Patron Name: mlewy@mit.edu
Status: STAFF
Dept: EECS

Journal Title: Transactions of the ... Prague Conference on Information Theory, Statistical Decision Functions, Random Processes.

Volume: Trans. 3rd Prague Conf. Inf. T
Issue:
Month/Year: 1962
Pages: 689–723.

Article Author: V. Strassen

Article Title: Asymptotische abschätzungen in Shannon's information-stheorie,

Imprint:

OCLC#:
Item #:

Request Type: Article

Notes: pp. 689–723.

Patron Information:
Hold for Pickup

Michael Lewy (mlewy@mit.edu)
77 Mass Ave, 32-D633
32-D633
Cambridge, MA 02139

email: mlewy@mit.edu

OCC

– 02020 1429

ASYMPTOTISCHE ABSCHÄTZUNGEN IN SHANNONS INFORMATIONSTHEORIE

V. STRASSEN

GÖTTINGEN — BERKELEY

1. EINLEITUNG

Sei X eine endliche Menge, p eine Wahrscheinlichkeitsverteilung (WV) in X, $X_n = \{x_n = (x^1, \ldots, x^n) \mid x^k \in X\}$ und p_n das unabhängige Produkt in X_n von n Exemplaren von p. Man nennt die Folge $(p_n)_{n \geqq 1}$ eine (Informations-)Quelle mit stationären unabhängigen Zeichen über dem Alphabet X. Für festes ε, $0 < \varepsilon < 1$, ist man an der kleinsten Mächtigkeit $\beta(n, \varepsilon)$ von Mengen $E \subseteq X_n$ mit $p_n(E) \geqq 1 - \varepsilon$ interessiert. Sei

$$h(x) = -\log p\{x\} .$$

Man nennt

$$H = \int h \, \mathrm{d}p = - \sum_{p\{x\} > 0} p\{x\} \log p\{x\}$$

die Entropie der Quelle $(p_n)_{n \geqq 1}$.

Bekanntlich gilt

$$\log \beta(n, \varepsilon) = nH + O(\sqrt{n}) \tag{1.1}$$

(einen Überblick über die Literatur zu diesem Resultat und seinen Verallgemeinerungen findet man etwa in [2]). Faßt man e^{nH} als Näherungswert für $\beta(n, \varepsilon)$ auf, so verhält sich der relative Fehler wie $e^{O(\sqrt{n})}$, geht also möglicherweise sehr rasch gegen ∞.

Sei S die Streuung von h und λ die durch

$$\Phi(\lambda) = 1 - \varepsilon \tag{1.2}$$

eindeutig bestimmte reelle Zahl, wobei Φ die gaußsche Verteilungsfunktion ist.

Juschkewitsch [3] hat folgende Verschärfung von (1.1) bewiesen (nicht nur für Quellen mit stationären, unabhängigen Zeichen, sondern allgemeiner für stationäre Markoffsche Quellen):

$$\log \beta(n, \varepsilon) = nH + \sqrt{n} \,.\, \lambda S + o(\sqrt{n}) .$$

Wir schließen nun den trivialen Fall $S = 0$ aus und definieren das reelle Polynom Q durch

$$Q(t) = \frac{\int (h - H)^3 \, \mathrm{d}p}{6S^2} (t^2 - 1) . \tag{1.3}$$

Für gegebenes $\varrho > 0$ sei $d(t)$ der absolut kleinste Rest von t mod ϱ, also

$$d(t) = t - k\varrho \qquad \left(1 - \frac{\varrho}{2} < t - k\varrho \leqq \frac{\varrho}{2}\right), \tag{1.4}$$

und w sei mit Hilfe von d so

$$w = -d + \log\left(d + \frac{\varrho}{2} \cdot \frac{\mathrm{e}^\varrho + 1}{\mathrm{e}^\varrho - 1}\right) \tag{1.5}$$

definiert. w ist eine stetige periodische (also beschränkte) Funktion mit der Periode ϱ.

Satz 1

$$\log \beta(n, \varepsilon) = nH + \sqrt{n} \,.\, \lambda S - \tfrac{1}{2} \log n + Q(\lambda) - \tfrac{1}{2}\lambda^2 - \log(\sqrt{2\pi} \,.\, S) + o(1) ,$$

falls die Verteilung von h nicht gitterförmig ist (siehe [4]) *und*

$$\log \beta(n, \varepsilon) = nH + \sqrt{n} \,.\, \lambda S - \tfrac{1}{2} \log n + Q(\lambda) - \tfrac{1}{2}\lambda^2 - \log(\sqrt{2\pi} \,.\, S) + \\ + w(-na + \sqrt{n} \,.\, \lambda S + Q(\lambda)) + o(1) ,$$

falls die Verteilung von h gitterförmig ist mit der Gitterkonstanten (span) ϱ, und wobei a eine reelle Zahl ist, die mit positiver Wahrscheinlichkeit von $h - H$ angenommen wird.

Aus Satz 1 folgt, daß der relative Fehler von

$$\frac{1}{\sqrt{2\pi n} \,.\, S} \exp\{nH + \sqrt{n} \,.\, \lambda S + Q(\lambda) - \tfrac{1}{2}\lambda^2\} \quad \text{bzw.}$$

$$\frac{1}{\sqrt{2\pi n} \,.\, S} \exp\{nH + \sqrt{n} \,.\, \lambda S + Q(\lambda) - \tfrac{1}{2}\lambda^2 + w(-na + \sqrt{n} \,.\, \lambda S + Q(\lambda))\}$$

als Näherungswert für $\beta(n, \varepsilon)$ mit wachsendem n gegen 0 strebt. Nun sei Y eine zweite endliche Menge, $Y_n = \{y_n = (y^1, \ldots, y^n) \mid y^k \in Y\}$ und P ein Markoffscher Kern (transition probability function) von Y nach X. Für jedes $y \in Y$ ist also $P(., y)$ eine WV in X. Anstatt $P(\{x\}, y)$ schreiben wir $P(x, y)$. Das unabhängige Produkt P_n von n Exemplaren von P ist durch

$$P_n(x_n, y_n) = P(x^1, y^1) \ldots P(x^n, y^n)$$

definiert. Man nennt die Folge $(P_n)_{n \geqq 1}$ einen stationären Kanal ohne Gedächtnis von Y nach X.

Ein ε-Code für P_n ist eine Abbildung f einer Teilmenge von X_n nach Y_n derart, daß

(1.6) $$P_n(f^{-1}\{y_n\}, y_n) \geqq 1 - \varepsilon \quad (y_n \in f(X_n))$$

gilt. Die Anzahl der Punkte von $f(X_n)$ heißt die Länge des Codes. Man interessiert sich für die maximale Länge $N(n, \varepsilon)$ von ε-Codes für P_n. Für beliebige WVen α in Y definieren wir die WVen $P\alpha$ in X und $P \times \alpha$ in $X \times Y$ durch

(1.7) $$(P\alpha)\{x\} = \int P(x, y)\, \alpha(\mathrm{d}y) = \sum_{y \in Y} P(x, y)\, \alpha\{y\}$$

(1.8) $$(P \times \alpha)\{(x, y)\} = P(x, y)\, \alpha\{y\}\,.$$

Sei

(1.9) $$i_\alpha(x, y) = \log \frac{P(x, y)}{(P\alpha)\{x\}}$$

und

(1.10) $$I_\alpha = \int i_\alpha \,\mathrm{d}(P \times \alpha) = \sum_{x,y:P(x,y)\alpha\{y\}>0} P(x, y)\, \alpha\{y\} \log \frac{P(x, y)}{(P\alpha)\{x\}}\,,$$

und sei

(1.11) $$C = \sup_\alpha I_\alpha\,.$$

Man nennt C die Kapazität des Kanals $(P_n)_{n \geqq 1}$. Bekanntlich gilt

$$nC + O(\sqrt{n}) < \log N(n, \varepsilon) < nC + \varepsilon\, O(n)$$

(Coding Theorem mit schwacher Umkehrung, einen Überblick über Literatur zu diesem Resultat und seinen Verallgemeinerungen findet man z. B. in [2], [5], [6], [8] und [16].)

Wolfowitz [5] hat $\varepsilon\, O(n)$ in der obigen Abschätzung durch $O(\sqrt{n})$ ersetzt (starke Umkehrung) also

$$\log N(n, \varepsilon) = nC + O\,(\sqrt{n})$$

bewiesen und bemerkt, daß für $\varepsilon > \frac{1}{2}$ das Glied $O(\sqrt{n})$ schließlich positiv wird.

Für eine spezielle Klasse von stationären Kanälen ohne Gedächtnis (symetrische Kanäle mit zweipunktigem Eingangsalphabet) hat Weiss [7] gezeigt, daß für $\varepsilon < \frac{1}{2}$ das Glied $O(\sqrt{n})$ schließlich negativ wird, genauer, daß

(1.12) $$\log N(n, \varepsilon) \leqq nC - \sqrt{n}\,.\,\lambda T + o(\sqrt{n})$$

gilt, wobei $T > 0$ die Streuung von i_α bezüglich $P \times \alpha$ mit dem durch $I_\alpha = C$ eindeutig bestimmten α ist (dabei wird von trivial ausgearteten Fällen abgesehen).

Für stationäre Kanäle ohne Gedächtnis ist

(1.13) $$\bar{A} = \{\alpha \mid \alpha \text{ WV in } Y, I_\alpha = C\}$$

im allgemeinen nicht einpunktig. Sei G_α die Streuung von i_α bezüglich $P \times \alpha$ und

$$T_1 = \min_{\alpha \in \bar{A}} G_\alpha , \tag{1.14}$$

$$T_{-1} = \max_{\alpha \in \bar{A}} G_\alpha .$$

SATZ 2. *Es gilt*

$$\log N(n, \varepsilon) = nC - \sqrt{n} \, . \, \lambda T_{\operatorname{sign}\lambda} + O(\log n) ,$$

genauer

$$\log N(n, \varepsilon) \leqq nC - \sqrt{n} \, . \, \lambda T_{\operatorname{sign}\lambda} + |Y| \log n \quad (n \text{ hinreichend groß}) , \tag{1.15}$$

$$\log N(n, \varepsilon) \geqq nC - \sqrt{n} \, \lambda T_{\operatorname{sign}\lambda} + O(1) \quad (T_{\operatorname{sign}\lambda} > 0) , \tag{1.16}$$

$$\log N(n, \varepsilon) \geqq nC - \tfrac{1}{2} \log n \quad (n \text{ hinreichend groß}, T_{\operatorname{sign}\lambda} = 0) , \tag{1.17}$$

wobei $|Y|$ *die Mächtigkeit von* Y *ist und* sign $\lambda = 1$ ($\lambda \geqq 0$), sign $\lambda = -1$ ($\lambda < 0$).*)

Satz 2 zeigt unter anderem, daß die obere Schranke für $\log N(n, \varepsilon)$ in (1.12) im Rahmen der dort angegebenen Genauigkeit die bestmögliche ist.

