Das Prinzip der maximalen Entropie und seine Anwendung in der Codierungs- und Suchtheorie

Von TH. FISCHER

1. Einführung

Das von E. T. Jaynes ursprünglich für die statistische Mechanik begründete Prinzip der maximalen Entropie hat zwar vielfältige Anwendungen gefunden, ist aber bis heute nicht unumstritten. Einen Überblick sowohl über die verschiedenen Anwendungsgebiete als auch über die Schwierigkeiten, die sich bei der Begründung dieses Prinzips ergeben, findet man in [13]. In dieser Arbeit wird nun eine Ungleichung bewiesen, durch die die Verwendung der entropiemaximierenden Verteilung für gewisse Anwendungsbereiche eine neue Rechtfertigung erfährt.

Es sei p eine Wahrscheinlichkeitsverteilung über der endlichen Menge X. Mit

$$H(p) =_{\mathrm{Df}} \sum_{x \in X} p(x) \log p(x)$$

sei die Entropie der Verteilung p bezeichnet. Aus der Codierungstheorie ist gut bekannt, daß für jede weitere Verteilung q über X die Funktion

$$H(p, q) =_{\mathrm{Df}} \sum_{x \in X} p(x) \log q(x)$$

stets größer oder gleich $H(p)$ ist. Im Gegensatz zur Entropie besitzt die Funktion $H(p, q)$ auch keine obere Schranke. Wenn man jedoch eine konvexe und kompakte Menge P von Verteilungen über X fest vorgibt, läßt sich stets eine Verteilung q so finden, daß für alle $p \in P$ die Abschätzung

$$H(p, q) \leqq \max_{p \in P} H(p) \tag{1}$$

gilt.

Diese Ungleichung besitzt nicht nur theoretische Bedeutung. In der Codierungstheorie interessiert man sich für die Existenz von Codes, die nicht nur für eine spezielle Verteilung, sondern für eine ganze Klasse von Verteilungen gleichermaßen geeignet sind, vgl. [2]. Die Ungleichung (1) eröffnet eine besonders einfache Möglichkeit, derartige universelle Codes zu erhalten. Wendet man nämlich bestimmte Konstruktionsalgorithmen nicht auf die Quellverteilung p, sondern auf eine andere

Verteilung q an, dann läßt sich die Güte des erhaltenen Codes mit Hilfe der Funktion $H(p, q)$ charakterisieren [3]. Durch Anwendung bekannter Konstruktionsalgorithmen auf eine entropiemaximierende Verteilung lassen sich also gute universelle Codes gewinnen. Ähnliches gilt für die Konstruktion universeller Suchbäume. Dabei ist besonders wichtig, daß dies auch für kleine Blocklängen gilt, für große Blocklängen liefern dagegen asymptotische Methoden bessere Ergebnisse [2].

Im nachfolgenden Abschnitt wird zunächst eine Verallgemeinerung der Ungleichung (1) für die von A. Rényi eingehend untersuchte Entropie der Ordnung α bewiesen. Anschließend werden dann die schon angedeuteten Anwendungen dieses Ergebnisses auf verschiedene Probleme der Codierungs- und Suchtheorie eingehend dargelegt.

2. Eine Ungleichung für die α-Entropie

Es sei X eine endliche nichtleere Menge und p eine Wahrscheinlichkeitsverteilung über X. Für jeden reellen Parameter α, $0<\alpha\leqq 1$, und jede Verteilung p sei die Entropie der Ordnung α (α-Entropie) als

$$H_\alpha(p) =_{\mathrm{Df}} \begin{cases} H(p) & \text{für} \quad \alpha = 1 \\ \dfrac{1}{1-\alpha} \log \sum\limits_{x\in X} p(x)^\alpha & \text{für} \quad \alpha\in(0,1) \end{cases}$$

erklärt. Als Logarithmenbasis ist dabei jede von 1 verschiedene positive reelle Zahl zugelassen. Wie üblich wird $0\log 0=0$ gesetzt. Die Eigenschaften der α-Entropie wurden in [12] eingehend untersucht.

