KOMPLEXITÄT UND ZUFÄLLIGKEIT

Inaugural - Dissertation

zur

Erlangung des Doktorgrades der Naturwissenschaften

vorgelegt beim Fachbereich Mathematik

der Johann Wolfgang Goethe Universität

zu Frankfurt am Main

von

Peter Gács aus Budapest

Gedruckt mit Genehmigung
des Fachbereichs Mathematik
Dekan: Prof. Dr. H. Dinges
Erster Gutachter: Prof. Dr. C.P.Schnorr
Zweiter Gutachter: Prof. Dr. H. Dinges
Tag der mündlichen Prüfung: 12.12.1978.

Motto: "Ainsi au jeu de croix ou pile,
l'arrivée de croix cent fois
de suite nous paraît extraordinaire,
parce que le nombre presque infini
des combinaisons qui peuvent arriver
en cent coups étant partagé
en séries régulières, ou dans
lesquelles nous voyons régner un
ordre facile à saisir, et en séries
irregulières, celles-ci sont
incomparablement plus nombreuses."

(Laplace)

§ 0. Einleitung

Die klassische Wahrscheinlichkeitstheorie hat wichtige Probleme, die den Begriff der Wahrscheinlichkeit betreffen, ungelöst gelassen. Die Grundsituation der mathematischen Statistik führt uns gleich zu einem der dadurch entstandenen Paradoxe.

Sei uns eine Folge $x=x_1 \ldots x_{100}$ von 0-en und 1-en gegeben und dabei behauptet, daß die 0-en und 1-en die Ergebnisse Kopf und Schrift in einem wirklich vorgegangenen Prozeß des Münzenwerfens widergeben. Wenn die Folge aus lauter 1-en besteht, würden wir die Wahrheit der Behauptung bezweifeln. Dieser Zweifel ist jedoch durch keine allgemeinen Prinzipien der klassischen Wahrscheinlichkeitstheorie begründbar: die Folge 11...1 (100 mal) ist genau so wahrscheinlich wie eine beliebige andere ($P(x)=2^{-100}$).

Dieses Problem hat auch schon Laplace beschäftigt, und er, viele seiner Nachfolger überholend, hat auch schon die qualitativ richtige Antwort gefunden: es gibt insgesamt so wenig "reguläre" Folgen, daß, wenn eine von Ihnen erscheint, wir mit Recht argwöhnisch sein können.

Anderthalb Jahrhunderte brauchte es, bis mit Hilfe der mittlerweile entstandenen Rekursionstheorie die Gedanken von Laplace einen streng mathematischen Sinn erhalten konnten. Kolmogorov, der gegenüber den Versuchen von Mises, die Wahrscheinlichkeitstheorie auf den Begriff der Zufälligkeit zu gründen, in den 30er Jahren noch eher skeptisch war, hat 1965 einen Begriff der Komplexität $K(x)$ der Folge x eingeführt, der die erwünschten Eigenschaften der "Irregularität" von Laplace hat. (Die Priorität gehört eigentlich dem Amerikaner Solomonoff). Gemessen mit $K(x)$ hat eine 0-1-Folge der Länge n eine Komplexität $\preccurlyeq n$ ($\preccurlyeq$ bezeichnet kleiner gleich innerhalb einer additiven Konstante), aber eine Komplexität kleiner als $n-k$ können nur 2^{n-k} Folgen haben. (Für die Definition von $K(x)$ siehe §4). Regularität kann also auch als ein Maß der Nichtzufälligkeit betrachtet werden, wenigstens in dem Fall der Gleichverteilung.

Der Begriff der Komplexität macht zwar schwierige philosophische Probleme lösbar, seine unmittelbare Benutzung stößt aber auf Hindernisse: $K(x)$ ist eine unberechenbare Funktion. Der Autor hat in Kapitel 4 gezeigt, daß $K(x)$ sogar in einem sehr starken Sinne unberechenbar ist: Für alle n gibt es 0-1-Folgen der Länge n mit $K(K(x)|x) \succcurlyeq \log_2 n - \log_2 \log_2 n$. Hier $K(K(x)|x)$ ist die bedingte Komplexität von $K(x)$, wenn x bekannt ist. (Bemerken wir, daß $K(x)$, selbst eine Zahl $\preccurlyeq n$, nie komplizierter als $\log_2 n$ sein kann).

In den folgenden Jahren wurden viele Varianten von der Beschreibungskomplexität empfohlen und untersucht. Eine von ihnen, die 1971 durch Levin und 1974 durch Chaitin vorgeschlagen wurde, ist asymptotisch gleich zu $K(x)$, scheint aber den informationstheoretischen und wahrscheinlichkeitstheoretischen Anwendungen mehr gerecht zu sein. Diese Größe, durch $H_N(x)$ bezeichnet, wurde in §4 untersucht. Wir haben in [12] die Identität

$$H_N(x,y) \asymp H_N(x) + H_N(y|x, H_N(x)) \quad , \quad (0,1)$$

die analog zur Identität

$$\mathcal{H}(X,Y) = \mathcal{H}(X) + \mathcal{H}(Y|X)$$

der Informationstheorie ist, festgestellt. Hier ist $\asymp$ Gleichheit innerhalb einer additiven Konstante, $\mathcal{H}$ die Entropie der Zufallsvariablen X,Y. Daß $H_N(x)$ von der Bedingung in (0.1) nicht weggelassen werden kann, ist in §4 gezeigt, gerade durch das früher erwähnte Ergebnis über $K(K(x)|x)$.

Der Schwede Martin-Löf, der 1965 ein Jahr bei Kolmogorov verbracht hatte, ist mit einem allgemeineren Zugang zur Zufälligkeit aufgekommen. Für ein beliebiges berechenbares Wahrscheinlichkeitsmaß P auf dem Raum der unendlichen 0-1-Folgen Ω führt er den Begriff der konstruktiven Nullmengen ein. Er zeigt, daß die Vereinigung aller konstruktiven Nullmengen selbst eine konstruktive Nullmenge ist. Die Folgen außerhalb dieser Menge sind dann als zufällig erklärt. Seine Konstruktion einer Nullmenge (als der Durchschnitt $\bigcap_m U_m$ von einer rekursiven Folge von konstruktiv offenen Mengen U_m mit $P(U_m) \leq 2^{-m}$) gestattet die Definition einer

Funktion $d_M(\omega|P) = \max\{m|\omega\in U_m\}$, die auch bei den zufälligen Elementen ω die Abweichung von der Zufälligkeit (den Defekt der Zufälligkeit) mißt. (Alle Elemente mit positiver Wahrscheinlichkeit sind zufällig, aber nicht in dem gleichen Maße: $d_M(\omega|P)$ zeigt diese Unterschiede). d_M kann Martin-Löfs Test genannt werden.

Es gibt eine Fülle von Untersuchungen, die zeigen, daß die Zufälligkeit im Sinne von Martin-Löf einer unendlichen 0-1-Folge auch durch das Verhalten der Komplexität der Anfangsstücke von ω charakterisiert werden kann. Das fünfte Kapitel der Arbeit faßt mehrere dieser Ergebnisse in einer einheitlichen Weise zusammen (d_M wird in Komplexitätstermen ausgedrückt), unter ihnen ältere (1971), bisher nicht publizierte, und neuere Ergebnisse des Autors.

Das sechste Kapitel ist einigen Fragen über Levins gleichmäßigen Test d_L gewidmet. Er wird für den Fall der berechenbaren Maße durch eine Formel angegeben, die eine Verallgemeinerung von Chaitins Zufälligkeitstest darstellt. Eine Variante ist weiterhin vorgeschlagen, für die alle gewünschten Eigenschaften von d_L bewiesen werden können.

Ich will meinen Dank an Professor Dr. C.P. Schnorr hier ausdrücken, der mich zu dieser Arbeit angespornt, unterstützt und mir mit wertvollen Bemerkungen geholfen hat. Dank gilt auch meinem Freund, L.A. Levin, für die Mitteilung vieler seiner Ideen, die seinen äußerst lakonischen veröffentlichten Arbeiten zugrunde liegen.

§ 1 Bezeichnungen und Definitionen

Bezeichnungen: Alle Logarithmen und exp-s sind zur Basis 2.

N = Die Menge der natürlichen Zahlen

$N_k = 0,1,\ldots,k-1$

Q = die Menge der rationellen Zahlen

R = die Menge der reellen Zahlen.

$R_+ = R\cap[0,\infty)$

$\bar{R} = R\cup\{\infty\}$, $\bar{R}_+ = R_+\cup\{\infty\}$.

$N_2^* = \bigcup_n N_2^n\cup\{\Lambda\}$ wo Λ das sogenannte leere Wort ist.

Wir setzen eine rekursive ein-eindeutige Korrespondenz $\kappa: N_2^* \to N$ mit $\kappa(x) \geq l(x)$ fest, wo $l(x)$ die Länge des Wortes x ist. $\Omega = N_2^\infty$ ist die Menge der unendlichen binären Folgen. Für $x,y\in N_2^*$, ist xy die Konkatenation von x und y, $x \subset y$ genau dann, wenn $\exists z(xz = y)$.