Aus Satz 1 und 2 zusammen folgt (Abschnitt 5), daß ein stationärer Kanal ohne Gedächtnis mit der Kapazität C die Nachrichten einer stationären Quelle mit unabhängigen Zeichen und der Entropie $H = C$ auf die Dauer nur mit großer Irrtumswahrscheinlichkeit übertragen kann (dabei nehmen wir an, daß wenigstens eine der beiden Zahlen S, T_1 positiv ist).

Satz 1 läßt sich mit $O(1)$ an Stelle von $o(1)$ auf nichtstationäre Quellen mit unabhängigen Zeichen verallgemeinern (Abschnitt 3). Entsprechendes gilt für Satz 2 (was allerdings hier nicht bewiesen wird: vgl. auch [8]).

Für stationäre Kanäle ohne Gedächtnis, die einer etwas strengeren Symmetriebedingung als der in [7] genügen, hat Dobruschin [6] ohne Beweis die Abschätzung

$$\log N'(n, \varepsilon) = nC - \sqrt{n} \, . \, \lambda T - \tfrac{1}{2} \log n + O(1) \tag{1.18}$$

angegeben, wobei $N'(n, \varepsilon)$ wie Abschnitt 5 (iv) definiert ist. Aus (5.5) und (1.16) folgt, daß (1.18) im allgemeinen nicht richtig ist.**)

---

*) Für symmetrische Kanäle kann man $|Y| \log n$ in (1.15) durch $O(1)$ (falls $T = 0$) oder $\frac{1}{2} \log n + O(1)$ (falls $T > 0$) ersetzen. Die Beweise vereinfachen sich in diesem Falle wesentlich.

**) Nach einer brieflichen Mitteilung von Herrn Prof. Dobruschin muß es in (1.18) $+\frac{1}{2} \log n$ anstatt $-\frac{1}{2} \log n$ heißen. Hieraus und aus (5.5), (5.7) ergibt sich für $T > 0$ genau das Ergebnis des vorangehenden Nachtrags und (1.16) (unter Dobruschins Symmetriebedingung). Falls Dobruschins Ergebnis auch für $T = 0$ richtig ist (was zweifelhaft erscheint), erhält man durch Vergleich mit dem vorangehenden Nachtrag und (5.7) sogar

$$\log N(n, \varepsilon) = nC + O(1)$$

für solche Kanäle.

Die Beweise von Satz 1 und Satz 3 verwenden wesentlich Ideen von Cramér [18] (siehe auch Feller [19]). Der Beweis von Satz 2 stützt sich auf Methoden von Feinstein und Wolfowitz. Eine Reihe von Arbeiten (Feinstein [20], Elias [21], Shannon, Fano [2]) befassen sich mit der Abschätzung von ε_n, wenn $N(n, \varepsilon_n)$ gegeben ist (beim stationären Kanal ohne Gedächtnis). Satz 2 läßt sich auf dieses Problem anwenden, vorausgesetzt, $N(n, \varepsilon_n)$ bewegt sich in der Nähe von e^{nC}. Gilt nämlich

$$N(n, \varepsilon_n) = e^{nC-(\sqrt{n})K_n},$$

wobei die K_n beschränkt sind (nicht notwendig $\geqq 0$) und nehmen wir etwa $T_1 > 0$ an, so folgt aus Satz 2 leicht

$$\varepsilon_n = \frac{1}{\sqrt{(2\pi)}} \int_{(K_n/T \operatorname{sign} K_n)}^{\infty} e^{-(t^2/2)}\, dt + O\left(\frac{\log n}{\sqrt{n}}\right)$$

(man bemerke hierzu, daß Satz 2 gleichmäßig für ε aus einem beliebigen kompakten Teil des offenen Einheitsintervalls gilt). [20] bis [22] liefern in diesem Fall asymptotisch ungenauere Abschätzungen, wonach z. B. ε_n stets gegen 0 gehen könnte (die dort angegebenen Resultate sind allerdings weitgehend für endliche n gültig und lassen sich auf beliebige Folgen $N(n, \varepsilon_n)$ anwenden).

2. FORMULIERUNG UND BEWEIS EINER VERALLGEMEINERUNG VON SATZ 1

Sei außer der WV p in X noch ein endliches Maß q in X gegeben, bezüglich dessen p totalstetig ist, sei q_n das Produktsmaß in X_n von n Exemplaren von q und sei

$$\beta(n, \varepsilon) = \min_{E \subseteq X_n, p_n(E) \geqq 1-\varepsilon} q_n(E). \tag{2.1}$$

Wir erklären die Funktion h auf X durch

$$h(x) = -\log \frac{p\{x\}}{q\{x\}} = -\log \frac{dp}{dq}(x) \tag{2.2}$$

und setzen

$$H = \int h\, dp = - \sum_{x: p\{x\} > 0} p\{x\} \log \frac{p\{x\}}{q\{x\}}. \tag{2.3}$$

Sei S die Streuung von h bzgl. p (wenn im folgenden von der Verteilung oder von Momenten einer Funktion auf X oder X_n die Rede ist, ist dies stets auf p oder p_n zu beziehen) und sei Q formal wie in (1.3) definiert, aber mit der jetzigen Bedeutung von h, H und S. Wir beweisen nun Satz 1 in der neuen Interpretation, wobei also $\beta(n, \varepsilon)$, h usw. durch (2.1), (2.2) usw. erklärt sind. Wählt man für q das Maß, das jeder Menge in X die Anzahl ihrer Punkte zuordnet, so erhält man den alten Satz 1 als Spezialfall des neuen.

Ist q eine WV, so ist $\beta(n, \varepsilon)$ die Irrtumswahrscheinlichkeit zweiter Art eines optimalen nicht randomisierten Tests zur Irrtumswahrscheinlichkeit erster Art ε zwischen p_n als Nullhypothese und q_n als Alternative.

Sei

$$h_n = -\log \frac{\mathrm{d}p_n}{\mathrm{d}q_n},$$

also

$$h_n(x_n) = \sum_{i=1}^{n} h(x^i) \tag{2.4}$$

p_n – fastüberall, und seien $\mu_n > 0$ so bestimmt, daß

$$p_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n \right\} \geqq 1 - \varepsilon \tag{2.5}$$

$$p_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n \right\} < 1 - \varepsilon$$

ist. Sei

$$\bar{\beta}(n, \varepsilon) = \min_{\int f \mathrm{d}p_n \geqq 1 - \varepsilon} \int f \, \mathrm{d}q_n , \tag{2.6}$$

wo f eine Variable für Funktionen von X_n in das abgeschlossene Einheitsintervall ist (ist q eine WV, so ist $\bar{\beta}(n, \varepsilon)$ also die Irrtumswahrscheinlichkeit zweiter Art eines optimalen randomisierten Tests). Aus dem Lemma von Neyman und Pearson [9] folgt*)

$$q_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n \right\} \leqq \bar{\beta}(n, \varepsilon) = q_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n \right\} + \tag{2.7}$$

$$+ \frac{1}{\mu_n} \left(1 - \varepsilon - p_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} < \mu_n \right\} \right) \leqq q_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n \right\}$$

und daraus

$$\bar{\beta}(n, \varepsilon) \leqq \beta(n, \varepsilon) \leqq \bar{\beta}(n, \varepsilon) + \frac{1}{\mu_n} \max_{X_n} p_n(x_n) .$$

Bemerkt man, daß wegen $S > 0$ $\max_{X_n} p_n(x_n)$ exponentiell gegen 0 strebt und zieht man die asymptotische Näherung (2.35) für μ_n zu Hilfe, so ergibt der Vergleich mit Satz 1, daß es genügt, diesen Satz für $\bar{\beta}(n, \varepsilon)$, anstatt $\beta(n, \varepsilon)$ zu beweisen.

Nun ist

$$q_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n \right\} = \int_{\mathrm{d}p_n/\mathrm{d}q_n > \mu_n} \frac{1}{\mathrm{d}p_n/\mathrm{d}q_n} \, \mathrm{d}p_n = \int_{h_n < -\log \mu_n} \exp\{h_n\} \, \mathrm{d}p_n =$$

$$= \exp\{nH\} \int_{t < \lambda_n} \exp\{t \sqrt{n} \, . \, S\} \, \mathrm{d}F_n(t)$$

*) In diesem Zusammenhang bin ich Herrn Dr. W. Fieger für eine kritische Bemerkung sehr dankbar.

mit

$$F_n(t) = p_n \left\{ \frac{h_n - nH}{\sqrt{n} \,.\, S} \leqq t \right\} \tag{2.8}$$

und

$$\lambda_n = \frac{-\log \mu_n - nH}{\sqrt{n} \,.\, S}\,. \tag{2.9}$$

Also

$$q_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n \right\} = \exp\{nH\} \int_{t < \lambda_n} \exp\{t \sqrt{n} \,.\, S\}\, \mathrm{d}F_n(t) = \tag{2.10}$$

$$= \exp\{nH + \lambda_n \sqrt{n} \,.\, S\} \int_{t < \lambda_n} \exp\{(t - \lambda_n) \sqrt{n} \,.\, S\}\, \mathrm{d}F_n(t) =$$

$$= \exp\{nH + \lambda_n \sqrt{n} \,.\, S\} \int_{z < 0} \mathrm{e}^z\, \mathrm{d}F_n \left(\frac{z}{\sqrt{n} \,.\, S} + \lambda_n \right).$$

Ebenso erhält man

$$q_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n \right\} = \exp\{nH + \lambda_n \sqrt{n} \,.\, S\} \int_{z \leqq 0} \mathrm{e}^z\, \mathrm{d}F_n \left(\frac{z}{\sqrt{n} \,.\, S} + \lambda_n \right), \tag{2.11}$$

also nach (2.7)

$$\int_{z < 0} \mathrm{e}^z\, \mathrm{d}F_n \left(\frac{z}{\sqrt{n} \,.\, S} + \lambda_n \right) \leqq \bar{\beta}(n, \varepsilon) \exp\{-nH - \lambda_n \sqrt{n} \,.\, S\} \leqq \tag{2.12}$$

$$\leqq \int_{z \leqq 0} \mathrm{e}^z\, \mathrm{d}F_n \left(\frac{z}{\sqrt{n} \,.\, S} + \lambda_n \right).$$

Aus (2.8) und (2.9) folgt

$$F_n(\lambda_n) = p_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n \right\} \tag{2.13}$$

$$F_n(\lambda_n - 0) = p_n \left\{ \frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n \right\},$$

also wegen (2.5)

$$F_n(\lambda_n) \geqq 1 - \varepsilon \tag{2.14}$$

$$F_n(\lambda_n - 0) < 1 - \varepsilon\,.$$

Wegen (2.4) ist F_n die Verteilungsfunktion einer auf Erwartungswert 0 und Streuung 1 normierten Summe von n unabhängigen und identisch verteilten Zufallsvariablen.