Von E.G. Nath stammt folgende Verallgemeinerung der α-Entropie für zwei Verteilungen p und q:

$$H_\alpha(p, q) =_{\mathrm{Df}} \begin{cases} H(p, q) & \text{für} \quad \alpha = 1 \\ \dfrac{1}{1-\alpha} \log \sum\limits_{x\in X} p(x)q(x)^{\alpha-1} & \text{für} \quad \alpha\in(0,1). \end{cases}$$

Offensichtlich ist die Funktion $H_\alpha(p, q)$ ebenso wie die α-Entropie stets nichtnegativ, im Gegensatz zu dieser aber unbeschränkt nach oben. Setzt man jedoch voraus, daß mit $q(x)$ auch stets $p(x)$ verschwindet, dann ist $H_\alpha(p, q)$ endlich und stetig sowohl in p als auch in q. Weiterhin gilt stets $H_\alpha(p, q)\geqq H_\alpha(p)$, vgl. [3, 11], und für $q=p$ erhält man $H_\alpha(p, p)=H_\alpha(p)$.

Es sei jetzt P eine konvexe und kompakte Menge von Wahrscheinlichkeitsverteilungen über X und α sei beliebig aber fest gewählt. Ferner sei p_0 eine Verteilung, die die α-Entropie in P maximiert, d. h. es gilt $H_\alpha(p_0)=\max\limits_{p\in P} H_\alpha(p)$.

Satz 1. Für alle $p\in P$ gilt $H_\alpha(p, p_0)\leqq H_\alpha(p_0)$.

Zum Beweis von Satz 1 wird der folgende Hilfssatz benötigt (R bezeichne dabei den Körper der reellen Zahlen):

Hilfssatz. G sei ein konvexes Gebiet im R^n und g sei eine in G definierte reellwertige Funktion. Dann ist $g(p)$ dann und nur dann eine konvexe Funktion des Variablenvektors $p\in G$, wenn die reelle Funktion $f(\lambda)=_{\mathrm{Df}} g(\lambda p'+(1-\lambda)p'')$, $0\leqq\lambda\leqq 1$, für alle $p', p''\in G$ konvex bezüglich λ ist.

Beweis. Es seien $\lambda_1, \lambda_2 \in [0, 1]$ und für beliebige $p', p'' \in G$ sei $p_i =_{\mathrm{Df}} \lambda_i p' + (1-\lambda_i)p''$, $i=1, 2$. Mit p' und p'' liegen auch p_1 und p_2 in G. Wenn g konvex in G ist, erhält man für jedes $\varrho \in [0, 1]$ unter Beachtung der Gleichung $1 = \varrho + (1-\varrho)$

$$
\begin{aligned}
f(\varrho\lambda_1+(1-\varrho)\lambda_2) &= g\big((\varrho\lambda_1+(1-\varrho)\lambda_2)p'+(1-\varrho\lambda_1-(1-\varrho)\lambda_2)p''\big) = \\
&= g(\varrho p_1+(1-\varrho)p_2) \leqq \varrho g(p_1)+(1-\varrho)g(p_2) = \\
&= \varrho g(\lambda_1 p'+(1-\lambda_1)p'')+(1-\varrho)g(\lambda_2 p'+(1-\lambda_2)p'') = \\
&= \varrho f(\lambda_1)+(1-\varrho)f(\lambda_2),
\end{aligned}
$$

d. h. f ist konvex. Wenn hingegen f als konvex vorausgesetzt wird, erhält man mit $\lambda_1=1$, $\lambda_2=0$ die Ungleichung $f(\varrho)\leqq \varrho f(1)+(1-\varrho)f(0)$, woraus sich unter Beachtung der Definition von f unmittelbar die Konvexität von g ergibt.

Beweis von Satz 1. Für beliebige Verteilungen $p\in P$ sei

$$f(\lambda) =_{\mathrm{Df}} H_\alpha(\lambda p+(1-\lambda)p_0).$$

Da die α-Entropie eine konkave Funktion des Variablenvektors p ist, ist auch f konkav bezüglich λ. Da überdies $f(0)=H_\alpha(p_0)=\max\limits_{p\in P} H_\alpha(p)$ gilt und alle Verteilungen $\lambda p+(1-\lambda)p_0$ wegen der Konvexität von P zu P gehören, erhält man für alle $\lambda\in[0, 1]$ die Ungleichung $f(\lambda)\leqq f(0)$. Da f konkav ist, ergibt sich daraus, daß f eine monoton fallende Funktion von λ ist. Für alle $\lambda\in[0, 1]$ gilt also $f'(\lambda)\leqq 0$.