Für $\omega \in N_2^\infty$, $\omega = \omega_1\omega_2\ldots(\omega_i \in N_2)$ und $\omega^n = \omega_1,\ldots\omega_n$.
Für zwei nichtnegative Funktionen f,g schreiben wir $f \preccurlyeq g$, wenn es eine Konstante $c > 0$ gibt mit $cf \preccurlyeq g$. $f \approx g \Leftrightarrow f \preccurlyeq g$ und $g \preccurlyeq f$, $f \prec g \Leftrightarrow \exp f \preccurlyeq \exp g$, $f \asymp g \Leftrightarrow f \preccurlyeq g$ und $g \preccurlyeq f$.
Wir betrachten N_2^∞ mit seiner gewöhnlichen Topologie bestimmt durch die Basis $\{xN_2^\infty | x\in N_2^*\}$, R mit seiner gewöhnlichen Topologie $\{(r_1,r_2)|r_1,r_2\in Q,\ r_1 < r_2\}$.

Sei $\mathcal{M}$ die Menge aller Wahrscheinlichkeitsmaße über Ω, mit ihrer (schwachen) Topologie bestimmt durch die Subbasis aus

Mengen der Form $\{\mu \in \mathcal{M} \mid r_1 < \mu(x) < r_2\}(r_1, r_2 \in Q)$. Hier ist $\mu(x) := \mu(xN_2^\infty)$. Wir werden N und N_2^* manchmal auch als diskrete topologische Räume betrachten (sie sind im wesentlichen äquivalent durch die Korrespondenz κ). In jedem dieser (lokalkompakten topologischen Räume X ist also eine Basis $\{U_i \mid i \in N\}$ mit einer festgelegten Aufzählung gegeben. Auch gewisse neue Räume, gebildet als Produkte, werden betrachtet.

Definition Ein Element x des Raumes X ist berechenbar, wenn $\{i \mid x \in U_i\}$ (rekursiv) aufzählbar ist. Eine offene Menge $G \subset X$ ist konstruktiv offen, wenn $\{i \mid \bar{U}_i \subset G\}$ aufzählbar ist.
$F \subset X$ ist konstruktiv abgeschlossen, wenn F^c, das Komplement von F konstruktiv offen ist.

Eine Funktion $f: X \to R$ ist halbberechenbar (von unten), wenn $\{(r,x) \mid r \in R, x \in X,\ f(x) > r\}$ konstruktiv offen ist. f ist berechenbar, wenn f und -f halbberechenbar sind.

Ein Maß μ ist berechenbar, wenn es als Element von $\mathcal{M}$ berechenbar ist. Es ist leicht zu sehen, daß diese Forderung äquivalent zur Bedingung ist, daß $\mu(x)$ als eine Funktion über N_2^* berechenbar sei.

§ 2 Martin-Löfs Tests

Definition [1] Sei μ ein berechenbares Maß. Ein Martin-Löf-Test ist eine halbberechenbare Funktion $d: N_2^\infty \to \bar{R}$ mit

$$\forall k \,.\, \mu\{\omega \mid d(\omega) > k\} \leq 2^{-k}$$

Satz 2.1 [Martin-Löf, 1] Unter den Martin-Löf-Tests gibt es einen maximalen im Sinne der Ordnung $\leq$.

Definition Wir legen einen maximalen ML-Test ein für allemal für jedes μ fest und nennen ihn $d_M(\omega \mid \mu)$, der universelle ML-Test. Eine Folge ω heißt zufällig, wenn $d_M(\omega \mid \mu) < \infty$.
Der Begriff der zufälligen Folge ist invarianter als Martin-Löf-Tests. Es ist leicht, verschiedene vernünftige Weisen vorzustellen, in denen man Nichtzufälligkeiten in Folgen messen kann. Die Frage aber, welche Folgen sind zufällig schlechthin, ist durch die meisten Zugänge äquivalent beantwortet.
Wir verallgemeinern ein wenig den Begriff eines Tests.

Definition Für ein berechenbares Maß μ, eine halbberechenbare Funktion $d: \Omega \to \bar{R}_+$ ist ein Test, wenn

$$\lim_m \mu\{\omega \mid d(\omega) > m\} = 0 \text{ „rekursiv“},$$

d.h. es gibt eine rekursive Funktion m(k) mit $\mu\{\omega \mid d(\omega) > m(k)\} \leq 2^{-k}$.
Für einen Test d gilt

$$\{\omega \mid d(\omega) = \infty\} \subset \{\omega \mid d_M(\omega) = \infty\}.$$

Im Falle der Gleichheit sagen wir, daß der Test universell ist. Alle universellen Tests d, mit denen wir zu tun haben werden, sind asymptotisch gleich zu d_M, d.h. es gilt

$$\lim_{d\to\infty} d_M/d = 1 .$$

§ 3 Apriori Wahrscheinlichkeit

Maße μ sind bestimmt durch ihre Werte $\mu(x)$ auf N_2^*. Untere Grenzen von Mengen von Maßen sind nicht mehr durch diese Werte bestimmt, wir interessieren uns aber nicht für ihre Werte anderswo. (Die allgemeine Definition von Halbmaßen findet man in [11]).

Definition [2] Ein Halbmaß über Ω ist eine Funktion $\varphi: N_2^* \to R_+$ mit

$$\varphi(x) \geq \varphi(x0) + \varphi(x1), \quad \varphi(\Lambda) \leq 1.$$

Wir bezeichnen die Menge der Halbmaße durch $\mathcal{S}$. $\mathcal{S}$ erhält dieselbe Topologie wie $\mathcal{M}$, so $\mathcal{M}$ ist ein Unterraum von $\mathcal{S}$. Es ist einfach zu zeigen, daß $\varphi(x) = \inf\{\mu(x) \mid \mu \geq \varphi, \mu \in \mathcal{M}\}$, und daß die untere Grenze von einer beliebigen Menge von Maßen ein Halbmaß ist.

Ein Halbmaß ist halbberechenbar, wenn es als eine Funktion $N_2^* \to R_+$ halbberechenbar ist. Es ist nicht schwer zu sehen, daß φ genau dann halbberechenbar ist, wenn die Menge $\{\mu \in \mathcal{M} \mid \mu \geq \varphi\}$ konstruktiv abgeschlossen in $\mathcal{M}$ ist.

Wir werden auch Halbmaße über die diskreten Räume N, N_2^* betrachten. (Ein Halbmaß über N ist durch eine Funktion $\varphi: N \to R^+$,

$$\Sigma_x \varphi(x) \leq 1$$

gegeben). Halbmaße über N_2^* entsprechen durch κ denen über N.

Satz 3.1 [Levin 2] In der Menge aller halbberechenbaren Halbmaße gibt es ein maximales Element im Sinne der Relation $\leq$.

Wir fixieren ein maximales Halbmaß $M = M_\Omega$ und nennen es die apriori Wahrscheinlichkeit über Ω. Die apriori Wahrscheinlichkeiten über N, N_2^* bezeichnen wir durch M_N (wir schreiben $M_N(x)$ auch für $M_N(\kappa(x))$ ohne Gefahr eines Mißverständnisses).
Wenn keine Mißverständnisse entstehen können, lassen wir die Subscript weg.
Der folgende Satz zeigt die Rolle der apriori Wahrscheinlichkeit bei der Bestimmung der Zufälligkeit.

Satz 3.2 [Levin, 3], siehe auch [Schnorr, 4].
$d_s(\omega|\mu) := \sup_n \log M_\Omega(\omega^n) - \log \mu(\omega^n)$ ist ein universeller Test für ein beliebiges berechenbares Maß μ.

Bemerkung 1. Für ein fixiertes berechenbares Maß μ ist $M_\Omega(\omega^n)/\mu(\omega^n)$ von unten beschränkt. ω ist also zufällig genau dann, wenn $\mu(\omega^n) \approx M_\Omega(\omega^n)$.

2. d_s ist auch für unberechenbare Maße definiert, und man hat $d_s(\omega|\mu) \leq 0$ für alle $\mu \geq M$. Wir können diese Tatsache so interpretieren, daß nach der apriori Wahrscheinlichkeit alle Folgen zufällig sind. In § 6 finden wir, daß diese Behauptung auch für eine breite Klasse von "gleichmäßigen Tests" wahr bleibt.

§ 4 Komplexität

Bezeichnung $H(x) = -\log M(x)$.

Die Zahl $H(x)$ kann aus mehreren Gründen als eine Art Komplexität betrachtet werden. Die erste Definition einer Beschreibungskomplexität stammt von Kolmogorov und Solomonov [5].

Definition Sei $A:N_2^* \times N_2^* \to N_2^*$ eine partielle rekursive Funktion. Wir schreiben

$$K_A(y|x) = \min\{l(p) \mid A(p,x) = y\},$$
$$K_A(y) = K_A(y|\Lambda).$$

Kolmogorov zeigte, daß es eine "optimale" partielle rekursive Funktion U gibt mit $K_U \preccurlyeq K_A$ für eine beliebige andere p.r. A. Wir legen ein solches U fest und schreiben K für K_U. $K(y|x)$ ist die Kolmogorovsche Komplexität von y gegeben x.

Die Kolmogorovsche Komplexität besitzt Eigenschaften, die man intuitiv von einer Komplexität verlangt, andere (z.B. die in der Einleitung erwähnte), die ihr bei der Begründung der Wahrscheinlichkeitstheorie eine Rolle geben, und wieder andere, die sie mit der Entropie der Informationstheorie in eine enge Beziehung setzen. Diese Eigenschaften, wenn sie die Form der Ungleichungen haben, enthalten oft Restglieder logarithmischer Ordnung. Um zu Ungleichungen innerhalb einer additiven Konstante zu kommen, hat es sich als nützlich erwiesen, andere Komplexitäten zu definieren, deren intuitive Bedeutung und numerischer Wert zu denen von $K(x)$ nah ist. Eine dieser Varianten, gefunden durch Levin [6], hat besonders günstige Eigenschaften.