Sei h nicht gitterförmig verteilt. Nach einem bekannten Satz von Cramér und Esseen ([11], [4]) ist dann

$$F_n(t) = \Phi(t) + \frac{1}{\sqrt{2\pi n}} \exp\left\{-\frac{t^2}{2}\right\} Q_1(t) + o\left(\frac{1}{\sqrt{n}}\right) \tag{2.15}$$

gleichmäßig in t, wobei

$$Q_1(t) = \frac{1}{6S^3}\int (h - H)^3 \, \mathrm{d}p(1 - t^2) = -\frac{1}{S} Q(t)\,.$$

Wir setzen

$$B(t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} Q_1(t) = \Phi'(t)\, Q_1(t) \tag{2.16}$$

und erhalten

$$\int_{z<0} \mathrm{e}^z \, \mathrm{d}F_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = \int_{z<0} \mathrm{e}^z \, \mathrm{d}\Phi\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) + \tag{2.17}$$

$$+ \frac{1}{\sqrt{n}} \int_{z<0} \mathrm{e}^z \, \mathrm{d}B\left(\frac{z}{\sqrt{n}} + \lambda_n\right) + o\left(\frac{1}{\sqrt{n}}\right) =$$

$$= \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{2\pi}\,.\,S}\int_{z<0} \exp\left\{z - \frac{1}{2}\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right)^2\right\} \mathrm{d}z + B(\lambda_n) -\right.$$

$$\left. - \int_{z<0} \mathrm{e}^z\, B\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) \mathrm{d}z + o(1)\right) =$$

$$= \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{2\pi}\,.\,S} \exp\left\{-\frac{\lambda^2}{2}\right\} + o(1)\right) = \frac{1}{\sqrt{2\pi n}\,.\,S} \exp\left\{-\frac{\lambda^2}{2} + o(1)\right\},$$

dann nach (1.2), (2.14) und dem zentralen Grenzwertsatz hat man

$$\lim_{n\to\infty} \lambda_n = \lambda\,. \tag{2.18}$$

Ebenso leitet man

$$\int_{z\leq 0} \mathrm{e}^z \, \mathrm{d}F_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = \frac{1}{\sqrt{2\pi n}\,.\,S} \exp\left\{-\frac{\lambda^2}{2} + o(1)\right\} \tag{2.19}$$

ab.

Aus (2.17), (2.19) und (2.12) folgt also

$$\bar{\beta}(n, \varepsilon) = \frac{1}{\sqrt{2\pi n}\,.\,S} \exp\left\{nH + \lambda_n \sqrt{n}\,.\,S - \frac{\lambda^2}{2} + o(1)\right\}. \tag{2.20}$$

Es ist noch λ_n abzuschätzen. Sei

$$\Delta\lambda_n = \lambda_n - \lambda$$

und

$$\Delta\Phi_n = \Phi(\lambda_n) - \Phi(\lambda) = \Phi(\lambda_n) - (1 - \varepsilon)\,.$$

Aus (2.15), (2.16) und (2.14), (1.2) folgt

$$\Phi(\lambda_n) + \frac{1}{\sqrt{n}} B(\lambda_n) + o\left(\frac{1}{\sqrt{n}}\right) = F_n(\lambda_n - o) < \Phi(\lambda) \leqq F_n(\lambda_n) =$$
$$= \Phi(\lambda_n) + \frac{1}{\sqrt{n}} B(\lambda_n) + o\left(\frac{1}{\sqrt{n}}\right),$$

also

$$\Delta\Phi_n = -\frac{1}{\sqrt{n}} B(\lambda_n) + o\left(\frac{1}{\sqrt{n}}\right) = -\frac{1}{\sqrt{n}} B(\lambda) + o\left(\frac{1}{\sqrt{n}}\right) = \tag{2.21}$$
$$= -\frac{1}{\sqrt{n}} \Phi'(\lambda)\, Q_1(\lambda) + o\left(\frac{1}{\sqrt{n}}\right)$$

wegen (2.18), der Stetigkeit von B und (2.16). Andererseits hat man

$$\Delta\Phi_n = \Phi'(\lambda)\, \Delta\lambda_n + o(\Delta\lambda_n) = \Phi'(\lambda)\, \Delta\lambda_n + o(\Delta\Phi_n)$$

und da nach (2.21) jedenfalls

$$\Delta\Phi_n = O\left(\frac{1}{\sqrt{n}}\right)$$

ist,

$$\Delta\Phi_n = \Phi'(\lambda)\, \Delta\lambda_n + o\left(\frac{1}{\sqrt{n}}\right).$$

Hieraus und aus (2.21) folgt

$$\Delta\lambda_n = -\frac{1}{\sqrt{n}} Q_1(\lambda) + o\left(\frac{1}{\sqrt{n}}\right),$$

also

$$\lambda_n = \lambda + \frac{1}{\sqrt{n}\,.\,S} Q(\lambda) + o\left(\frac{1}{\sqrt{n}}\right). \tag{2.22}$$

Zusammen mit (2.20) ergibt sich schließlich

$$\bar{\beta}(n, \varepsilon) = \frac{1}{\sqrt{2\pi n}\,.\,S} \exp\left\{nH + \sqrt{n}\,.\,\lambda S + Q(\lambda) - \frac{\lambda^2}{2} + o(1)\right\}.$$

Das liefert die Behauptung des Satzes für nicht gitterförmig verteiltes h.

Sei nun h bezüglich p gitterförmig verteilt mit der maximalen Schrittweite ϱ, und sei a reell mit $p\{h - H = a\} > 0$. Nach einen bekannten Satz von Esseen ([11], [4]) ist dann

$$F_n(t) = \Phi(t) + \frac{1}{\sqrt{n}} B(t) + \frac{1}{\sqrt{n}} G_n(t) + o\left(\frac{1}{\sqrt{n}}\right) \tag{2.23}$$

gleichmäßig in t, wobei B dieselbe Bedeutung wie vorhin hat und

$$G_n(t) = \frac{\varrho}{S}\,\Phi'(t)\left(\left[\left(t - \frac{a\sqrt{n}}{S}\right)\frac{\sqrt{n}\,.\,S}{\varrho}\right] - \left(t - \frac{a\sqrt{n}}{S}\right)\frac{\sqrt{n}\,.\,S}{\varrho} + \frac{1}{2}\right)$$

ist ($[t]$ bedeutet die größte ganze Zahl $\leqq t$).

Wir knüpfen an die rechte Seite von (2.10) an. Aus (2.23) und (2.18) folgt

$$\int_{z<0} e^z\, dF_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = \frac{1}{\sqrt{n}}\left(\frac{1}{S}\,\Phi'(\lambda) + o(1) + \right. \tag{2.24}$$

$$\left. + \int_{z<0} e^z\, dG_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right)\right).$$

Ferner ist

$$\int_{z<0} e^z\, dG_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = G_n(\lambda_n - 0) - \int_{z<0} G_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) e^z\, dz\,. \tag{2.25}$$

Nach Voraussetzung nimmt die Zufallsvariable h mit positiver Wahrscheinlichkeit höchstens die Werte

$$H + a + k\varrho \quad (k \text{ ganz}),$$

die Zufallsvariable h_n also höchstens die Werte

$$n(H + a) + k\varrho \quad (k \text{ ganz})$$

an. Nach (2.5) ist $-\log \mu_n$ eine dieser Zahlen, so daß

$$\frac{1}{\varrho}(\lambda_n\sqrt{n}\,.\,S - na) = \frac{1}{\varrho}(-\log \mu_n - n(H + a)) \tag{2.26}$$

ganz ist. Wir erhalten

$$G_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = \frac{\varrho}{S}\,\Phi'\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right)\left(\left[\frac{z}{\varrho} + \frac{1}{\varrho}\left((\lambda_n\sqrt{n}\,.\,S - an)\right]\right. - \right. \tag{2.27}$$

$$\left. - \left(\frac{z}{\varrho} + \frac{1}{\varrho}(\lambda_n\sqrt{n}\,.\,S - an)\right) + \frac{1}{2}\right) = \frac{\varrho}{S}\,\Phi'\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right)\left(\left[\frac{z}{\varrho}\right] - \frac{z}{\varrho} + \frac{1}{2}\right),$$

speziell

$$G_n(\lambda_n) = \frac{\varrho}{2S}\,\Phi'(\lambda_n) \tag{2.28}$$

und

$$G_n(\lambda_n - 0) = -\frac{\varrho}{2S}\,\Phi'(\lambda_n), \tag{2.29}$$

also aus (2.25) nach dem Satz von Lebesgue

$$\int_{z<0} e^z \, dG_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = -\frac{\varrho}{2S}\,\Phi'(\lambda) - \frac{\varrho}{S}\,\Phi'(\lambda)\int_{z<0} e^z\left(\left[\frac{z}{\varrho}\right] - \frac{z}{\varrho} + \frac{1}{2}\right) dz +$$

$$+ o(1) = \frac{1}{S}\,\Phi'(\lambda)\left(-\varrho - \int_{z<0} e^z\left[\frac{z}{\varrho}\right]\varrho \, dz + \int_{z<0} e^z z \, dz + o(1)\right) =$$

$$= \frac{1}{S}\,\Phi'(\lambda)\left(-\varrho - \varrho\sum_{k=-\infty}^{-1} k\int_{k\varrho}^{(k+1)\varrho} e^z \, dz - 1 + o(1)\right) =$$

$$= \frac{1}{S}\,\Phi'(\lambda)\left(-\varrho + \frac{\varrho}{1 - e^{-\varrho}} - 1 + o(1)\right) =$$

$$= \frac{1}{S}\,\Phi'(\lambda)\left(\frac{\varrho}{e^\varrho - 1} - 1 + o(1)\right).$$

Zusammen mit (2.24) folgt hieraus

$$\int_{z<0} e^z \, dF_n\left(\frac{z}{\sqrt{n}\,.\,S} + \lambda_n\right) = \frac{1}{\sqrt{n}\,.\,S}\,\Phi'(\lambda)\left(\frac{\varrho}{e^\varrho - 1} + o(1)\right),$$

also wegen (2.10)

$$q_n\left\{\frac{dp_n}{dq_n} > \mu_n\right\} = \frac{1}{\sqrt{n}\,.\,S}\,\Phi'(\lambda)\exp\{nH + \lambda_n\sqrt{n}\,.\,S\}\left(\frac{\varrho}{e^\varrho - 1} + o(1)\right). \tag{2.30}$$

Sei nun $0 < \Theta_n \leqq 1$, so gewählt, daß

$$\Theta_n F_n(\lambda_n) + (1 - \Theta_n) F_n(\lambda_n - 0) = 1 - \varepsilon$$

ist (wegen (2.14) geht das). Setzt man

$$M_n = \Phi + \frac{1}{\sqrt{n}} B, \tag{2.31}$$

so haben die Gleichungen

$$M_n(t) = 1 - \varepsilon$$

für hinreichend großes n eindeutig bestimmte Lösungen λ_n^0 Analog wie (2.22) beweist man

$$\lambda_n^0 = \lambda + \frac{1}{\sqrt{n}\,.\,S}\,Q(\lambda) + o\left(\frac{1}{\sqrt{n}}\right). \tag{2.32}$$

