathcal{U})$ gibt (man kann dies als die Definition von 'offen' ansehen).

Bei maßtheoretischen Betrachtungen in Ω wollen wir stets das durch

$$\pi([x_1, \ldots, x_n]) = 2^{-n} \qquad (n = 1,2,\ldots)$$

definierte (Bernoulli)-Maß π zugrundelegen. Dies Maß wird vorerst nur zur leichteren Formulierung von Sachverhalten, die man auch anders ausdrücken könnte, benutzt. Erst im Beweis von Satz 4.4 kommt die Maßtheorie wirklich zum Tragen.

Eine offene Menge $\mathcal{U} \subseteq \Omega$ kann als kritische Region für einen Sequentialtst in Ω zur Sicherheitswahrscheinlichkeit 2^{-m} dienen, falls

$$\pi(\mathcal{U}) \leq 2^{-m}$$

gilt. Ist für jedes $m = 0,1,2,\ldots$ eine solche kritische Region $\mathcal{U}_m$ mit $\mathcal{U}_0 = \Omega$,

$$\pi(\mathcal{U}_m) \leq 2^{-m}$$

und $\mathcal{U}_0 \supseteq \mathcal{U}_1 \supseteq \mathcal{U}_2 \supseteq \ldots$ gegeben, so können wir die Folge $(\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \ldots)$ als einen Sequentialtest in Ω ansehen (vgl. die verschärfte Fassung dieses Begriffs in Definition 4.1).

Aus den sequentiellen Teilmengen

$$U_m = U(\mathcal{U}_m)$$

der Menge X aller endlichen Folgen bilden wir die Teilmenge

$$U = \{(m, (x_1,\ldots,x_n)) \mid (x_1,\ldots,x_n) \in U_m,\ m = 0,1,2,\ldots\}$$

von N x X, (daß der Buchstabe U vorhin auch für Teilmengen von X verwendet wurde, wird hoffentlich nicht stören).

Sie ist eine <u>sequentielle</u> Teilmenge von N x X in dem Sinne, daß ihre Schnitte

$$U_m = \{(x_1,\ldots,x_n) \mid (m,(x_1,\ldots,x_n)) \in U\}$$

sequentielle Teilmengen von X sind und $U_o \supseteq U_1 \supseteq U_2 \supseteq \ldots$ erfüllen. Wir transponieren jetzt unsere Begriffe naxh N x X und treffen die

Definition 4.1. Eine Teilmenge U von N x X heißt <u>sequentiell</u>, falls ihre Schnitte

$$U_m = \{(x_1,\ldots,x_n) \mid (m, (x_1,\ldots,x_n)) \in U\}$$

sequentielle Teilmengen von X sind und $X = U_o \supseteq U_1 \supseteq U_2 \supseteq \ldots$

erfüllen. U heißt ein <u>Sequentialtest</u> (in X), wenn die Mengen $\mathcal{U}_m = \mathcal{U}(U_m)$ die Bedingung

$$\pi(\mathcal{U}_m) \leq 2^{-m} \qquad (m = 0,1,\ldots)$$

erfüllen, und U eine rekursiv aufzählbare Teilmenge von N x X ist.

Zu jedem Sequentialtest U in N x X gewinnt man sofort einen Sequentialtest $(\mathcal{U}_o, \mathcal{U}_1, \ldots)$ in Ω, indem man $\mathcal{U}_m = \mathcal{U}(U_m)$

$(m = 1,2,\ldots)$ setzt. Die Beziehung zwischen U und $(\mathcal{U}_0, \mathcal{U}_1, \ldots)$ ist umkehrbar eindeutig; deshalb wollen wir U und $(\mathcal{U}_0, \mathcal{U}_1, \ldots)$ als synonym betrachten und so z.B. von Sequentiealtests in Ω sprechen.

Alle interessanten Tests aus der Wahrscheinlichkeitstheorie des Raumes Ω (mit π) sind Sequentialtests in diesem Sinne.

Definition 4.2. Ein Sequentialtest U heißt universell, wenn es zu jedem Sequentialtest V eine ganze Zahl $c \geq 0$ gibt, derart, daß

$$V_{m+c} \subseteq U_m \qquad (m = 1,2,\ldots)$$

gilt.

Theorem 4.2 (Martin-Löf [5]).

Es gibt universelle Sequentialtests.

Beweis: Der Beweis beruht auf ähnlichen Überlegungen wie der Beweis von Theorem 2.2 (wie schon der Beweis des analogen Theorems 3.4). Wir begnügen uns daher mit einer Skizze.