Eine Menge $E \subset N_2^*$ heißt präfixfrei, wenn $x,y \in E \Rightarrow x \not\subset y$.
Eine partielle Funktion $A:N_2^* \times N_2^* \to N_2^*$ ist präfixfrei, wenn für alle y $\{p|A(p,y)$ ist definiert$\}$ eine präfixfreie Menge ist.
Sei $\mathcal{F}$ die Menge aller präfixfreien p.r. Funktionen $A:N_2^* \times N_2^* \to N_2^*$.
In der Menge $\mathcal{F}$ gibt es auch eine "optimale" p.r. T, die die Eigenschaft hat, daß für alle anderen $A \in \mathcal{F}$, $K_T \preccurlyeq K_A$.
K_T ist die neue Komplexität, und ihre wichtigste Eigenschaft ist, daß sie mit dem Logarithmus H_N der apriori Wahrscheinlichkeit M_N über N beinahe übereinstimmt.

Bezeichnung Wir benutzen manchmal statt T die Funktion T' definiert durch $T'(p,x) = y \Leftrightarrow \exists q \subset p.\ T(q,x) = y$.

Satz 4.1 (Levin) $H_N(x) \asymp K_T(x)$.

Satz 4.1 würde auch für bedingte Komplexitäten gelten, wenn wir den Begriff eines bedingten Halbmaßes definiert hätten.

Wir nennen die halbberechenbare Funktion $\varphi(x \mid y)$ ein bedingtes Halbmaß, wenn sie für jedes y ein Halbmaß ist. Unter den bedingten, halbberechenbaren Halbmaßen gibt es ein maximales im Sinne von $\leq$. Wir nennen ein fixiertes maximales bedingtes halbberechenbares Halbmaß $M(x|y)$ die bedingte apriori Wahrscheinlichkeit, $H(x|y) := -\log M(x|y)$. Dann gilt auch $H(x|y) \asymp K_T(x|y)$.

Die drei verschiedenen Arten von Komplexität sind numerisch nicht sehr verschieden. Es gilt bekanntlich

Satz 4.2 a) $K \lesssim H_N \lesssim K + 2 \log K$,

b) $H_\Omega \lesssim H_N$.

c) Für eine beliebige präfixfreie rekursiv aufzählbare Menge $E \subset N_2^*$ gibt es eine Konstante c mit

$$\forall x \in E,\ y \in N_2^*.\ H_N(x|y) \leq H_\Omega(x|y) + c .$$

Die Komplexitäten, die wir bisher definiert haben, sind unberechenbar. Es gilt sogar, wie Kolmogorov gezeigt hat, daß alle partiell rekursive untere Abschätzungen zu K(x) von oben durch eine Konstante beschränkt sind. Chaitin ist in zwei Arbeiten den beweistheoretischen Folgen dieser Ergebnisse nachgegangen.

Da die Beschreibungskomplexitäten immer von oben halbberechenbar sind, sind sie die untere Grenze ihrer berechenbaren oberen Abschätzungen. Die berechenbaren oberen Abschätzungen

von K sind in [2] charakterisiert, die von H_N sind genau die Funktionen f(x) mit $\Sigma_x 2^{-f(x)} < \infty$. (Abschätzung im Sinne von $\lesssim$).

Von nun an werden wir in diesem Kapitel nur mit $H_N(x)$ zu tun haben, darum kann der Index N weggelassen werden.

Betrachten wir jetzt nur die Komplexitäten von natürlichen Zahlen. Können wir etwas mehr über die monotonen oberen Abschätzungen sagen? Ja, wir können sogar die kleinsten bis auf $\asymp$ bestimmen, aber für H wird das wieder keine berechenbare Funktion. Bemerken wir, daß für eine beliebige Funktion f(n) ihre kleinste monotone Abschätzung durch

$$f^*(n) = \max_{k \leq n} f(k)$$

gegeben ist.

Satz 4.3 [5,7] $K^*(n) \asymp \log n$,

$H^*(n) \asymp \log n + H([\log n])$

Bemerkung Die Abschätzung H^* ist zwar nicht berechenbar, aus ihr folgen aber sofort einfache berechenbare Abschätzungen, wie

$$\log n + \log \log n + 2 \log \log \log n.$$

Diese sind leider für große n immer ungenauer.
Der Autor, wie auch R. Solovay (siehe [8]) haben, unveröffentlicht, berechenbare obere Abschätzungen für H gefunden, die auf unendlich vielen Stellen scharf sind.

Satz 4.4 Es gibt eine berechenbare obere Abschätzung f(x) von H(x), für die auf unendlich vielen Stellen die Ungleichung

$$f(x) \leq H(x) + c$$

gilt.

Beweis Sei T der optimale Algorithmus mit $K_T \asymp H_N$.
Es gibt ein berechenbares Prädikat Q(p,x,t) über $N_2^* \times N_2^* \times N$ mit folgenden Eigenschaften.

(i) $T(p) = x \Leftrightarrow \exists t . Q(p,x,t)$

(ii) $Q(p,x,t) \wedge Q(p,x,t') \Rightarrow t = t'$.

Sei z.B. $\mathcal{T}$ eine Turingmaschine, die T berechnet, dann definieren wir Q als " $\mathcal{T}$ hält auf p in t Schritten, und gibt x aus".
Wir definieren die beschränkte Komplexität

$$H^t(x) = \min\{l(p) \leq 2l(x) \mid \exists t' \leq t . Q(p,x,t')\}.$$

Es gibt eine rekursive Funktion $g: N \times N_2^* \to N$ und $c < \infty$ mit

$$H^{t-1}(x) > H^t(x) \Rightarrow H^{g(t,x)}(\langle x,t\rangle) \leq H^t(x) + c. \qquad (4.4)$$

(Hier ist $\langle\ ,\ \rangle$ eine rekursive Paarungsfunktion, $\langle\ \rangle_1$, $\langle\ \rangle_2$ ihre Umkehrung). Bei $H^{t-1}(x) > H^t(x)$ kann nämlich aus einem kürzesten t-beschränkten Programm auch t bestimmt werden.
Sei nun

$$f(z) = H^{g(\langle z\rangle_1, z)}(z).$$

f ist die Länge eines Programms für z, darum eine obere Abschätzung für H. Sei jetzt z von der Form $\langle x,t\rangle$, mit $H(x) = H^t(x) < H^{t-1}(x)$. Dann folgt aus (4.4) $f(z) \leq H(z) + c$,

Q.E.D.

Wie hoffnungslos H(x) (wie auch K(x) unberechenbar ist, ist auch an dem folgenden Satz zu sehen.

Satz 4.5 [12] Für alle n gibt es ein $x \in N_2^n$ mit

$$H(H(x) \mid x) \geq \log n - \log\log n - c$$

für eine Konstante $c < \infty$.

Bemerkung H(x), als eine Zahl kleiner als n + 2 log n, hat eine Komplexität, die sogar ohne Bedingung nicht größer als log n + 2 log log n ist. Das zeigt, daß x bei der Berechnung von H(x) manchmal fast gar nichts hilft.

Beweis Sei $s \leq \log n$ eine Zahl mit der Eigenschaft

$$\forall x \in N_2^n \exists p \in N_2^s . T'(p,x) = H(x).$$

Wir müssen die Ungleichung

$$s \geq \log n - \log\log n \qquad (4.5)$$

beweisen.

Wir sagen, daß $p \in N_2^s$ annehmbar für $x \in N_2^n$ ist, wenn es ein $k = T'(p,x)$ und ein $q \in N_2^k$ gibt mit $T'(q,\Lambda) = x$.
Bezeichnen wir durch $M_i(n,s) = M_i$ die Menge aller $x \in N_2^n$ mit wenigstens i verschiedenen $p \in N_2^s$, die annehmbar für x sind.
Betrachten wir die Folge

$$N_2^n = M_0 \supset M_1 \supset \ldots \supset M_j \supset M_{j+1} = \emptyset$$

wo $M_j \neq \emptyset$. Offenbar gilt

$$j \leq 2^s \qquad (4.6)$$

Wir zeigen jetzt, daß

$$\log \#M_i \preccurlyeq \log \#M_{i+1} + 5 \log n. \qquad (4.7)$$

Dazu finden wir ein $x \in M_i - M_{i+1}$ mit

$$\log \#M_i - 1 \leq H(x) \preccurlyeq \log \#M_{i+1} + 5 \log n. \qquad (4.8)$$

Die zweite Ungleichung in (4.8) wird gelten, weil wir für x ein Programm mit der Länge auf der rechten Seite von (4.8) schreiben. Wir schicken einige Bemerkungen voraus.

1. Die Menge

   $\{(i,x,n,s) \mid x \in M_i(n,s)\}$

   ist aufzählbar. Aus den Zahlen $i,n,s,\#M_{i+1}$ läßt sich darum die Menge M_{i+1} bestimmen.