Ferner haben wir

$$0 = \Theta_n F_n(\lambda_n) + (1 - \Theta_n) F_n(\lambda_n - 0) - M_n(\lambda_n^0) =$$

$$= \Theta_n M_n(\lambda_n) + \Theta_n \frac{\varrho}{2\sqrt{n}\,.\,S}\,\Phi'(\lambda_n) + o\left(\frac{1}{\sqrt{n}}\right) + (1 - \Theta_n) M_n(\lambda_n) +$$

$$+ (\Theta_n - 1)\frac{\varrho}{2\sqrt{n}\,.\,S}\,\Phi'(\lambda_n) + o\left(\frac{1}{\sqrt{n}}\right) - M_n(\lambda_n^0)$$

(wegen (2.23), (2.31), (2.28) und (2.29))

$$= \Theta_n M_n(\lambda_n) + \Theta_n \frac{\varrho}{2\sqrt{n}\,.\,S} \Phi'(\lambda) + (1 - \Theta_n) M_n(\lambda_n) +$$

$$+ (\Theta_n - 1) \frac{\varrho}{2\sqrt{n}\,.\,S} \Phi'(\lambda) - M_n(\lambda_n) - (\lambda_n^0 - \lambda_n) M_n'(\lambda_n) + o(\lambda_n^0 - \lambda_n) + o\left(\frac{1}{\sqrt{n}}\right)$$

(wegen (2.18))

$$= (2\Theta_n - 1) \frac{\varrho}{2\sqrt{n}\,.\,S} \Phi'(\lambda) - (\lambda_n^0 - \lambda_n) \Phi'(\lambda) + o(\lambda_n^0 - \lambda_n) + o\left(\frac{1}{\sqrt{n}}\right),$$

also

$$\lambda_n^0 - \lambda_n = (2\Theta_n - 1) \frac{\varrho}{2\sqrt{n}\,.\,S} + o\left(\frac{1}{\sqrt{n}}\right),$$

d. h.

$$- \frac{\varrho}{2\sqrt{n}\,.\,S} + o\left(\frac{1}{\sqrt{n}}\right) < \lambda_n^0 - \lambda_n \leqq \frac{\varrho}{2\sqrt{n}\,.\,S} + o\left(\frac{1}{\sqrt{n}}\right)$$

oder auch

$$-\frac{\varrho}{2} < (-na + \sqrt{n}\,.\,S\lambda_n^0 + o(1)) - (\sqrt{n}\,.\,S\lambda_n - na) \leqq \frac{\varrho}{2}.$$

Hieraus folgt nach (1.4) und (2.26)

$$d(-na + \sqrt{n}\,.\,S\lambda_n^0 + o(1)) = \sqrt{n}\,.\,S(\lambda_n^0 - \lambda_n) + o(1),$$

zusammen mit (2.32) also

(2.33)

$$\lambda_n = \lambda + \frac{1}{\sqrt{n}\,.\,S} Q(\lambda) - \frac{1}{\sqrt{n}\,.\,S} d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + o(1)) + o\left(\frac{1}{\sqrt{n}}\right).$$

Dies setzen wir in (2.30) ein und erhalten

(2.34)
$$q_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n\right\} = \frac{1}{\sqrt{n}\,.\,S} \Phi'(\lambda) \exp\{nH + \lambda\sqrt{n}\,.\,S + Q(\lambda) -$$

$$- d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + o(1)) + o(1)\} \frac{\varrho}{\mathrm{e}^{\varrho} - 1}.$$

Zur Bestimmung von $\bar{\beta}(n, \varepsilon)$ durch die Gleichung (2.7) fehlt uns noch die Berechnung von μ_n und von

$$\left(1 - \varepsilon - p_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n\right\}\right).$$

Wegen (2.9) und (2.33) ist

$$\mu_n = \exp\{-nH - \lambda_n\sqrt{n}\,.\,S\} = \exp\{-nH - \lambda\sqrt{n}\,.\,S + {} + d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + o(1)) + o(1)\}\,. \tag{2.35}$$

Nach (2.13) gilt

$$1 - \varepsilon - p_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n\right\} = 1 - \varepsilon - F_n(\lambda_n - 0) = \tag{2.36}$$

$$= \Phi(\lambda) - \Phi(\lambda_n) - \frac{1}{\sqrt{n}}B(\lambda_n) - \frac{1}{\sqrt{n}}G_n(\lambda_n - 0) + o\left(\frac{1}{\sqrt{n}}\right)$$

(nach (1.2) und (2.23))

$$= \Phi(\lambda) - \Phi(\lambda) - \Phi'(\lambda)\left(\frac{1}{\sqrt{n}\,.\,S}Q(\lambda) - \frac{1}{\sqrt{n}\,.\,S}d(-na + \sqrt{n}\,.\,S\lambda + {} + Q(\lambda) + o(1))\right) - \frac{1}{\sqrt{n}}B(\lambda) + \frac{1}{\sqrt{n}}\frac{\varrho}{2S}\Phi'(\lambda) + o\left(\frac{1}{\sqrt{n}}\right)$$

(nach (2.33) und (2.29))

$$= \frac{1}{\sqrt{n}\,.\,S}\Phi'(\lambda)\left(d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + o(1)) + \frac{\varrho}{2} + o(1)\right)$$

(wegen (2.16)).

Wir fassen zusammen: Aus (2.7), (2.34), (2.35) und (2.36) folgt

$$\bar{\beta}(n, \varepsilon) = \frac{1}{\sqrt{n}\,.\,S}\Phi'(\lambda)\exp\{nH + \sqrt{n}\,.\,\lambda S + Q(\lambda) - d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + {} + o(1))\}\left(\frac{\varrho}{e^\varrho - 1}e^{o(1)} + \left(d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + o(1)) + \frac{\varrho}{2} + o(1)\right)e^{o(1)}\right)$$

und dies ist wegen $0 \leqq d + \frac{1}{2}\varrho \leqq \varrho$

$$= \frac{1}{\sqrt{n}\,.\,S}\Phi'(\lambda)\exp\{nH + \sqrt{n}\,.\,\lambda S + Q(\lambda) - d(-na + \sqrt{n}\,.\,S\lambda + {} + Q(\lambda) + o(1))\}\,e^{o(1)}\,d(-na + \sqrt{n}\,.\,S\lambda + Q(\lambda) + o(1)) + \frac{\varrho}{2}\frac{e^\varrho + 1}{e^\varrho - 1}$$

und nach (1.5)

$$= \frac{1}{\sqrt{n}\,.\,S}\Phi'(\lambda)\exp\{nH + \sqrt{n}\,.\,\lambda S + Q(\lambda) + w(-na + \sqrt{n}\,.\,S\lambda + {} + Q(\lambda) + o(1)) + o(1)\}\,.$$

Im Gegensatz zu d ist w eine stetige Funktion, so daß wir $o(1)$ aus dem Argument von w herausziehen können. Wegen (2.1) erhalten wir schließlich

$$\bar{\beta}(n, \varepsilon) = \frac{1}{\sqrt{n} \,.\, S} \Phi'(\lambda) \exp \{nH + \sqrt{n} \,.\, \lambda S + Q(\lambda) + \\ + w(-na + \sqrt{n} \,.\, S\lambda + Q(\lambda)) + o(1)\}$$

und damit die Behauptung des Satzes für gitterförmig verteiltes h.

3. NICHTSTATIONÄRE QUELLEN MIT UNABHÄNGIGEN ZEICHEN

Sei $n \geqq 1$, seien X^k $(1 \leqq k \leqq n)$ endliche Mengen und seien p^k WVen, q^k endliche Maße in X^k derart, daß p^k totalstetig bezüglich q^k ist $(1 \leqq k \leqq n)$. Seien ferner

$$X_n = X^1 \times \ldots \times X^n$$
$$p_n = p^1 \times \ldots \times p^n$$
$$q_n = q^1 \times \ldots \times q^n \,.$$

Wir setzen

$$H^k = \int h^k \, \mathrm{d}p^k \,, \quad h^k = -\log \frac{\mathrm{d}p^k}{\mathrm{d}q^k} \,, \tag{3.1}$$

$$H_n = \frac{1}{n} \sum_{k=1}^{n} H^k \,, \tag{3.2}$$

$$S_n = \left(\frac{1}{n} \sum_{k=1}^{n} \int (h^k - H^k)^2 \, \mathrm{d}p^k\right)^{1/2} \tag{3.3}$$

$$R_n = \left(\frac{1}{n} \sum_{k=1}^{n} \int |h^k - H^k|^3 \, \mathrm{d}p^k\right)^{1/3} \tag{3.4}$$

und

$$\beta(n, \varepsilon) = \min_{p_n(E) \geqq 1 - \varepsilon} q_n(E) = q_n(E_{n,\varepsilon}) \text{ (etwa)} \tag{3.5}$$

mit $p_n(E_{n,\varepsilon}) \geqq 1 - \varepsilon$.

SATZ 3. *Es gilt*

$$|\log \beta(n, \varepsilon) - nH_n - \sqrt{n} \,.\, \lambda S_n + \tfrac{1}{2} \log n| < \frac{140}{\delta^8}$$

$$\left(S_n \geqq \delta,\ R_n \leqq \frac{1}{\delta},\ \delta \leqq \varepsilon \leqq 1 - \delta,\ \sqrt{n} \geqq \frac{140}{\delta^8}\right).$$

Beweis. Sei $\delta > 0$ gegeben, $S_n \geqq \delta$, $R_n \leqq 1/\delta$ und $\delta \leqq \varepsilon \leqq 1 - \delta$. Ohne Beschränkung der Allgemeinheit nehmen wir $\delta \leqq \frac{1}{2}$ an. Wir bestimmen wieder μ_n so, daß

$$(3.6) \qquad p_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n\right\} \geqq 1 - \varepsilon$$

$$p_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n\right\} < 1 - \varepsilon$$

ist und setzen

$$(3.7) \qquad g_n = \frac{1}{\sqrt{n}\,.\,S_n}\left(-\log\frac{\mathrm{d}p_n}{\mathrm{d}q_n} - nH_n\right).$$

$$(3.8) \qquad \lambda_n = \frac{1}{\sqrt{n}\,.\,S_n}\left(-\log\mu_n - nH_n\right).$$

Die Verteilungsfunktion von g_n bezüglich p_n nennen wir F_n. g_n ist eine normierte Summe von n unabhängigen Zufallsvariablen mit endlichen dritten Momenten derart, daß die Summe der Varianzen nicht verschwindet. Ein bekannter Satz von Berry und Esseen ([11] S. 43) lautet mit unseren Bezeichnungen

$$(3.9) \qquad |F_n(t) - \Phi(t)| \leqq 7{,}5\,\frac{R_n^3}{S_n^3}\frac{1}{\sqrt{n}} \leqq \frac{7{,}5}{\delta^6}\frac{1}{\sqrt{n}} \quad (t \text{ reell}).$$