Für $\alpha\in(0, 1)$ erhält man

$$f'(\lambda) = \frac{\alpha}{1-\alpha}\sum_{x\in X}\frac{(\lambda p(x)+(1-\lambda)p_0(x))^{\alpha-1}(p(x)-p_0(x))\log e}{\sum\limits_{x'\in X}(\lambda p(x')+(1-\lambda)p_0(x'))} \leqq 0.$$

Setzt man in dieser Ungleichung $\lambda=0$, erhält man

$$\frac{\alpha}{1-\alpha}\sum_{x\in X}\frac{p_0(x)^{\alpha-1}}{\sum\limits_{x'\in X}p_0(x')^\alpha}(p(x)-p_0(x))\leqq 0,$$

wegen $\alpha/(1-\alpha)>0$ and $\sum\limits_{x'\in X}p_0(x')^\alpha>0$ ergibt sich daraus

$$\sum_{x\in X}p(x)p_0(x)^{\alpha-1}\leqq\sum_{x\in X}p_0(x)^\alpha$$

und durch Logarithmierung und anschließende Multiplikation mit dem Faktor $1/(1-\alpha)$ erhält man die Behauptung von Satz 1.

Zum Beweis für den Fall $\alpha=1$ bildet man ebenfalls die Ableitung von f. Es gilt

$$f'(\lambda) = -\sum_{x\in X}(p(x)-p_0(x))\log((1-\lambda)p_0(x)+\lambda p(x))\leqq 0.$$

Für $\lambda=0$ folgt daraus $H(p, p_0)\leqq H(p_0)$.

Es bleibt noch zu zeigen, daß die Ableitung an der Stelle 0 auch tatsächlich existiert. Man sieht leicht ein, daß $f'(\lambda)$ für $\lambda=0$ nur dann nicht definiert ist, wenn es ein $\bar{x}$ mit $p(\bar{x})>0$ und $p_0(\bar{x})=0$ gibt. O. B. d. A. sei $\bar{x}$ das einzige Element mit dieser Eigenschaft. Während alle übrigen Summanden einen endlichen Wert besitzen, strebt dann der in $f'(\lambda)$ vorkommende Ausdruck $-p(\bar{x})\log \lambda p(\bar{x})$ für $\lambda \to 0$ gegen ∞. Es gilt daher $\lim_{\lambda\to 0} f'(\lambda)=\infty$. Dies widerspricht aber der Tatsache, daß die Ableitung von f nicht positiv ist.

Wie in früheren Arbeiten gezeigt wurde, spielt auch die Funktion

$$H'_\alpha(p, q) =_{\mathrm{Df}} \alpha H_\alpha(p, q)+(1-\alpha)H_\alpha(q)$$

im Zusammenhang mit Codierungsproblemen eine wichtige Rolle [3, 4].

Folgerung 1. Für alle $p\in P$ gilt $H'_\alpha(p, p_0)\leqq H_\alpha(p_0)$.

Der Beweis dieser Ungleichung folgt unmittelbar aus Satz 1 und der Definition von $H'_\alpha(p, q)$.

3. Universelle längenvariable Codierungen

Mit $[X, p]$ sei jetzt eine diskrete gedächtnislose Quelle bezeichnet und c_n: $X^n \to W(Y)$ sei eine Codierung aller Wörter der Länge n in Wörter über einem anderen Alphabet Y. Die Länge eines Codeworts $c_n(u)$ wird mit $l(u)$ bezeichnet. In [1] wurde als Maß für den Aufwand einer Codierung c_n die (exponentielle) mittlere Codewortlänge

$$L_p^t(c_n) =_{\mathrm{Df}} \frac{1}{t}\log_b \sum_{u\in X^n} p(u)\, b^{tl(u)} \tag{2}$$

vorgeschlagen. Dabei bezeichnet b die Kardinalzahl von Y und t ist eine beliebige positive reelle Zahl. Für $t\to 0$ geht (2) in die lineare mittlere Codewortlänge

$$L_p(c_n) =_{\mathrm{Df}} \sum_{u\in X^n} p(u)\, l(u)$$

über, es ist daher gerechtfertigt $L_p^0(c_n) =_{\mathrm{Df}} L_p(c_n)$ zu setzen.