2. Für Elemente x der Menge $M_i - M_{i+1}$ ist H(x) aus i, n, s und x berechenbar: das kürzeste Programm zu x ist unter denen, die durch die genau i verschiedenen annehmbaren p geliefert werden. Wir suchen nach annehmbaren p, bis wir schon i Stück gefunden haben.

3. Wir können die Ungleichung

   $$\log(\#M_i - \#M_{i+1}) \geq \log \# M_i - 1 \qquad (4.9)$$

   voraussetzen, sonst ist (4.7) trivial.

Das Programm arbeitet, wie folgt.

1. Es benutzt die Zahlen $n,i,s,[\log \#M_i]$ und $\#M_{i+1}$. Dazu werden der Reihe nach folgende Programmlängen gebraucht: $\log n + 2 \log\log n$, $\log n + 2 \log\log n$, $2 \log\log n$, $\log n + 2 \log\log n$, $\log \#M_{i+1} + \log n + 2 \log\log n$.

2. Es beginnt mit der Aufzählung von $M_i - M_{i+1}$. Für alle x aus dieser Menge berechnet es H(x). Wegen (4.9) kommt früher oder später ein erstes Element x mit $H(x) \geq [\log \#M_i] - 1$ vor. Es wird als Ergebnis ausgegeben.

Die Gesamtlänge unseres Programms beträgt also

$$\text{Konstante} + \log \# M_{i+1} + 4 \log n + 10 \log\log n \preccurlyeq$$
$$\preccurlyeq \log \# M_{i+1} + 5 \log n.$$

Damit ist (4.7) bewiesen. Aus ihr folgt

$$j \gtrsim \frac{n}{5 \log n},$$

das gibt mit (4.6) die gewünschte Ungleichung (4.5).

Q.E.D.

Die Komplexitäten zeigen eine weitgehende Analogie und auch interessante numerische Beziehungen zur Entropie $\mathcal{H}(X)$ eines Zufallsvariablen X. Die vorhandenen Aussagen sind genauer und inhaltlicher für H(x) als für K(x), darum beschränken wir uns in bezug auf K(x) auf zusätzliche Bemerkungen.

Definieren wir

$$I(x;y) = H(y) - H(y|x).$$

Es gilt die Ungleichung wie in der Informationstheorie

$$H(x,y) \lesssim H(x) + H(y|x) \qquad (4.10).$$

Bemerken wir, daß für K diese Ungleichung nicht mit dieser Genauigkeit gilt.

Die volle Analogie hört auf, wenn wir die Gleichung

$$\mathcal{H}(X) + \mathcal{H}(Y|X) = \mathcal{H}(X,Y)$$

betrachten. Stattdessen haben wir folgenden Satz

Satz 4.6 [Gács - Levin, 12]

$$H(x,y) \asymp H(x) + H(y|x,H(x)). \qquad (4.11)$$

Beweis

a) ($\lesssim$). Wenn wir das kürzeste Programm für x haben, ist uns auch seine Länge, H(x) bekannt. Darum, um y zu bestimmen, genügt noch ein zusätzlicher Programmteil der Länge H(y|x,H(x)).

b) ($\gtrsim$).
Die Funktion

$$\varphi(y|x,k) = \exp(k-H(x,y))$$

ist halbberechenbar von unten, und es gilt

$$\forall x.\Sigma_y \varphi(y|x,H(x)) \leq 1. \qquad (4.12)$$

(4.12) folgt aus

$$M(x) \leq M(x,\Lambda) \leq \Sigma_y M(x,y).$$

Darum gibt es ein bedingtes halbberechenbares Halbmaß $\overline{\varphi} \leq \varphi$ mit $\forall x.\, \overline{\varphi}(y|x,H(x)) \approx \varphi(y|x,H(x))$. Wir haben

$$\overline{\varphi}(y|x,k) \leq M(y|x,k).$$

Deshalb

$$H(y|x,H(x)) \lesssim H(x,y) - H(x). \qquad \text{Q.E.D.}$$

Korollar 1 $H(x,y) \asymp H(x,H(x),y)$

Korollar 2 $H(x) + H(y|x) - H(x,y) \asymp I(H(x);y|x) \lesssim$

$$\lesssim H(H(x)|x) \lesssim \log H(x) + 2 \log\log H(x),$$

mit $I(x;y|z) := H(y|z) - H(y|x,z)$.

Korollar 3 $I(x;y) - I(y;x) \asymp I(H(y);x|y) - I(H(x);y|x)$

Wie wir sehen, werden die Identitäten der Informationstheorie nicht genau erfüllt. Wir können aber allen Beziehungen auch die übliche informationstheoretische Form geben, wenn wir, wie Chaitin, auf die Halbberechenbarkeit der bedingten Komplexität verzichten und die neuen Größen

$\overline{x} = \langle x,H(x)\rangle$, $\overline{H}(y|x) = H(y|\overline{x})$, $\overline{I}(x;y) = \overline{H}(y) - \overline{H}(y|x) =$
$= H(x) + H(y) - H(x,y)$

einführen. Dann haben wir

$$\overline{H}(x) + \overline{H}(y|x) \asymp \overline{H}(x,y),$$
$$\overline{I}(x;y) \asymp \overline{I}(y;x).$$

$\bar{H}(y|x)$ ist auch $\asymp$ der Länge des kürzesten Programms, das y aus einem kürzesten Programm von x berechnet.

Daß die linke Seite in Korollar 2 nicht $\asymp 0$ ist, ist viel weniger trivial zu beweisen als die entsprechende Tatsache für K(x). (Siehe [2] für Beispiele). Wir müssen Satz 4.5 benutzen. Wir zeigen jetzt, daß auf einer unendlichen Menge von Stellen die linken Seiten in Korollar 2 und 3 asymptotisch gleich log H(x,y) sind.
Wir haben

$$H(x) + H(\bar{x}|x) - H(x,\bar{x}) \asymp H(H(x)|x).$$
$$I(x;\bar{x}) - I(\bar{x};x) \asymp -H(H(x)|x).$$

Daß $H(H(x)|x)$ manchmal die Größe von $\log H(x,\bar{x})$ asymptotisch erreicht, folgt aus Satz 4.5.

Korollar 4 [Levin, 12] $I(x;y)$ ist sogar asymptotisch nicht symmetrisch.

Beweis Setzen wir voraus: $l(x) = n$, n groß genug, c eine Konstante

$$H(H(x)|x) \geqslant \log n - \log\log n - c .$$

Wir betrachten x, $\bar{x}$ und H(x). Folgende Beziehungen sind unmittelbar evident.

$$I(x;H(x)) \preccurlyeq 3 \log\log n \prec \log n - \log\log n \preccurlyeq I(\bar{x};H(x)),$$
$$I(H(x);x) \asymp I(H(x);\bar{x})$$

Die Symmetrie wird darum entweder auf dem Paar x,H(x) oder auf dem Paar $\bar{x}$,H(x) stark verletzt.

Q.E.D.

Die Komplexitäten H(x) und K(x) sind also nicht nur unberechenbar, sondern tragen manchmal auch noch viel zusätzliche Information, die aus x nicht ermittelbar ist. Das ist die einzige Ursache der Nichterfüllung der genauen Analogien der informationstheoretischen Identitäten für H(x). Ein Umstand kann uns aber einigermaßen trösten: daß der Anteil der Folgen x mit großem $H(H(x)|x)$ sehr gering ist, sogar egal, mit welchem halbberechenbaren Halbmaß wir es messen.

Sei ν die charakteristische Funktion einer rekursiv aufzählbaren Menge, die die Eigenschaft

$$H_N(\nu^{2^n}|n) \asymp n \qquad (4.13)$$

hat. (Siehe [2].)
Bemerken wir, daß die Information, die ν über eine endliche Folge x liefert,

$$I(\nu;x) := H(x) - H_N(x|\nu),$$

wegen der Bemerkung nach Satz 4.1 definiert ist.

Satz 4.7 [Levin, 12]

$$H(H(x)|x) \preccurlyeq I(\nu;x) + 3 \log I(\nu;x)$$

Beweis Im Beweis des Satzes 4.4 haben wir die beschränkte Komplexität $H^t(x)$ eingeführt.

Sei $\nu_n(t)$ eine rekursive Funktion von n und t mit Werten aus $\{0,1\}$, monoton wachsend in t mit $\lim_t \nu_n(t) = \nu_n$.
Sei $t_0(x) = \min\{t | H^t(x) = H(x)\}$,
$t_1(n) = \min\{t | \nu^n(t) = \nu^n\}$,
$k = k(x) = \max\{i | t_1(2^i) < t_0(x)\}$, $\nu(i) = \nu^{2^i}$.

Bezeichnung Wir führen folgende anschaulichen Bezeichnungen ein

$$x \sqsubseteq y \bmod z \Leftrightarrow H(y|x,z) \asymp 0$$

$$x \equiv y \bmod z \Leftrightarrow x \sqsubseteq y \text{ und } y \sqsubseteq x \bmod z .$$

Sie sind natürlich nur für Funktionen $x(t)$, $y(t)$, $z(t)$ sinnvoll.

Lemma 4.1

$$H(x) \equiv \nu(k(x)) \bmod (x, k(x)) \tag{4.14}$$

Beweis Für $\sqsubseteq$ zeigen wir zunächst, daß

$$H(H(x)|\, x, \nu(k+1)) \asymp 0 \tag{4.15}$$

In der Tat, man kann aus $\nu(k+1)$ zunächst

$$t_1(2^{k+1}) \gtrsim t_0(x),$$

weiter daraus und x auch $H(x)$ bestimmen.