Aus (3.6) bis (3.8) folgt

$$(3.10) \qquad F_n(\lambda_n) \geqq 1 - \varepsilon$$

$$F_n(\lambda_n - 0) < 1 - \varepsilon\,.$$

Wir bezeichnen nun allgemein die Lösung λ der Gleichung

$$\Phi(\lambda) = 1 - \eta$$

mit $\lambda(\eta)$ $(0 < \eta < 1)$. Wie man leicht sieht, gilt

$$(3.11) \qquad \Phi'(\lambda(\eta)) \geqq \eta\sqrt{\frac{2}{\pi}} \qquad (0 < \eta \leqq \tfrac{1}{2}).$$

Aus (3.9) und (3.10) ergibt sich

$$|\Phi(\lambda(\varepsilon)) - \Phi(\lambda_n)| \leqq \frac{7{,}5}{\delta^6}\frac{1}{\sqrt{n}},$$

und daraus nach dem Mittelwersatz und wegen (3.11)

$$(3.12) \qquad |\lambda_n - \lambda(\varepsilon)| \leqq \frac{7{,}5}{\delta^6\sqrt{n}}\frac{\sqrt{2\pi}}{\delta} \leqq \frac{20}{\delta^7\sqrt{n}} \qquad \left(\frac{7{,}5}{\delta^6\sqrt{n}} < \frac{\delta}{2}\right).$$

Analog wie (2.7), (2.11) und (2.10) beweist man

$$(3.13)\quad q_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} > \mu_n\right\} \leqq \beta(n, \varepsilon) \leqq q_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n\right\}$$

$$(3.14)\quad q_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n\right\} = \exp\{nH_n + \sqrt{n}\,.\,\lambda_n S_n\}\int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}F_n\left(\frac{z}{\sqrt{n}\,.\,S_n} + \lambda_n\right)$$

und

$$(3.15)\quad q_n\left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \geqq \mu_n\right\} = \exp\{nH_n + \sqrt{n}\,.\,\lambda'_n S_n\}\int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}F_n\left(\frac{z}{\sqrt{n}\,.\,S_n} + \lambda'_n\right),$$

wobei über $\lambda'_n < \lambda_n$ noch verfügt werden kann. Aus (3.9) ergibt sich ähnlich wie (2.17) im Beweis von Satz 1

$$\int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}F_n\left(\frac{z}{\sqrt{n}\,.\,S_n} + \lambda_n\right) \leqq \frac{1}{\sqrt{n}}\left(\frac{1}{S_n} + \frac{15}{\delta^6}\right) \leqq \frac{16}{\delta^6\sqrt{n}},$$

zusammen mit (3.13), (3.14) und (3.12) und wegen

$$(3.16)\qquad S_n \leqq R_n \leqq \frac{1}{\delta}$$

also

$$(3.17)\quad \log\beta(n, \varepsilon) \leqq nH_n + \sqrt{n}\,.\,\lambda(\varepsilon)\,S_n - \tfrac{1}{2}\log n + \frac{24}{\delta^8} \qquad \left(\frac{7{,}5}{\delta^6\sqrt{n}} < \frac{\delta}{2}\right).$$

Wegen (3.16) und (3.9) ist

$$(3.18)\quad \int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}F_n\left(\frac{z}{\sqrt{n}\,.\,S_n} + \lambda'_n\right) \geqq \int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}F_n\left(\frac{z\delta}{\sqrt{n}} + \lambda'_n\right)$$
$$= \frac{1}{\sqrt{n}}\left(\sqrt{n}\int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}\Phi\left(\frac{z\delta}{\sqrt{n}} + \lambda'_n\right) + \frac{7{,}5}{\delta^6}\int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}V_n\left(\frac{z\delta}{\sqrt{n}} + \lambda'_n\right)\right),$$

wobei V_n eine Funktion beschränkter Variation mit

$$(3.19)\qquad |V_n(t)| \leqq 1 \quad (t \text{ reell})$$

ist. Wir haben

$$(3.20)\quad \sqrt{n}\int_{z\leqq 0} \mathrm{e}^z\,\mathrm{d}\Phi\left(\frac{z\delta}{\sqrt{n}} + \lambda'_n\right) = \delta\int_{z\leqq 0} \mathrm{e}^z\Phi'\left(\frac{z\delta}{\sqrt{n}} + \lambda'_n\right)\mathrm{d}z \geqq$$
$$\geqq \delta\int_{z\leqq 0} \mathrm{e}^z\Phi'\left(\frac{z\delta}{\sqrt{n}} - |\lambda'_n|\right)\mathrm{d}z\,.$$

Nun sei

$$(3.21) \qquad \lambda_n - \frac{K}{\sqrt{n}} \leqq \lambda_n' < \lambda_n\,,$$

$$K = \frac{91}{\delta^6 \Phi'(\lambda(\delta))}\,.$$

Wegen (3.12) und (3.11) gilt dann

$$(3.22) \quad |\lambda_n' - \lambda(\varepsilon)| \leqq \frac{91}{\delta^6 \Phi'(\lambda(\delta))\sqrt{n}} + \frac{20}{\delta^7 \sqrt{n}} \leqq \frac{139}{\delta^7 \sqrt{n}} \qquad \left(\frac{7{,}5}{\delta^6 \sqrt{n}} < \frac{\delta}{2}\right),$$

also

$$\delta \int_{z \leqq 0} \Phi'\left(\frac{z\delta}{\sqrt{n}} - |\lambda_n'|\right) \mathrm{e}^z \, \mathrm{d}z \geqq \delta \int_{z \geqq 0} \Phi'\left(\frac{z\delta}{\sqrt{n}} + \frac{139}{\delta^7 \sqrt{n}} + \lambda(\delta)\right) \mathrm{e}^{-z} \, \mathrm{d}z >$$

$$> \delta \int_0^1 \Phi'\left(\frac{140}{\delta^7 \sqrt{n}} + \lambda(\delta)\right) \mathrm{e}^{-z} \, \mathrm{d}z \qquad \left(\frac{7{,}5}{\delta^6 \sqrt{n}} < \frac{\delta}{2}\right)$$

$$> \delta \int_0^1 \Phi'(\delta + \lambda(\delta)) \, \mathrm{e}^{-z} \, \mathrm{d}z \qquad \left(\sqrt{n} > \frac{140}{\delta^8}\right)$$

$$> \frac{\delta}{2} \Phi'(\lambda(\delta)) \int_0^1 \mathrm{e}^{-z} \, \mathrm{d}z > \frac{\delta}{4} \Phi'(\lambda(\delta))\,,$$

wie man leicht sieht. Hieraus und aus (3.20) folgt

$$(3.23) \qquad \sqrt{n} \int_{z \leqq 0} \mathrm{e}^z \, \mathrm{d}\Phi\left(\frac{z\delta}{\sqrt{n}} + \lambda_n'\right) \geqq \frac{\delta}{4} \Phi'(\lambda(\delta)) \qquad \left(\sqrt{n} > \frac{140}{\delta^8}\right).$$

Wir schätzen nun den zweiten Summanden der rechten Seite von (3.18) nach unten *ab*. Zunächst ist

$$\int_{z \leqq 0} \mathrm{e}^z \, \mathrm{d}V_n\left(\frac{z\delta}{\sqrt{n}} + \lambda_n'\right) = V_n(\lambda_n') - \int_{z \leqq 0} V_n\left(\frac{z\delta}{\sqrt{n}} + \lambda_n'\right) \mathrm{e}^z \, \mathrm{d}z\,.$$

Behauptung: Für mindestens ein λ_n' in dem durch (3.21) angegebenen Intervall gilt

$$V_n(\lambda_n') - \int_{z \leqq 0} V_n\left(\frac{z\delta}{\sqrt{n}} + \lambda_n'\right) \mathrm{e}^z \, \mathrm{d}z > -\frac{\delta^7}{45} \Phi'(\lambda(\delta)) = -b \quad \text{(etwa)}\,.$$

Andernfalls haben wir nämlich

$$V_n(\lambda_n') - \int_{z \leqq 0} V_n\left(\frac{z\delta}{\sqrt{n}} + \lambda_n'\right) \mathrm{e}^z \, \mathrm{d}z \leqq -b \quad \left(\lambda_n - \frac{K}{\sqrt{n}} \leqq \lambda_n' < \lambda_n\right),$$

nach Einführung von U_n und u vermöge

$$V_n(t) = U_n\left(\frac{\sqrt{n}}{\delta}(t - \lambda_n)\right)$$

$$u = \frac{\sqrt{n}}{\delta}(\lambda_n' - \lambda_n)$$

also

$$U_n(u) - \int_{z \leqq 0} U_n(z + u)\, \mathrm{e}^z\, \mathrm{d}z \leqq -b \quad \left(\frac{K}{\delta} \leqq u < 0\right), \tag{3.25}$$

oder

$$U_n(u) - \mathrm{e}^{-u} \int_{t \leqq u} U_n(t)\, \mathrm{e}^t\, \mathrm{d}t \leqq -b \quad \left(\frac{K}{\delta} \leqq u < 0\right). \tag{3.26}$$

Aus (3.19) folgt

$$|U_n(t)| \leqq 1 \quad (t \text{ reell}).$$

Wir definieren zu festem n eine Folge von Funktionen U_n^i $(i = 1, 2, \ldots)$ durch

$$U_n^1 = U_n,$$

$$U_n^{1+i}(u) = \begin{cases} \mathrm{e}^{-u} \displaystyle\int_{b \leqq u} U_n^i(t)\, \mathrm{e}^t\, \mathrm{d}t - b & \left(-\dfrac{K}{\delta} \leqq u < 0\right) \\ U_n & (\text{sonst}). \end{cases}$$

Für $i > 1$ sind die U_n^i stetig im Intervall $\langle -K/\delta, 0)$ und erfüllen dort (3.26). Die Folge $U_n^i(t)$ ist bei festem n und t monoton nicht abnehmend und es gilt

$$|U_n^i(t)| \leqq 1 \quad (t \text{ reell}).$$

Hieraus ergibt sich, daß die Folge U_n^i mit wachsendem i gegen eine Funktion $\overline{U}_n$ mit

$$|\overline{U}_n(t)| \leqq 1 \quad (t \text{ reell})$$

konvergiert, die im Intervall $\langle -K/\delta, 0)$ der Integralgleichung

$$\overline{U}_n(u) - \mathrm{e}^{-u} \int_{t \leqq u} \overline{U}_n(t)\, \mathrm{e}^t\, \mathrm{d}t = -b$$

genügt. Deren Lösungen haben die Gestalt:

$$\overline{U}_n(t) = -bt + c \quad \left(-\frac{K}{\delta} \leqq t < 0\right).$$

Hieraus und aus

$$|\overline{U}_n(t)| \leqq 1 \quad (t \text{ reell})$$

folgt

$$\frac{K}{\delta} \leqq \frac{2}{b}$$

d. h. nach (3.21) und der Definition von b

$$\frac{91}{\delta^7 \Phi'(\lambda(\delta))} \leqq \frac{90}{\delta^7 \Phi'(\lambda(\delta))}$$

was nicht möglich ist. Damit ist die Behauptung (3.24) bewiesen.