Es sei jetzt $t\geqq 0$ fest gewählt und $\alpha =_{\mathrm{Df}} 1/(1+t)$. Weiterhin sei eine beliebige Verteilung q über X vorgegeben, die durch die Festlegung $q(u) =_{\mathrm{Df}} q(x_1)\cdot \ldots \cdot q(x_n)$ für alle $u=x_1\ldots x_n\in X^n$ zu einer Wahrscheinlichkeitsverteilung über X^n erweitert wird. Mit q_α wird die durch $q_\alpha(x) =_{\mathrm{Df}} q(x)^\alpha / \sum_{x\in X} q(x)^\alpha$, $x\in X$, definierte Hilfsverteilung bezeichnet, die sich ganz analog auf X^n ausdehnen läßt. Bekanntlich läßt sich stets eine eindeutig decodierbare Codierung c_n: $X^n \to W(Y)$ so finden, daß für alle Codewörter $c_n(u)$ die Beziehung

$$l(u) \leqq -\log_b q_\alpha(u)+1 \tag{3}$$

erfüllt ist. Dies kann zum Beispiel durch Anwendung des bekannten Shannonschen Algorithmus auf die Verteilung q_α erreicht werden. Für die mittlere Codewortlänge (gemittelt mit der Quellverteilung p) erhält man dann

$$L_p^t(c_n) \leqq nH'_\alpha(p, q)+1, \tag{4}$$

vgl. [5]. Für $t=0$ gilt offenbar $q_\alpha = q$ und man erhält

$$L_p(c_n) \leqq nH(p, q)+1.$$

Es sei jetzt P eine konvexe und kompakte Menge von Wahrscheinlichkeitsverteilungen über X. Durch P wird eine ganze Klasse diskreter gedächtnisloser Quellen mit dem gemeinsamen Alphabet X festgelegt. Ferner sei eine Verteilung $p_0 \in P$ so gewählt, daß

$$H_\alpha(p_0) = \max_{p \in P} H_\alpha(p)$$

gilt. Ersetzt man nun die Verteilung q durch p_0, dann gilt wegen (4) für jede Verteilung p

$$L_p^t(c_n) \leqq nH'_\alpha(p, p_0)+1$$

und für alle $p \in P$ ergibt sich aus der Folgerung 1

$$L_p^t(c_n) \leqq nH_\alpha(p_0)+1.$$

Satz 2. Für jede konvexe und kompakte Klasse P von diskreten gedächtnislosen Quellen $[X, p]$ und jedes $t \geqq 0$ läßt sich eine Folge eindeutig decodierbarer Codierungen c_n: $X^n \to W(Y)$ derart konstruieren, daß für jede Verteilung $p \in P$ die Ungleichung

$$L_p^t(c_n) \leqq n \max_{p \in P} H_\alpha(p)+1$$

erfüllt ist.

Da für die Konstruktion der Codierung c_n nur die Kenntnis von P benutzt wurde, läßt sich c_n als universelle Codierung für P auffassen. Im Unterschied zu den üblicherweise betrachteten asymptotisch universellen Codierungen [2] gilt diese Eigenschaft für jedes n, insbesondere also auch für $n=1$. Dies ist wichtig, weil mit wachsendem n auch der Umfang des Codebuchs exponentiell anwächst. Überdies lassen sich die erwähnten asymptotischen Methoden auch nicht für die Konstruktion von Suchbäumen nutzen.

4. Universelle binäre Suchbäume

Als Maß für den Suchaufwand in einem binären Suchbaum B wird gewöhnlich die mittlere Weglänge

$$L_p(B) =_{\mathrm{Df}} \sum_{x \in X} p(x)\, l(x)$$

betrachtet. Dabei ist X eine den Endknoten von B zugeordnete endliche linear geordnete Menge, $l(x)$ bezeichnet die Weglänge (Anzahl der Kanten) von der Wurzel des Baums bis zu dem mit x markierten Endknoten und p ist eine Wahrscheinlichkeitsverteilung über X. Bekanntlich läßt sich zu jeder Zugriffsverteilung p ein binärer Suchbaum so konstruieren, daß für alle $x \in X$ die Ungleichung

$$l(x) \leqq -\log_2 p(x)+2 \tag{5}$$

gilt. Dies kann zum Beispiel mit den Algorithmen von Gilbert und Moore [9] oder Mehlhorn [10] geschehen.