Der erste Schritt besteht in einer Gödel-Numerierung der Menge aller Sequentialtests in $X = U^{(0)}, U^{(1)}, \ldots$. Sie gibt Anlaß zu einer rekursiv aufzählbaren Teilmenge

$$T \subseteq N \times N \times X = \{(i,m,x) \mid i,m \in N,\ x \in X\}$$

(wir schreiben x statt $(x_1,\ldots,x_n)$), deren 'Schnitte'

$$T_i = \{(m, x) \mid (i, m, x) \in T\}$$

gerade die $U^{(i)}$ sind:

$$T_i = U^{(i)} \quad (i = 0, 1, \ldots) .$$

Die berechenbare (nicht überall definierte) Funktion

$$f: \quad (i, m+i, x) \longrightarrow (m, x)$$

bildet die Menge T in eine rekursiv aufzählbare Menge $U \subseteq N \times X$ ab. Sie ist ein Sequentialtest, wie man aus der Relation

$$U_m = \bigcup_{i=0}^{\infty} U_{m+i}^{(i)}$$

für ihre 'Schnitte' leicht entnimmt. Nun sieht man unmittelbar, daß U ein universeller Sequentialtest ist: Jeder Sequentialtest kommt als ein $U^{(i)}$ mit einer passenden Gödel-Nummer i vor. Die Konstante c aus Definition 4.2 ist einfach diese Gödel-Nummer.

Definition 4.3. Eine Teilmenge $\mathcal{N}$ von Ω heißt eine konstruktive Nullmenge, wenn es einen Sequentialtest $(\mathcal{U}_0, \mathcal{U}_1, \ldots)$ in Ω gibt, derart, daß

$$\mathcal{N} \subseteq \bigcap_{m=1}^{\infty} \mathcal{U}_m$$

gilt. Die konstruktive Nullmenge $\bigcap_{m=1} \mathcal{U}_m$ heißt die Nullmenge des Tests $(\mathcal{U}_0, \mathcal{U}_1, \ldots)$.

Theorem 4.3. Die Nullmengen sämtlicher universellen Sequentialtests stimmen überein; wir nennen diese Nullmenge $\mathcal{N}_{univ}$ die universelle konstruktive Nullmenge.

Sie umfaßt jede konstruktive Nullemenge, ist also die größte konstruktive Nullmenge.

Der Beweis ist unmittelbar aus Theorem 4.2 bzw. Definition 4.2 abzulesen.

Definition 4.4. Eine unendliche Folge $(x_1, x_2, \ldots)$ heißt zufällig, wenn sie nicht zur universellen konstruktiven Nullmenge $\mathcal{N}_{univ}$ gehört.
Eine zufällige Folge ist also dadurch definiert, daß sie jeden Sequentialtest 'überlebt'. Da jedes vernünftige Fastüberallgesetz der Wahrscheinlichkeitstheorie von π in Ω einem Sequentialtest entspricht, folgt: Eine zufällige Folge erfüllt alle Fastüberall-Gesetze der Wahrscheinlichkeitstheorie (z.B. den Satz vom iterierten Logarithmus). Sie ist insbesondere ein Kollektiv im Sinne von von Mises.

Theorem 4.4. Es gibt unendliche zufällige Folgen. Sie bilden eine Menge vom π-Maß 1.

Beweis. Die universelle Nullmenge $\mathcal{N}_{univ}$ ist eine π-Nullmenge.

Wir haben also die Maßtheorie benützt, um die Existenz von unendlichen zufälligen Folgen sicherzustellen. Das ist kein Wunder:

Theorem 4.5. Sowohl $\mathcal{N}_{univ}$ als auch $\Omega - \mathcal{N}_{univ}$ sind überabzählbar.

Beweis. Sei $\mathcal{U}_0 = 0$ und

$$\mathcal{U}_m = \{\omega = (x_1, x_2, \ldots) \mid x_{2k} = 0 \ (k = 0, \ldots, m)\}$$

Dann ist $(\mathcal{U}_0, \mathcal{U}_1, \ldots)$ ein Sequentialtest mit der Nullmenge

$$\mathcal{N} = \{\omega = (x_1, x_2, \ldots) \mid x_{2k} = 0 \ (k = 0, 1, \ldots)\}.$$

Sie ist überabzählbar und in $\mathcal{N}_{univ}$ enthalten.- Die Überabzählbarkeit von $\Omega - \mathcal{N}_{univ}$ folgt aus Satz 4.4.

Theorem 4.6. Ist $(x_1, x_2, \ldots) \in \Omega$ rekursiv in dem Sinne, daß die Folge $x_1, x_1x_2, x_1x_2x_3, \ldots$ ihrer Abschnitte rekursiv aufzählbar ist, so ist $(x_1, x_2, \ldots)$ keine zufällige Folge.

Beweis. $(x_1, x_2, \ldots)$ gehört zur Nullmenge des Sequentialtests $(\mathcal{U}_0, \mathcal{U}_1, \ldots)$ mit $\mathcal{U}_0 = \Omega$, $\mathcal{U}_m = \overline{[x_1, \ldots, x_m]}$, also zu $\mathcal{N}_{univ}$.