Wir können also für jedes n im Intervall $\langle \lambda_n - K/\sqrt{n}, \lambda_n)$ ein λ_n' so wählen, daß

$$\int_{z \leqq 0} e^z \, dV_n\left(\frac{z\delta}{\sqrt{n}} + \lambda_n'\right) > -\frac{\delta^7}{45} \Phi'(\lambda(\delta)) \tag{3.27}$$

gilt.

Nach (3.18), (3.23) und (3.27) ist nun

$$\int_{z \leqq 0} e^z \, dF_n\left(\frac{z}{\sqrt{n} \, . \, S_n} + \lambda_n'\right) \geqq \frac{1}{\sqrt{n}}\left(\frac{\delta}{4} \Phi'(\lambda(\delta)) - \frac{\delta}{6} \Phi'(\lambda(\delta))\right) \quad \left(\sqrt{n} > \frac{140}{\delta^8}\right)$$

$$\geqq \frac{1}{\sqrt{n}} \frac{\delta}{12} \Phi'(\lambda(\delta))$$

$$> \frac{1}{\sqrt{n}} \frac{\delta^2}{18}$$

(wegen (3.11)).

Hieraus und aus (3.15) folgt

$$q_n\left(\frac{dp_n}{dq_n} > \mu_n\right) \geqq \exp\{nH_n + \sqrt{n} \, . \, \lambda_n' S_n - \tfrac{1}{2}\log n\} \frac{\delta^2}{18} \cdot \left(\sqrt{n} > \frac{140}{\delta^8}\right) \tag{3.28}$$

also wegen (3.22) und (3.16)

$$\log q_n\left\{\frac{dp_n}{dq_n} > \mu_n\right\} \geqq nH_n + \sqrt{n} \, . \, \lambda(\varepsilon) \, S_n - \tfrac{1}{2}\log n - \frac{140}{\delta^8} \quad \left(\sqrt{n} > \frac{140}{\delta^8}\right) \tag{3.29}$$

und wegen (3.13) schließlich

$$\log \beta(n, \varepsilon) \geqq nH_n + \sqrt{n} \, . \, \lambda S_n - \tfrac{1}{2}\log n - \frac{140}{\delta^8} \quad \left(\sqrt{n} > \frac{140}{\delta^8}\right). \tag{3.30}$$

Hieraus und aus (3.17) folgt der Satz.

LEMMA 1. *Sei A die Menge aller WVen in Y. I_α (siehe (1.10)) ist eine konkave Funktion auf A. Für $\alpha_0, \alpha_1 \in A$ und $0 < t < 1$ sind*

$$I_{(1-t)\alpha_0 + t\alpha_1} = (1 - t) I_{\alpha_0} + t I_{\alpha_1}$$

und

$$P\alpha_0 = P\alpha_1$$

äquivalent. Es gilt

$$\sum_{x:P(x,y)>0} P(x, y) \log \frac{P(x, y)}{(P\alpha)\{x\}} \leqq C \quad (y \in Y, \alpha \in \bar{A}) \tag{3.31}$$

und

$$\sum_{x:P(x,y)>0} P(x, y) \log \frac{P(x, y)}{(P\alpha)\{x\}} = C \quad (\alpha\{y\} > 0, \alpha \in \bar{A}) . \tag{3.32}$$

Umgekehrt folgt $\alpha \in \bar{A}$ *aus*

$$\sum_{x:P(x,y)>0} P(x, y) \log \frac{P(x, y)}{(P\alpha)\{x\}} \geqq \tag{3.33}$$

$$\geqq \sum_{x:P(x,z)>0} P(x, z) \log \frac{P(x, z)}{(P\alpha)\{x\}} \quad (y, z \in Y; \alpha\{y\} > 0) .$$

Sei $\bar{\alpha} \in \bar{A}$ *beliebig. Es gilt*

$$\bar{A} = \{\alpha \mid \alpha \in A, P\alpha = P\bar{\alpha}; \alpha\{y\} = 0 \quad \text{für alle } y \text{ mit} \tag{3.34}$$

$$\sum_{x:P(x,y)>0} P(x, y) \log \frac{P(x, y)}{(P\bar{\alpha})\{y\}} < C\} .$$

Den elementaren Beweis dieses im wesentlichen wohlbekannten Lemmas lassen wir weg. (vgl. [12]).

4. BEWEIS VON SATZ 2

Da I_α stetig von α abhängt, ist $\bar{A}$ (siehe (1.13)) eine kompakte und nicht-leere Teilmenge von A.

$$G_\alpha = \left(\sum_{x,y:P(x,y)\alpha\{y\}>0} \left(\log \frac{P(x, y)}{(P\alpha)\{x\}} - I_\alpha \right)^2 P(x, y)\, \alpha\{y\} \right)^{1/2} , \tag{4.1}$$

ist ebenfalls eine stetige Funktion von α, so daß die Definitionen (1.14) sinnvoll sind. Wir wählen ein

$$\bar{\alpha} \in \bar{A} \tag{4.2}$$

mit

$$G_{\bar{\alpha}} = T_{\mathrm{sign}\,\lambda} \tag{4.3}$$

und bezeichnen mit $\bar{\alpha}_n$ das unabhängige Produkt von n Exemplaren von $\bar{\alpha}$. Jedem $y_n \in Y_n$ ordnen wir die WVen

$$p^k = P(\;, y^k) \tag{4.4}$$

$$q^k = P\bar{\alpha} \qquad (1 \leqq k \leqq n) \tag{4.5}$$

zu und schreiben E_{y_n} für $E_{n,\varepsilon-2/\sqrt{n}}$, β_{y_n} für $\beta(n, \varepsilon - 2/\sqrt{n})$ (siehe (3.5)). Wir haben also

$$P_n(E_{y_n}, y_n) \geqq 1 - \varepsilon + \frac{2}{\sqrt{n}} \tag{4.6}$$

und

$$(P\bar{\alpha})_n\,(E_{y_n}) = \min_{P_n(E,y_n)\geqq 1-\varepsilon+2/\sqrt{n}} (P\bar{\alpha})_n\,(E) = \beta_{y_n}\,. \tag{4.7}$$

Ersetzen wir entsprechend H_n, S_n und R_n (siehe (3.2) bis (3.4)) durch H_{y_n}, S_{y_n} und R_{y_n}, so folgt aus Lemma 1

$$H_{y_n} = -C \quad (\bar{\alpha}_n\text{-fastüberall})\,. \tag{4.8}$$

Die Funktion

$$S^2(y) = \sum_{x:P(x,y)>0} \left(\log \frac{P(x, y)}{(P\bar{\alpha})\,\{x\}} - C\right)^2 P(x, y) \tag{4.9}$$

hat wegen (4.1) und (4.3) bezüglich $\bar{\alpha}$ den Erwartungswert $T^2_{\mathrm{sign}\lambda}$ und endliche Varianz. Ferner ist

$$S^2_{y_n} = \frac{1}{n} \sum_{k=1}^{n} S^2(y^k) \quad (\bar{\alpha}_n\text{-fastüberall})\,, \tag{4.10}$$

so daß für hinreichend großes K nach der Tschebyscheffschen Ungleichung

$$\bar{\alpha}_n(D_n) > \tfrac{1}{2} \quad (n \geqq 1) \tag{4.11}$$

mit

$$D_n = \left\{y_n \middle|\, |S^2_{y_n} - T^2_{\mathrm{sing}\lambda}| < \frac{K}{\sqrt{n}},\quad \bar{\alpha}_n\{y_n\} > 0\right\} \tag{4.12}$$

gilt.

Ein ε-Code f für P_n heiße zulässig, wenn er

$$f(X_n) \subseteq D_n \tag{4.13}$$

und

$$f^{-1}\{y_n\} \subseteq E_{y_n} \quad (y_n \in (f(X_n)) \tag{4.14}$$

erfüllt.

Sei der ε-Code f zulässig mit maximaler Länge N unter allen zulässigen ε-Codes für P_n. Behauptung:

$$(P\bar{\alpha})_n\,(f^{-1}Y_n) \geqq \frac{1}{\sqrt{n}} \quad (n \text{ hinreichend groß})\,. \tag{4.15}$$

Andernfalls haben wir nämlich

$$\frac{1}{\sqrt{n}} > (P\bar{\alpha})_n (f^{-1} Y_n) = (P_n \bar{\alpha}_n)(f^{-1} Y_n) \geqq$$

$$\geqq \int_{D_n} P_n(f^{-1} Y_n, y_n)\, \bar{\alpha}_n(\mathrm{d}y_n) \geqq \tfrac{1}{2} \min_{y_n \in D_n} P_n(f^{-1} Y_n, y_n) .$$

Es gibt also ein $y_n^0 \in D_n$ mit

$$P_n(f^{-1} Y_n, y_n^0) < \frac{2}{\sqrt{n}} . \tag{4.16}$$

Hieraus und aus (4.6) folgt

$$P_n(E_{y_n^0} - f^{-1} Y_n, y_n^0) \geqq 1 - \varepsilon . \tag{4.17}$$

Sei

$$f_0(x_n) = \begin{cases} f(x_n) & (x_n \in f^{-1} Y_n) \\ y_n^0 & (x_n \in E_{y_n^0} - f^{-1} Y_n) . \end{cases}$$

Wegen (4.17) und $y_n^0 \in D_n$ ist f_0 ein zulässiger ε-Code für P_n, wegen (4.16) und $\varepsilon < 1$ hat f_0 für hinreichend große n die Länge $N + 1$. Dies verträgt sich nicht mit der Maximalität von N, (4.15) ist also richtig.