Bemerken wir dann, daß

$$H(\nu(k+1)|\, \nu(k), k) \asymp 0. \tag{4.16}$$

Nach einer Variante unserer informationstheoretischen Identität (4.11), $H(x|y, H(y|z)$, $H(x,y|z) - H(y|z)$, und (4.13) hat man nämlich

$$H(\nu(k+1)|\, \nu(k), k) \asymp H(\nu(k+1)|\nu(k), H(\nu(k)|k)) \asymp$$

$$\asymp H(\nu(k+1), \nu(k)|k) - H(\nu(k)|k) \asymp k - k \asymp 0$$

Aus (4.15) und (4.16) folgt $\sqsubseteq$.

$\sqsupseteq$ gilt, weil aus x und $H(x)$ auch

$$t_0(x) > t_1(2^k)$$

bestimmbar ist. Q.E.D.

Wir setzen den Beweis von Satz 4.7 fort.

Aus Lemma 4.1 und (4.13)

$$H(H(x)|\, x) \lesssim H(\nu(k)|\, x) + H(H(x)|\, \nu(k), x) \lesssim H(\nu(k)) \tag{4.17}$$

Andererseits folgt, daß $\overline{x}$ eine fast volle Information über $\nu(k)$ hat:

$$I(\overline{x}; \nu(k)) \asymp H(\nu(k)) - H(\nu(k)|\, \overline{x}) \gtrsim$$

$$\gtrsim H(\nu(k)) - H(\nu(k)|\, \overline{x}, k) - H(k) \asymp$$

$$\asymp H(\nu(k)) - H(k)$$

Sei $\Delta = H(\nu(k)) - H(\nu(k)|k) = I(k; \nu(k))$. Dann $\Delta \lesssim H(k), H(\Delta) \lesssim 2 \log \log k$.

Weitere Umformungen benutzen die Identität $I(\overline{x};y) \asymp I(\overline{y};x)$.

$$H(\nu(k)) - H(k) \lesssim I(\overline{x}; \nu(k)) \asymp I(\overline{\nu(k)}; x) \lesssim I(\nu; x) + H(k) + H(\Delta).$$

Daher $I(\nu;x) \gtrsim H(\nu(k)) - 2H(k) - H(\Delta) \gtrsim H(\nu(k)) - 3 \log k$ (4.18)

Aus (4.17) und (4.18),

$$H(H(x)|x) \lesssim H(\nu(k)) \lesssim I(\nu;x) + 3 \log I(\nu;x).$$

Q.E.D.

Bemerkung Eine einfache Umformung gibt auch

$$H(H(x)|\, x) - 3 \log H(H(x)|\, x) \lesssim I(\nu;x). \tag{4.19}$$

Korollar Für alle $i > 0$ gilt

$$M\{x | H(H(x)|\, x) \gtrsim i\} \lesssim i^3 \cdot 2^{-i} \tag{4.20}$$

Beweis $I(\nu;x) = H(x) - H(x|\nu) \asymp \log(M(x|\nu)/M(x))$.

Da $M(\cdot|\nu)$ ein Halbmaß ist, haben wir

$$M\{x|M(x|\nu)/M(x) \geq j\} < j^{-1} \qquad (4.21)$$

für alle j. (4.21) folgt jetzt aus (4.19) und (4.21).

Dieses Korollar besagt, daß die apriori Wahrscheinlichkeit all derjenigen x, die eine schwer bestimmbare Komplexität haben, gering ist. Die Wahrscheinlichkeit wird dann klein auch durch ein beliebiges berechenbares Maß μ gemessen, weil M dieses Maß μ im Sinne von $\leq$ majoriert.

§5 Darstellung der Tests durch Komplexitäten

Satz 3.2 gibt zwar eine höchst befriedigende Charakterisierung der Zufälligkeit durch das Verhalten der apriori Wahrscheinlichkeit (ihr Logarithmus ist eine Art Komplexität), es bleiben aber noch technisch interessante Fragen zurück

a) auch Martin Löfs Tests selbst durch Komplexität auszudrücken,
b) zu klären, wieweit andere Komplexitäten zur Charakterisierung der Zufälligkeit geeignet sind.
   Wir beginnen mit einem Ausdruck für die Kolmogorovsche Komplexität durch H.

Satz 5.1 (Levin)

$$K(x) \prec \min\{i|H_N(x|i) \leq i\} \quad \text{und} \qquad (5.1)$$

$$K(x) \prec H_N(x|K(x)) \qquad (5.2)$$

Bemerkung $K(x|y)$ ist durch die natürliche Verallgemeinerung dieser Formel gegeben.

Beweis $H_N(x|K(x)) \preccurlyeq K(x)$
ist offensichtlich aus der Definition dieser Größen. Wenn wir ein i mit $H_N(x|i) \leq i$ haben, dann gilt auch $K(x) \preccurlyeq i$.
Sei nämlich

$$A(p) = T'(p,l(p)).$$

Für beliebige i, p, x mit $l(p) \leq i$, $T(p,i) = x$

hat man

$$A(p0^{i-l(p)}) = x.$$

Darum gilt $K(x) \preccurlyeq K_A(x) \preccurlyeq i$.

(5.2) ist leicht zu sehen. Q.E.D.

Dieser Satz gibt uns die Idee zu der folgenden Definition einer Hilfskomplexität.

Definition $\tilde{H}(x;k) = \min\{i|H(x|k - i) \leq i\}$.

Die Definition ist sinnvoll für sowohl H_N als auch H_Ω.
Ähnlich zu (5.2) haben wir die Identität

$$\tilde{H}(x;k) \prec H(x|k - \tilde{H}(x;k)). \qquad (5.3)$$

Bemerkung 1. $\tilde{H}(x;k|y)$ kann analog definiert werden. Satz 5.1 zeigt uns dann, daß

$$\tilde{H}_N(x;k|k) \prec K(x|k) \qquad (5.4)$$

gilt. $\tilde{H}_N$ kann also als eine Verallgemeinerung von K betrachtet werden.

2. $\tilde{H}_\Omega \lesssim \tilde{H}_N$. (5.5)

In der Tat haben wir

$$H_\Omega(x|k-i) \lesssim H_N(x|k-i),$$

aus $H_N \leq i$ folgt somit $H_\Omega \lesssim i$.

3. $\tilde{H}$ ist offenbar halbberechenbar von oben.

Wir haben in Satz 3.2 gesehen, daß die Prüfung der Zufälligkeit natürlich mit einer Vergleichung von $-\log \mu(\omega^n)$ und irgend einer Komplexität von ω^n verbunden ist.

Bezeichnung $l_\mu(\omega^n) = [-\log \mu(\omega^n)]$.

Satz 5.2 Für ein festgelegtes berechenbares Maß μ hat man

$$d_M(\omega|\mu) \asymp \sup_n l_\mu(\omega^n) - \tilde{H}_\Omega(\omega^n; l_\mu(\omega^n)) \asymp$$
$$\asymp \sup_n l_\mu(\omega^n) - \tilde{H}_N(\omega^n; l_\mu(\omega^n)).$$

Beweis Seien die Ausdrücke in Satz 5.2 entsprechend mit d_{ML} und d_{MC} bezeichnet. Nach (5.5) haben wir $d_{MC} \lesssim d_{ML}$.

Wir müssen zeigen, daß $d_{ML} \lesssim d_M$ und $d_M \lesssim d_{MC}$. $d_{ML} \lesssim d_M$ wird bewiesen, wenn wir zeigen, daß d_{ML} halbberechenbar ist (das ist klar von der Definition) und daß $\forall m.\ \mu\{\omega | d_{ML}(\omega|\mu) > m\} \leq 2^{-m}$ gilt.

Wir benutzen die Identität (5.3). Dann haben wir mit einem $c > 0$:

$$\{\omega \mid d_{ML}(\omega|\mu) = m\} \subset$$
$$\subset \{\omega | \exists n, -\log \mu(\omega^n) - H(\omega^n|m) \geq m - c\} =$$
$$= \{\omega | \exists n.\ M_\Omega(\omega^n|m)/\mu(\omega^n) \geq 2^{m-c}\}.$$

Folgendes einfache Lemma gilt für alle Halbmaße φ und Maße μ.

Lemma 5.1 $\mu\{\omega | \exists n.\ \varphi(\omega^n)/\mu(\omega^n) > \alpha\} < \alpha^{-1}$.

Beweis Sei $n(\omega)$ das erste n (wenn es existiert) mit $\varphi(\omega^n)/\mu(\omega^n) > \alpha$.

Die Menge

$$A = \{x | \exists \omega.\ x = \omega^{n(\omega)}\}$$

ist präfix-frei. Darum gilt

$$\Sigma_{x \in A} \varphi(x) \leq 1$$

für ein beliebiges Halbmaß φ. Andererseits, wegen der Definition von A, hat man

$$x \in A \Rightarrow \mu(x) \leq \varphi(x) \cdot \alpha^{-1},$$

$$\mu\{\omega | \exists n.\ \varphi(\omega^n)/\mu(\omega^n) > \alpha\} = \Sigma_{x \in A} \mu(x) \leq \alpha^{-1}.$$ Q.E.D.