Aus Theorem 4.5 und Theorem 4.6 folgt, daß es nicht-zufällige Folgen gibt, die nicht rekursiv sind, denn an rekursiven gibt es nur abzählbar viele.
Eine feinere Klassifikation der nicht zufälligen Folgen erhält man, indem man einen festen universellen Sequentialtest U wählt und für jedes $\omega = (x_1, x_2, \ldots) \in \Omega$

$$m_n(\omega) = \sup_{(x_1, \ldots, x_n) \in U_m} m$$

setzt. Eine Folge ω ist genau dann zufällig, wenn die Folge

(2) $\quad m_1(\omega), m_2(\omega), \ldots$

beschränkt bleibt. Man kann die nichtzufälligen Folgen nach dem Wachstum der Folge (2) klassifizieren, (derartige Ideen finden sich auch bei Borel [2]).

Wir beweisen nun folgendes Gegenstück zu Theorem 4.1:

Theorem 4.7. Sei $f(n)$ $(n = 1,2,\ldots)$ eine berechenbare Funktion, derart, daß die Reihe $\sum_{n=1}^{\infty} 2^{-f(n)}$ konstruktiv schnell konvergiert in dem Sinne, daß es eine berechenbare Folge $N_1, N_2, \ldots$ von ganzen Zahlen > 0 mit

$$\sum_{n=N_m}^{\infty} 2^{-f(n)} < 2^{-m} \quad (m = 1, 2, \ldots)$$

gibt. Dann gibt es zu jeder zufälligen Folge $(x_1, x_2, \ldots)$ ein $n_o > 0$, derart, daß

$$K(x_1 \ldots x_n | n) \geq n - f(n) \qquad (n \geq n_o)$$

gilt.

Beweis. Wir konstruieren den Sequentialtest $(\mathfrak{W}_0, \mathfrak{W}_1, \ldots)$ mit $\mathfrak{W}_0 = \Omega$, und

$$\mathfrak{W}_m = \{\omega = (x_1, x_2, \ldots) \mid \text{es gibt ein } n \geq N_m \text{ mit } K(x_1 \ldots x_n | n) < n - f(n)\}$$

Die zugehörige Menge $V \subseteq \mathbb{N} \times X$ läßt sich in der Gestalt

$$V = \{(m, (y_1 \ldots y_n)) \mid \text{es gibt ein } n \text{ und ein Programm } p \text{ mit } U(p,m) = x,\ l(p) < n - f(n),\ n \geq N_m\}$$

schreiben und ist daher rekursiv aufzählbar. Nach Satz 4.2 gibt es eine ganze Zahl $c \geq 0$ derart, daß

$$V_{m+c} \subseteq U_m \qquad (m = 0,1,\ldots)$$

gilt. Wenn also $(x_1, x_2, \ldots)$ eine zufällige Folge, und etwa nicht in $\mathcal{U}_m = \mathcal{U}(U_m)$, somit also auch nicht in $\mathcal{U}_{m+c}$ ist, so gibt es nach der Definition der $\mathcal{U}_m$ kein $n \geq N_{m+c}$ mit $K(x_1 \ldots x_n | n) < n - f(n)$.
Es gilt also

$$K(x_1 \ldots x_n | n) \geq n - f(n) \qquad (n \geq N_{m+c}) .$$

Wir bemerken noch, daß man auch eine vernünftige Definition des Begriffs 'zufällige Folge' geben kann, die nicht auf ein bestimmtes Maß wie π festgelegt ist (vgl. Martin-Löf [6]).

Literatur zu §§3,4

[1] B o r e l, E.
Méthodes et problèmes de Théorie des Fonctions
Gauthier-Villars, Paris 1922, 38 - 66
('Sur la classification des ensembles de mesure nulle')

[2] ----, Note VI zu: Leçons sur le Théorie des Fonctions
(Gauthier-Villars, Paris 1928)

[3] C h a i t i n, G.J.
On the length of programs for computing finite binary sequences (1966) (unveröffentlicht)

[4] K l e e n e, S.C.
Introduction to Metamathematics
Amsterdam - Groningen 1952

[5] K o l m o g o r o f f, A.N.
Три подхода к определению понятий "количество информации"
(Drei Vorschläge zur Definition des Begriffs 'Informationsinhalt')
Problemy peredači informacii 1 (1965), 3 - 11

[6] M a r t i n - L ö f, P.
The definition of random sequences
(Forskningsrapport, Institutionen för Försäkringsmatematik och Matematisk Statistik, Universitet Stockholms)
1965

[7] ----, О колебании сложности бесконечных двоичных последовательностей
(Über Schwankungen der Komplexität unendlicher binärer Folgen) 1965 (unveröffentlicht)

[8] S o l o m o n o f f, R.J.
A formal theory of inductive inference, Part I
Information and Control 7 (1964), 1 - 22