Wir erhalten für hinreichend große n

$$\begin{aligned} \frac{1}{\sqrt{n}} &\leqq (P\bar{\alpha})_n (f^{-1} Y_n) = \sum_{y_n \in f(X_n)} (P\bar{\alpha})_n (f^{-1}\{y_n\}) \leqq \\ &\leqq \sum_{y_n \in f(X_n)} (P\bar{\alpha})_n (E_{y_n}) \quad (\text{wegen } (4.14)) \\ &= \sum_{y_n \in f(X_n)} \beta_{y_n} \quad (\text{wegen } (4.7)) \\ &\leqq N \max_{y_n \in D_n} \beta_{y_n} \quad (\text{wegen } (4.13)) . \end{aligned} \tag{4.18}$$

Ist

$$T_{\mathrm{sign}\lambda} > 0 \tag{4.19}$$

so können wir wegen (4.12) für hinreichend große n Satz 3 anwenden und erhalten aus (4.8) und (4.12)

$$\max_{y_n \in D_n} \beta_{y_n} \leqq \exp\{-nC + \sqrt{n}\,.\,\lambda T_{\mathrm{sign}\lambda} - \tfrac{1}{2}\log n + O(1)\} . \tag{4.20}$$

Aus (4.18) und (4.20) ergibt sich

$$\log N(n, \varepsilon) \geqq \log N \geqq nC - \sqrt{n}\,.\,\lambda T_{\mathrm{sign}\lambda} + O(1) . \tag{4.21}$$

Ist

$$T_{\mathrm{sign}\lambda} = 0 , \tag{4.22}$$

so folgt aus (4.1) und (4.2) für y mit $\bar{\alpha}\{y\} > 0$

$$\frac{P(x, y)}{(P\bar{\alpha})_n \{x\}} = \mathrm{e}^C \quad (P(\ , y)\text{-fast alle } x)\,,$$

also für $y_n \in D_n$ (z. B.)

$$\frac{P_n(x_n, y_n)}{(P\bar{\alpha})_n \{x_n\}} = \mathrm{e}^{nC} \quad (P_n(\ , y_n)\text{-fast alle } x_n)\,.$$

Daraus ergibt sich

$$\beta_{y_n} \leqq \mathrm{e}^{-nC} \quad (y_n \in D_n)\,,$$

also

$$\max_{y_n \in D_n} \beta_{y_n} \leqq \mathrm{e}^{-nC}\,,$$

zusammen mit (4.18)

(4.23) $$\log N(n, \varepsilon) \geqq \log N \geqq nC - \tfrac{1}{2} \log n \quad (n \text{ hinreichend groß})\,.$$

Um (1.15) zu beweisen, nehmen wir zunächst $\varepsilon \leqq \frac{1}{2}$, also $\lambda > 0$ an. Jedem $y_n \in Y_n$ ordnen wir eine WV $\alpha_{y_n} \in A$ zu:

(4.24) $$\alpha_{y_n} = \frac{1}{n} \sum_{k=1}^{n} \delta_{y^k}\,,$$

wobei δ_y die Einheitsmasse im Punkte y bedeutet. Ferner setzen wir

(4.25) $$A_n = \{\alpha_{y_n} \mid y_n \in Y_n\}\,.$$

Sei f ein ε-Code maximaler Länge für P_n. Indem wir f mit seinem Graphen, also einer Teilmenge von $X_n \times Y_n$, identifizieren, ist es sinnvoll, f in Teilmengen f_i zu zerlegen. Sei

(4.26) $$f_\alpha = \{(x_n, y_n) \mid (x_n, y_n) \in f, \alpha_{y_n} = \alpha\} \quad (\alpha \in A_n)\,.$$

Alle f_α sind ε-Codes, etwa der Länge N_α. Da f maximale Länge hat, ist

(4.27) $$N(n, \varepsilon) = \sum_{\alpha \in A_n} N_\alpha\,.$$

Sei $\alpha \in A_n$. Wir wählen ein $y_n \in Y_n$ mit $\alpha_{y_n} = \alpha$ und setzen

(4.28) $$\begin{aligned} p^k &= P(\ , y^k) \\ q^k &= P_\alpha \end{aligned} \quad (1 \leqq k \leqq n)\,,$$

so daß wir $\beta(n, \varepsilon)$ gemäß (3.5) bilden können. Es ist leicht zu sehen, daß $\beta(n, \varepsilon)$ in Wahrheit nur von α (und nicht von dem speziell gewählten $y_n \in f_\alpha$) abhängt. Wir bezeichnen $\beta(n, \varepsilon)$ genauer mit $\beta_\alpha(n, \varepsilon)$.

Ist $y_n \in f_\alpha(X_n)$, so haben wir wegen

$$p_n(f_\alpha^{-1}\{y_n\}) = P_n(f_\alpha^{-1}\{y_n\}, y_n) \geqq 1 - \varepsilon$$

nach (4.28)

$$(P\alpha)_n\,(f_\alpha^{-1}\{y_n\}) = q_n(f_\alpha^{-1}\{y_n\}) \geqq \min_{p_n(E)\geqq 1-\varepsilon} q_n(E) = \beta_\alpha(n, \varepsilon)\,,$$

also auch

$$1 \geqq (P\alpha)_n\,(f_\alpha^{-1} Y_n) = \sum_{y_n \in f_\alpha(X_n)} (P_\alpha)_n\,(f^{-1}\{y\}) \geqq N_\alpha\, \beta_\alpha(n, \varepsilon)\,.$$

Bezeichnen wir die Mächtigkeit einer Menge M mit $|M|$, so folgt hieraus und aus (4.27)

$$(n+1)^{|Y|-1} \geqq |A_n| \geqq \sum_{\alpha\in A_n} N_\alpha\, \beta_\alpha(n, \varepsilon) \geqq N(n, \varepsilon) \min_{\alpha\in A_n} \beta_\alpha(n, \varepsilon)\,. \tag{4.29}$$

Sei nun zunächst

$$T_1 > 0\,. \tag{4.30}$$

Jedem $y_n \in Y_n$ können wir mittels (4.28) das Tripel H_n, S_n, R_n zuordnen. Ist $\alpha = \alpha_{y_n}$, so haben wir

$$\begin{aligned} H_n &= -\frac{1}{n} \sum_{1\leqq k\leqq n}\ \sum_{x:P(x,y^k)>0} P(x, y^k) \log \frac{P(x, y^k)}{(P_\alpha)\,\{x\}} = \\ &= -\sum_{x,y:P(x,y)\alpha\{y\}>0} P(x, y)\,\alpha\{y\} \log \frac{P(x, y)}{(P\alpha)\,\{x\}} = -I_\alpha\,, \end{aligned} \tag{4.31}$$

$$\begin{aligned} S_n^2 &= \frac{1}{n} \sum_k \sum_{x:P(x,y^k)>0} \left(-\log \frac{P(x, y^k)}{(P\alpha)\,\{x\}} + \right. \\ &\left. + \sum_{u:P(u,y^k)>0} P(u, y^k) \log \frac{P(u, y^k)}{(P\alpha)\,\{x\}}\right)^2 P(x, y^k) = \sum_{x,y:P(x,y)\alpha\{y\}>0} P(x, y)\,\alpha\{y\}\,\cdot \\ &\cdot \left(\log \frac{P(x, y)}{(P\alpha)\,\{x\}} - \sum_{u:P(u,y)>0} P(u, y) \log \frac{P(u, y)}{(P\alpha)\,\{x\}}\right)^2 = M_\alpha^2 \text{ (etwa)} \end{aligned} \tag{4.32}$$

und

$$\begin{aligned} R_n^3 &= \sum_{x,y:P(x,y)\alpha\{y\}>0} P(x, y)\,\alpha\{y\} \left|\log \frac{P(x, y)}{(P\alpha)\,\{x\}} - \right. \\ &\left. - \sum_{u:P(u,y)>0} P(u, y) \log \frac{P(u, y)}{(P\alpha)\,\{x\}}\right|^3 = L_\alpha^3 \text{ (etwa)}\,. \end{aligned} \tag{4.33}$$

H_n, S_n und R_n sind also Funktionen von α (anstatt von y_n). Darüberhinaus bilden $-I_\alpha$, M_α und L_α Fortsetzungen von H_n, S_n und R_n auf ganz A. Aus $M_\alpha > 0$ folgt natürlich $L_\alpha > 0$. Die Menge der $\alpha \in A$, für die $\varphi(\alpha)$ in strikt positiven Quadranten liegt, ist eine Umgebung von $\bar{A}$, denn für $\alpha \in \bar{A}$ gilt nach (4.32), (3.32), (4.1), (1.14) und (4.30)

$$M_\alpha = G_\alpha \geqq T_1 > 0\,. \tag{4.34}$$

Wie man leicht sieht, sind für $\alpha \in \bar{A}$ alle $P(, y)$ totalstetig bezüglich $P\alpha$. Dies gilt dann auch für die α aus einer geeigneten Umgebung von $\bar{A}$. In dieser Umgebung sind deshalb I_α und M_α^2 beliebig oft differenzierbare Funktionen von α. Ferner ist $\bar{A}$ konvex (siehe (3.34)). Sei nun $\eta > 0$ und U die abgeschlossene η-Umgebung von $\bar{A}$ (in A) derart, daß $\varphi(U)$ ganz im strikt positiven Quadranten liegt und I_α und M_α^2 in U beliebig oft differenzierbar sind. Nach den vorangehenden Bemerkungen gibt es ein solches η, und wegen $M_\alpha^2 > 0$ für $\alpha \in U$ ist dann auch $M_\alpha > 0$ in U beliebig oft differenzierbar. U ist (konvex und) kompakt. $\varphi(U)$ ist also auch kompakt, und es gibt ein $\delta > 0$ mit

$$M_\alpha \geqq \delta\,, \quad L_\alpha \leqq \frac{1}{\delta}\,, \quad \delta \leqq \varepsilon \leqq 1 - \delta \quad (\alpha \in U)\,.$$

Aus Satz 3 und (4.31) bis (4.33) folgt

$$\beta_\alpha(n, \varepsilon) = \exp\{-nI_\alpha + \sqrt{n}\,.\,\lambda M_\alpha - \tfrac{1}{2}\log n + O(1)\} \tag{4.35}$$

gleichmässig für $\alpha \in U \cap A_n$. Da A kompakt ist und der offene Kern $\mathring{U}$ von U die Menge $\bar{A}$ enthält, gibt es ein $C' < C$ mit

$$I_\alpha \leqq C' \quad (\alpha \in A - \mathring{U})\,.$$

Sei $C'' = \frac{1}{2}(C + C')$. Wegen (4.31), (4.32) und weil M_α beschränkt ist, haben wir gleichmässig für y_n mit $\alpha = \alpha_{y_n} \in A_n - \mathring{U}$

$$\begin{aligned} p_n\left\{\log\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \leqq nC''\right\} &\geqq \\ &\geqq 1 - p_n\left\{\left(\log\frac{\mathrm{d}p_n}{\mathrm{d}q_n} - nI_\alpha\right)^2 > n^2(C'' - C')^2\right\} > \\ &> 1 - \frac{M_\alpha^2}{n(C'' - C')^2} \geqq \frac{1 + \varepsilon}{2} \quad (n \text{ hinreichend groß})\,. \end{aligned} \tag{4.36}$$

Ist nun $E_{n,\varepsilon}$ mit Hilfe von (4.28) wie in (3.5) definiert ($E_{n,\varepsilon}$ hängt natürlich noch von y_n ab), so folgt aus (4.36) gleichmäßig für y_n mit $\alpha = \alpha_{y_n} \in A_n - \mathring{U}$

$$\beta_\alpha(n, \varepsilon) = q_n(E_{n,\varepsilon}) \geqq q_n\left(E_{n,\varepsilon} \cap \left\{\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \leqq \mathrm{e}^{nC''}\right\}\right) \geqq$$

$$\geqq \mathrm{e}^{-nC''}\, p_n\left(E_{n,\varepsilon} \cap \left\{\log\frac{\mathrm{d}p_n}{\mathrm{d}q_n} \leqq nC''\right\}\right) \geqq \frac{1 - \varepsilon}{2}\,\mathrm{e}^{-nC''} \quad (n \text{ hinreichend groß})\,,$$

also

$$\beta_\alpha(N, \varepsilon) \geqq \exp\{-nC'' + O(1)\}$$

gleichmäßig für $\alpha \in A_n - \mathring{U}$.