Wenn wir jetzt das Lemma auf das Halbmaß $M_\Omega(\omega^n|m)$ anwenden, bekommen wir

$$\mu\{\omega \mid d_{ML}(\omega|\mu) = m\} \leq 2^{-m+c}.$$

Jetzt zeigen wir $d_M \lesssim d_{MC}$.

Wegen der Halbberechenbarkeit von d_M gibt es eine rekursive Folge $\{(m_t, x_t)\}_{t\in N}$ von Elementen von $N\times N_2^*$ mit

$$\{(m_t,x_t) \mid t\in N\} = \{(m,x) \mid \forall \omega \in xN_2^\infty . d_M(\omega|\mu) > m\}.$$

Sei $t(m,\omega)$ das erste $t\in N$ (wenn es existiert) mit $\omega \in x_t N_2^\infty$ und $m \le m_t$. Die Menge

$$U = \{(m,x) \mid \exists \omega .\ x = x_{t(m,\omega)}\}$$

ist rekursiv aufzählbar,

$U_m = \{x \mid (m,x)\in U\}$ ist eine präfixfreie Menge von Folgen,

$U_m N_2^\infty = \{\omega \mid d_{ML}(\omega|\mu) \ge m\}$.

Darum gilt

$$\Sigma_{x\in U_m} \mu(x) \le 2^{-m}.$$

Dann ist aber die Funktion

$$\varphi(x|m) = \begin{cases} \mu(x)\cdot 2^m \text{ für } (m,x)\in U \\ 0 \quad \text{sonst} \end{cases}$$

ein bedingtes halbberechenbares Halbmaß über N_2^*, und wir haben $M_N(x|m) \stackrel{\times}{\ge} \varphi(x|m)$.

Darum gibt es für alle ω mit $d_M(\omega|\mu) > m$ ein $n(= l(x_{t(m,\omega)}))$ mit

$$H_N(\omega^n|m) \stackrel{+}{\le} -\log \mu(\omega^n) - m. \qquad (5.6)$$

Setzen wir $i = -\log \mu(\omega^n) - m$ in der Definition von $\tilde{H}$, dann folgt weiter aus (5.6) $\tilde{H}_N(\omega^n;[-\log \mu(\omega^n)]) \le -\log \mu(\omega^n) - m$.

Q.E.D.

Dieser Satz enthält als Spezialfall die erste bekannte exakte Relation zwischen Tests und Komplexitäten [10]. Sei π_n die Gleichverteilung über N_2^n. Ein universaler ML-Test $d_M(x|\pi_n)$ ist definierbar als maximale (in bezug auf $\stackrel{+}{\le}$), halbberechenbare Funktion $d(x|\pi_n)$ von x und n mit

$$\forall n,k.\ \pi_n\{x \mid d(x|\pi_n) \ge k\} \le 2^{-k}.$$

Korollar 1 [10] $d_M(x|\pi_n) \stackrel{+}{=} n - K(x|n)$.

Beweis Satz 5.2 bleibt natürlich wahr, auch für N_2 statt Ω und sogar, wenn wir n als Parameter überall in den Gleichungen lassen. Wir haben dann für ein beliebiges $x\in N_2^n$

$d_M(x|\pi_n) \stackrel{+}{<} -\log \pi_n(x) - H_N(\omega^n;[-\log \pi_n(x)]|n) =$

$= n - H_N(x;n|n) \stackrel{+}{<} n - K(x|n)$ wegen (5.4). Q.E.D.

Vom ersten Teil des Satzes 5.2 können wir auch Satz 3.2 unmittelbar beweisen.

Korollar 2 d_s ist asymptotisch gleich d_M.

Beweis Wir benutzen die bekannte Ungleichung

$H(x) \preccurlyeq H(x|j) + H_N(j) \preccurlyeq H(x|j) + 2 \log j$

Setzen wir $k = l_\mu(\omega^n)$, $\Delta = k - \tilde{H}(\omega^n;k)$.

Wir haben wegen (5.3) $k - H(\omega^n) \preccurlyeq k - H(\omega^n|\Delta) \asymp k - H(\omega^n;k)$,

folglich $d_S \preccurlyeq d_M$.

Aber

$k - \tilde{H}(\omega^n;k) \asymp k - H(\omega^n|\Delta) \preccurlyeq k - H(\omega^n) + 2 \log\Delta$,

also $\Delta - 2 \log\Delta \preccurlyeq k - H(\omega^n)$

$d_M - 2 \log d_M \preccurlyeq d_S \preccurlyeq d_M$,

was auch die asymptotische Gleichheit von d_S und d_M zeigt.

Q.E.D.

Der Beweis macht keinen Unterschied zwischen H_Ω und H_N, somit haben wir auch folgendes bewiesen.

Korollar 3 $d_C(\omega|\mu) = \sup_n l_\mu(\omega^n) - H_N(\omega^n)$

ist ein universeller Test, asymptotisch gleich zu d_M.

Dieser Test wurde für den Fall der Gleichverteilung durch Chaitin [7] vorgeschlagen. Schnorr hat seine Universalität bewiesen. Eine andere Charakterisierung von d_C ist in dem folgenden Satz enthalten.

Satz 5.3 Sei $t_C(\omega|\mu) = 2^{d_C(\omega|\mu)} = \sup_n \frac{M_N(\omega^n)}{\mu(\omega^n)}$.

Für ein beliebiges festes berechenbares Maß μ, $t_C(\omega|\mu)$ ist $\leq$-maximal unter den halbberechenbaren Funktionen $t(\omega)$ mit der Eigenschaft

$$\int t(\omega)\mu(d\omega) \leq 1.$$

Bemerkung Die Konstante in $\leq$ hängt von μ ab.

Beweis Wir haben

$$\sup_n \frac{M_N(\omega^n)}{\mu(\omega^n)} \leq \Sigma_n \frac{M_N(\omega^n)}{\mu(\omega^n)} = \Sigma_{x \in N_2^*} M_N(x) \cdot \frac{g_x(\omega)}{\int g_x(\omega)\mu(d\omega)} .$$

Hier $g_x(\omega) = \begin{cases} 1 \text{ für } x \subset \omega \\ 0 \text{ sonst.} \end{cases}$

Darum

$$\int t_C(\omega|\mu)\mu(d\omega) \leq \Sigma_x M_N(x) \leq 1.$$

Sei jetzt $t(\omega)$ eine beliebige halbberechenbare Funktion mit $\int t(\omega)\mu(d\omega) \leq 1$.

Es gibt dann eine rekursive Folge (x_i,k_i) mit

$$\sup_i 2^{k_i} \cdot g_{x_i}(\omega) \asymp \Sigma_i 2^{k_i} g_{x_i}(\omega) \asymp t(\omega).$$

Darum $2^{k_i}\mu(x_i) \leq M_N(i)$

$$2^{k_i} \leq \frac{M_N(i)}{\mu(x_i)} .$$

Wir haben

$$t(\omega) \leq \sup_i \frac{M_N(i)}{\mu(x_i)} \cdot g_{x_i}(\omega) \leq \sup_x \frac{M_N(x)}{\mu(x)} \cdot g_x(\omega) = t_C(\omega|\mu).$$

Q.E.D.

In dem nächsten Satz benutzen wir Kolmogorovs Komplexität, um das Hauptglied in Martin-Löfs Test auszudrücken.

Satz 5.4 Sei $\Delta(\omega^n|\mu) = l_\mu(\omega^n) - K(\omega^n|n, l_\mu(\omega^n))$.
Dann gilt

$$d_M(\omega|\mu) \asymp \sup_n \Delta(\omega^n|\mu) - \tilde{H}(n, l_\mu(\omega^n); \Delta(\omega^n|\mu)).$$

Wie im Korollar zu Satz 5.2 bekommen wir einen einfacheren universellen Test, in dem K auftritt, wenn wir nicht fordern, daß er gerade Martin-Löfs Test sei.

Korollar 1 $d_K(\omega|\mu) := \sup_n \Delta(\omega^n|\mu) - H(n, l_\mu(\omega^n))$
ist ein universeller Test, asymptotisch gleich zu d_M.

Die Form dieses Testes ist besonders einfach für die Gleichverteilung.

Korollar 2 $\sup_n n - K(\omega^n|n) - H(n)$
ist ein Test für die Gleichverteilung. Eine Folge ω ist also gerade dann zufällig nach dem Lebesgue-Maß, wenn
$n \preccurlyeq K(\omega^n|n) + H(n)$
gilt.

Beweis des Satzes 5.4

Zuerst beweisen wir, daß der Ausdruck d_{MK} in dem Satz einen Martin-Löf-Test definiert. Seine Halbberechenbarkeit ist klar aus seiner Definition. Wir müssen zeigen, daß

$$\mu\{\omega \mid d_{MK}(\omega|\mu) > m\} \leq 2^{-m}.$$

Mit einer Konstante c haben wir, wie im Beweis des Satzes 5.2,

$$\{\omega \mid d_{MK}(\omega|\mu) = m\} \subseteq$$
$$\subseteq \{\omega \mid \exists n\ \Delta(\omega^n|\mu) - H(n, l_\mu(\omega^n)|m) \geq m - c\} \subseteq$$
$$\subseteq \bigcup_{n,k} \{\omega \mid l_\mu(\omega^n) = k,\ k - K(\omega^n|n,k) - H(n,k|m) \geq m - c\}.$$

Wir können weiter so abschätzen:

$$\mu\{\omega \mid l_\mu(\omega^n) = k,\ K(\omega^n|n,k) \leq k - m - H(n,k|m) + c\} \leq$$
$$\leq 2^{-k} \#\{x \in N_2^n \mid K(x|n,k) \leq k - m - H(n,k|m) + c\} \leq 2^{-m+c} M(n,k|m)$$

wegen bekannter Eigenschaften von Kolmogorovs Komplexität.
Darum endlich

$$\mu\{\omega \mid d_{MK}(\omega|\mu) = m\} \leq \Sigma_{n,k} 2^{-m+c} \cdot M(n,k|m) \lesssim 2^{-m+c}.$$

Damit ist $d_{MK} \preccurlyeq d_M$ bewiesen.