Es sei nun wieder angenommen, daß von der konkret vorliegenden Zugriffsverteilung nur ihre Zugehörigkeit zu einer konvexen und kompakten Klasse P bekannt ist. Ferner sei p_0 so gewählt, daß $H(p_0)=\max_{p \in P} H(p)$ gilt. Wendet man die oben genannten Verfahren auf die Verteilung p_0 an, dann erhält man wegen [5] einen

Suchbaum B mit

$$L_p(B) \leqq H(p, p_0)+2.$$

Durch Anwendung von Satz 1 für $\alpha=1$ erhält man daraus unmittelbar das folgende Ergebnis:

Satz 3. Zu jeder konvexen und kompakten Klasse P von Zugriffsverteilungen läßt sich ein binärer Suchbaum B so konstruieren, daß für alle $p \in P$

$$L_p(B) \leqq \max_{p \in P} H(p)+2$$

gilt.

Es soll hier nur kurz erwähnt werden, daß Satz 3 auch für den Fall verallgemeinert werden kann, daß mit positiver Wahrscheinlichkeit nach Informationen gesucht wird, die nicht im Suchbaum abgespeichert sind. In diesem Fall kann dann nur der ohnehin günstigere Algorithmus von Mehlhorn verwendet werden, vgl. [5].

5. Universelle Blockcodierungen

In der klassischen Informationstheorie werden als Blockcodes Mengen der Form

$$C_n(p) =_{\mathrm{Df}} \{u/u \in X^n \wedge p(u) > \varepsilon\}$$

betrachtet, vgl. [8]. Dabei ist $\varepsilon > 0$ eine vorgegebene Schranke. Für konvexe und kompakte Klassen P von diskreten gedächtnislosen Quellen $[X, p]$ läßt sich ein für P universell geeigneter Blockcode ebenfalls mit Hilfe einer entropiemaximierenden Verteilung bestimmen. Es sei p_0 also wieder so gewählt, daß $H(p_0)=\max_{p \in P} H(p)$ gilt. Für jede reelle Zahl $R>0$ sei dann

$$C_n(p_0) =_{\mathrm{Df}} \{u/u \in X^n \wedge p_0(u) > 2^{-nR}\}.$$

Mit der durch

$$\varphi(u) =_{\mathrm{Df}} \begin{cases} 0 & \text{für} \quad u \in C_n(p_0) \\ 1 & \text{für} \quad u \in X^n \setminus C_n(p_0) \end{cases}$$

definierten Funktion φ läßt sich die Fehlerwahrscheinlichkeit bei Verwendung des Codes $C_n(p_0)$ durch

$$P_n =_{\mathrm{Df}} \sum_{u \in X^n} p(u)\varphi(u)$$

ausdrücken. Offenbar gilt für alle $\alpha \in (0, 1)$ die Abschätzung

$$\varphi(u) \leqq \left[\frac{p_0(u)}{2^{-nR}}\right]^{\alpha-1}.$$

Damit erhält man

$$P_n \leqq 2^{-n(1-\alpha)R} \sum_{u \in X^n} p(u) p_0(u)^{\alpha-1} = \exp\left\{-n(1-\alpha)R+\log \sum_{u \in X^n} p(u) p_0(u)^{\alpha-1}\right\} =$$

$$= \exp\{-n(1-\alpha)[R-H_\alpha(p, p_0)]\} \leqq \exp\left\{-n \sup_{\alpha \in (0,1)} (1-\alpha)[R-H_\alpha(p, p_0)]\right\}.$$

Ähnlich wie in [8] läßt sich zeigen, daß der Exponent

$$E(R, p, p_0) = \sup_{\alpha \in (0,1)} (1-\alpha)[R - H_\alpha(p, p_0)]$$

positiv ist, wenn der Parameter R die Bedingung

$$R > H(p, p_0)$$

erfüllt. Wegen $H(p_0) \geqq H(p, p_0)$ für alle $p \in P$ ist dann für $R > H(p_0)$ erst recht $E(R, p, p_0) > 0$. Damit ist der folgende Satz bewiesen:

Satz 4. Für jede konvexe und kompakte Klasse P von diskreten gedächtnislosen Quellen $[X, p]$ und jedes n läßt sich ein Blockcode $C_n \subseteq X^n$ derart bestimmen, daß für die bei der Verwendung von C_n entstehende Fehlerwahrscheinlichkeit P_n

$$P_n \leqq 2^{-nE(R, p, p_0)}$$

gilt. Wenn $R > \max_{p \in P} H(p)$ ist, gilt für alle $p \in P$

$$\lim_{n \to \infty} P_n = 0.$$

Damit ist eine Folge von Blockcodes gefunden, deren Fehlerwahrscheinlichkeit für alle Quellen $[X, p]$ mit $p \in P$ verschwindet. Die Folge $(C_n)_{n=1,2,\ldots}$ kann daher ebenso wie die einzelnen Blockcodes als universell geeignet für P angesehen werden.