Der Beweis von $d_M \preccurlyeq d_{MK}$ beruht auf der folgenden Abschätzung, die gleich aus der Definition der Kolmogorov-Komplexität folgt.

Für beliebige $x \in N_2^*$, $\omega \in xN_2^\infty$, $n \geq l(x)$
$$K(\omega^n|n,x,l_\mu(\omega^n)) \preccurlyeq l_\mu(\omega^n) + \log \mu(x). \qquad (5.7)$$

Benutzen wir jetzt die Aufzählbarkeit der Menge U, die wir im Beweis des Satzes 5.2 definiert haben. Wir definieren das bedingte halbberechenbare Halbmaß φ über N^2 durch

$$\varphi(n,k|m) = \begin{cases} 2^{m-1} \cdot \mu(\kappa^{-1}(n)), \text{ falls } (m, \kappa^{-1}(n)) \in U, \text{ und} \\ \qquad 2^{-k-1} < \mu(\kappa^{-1}(n)) < 2^{-k+1} \\ 0 \text{ sonst.} \end{cases}$$

Wenn $d_M(\omega|\mu) > m$, dann für ein $x \in U_m$ mit $n = \kappa(x)$, $k = l_\mu(\omega^n)$ haben wir wegen (5.7):

$$K(\omega^n|n,k) \preccurlyeq k + \log \varphi(n,k|m) - m \preccurlyeq k - H(n,k|m) - m$$
$$H(n,k|m) \preccurlyeq \Delta(\omega^n|\mu) - m$$
$$H(n,k;\Delta(\omega^n|\mu)) \preccurlyeq \Delta(\omega^n|\mu) - m,$$

durch die Schlußweise, die auch am Ende des Beweises von Satz 5.2 benutzt war. Q.E.D.

§6. Gleichmäßige Tests

In den vorangehenden Kapiteln haben wir über Tests $d(\omega|\mu)$ gesprochen, die von ω und μ abhängen. μ war aber dabei immer ein festes berechenbares Maß und an d waren Forderungen gestellt nur als an eine Funktion von ω. Viele Wahrscheinlichkeitstheoretiker würden die Einschränkung zu berechenbaren Maßen für ungerechtfertigt halten. Für ein Experimentalergebnis z.B. wäre es unnatürlich, die passenden Verteilungen nur unter den berechenbaren zu suchen.

Levin hat in [11] einen allgemeinen gleichmäßigen Test $d_L(\omega|\mu)$ definiert, der für ein jedes berechenbares Maß ein universeller Test ist. Bemerken wir, daß ein jeder Ausdruck durch Komplexitäten für Tests, die wir in §5 getroffen haben, im Prinzip als eine Definition für gleichmäßige Tests betrachtet werden kann. Sie sind halbberechenbar in (ω,μ), haben gewisse Normiertheitseigenschaften für festes μ und sind sogar in gewissem Sinn equivalent, solange nur berechenbare Maße betrachtet werden. Das ist nicht mehr offensichtlich für unberechenbare μ und in der Tat, Levins Test $d_L(\omega|\mu)$ entdeckt mehr Unzufälligkeiten als z.B. $d_S(\omega|\mu)$.

Wir wollen nicht alle Definitionen aus [11] wiederholen, die für eine selbsttragende Definition eines gleichmäßigen Tests (L-Test) notwendig sind. In [11] ist ein Halbmaß als ein konkaves Funktional über $C(\Omega)$ definiert, und in diesem Kapitel benutzen wir auch diese Definition, die verschieden von dem in §2 ist. Ein Halbmaß im alten Sinne ist eine Beschränkung eines Halbmaßes im neuen Sinne.

Levins gleichmäßige Tests d haben unter anderem Eigenschaften, die von allen gleichmäßigen Tests erwünscht wären:
(wir stellen sie für t = exp d auf):

(i) $t(\omega|\mu)$ ist halbberechenbar in (ω,μ),

(ii) $\int t(\omega|\mu)\,\mu(d\omega) \leq 1$ für alle μ.

Diese Eigenschaften haben zur Folge, daß für eine jede berechenbare Verteilung μ eine Konstante $c_\mu < \infty$ existiert mit $d_L(\omega|\mu) \leq d_C(\omega|\mu) + c_\mu$. Hier ist d_C Chaitins Test, den wir im Korollar 3 des Satzes 5.2 definiert haben. Andererseits ist es einfach zu zeigen, daß der Test

$$d_C(\varphi|\psi) = \sup_{x \in N_2^*} M_N(x) \cdot \frac{\varphi(g_x)}{\psi(g_x)}$$

für Halbmaße φ,ψ $(\frac{0}{0} := 0)$ ein gleichmäßiger L-Test ist.
Wir haben folglich

Satz 6.1 Für ein jedes berechenbares Maß μ gibt es eine Konstante c_μ mit

$$|d_L(\omega|\mu) - d_C(\omega|\mu)| \leq c_\mu.$$

Wir führen jetzt ein modifiziertes Konzept des gleichmäßigen Tests ein. Es hat die Vorteile einer durchsichtigeren Definition, in der Halbmaße vermieden sein können. Dann zeigen wir, daß dieser Test auch die Eigenschaften besitzt, die in [11] über d_L ausgesagt wurden.

Definition Die Funktion $t(\omega|\mu)$ ist ein P-Test, wenn (i), (ii) erfüllt sind, und weiter

(iii) für alle $c > 0$,

$$c\mu > \nu \Rightarrow \quad t(\omega|\mu) < ct(\omega|\nu).$$

Mit anderen Worten, $1/t(\omega|\mu)$ kann auf die Menge aller endlichen unnormierten Maße zu einer homogenen, monotonen Funktion erweitert werden.

Bemerkung Diese Bedingung sichert es, daß $t(\omega|\mu)$ nicht dadurch vergrößert werden kann, daß man ω (z.B. durch Verschiebung einer Masse zu einem Punkt, der zu ω in einem einfachen Verhältnis steht), durch Veränderungen des Maßes, die ω eigentlich gar nicht betreffen, auffallender macht.

(iv) $1/t(\omega|\mu)$ ist konkav in μ.

Bemerkung Diese Bedingung scheint noch weniger motiviert zu sein. Eine einfache Folge ist jedenfalls

$$t(\omega|\mu) \leq c, t(\omega|\nu) \leq c \Rightarrow t(\omega|\tfrac{1}{2}\mu + \tfrac{1}{2}\nu) \leq c.$$

Sie schließt solche sonderbaren Tests aus, die ω dann als unzufällig ertappen, wenn in einer gewissen Reihe $U_1, \ldots, U_n$ von Mengen mit $\omega \in U_1 \cap \ldots \cap U_n$ alle U_i kleine Wahrscheinlichkeit haben.

Wir definieren für einen P-Test

$$t(\varphi|\mu) = \varphi(t(\cdot|\mu)) \quad (= \int t(\omega|\mu)\,\mu(d\omega))$$

$$t(\varphi|\psi) = \sup_{\mu \gtrsim \psi} t(\varphi|\mu).$$

Bemerken wir, daß das Integral $\varphi(f)$ einer halbstetigen Funktion f durch ein Halbmaß φ eine natürliche Definition hat.

Satz 6.2 Es gibt einen P-Test $t_p(\omega|\mu)$, der maximal im Sinne von $\leq$ ist.

Der Beweis dieses Satzes ist nicht verschieden von den Beweisen aller anderen Sätze dieser Art, er kann darum weggelassen werden.

Die Relation zwischen t_L und t_p ist klar für Maße:

$t_L(\mu|\nu) \leq t_p(\mu|\nu)$,

weil t_L eine strengere Bedingung der Normiertheit erfüllen muß. Die apriori Wahrscheinlichkeit behält ihre merkwürdige Eigenschaft auch für t_p.

Satz 6.3 $t_p(\omega|M_\Omega) < c$ mit einer universellen Konstante $c < \infty$.

Der Beweis ist analog zum Beweis von Satz 2 in [11].

Wir zeigen endlich, daß t_p auch die Erhaltungseigenschaft hat.

Ein stochastischer Operator ist ein monotoner linearer Operator $A: C(\Omega) \to C(\Omega)$ mit $A(1)=1$. Stochastische Operatoren, ursprünglich auf stetigen Funktionen definiert, können auch über alle halbstetigen Funktionen $f: \Omega \to \bar{R}$ erweitert werden.

Satz 6.4 Für einen beliebigen berechenbaren (siehe [11]) stochastischen Operator A hat man

$$t_p(\varphi A|\psi A) \leq t_p(\varphi|\psi).$$

Beweis Definieren wir für alle ω und μ

$$t^A(\omega|\mu) = (At_p(\cdot|\mu A)\ (\omega).$$

Wir zeigen zunächst, daß t^A ein P-test ist. Die einzige Eigenschaft, die nicht trivial verifizierbar ist, ist die Konkavität von $1/t^A(\omega|\mu)$ in μ. Sie folgt aus

Lemma 6.1 Sei $f(\omega,\mu)$ eine stetige Funktion, für die $1/f(\omega,\mu)$ konkav in μ ist, ν ein positives Maß über Ω. Dann ist auch

$$1/(P_\nu f)(\mu) := 1/\int f(\omega,\mu)\nu(d\omega)$$

konkav in μ.

Beweis 1. Die Funktion

$$g(x_o,x_1) = \frac{1}{1/x_o+1/x_1}$$

is konkav für $x_o,x_1 > 0$. Sie ist nämlich homogen und für $x_o + x_1 = 1$ konkav.

2. Das Funktional $h \to 1/P_\nu(1/h)$ ist konkav. Konkavität bedeutet nämlich

$$\frac{\lambda_o}{P_\nu(1/h_o)} + \frac{\lambda_1}{P_\nu(1/h_1)} \leq 1/P_\nu\left(\frac{1}{\lambda_o h_o + \lambda_1 h_1}\right)$$

oder mit $g_i := \frac{1}{\lambda_i h_i}$

$$P_\nu\left(\frac{1}{1/g_o+1/g_1}\right) \leq \frac{1}{1/P_\nu g_o + 1/P_\nu g_1}$$

Das ist aber eben die Konkavität von $g(x_o,x_1)$. Q.E.D.

Wir haben gezeigt, daß $t^A \leq t_p$ ist.

Wenn wir φ zu beiden Seiten dieser Ungleichung anwenden, bekommen wir

$$t_p(\varphi A|\mu A) = t^A(\varphi|\mu) \leq t_p(\varphi|\mu).$$

Wir sind fertig, wenn wir zeigen können, daß

$$t_p(\varphi A|\psi A) = t^A(\varphi|\psi).$$

Per definitionem

$$t_p(\varphi A|\psi A) = \sup_{\mu \geq \psi A} t_p(\varphi A|\mu)$$

$$t^A(\varphi|\psi) = \sup_{\mu \geq \psi} t_p(\varphi A|\mu A).$$

Es ist darum genug zu zeigen, daß aus $\mu \geq \psi A$ die Existenz eines $\nu \geq \psi$ folgt mit $\mu \geq \nu A$. Setzen wir $\nu g = \mu f$ für alle g der Form Af. Dann haben wir $\mu = \nu A$. Die Definition von ν ist eindeutig, denn $Ah = 0 \Rightarrow \mu h \geq 0$, darum $\mu = 0$ auf KerA. ν ist positiv (wegen $\nu \geq \psi$) und linear auf ImA, $\nu(1) = 1$. Darum ist es auch stetig und kann zu einem Maß über Ω bei der Erhaltung der Eigenschaften $\nu \geq \varphi$, $\mu = \nu A$ erweitert werden.

Q.E.D.

Bemerkung $t_L(\varphi|\psi)$ hat auch die Eigenschaft, daß $1/t_L(\varphi|\psi)$ für jedes Halbmaß φ in ψ konkav ist. Über t_p können wir das gleiche nur für φ behaupten, die Maße sind.

Wir geben endlich einen besonders explizit definierten (nicht unbedingt maximalen) P-Test an, der auch die Erhaltungseigenschaft hat.

Sei $\{f_i\}_{i \in N}$ eine rekursive Aufzählung aller positiven stetigen Funktionen $f_i : \Omega \to Q$ mit einer endlichen Menge von Werten. Sei $f_i > 2^{-i}$. Der Test ist definiert als

$$t_o(\omega|\mu) = \sum_i M_N(i) \cdot \frac{f_i(\omega)}{\mu(f_i)} .$$

Die Erhaltungseigenschaft kann einfach und direkt bewiesen werden.

Literatur

[1] Martin-Löf, Per: The definition of random sequences. Information and Control 6 (1966) 602-619

[2] Zvonkin, A.K., Levin, L.A.: The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. Russ. Math. Surv. 25/6 (1970) 83

[3] Levin, L.A.: On the notion of a random sequence. Soviet Math. Dokl. 14 (1973) 1413

[4] Schnorr, C.P.: Process complexity and effective random tests. J. Comput. Syst. Sci. 7 (1973) 376

[5] Kolmogorov, A.N.: Three approaches to the quantitative definition of information. Prob. Info. Transmission 1/1 (1965) 1

[6] Levin, L.A.: Some theorems on the algorithmic approach to Probability Theory and Information Theory. (Ph. D. thesis, 1971)

[7] Chaitin, G.J.: A theory of program-size formally identical to Information Theory. J.A.C.M. 22 (1975) 329

[8] Chaitin, G.J.: Algorithmic Information Theory. IBM J. Res. Dev. (July 1977) 350-359

[9] Schnorr, C.P.: Unpublished manuscript

[10] Martin-Löf, Per: Algorithmen und zufällige Folgen. Lecture Notes, University of Erlangen, 1966

[11] Levin, L.A.: Uniform tests of randomness. Soviet Math. Doklady 17 (1976) 337

[12] Gács, P.: On the symmetry of algorithmic information. Soviet Math. Dokl.15 (1974) 1477-1480, Corrections ibid. No.6, V

Lebenslauf von Péter Gács

Geboren am 9. Mai 1947 in Budapest, Ungarn.
Vater: László Gács, Mutter: Magda Ságvári.

Einschulung 1953 in die Volksschule Felsöerdösor in Budapest. Von 1961 bis 1965 Besuch des Gymnasiums Ferencz Rákóczi II. in Budapest. Dort Abitur im Jahre 1965.

Von 1965 bis 1970 Studium der Mathematik an der Roland Eötvös-Universität Budapest.

Im Juli 1970 Diplom in Mathematik. Diplom bei Prof. Dr. Imre Csiszár über "Die Hausdorffsche Dimension der Wahrscheinlichkeitsverteilungen".

Ab September 1970 wissenschaftlicher Mitarbeiter im Institut für Mathematik der Ungarischen Akademie der Wissenschaften.

Im Studienjahr 1972-73 Stipendiat der Ungarischen Akademie der Wissenschaften an der Lomonosow Staatsuniversität Moskau.

1975 Januar - Februar: Oberassisitent im Institut für Mathematische Statistik und Wirtschaftsmathematik der Georg August-Universität Göttingen.

Im Wintersemester 1977/78 Gastdozent im Fachbereich Mathematik der Johann Wolfgang Goethe-Universität Frankfurt am Main.

Ab Juni 1978 wissenschaftlicher Mitarbeiter bei der Arbeitsgruppe 7.2 ebendort.

Akademische Lehrer (in alphabetischer Reihenfolge):

Prof. Dr. P. Bod, Prof. Dr. M. Bognár, Frau Prof. K. Bognár, Prof. Dr. I. Csiszár, Prof. Dr. L. Czách, Prof. Dr. E. Fried, Prof. Dr. Gy. Hajós, Prof. Dr. L. Hársing, Prof. Dr. O. Kis, Prof. Dr. E. Kósa, Prof. Dr. A. Mogyoródy, Prof. Dr. B. Nagy, Prof. Dr. L. Paál, Prof. Dr. A. Prékopa, Prof. Dr. A. Rényi, Frau Prof. Dr. K. Rényi, Prof. Dr. C.P. Schnorr, Prof. Dr. V.A. Uspenskij, Prof. Dr. T. Varró, Prof. Dr. J. Vince.

Gács Péter

Péter Gács

Publikationen , Forschungsergebnisse

1. Hausdorff-dimension and probability distributions, Periodica Math. Hung.3 (1973) 59-71

2. Packing of convex sets in the plane with a large number of neighbours, Acta Sci. Hung.23 (1972) 383-388

3. Common information is far less than mutual information, Problems of Control and Information Theory 2 (1973) 149-162 (mit J. Körner)

4. On the symmetry of algorithmic information DAN.218 (1974) 1265-67 (in Russisch) (in Englisch in Sovieth Math. Dokl.)

5. On a problem of Cox concerning point processes in of "controlled variability". Annals of Probability 3N°4(1975) 597-607 (mit D. Szász)

6. Bounds on conditional probabilities with applications in multi-user communication, Z. Wahrscheinlichkeitstheorie 34 (1976) 157-177 (mit R. Ahlswede und J. Körner)

7. Spreading of sets in product spaces and hyper-contraction of the Markov operator, Annals of Probability, 4 (1976) 925-939 (mit R. Ahlswede)

8. Some remarks on generalized spectra, Zeitschrift für Logik und Grundl.Math. (to appear) (mit L. Lovász)

9. Two contributions to information theory Colloquia Societatis J. Bolyai, Topics in Inf. The. Keszthely 1975, 17-40 (mit R. Ahlswede)

10. One-dimensional homogenuous media, dissolving finite islands, Problems of Information Transmission (to appear) (mit L. Levin und G. Kurdiumov)

Sonstiges

1. On the concept of randomness (in Ungarisch) Magyar Tudomany 18/3 (1973) 152-159

2. On problems solvable by search (in Ungarisch) Szigma (1974) 7/1-2, 89-110

3. Algorithms (Buch in Ungarisch) Müszaki Könyvkiadó, 1978 (mit L. Lovász)