6. Abschließende Bemerkungen

1. Im Zusammenhang mit den Sätzen 2 und 3 muß auf das folgende offene Problem hingewiesen werden: Es ist nicht bekannt, was die optimalen Algorithmen bei der Anwendung auf eine entropiemaximierende Verteilung leisten. Die weiter oben benutzte Methode gibt hierüber keinen Aufschluß, da die optimalen Algorithmen nicht die für den Beweis der Sätze 2 und 3 wesentlichen Ungleichungen (3) bzw. (5) erfüllen. Ähnlich verhält es sich auch mit einigen anderen Konstruktionsverfahren, vgl. z. B. [6].

2. Die Fehlerabschätzung im Satz 4 ist etwas schwächer als die von Jelinek für den Fall einer bekannten Verteilung p angegebene. Eine entsprechende Abschätzung, die den Fehlerexponenten von Jelinek als Spezialfall enthält, wurde in einer früheren Arbeit bewiesen [3].

3. Abschließend sei noch darauf verwiesen, daß das Prinzip der maximalen Entropie hier gegenüber dem üblichen Gebrauch etwas erweitert wurde. Einerseits wurde es auf die Entropie der Ordnung α bezogen, andererseits wurde es durch die Einbeziehung der Ungleichung von Satz 1 ergänzt. Durch Satz 1 erhält das Prinzip der maximalen Entropie hier zugleich seine eigentliche Rechtfertigung.

Zusammenfassung

In dieser Arbeit werden universelle Codier- und Suchverfahren beschrieben, die auf der Verwendung des Prinzips der maximalen Entropie beruhen. Hierzu wird eine Ungleichung für verallgemeinerte Entropiemaße bewiesen, durch die das Prinzip der maximalen Entropie zugleich eine neue Rechtfertigung erhält.

AKADEMIE DER WISSENSHAFTEN DER DDR
INSTITUT FÜR MATHEMATIK
MOHRENSTRAßE 39, PF: 1304
DDR 1086 BERLIN

Literatur

[1] CAMPBELL, L. L., A coding theorem and Rényi's entropy, *Inform. and Control,* v. 8, 1965, pp. 423—429.
[2] DAVISSON, L. D., Universal noiseless coding, *IEEE Trans. Inform. Theory,* IT—19, 1973, pp. 783—795.
[3] FISCHER, TH., Über Verallgemeinerungen der Bongard-Entropie und Kodierungssätze für Quellen mit unbekanntem statistischen Verhalten, *Elektron. Informationsverarb. Kybernet.,* v. 13, 1977, pp. 125—135.
[4] FISCHER, TH., Some remarks on the role of inaccuracy in Shannon's theory of information transmission, Trans. Eighth Prague Conf. on Information Theory etc., Academia, Prague, 1978, Vol. A, pp. 211—226.
[5] FISCHER, TH., On the weighted path length of binary search trees for unknown access probabilities, *Lecture Notes in Computer Science,* v. 74, 1979, pp. 284—291.
[6] HORIBE, Y., An improved bound for weight-balanced tree, *Inform. and Control,* v. 34, 1977, pp. 148—151.
[7] JAYNES, E. T., Information theory and statistical mechanics, *Phys. Rev.,* v. 106, 1957, pp. 620—630; v. 108, 1957, pp. 171—190.
[8] JELINEK, F., *Probabilistic information theory,* McGraw Hill, New York, 1968.
[9] KNUTH, D. E., The art of computer programming, Vol. 3; *Sorting and searching,* Addison — Wesley, Reading, 1973.
[10] MEHLHORN, K., A best possible bound for weighted path length of binary search trees, *SIAM J. Comput.,* v. 6., 1977, pp. 235—239.
[11] NATH, P., Inaccuracy and coding theory, *Metrika,* v. 13, 1968, pp. 123—135.
[12] RÉNYI, A., Wahrscheinlichkeitsrechnung, Mit einem Anhang über Informationstheorie, DVW, Berlin, 1963.
[13] SHORE, J. E., R. W. JOHNSON, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inform. Theory,* IT—26, 1980, pp. 26—37.

(Eingegangen am 14. Sept. 1981)