Hieraus und aus (4.35) ergibt sich

$$(4.37)\qquad \min_{\alpha\in A_n}\beta_\alpha(n,\varepsilon)\geqq\min\{\min_{\alpha\in U\cap A_n}\beta_\alpha(n,\varepsilon),\ \min_{\alpha\in A_n-\mathring{U}}\beta_\alpha(n,\varepsilon)\}\geqq$$

$$\geqq\min\{\inf_{\alpha\in U}\exp\{-nI_\alpha+\sqrt{n}\,.\,\lambda M_\alpha-\tfrac{1}{2}\log n+O(1)\},\exp\{-nC''+O(1)\}\}=$$

$$=\inf_{\alpha\in U}\exp\{-nI_\alpha+\sqrt{n}\,.\,\lambda M_\alpha-\tfrac{1}{2}\log n+O(1)\}=$$

$$=\exp\{\inf_{\alpha\in U}(-nI_\alpha+\sqrt{n}\,.\,\lambda M_\alpha-\tfrac{1}{2}\log n)+O(1)\}\,.$$

Wir ordnen nun jedem $\alpha\in U$ dasjenige $\bar{\alpha}\in\bar{A}$ zu, für das

$$\inf_{\bar{\alpha}\in\bar{A}}\|\alpha-\bar{\alpha}\|$$

angenommen wir ($\|\ \|$ = euklidischer Abstand).

Seien I'_α und M'_α die ersten Ableitungen von I_α und M_α, I''_α die zweite Ableitung von I_α an der Stelle α (I''_α ist also eine Bilinearform über dem in der von A aufgespannten Hyperebene liegenden Vektorraum, siehe etwa [13]). Nach Definition von $\bar{A}$ und wegen $\bar{\alpha}\in\bar{A}$ haben wir

$$(4.38)\qquad I'_{\bar{\alpha}}(\alpha-\bar{\alpha})\leqq 0\,.$$

Weil I_α und M_α stetige Ableitungen beliebiger Ordnung auf dem Kompaktum U haben, gibt es nach der Taylorschen Formel ([13], Seite 94) ein $b>0$ mit

$$(4.39)\qquad -nI_\alpha+\sqrt{n}\,.\,\lambda M_\alpha\geqq-n(I_{\bar{\alpha}}+I'_{\bar{\alpha}}(\alpha-\bar{\alpha})+\tfrac{1}{2}I''_{\bar{\alpha}}(\alpha-\bar{\alpha})(\alpha-\bar{\alpha})+$$
$$+\,b\|\alpha-\bar{\alpha}\|^3+\sqrt{n}\,.\,\lambda(M_{\bar{\alpha}}-b\|\alpha-\bar{\alpha}\|)\quad(\alpha\in U)\,.$$

Nun ist

$$(4.40)\qquad \begin{aligned} I_{\bar{\alpha}}&=C\\ M_{\bar{\alpha}}&=G_{\bar{\alpha}}\geqq T_1\quad(\alpha\in U)\end{aligned}$$

(siehe (1.14)). Man kann ein $a>0$ mit

$$(4.41)\qquad I'_{\bar{\alpha}}(\alpha-\bar{\alpha})+\tfrac{1}{2}I''_{\bar{\alpha}}(\alpha-\bar{\alpha})(\alpha-\bar{\alpha})\leqq-a\|\alpha-\bar{\alpha}\|^2\quad(\alpha\in U)$$

finden, denn angenommen, dies ist falsch, so gibt es eine Folge $\alpha_j\in U$ ($j=1,2,\ldots$) mit

$$\liminf_j\,(I'_{\bar{\alpha}_j}(\alpha_j-\bar{\alpha}_j)+\tfrac{1}{2}I''_{\bar{\alpha}_j}(\alpha_j-\bar{\alpha}_j)(\alpha_j-\bar{\alpha}_j))\geqq 0\,,$$

und wegen (4.38)

$$\|\alpha_j-\bar{\alpha}_j\|=\eta\quad(j=1,2,\ldots)\,.$$

Sei α ein Häufungspunkt dieser Folge. Dann gilt

$$(4.42)\qquad \alpha\in U\,,\quad\|\alpha-\bar{\alpha}\|=\eta>0\,,$$

$$(4.43)\qquad I'_{\bar{\alpha}}(\alpha-\bar{\alpha})+\tfrac{1}{2}I''_{\bar{\alpha}}(\alpha-\bar{\alpha})(\alpha-\bar{\alpha})\geqq 0\,.$$

Sei $\alpha^t = t\alpha + (1-t)\bar{\alpha}$ und

$$I^t = \sum_{x,y:P(x,y)\alpha^t\{y\}>0} P(x,y)\,\alpha^t\{y\}\log\frac{P(x,y)}{(P\alpha^t)\{x\}}.$$

Dann ist

$$I''_{\bar{\alpha}}(\alpha-\bar{\alpha})(\alpha-\bar{\alpha}) = \left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}I^t\right)_{t=0} = -\sum_{(P\bar{\alpha})\{x\}>0}\frac{((P(\alpha-\bar{\alpha}))\{x\})^2}{(P\bar{\alpha})\{x\}}.$$

Hieraus und aus (4.38), (4.43) folgt

$$I'_{\bar{\alpha}}(\alpha-\bar{\alpha}) = 0$$

und

$$(P\alpha)\{x\} = (P\bar{\alpha})\{x\} \quad ((P\alpha)\{x\} > 0),$$

also

$$P\alpha = P\bar{\alpha}.$$

Nach Lemma 1 haben wir also

$$\alpha \in \bar{A}$$

im Widerspruch zu (4.42).

Damit ist (4.41) bewiesen, und aus (4.39) bis (4.41) folgt

$$-nI_\alpha + \sqrt{n}\,.\,\lambda M_\alpha \geqq -nC + n\|\alpha-\bar{\alpha}\|^2(a - b\|\alpha-\bar{\alpha}\|) + + \sqrt{n}\,.\,\lambda T_1 - \sqrt{n}\,.\,\lambda\|\alpha-\bar{\alpha}\|\,b.$$

Indem wir U, (d. h. η) nötigenfalls noch verkleinern, können wir $\|\alpha-\bar{\alpha}\| \leqq a/2b$ annehmen, so daß

(4.44)
$$-nI_\alpha + \sqrt{n}\,.\,\lambda M_\alpha \geqq -nC + \sqrt{n}\,.\,\lambda T_1 + n\|\alpha-\bar{\alpha}\|^2\frac{a}{2} - - \sqrt{n}\,.\,\lambda\|\alpha-\bar{\alpha}\|\,b \geqq -nC + \sqrt{n}\,.\,\lambda T_1 - \frac{\lambda^2 b^2}{2a}$$

folgt.

Aus (4.29), (4.37) und (4.44) ergibt sich schließlich

$$(n+1)^{|Y|-1} \geqq N(n,\varepsilon)\exp\{-nC + \sqrt{n}\,.\,\lambda T_1 - \tfrac{1}{2}\log n + O(1)\},$$

und daraus

$$N(n,\varepsilon) < \exp\{nC - \sqrt{n}\,.\,\lambda T_1 + (|Y| - \tfrac{1}{2})\log n + O(1)\}.$$

Nun sei

(4.45)
$$T_1 = 0.$$

Wir wählen ein $\bar{\alpha} \in \bar{A}$ und setzen

(4.46)
$$Z = \left\{y \mid y \in Y, \log\frac{\mathrm{d}P(\ ,y)}{\mathrm{d}(P\bar{\alpha})} \text{ hat positive Streuung bzgl. } P(\ ,y)\right\}.$$

Für $y \in Y - Z$ ist also $\mathrm{d}P(\,,y)/\mathrm{d}(P\alpha)$ $P(\,,y)$-fastüberall konstant und zwar gilt wegen $\bar{\alpha} \in \bar{A}$ und Lemma 1

$$0 \leqq \frac{P(x, y)}{(P\bar{\alpha})\{x\}} \leqq e^C \quad (P(\ , y)\text{-fastalle } x;\ y \in Y - Z)\,. \tag{4.47}$$

Ist $y_m \in Y_m$ mit

$$y^k \in Z \quad (1 \leqq k \leqq m)\,, \tag{4.48}$$

so bilden wir

$$\begin{aligned} p^k &= P(\ , y^k) \\ q^k &= P\bar{\alpha} \end{aligned} \qquad (1 \leqq k \leqq m)\,.$$

Nach (4.46) gibt es ein $\delta > 0$ mit $S_m \geqq \delta$, $R_m \leqq 1/\delta$ und $\delta \leqq \varepsilon \leqq 1 - \delta$ (unabhängig von y_m mit (4.48)). Aus dem Beweis von Satz 3 ((3.6), und (3.29)) folgt

$$p_m\left\{\frac{\mathrm{d}p_m}{\mathrm{d}q_m} > \mu_m\right\} < 1 - \varepsilon\,, \tag{4.49}$$

$$\log q_m\left\{\frac{\mathrm{d}p_m}{\mathrm{d}q_m} > \mu_m\right\} \geqq mH_m + \sqrt{m}\,.\,\lambda S_m - \tfrac{1}{2}\log m - \frac{140}{\delta^8} \quad \left(m > \frac{140}{\delta^8}\right). \tag{4.50}$$

Nun ist aber $H_m \geqq -C$ (siehe etwa (3.31)), so daß sich aus (4.50)

$$\log q_m^m\left\{\frac{\mathrm{d}p_m}{\mathrm{d}q_m} > \mu_m\right\} \geqq -mC - \tfrac{1}{2}\log m - \frac{140}{\delta^8} \quad \left\{m > \frac{140}{\delta^8}\right\} \tag{4.51}$$

unabhängig von y_m mit (4.48) ergibt.

Sei nun $y_n \in Y_n$ beliebig. Wir setzen wie oben

$$\begin{aligned} p^k &= P(\ , y^k) \\ q^k &= P\bar{\alpha} \end{aligned} \qquad (1 \leqq k \leqq n)$$

und bezeichnen das in (3.5) definierte $\beta(n, \varepsilon)$ mit $\beta_\alpha(n, \varepsilon)$, wobei $\alpha = \alpha_{y_n} \in A_n$ ist (diese $\beta_\alpha(n, \varepsilon)$ unterscheiden sich von den in (4.29) auftretenden $\beta_\alpha(n, \varepsilon)$, die Schreibweise $\beta_\alpha(n, \varepsilon)$ anstatt $\beta_{y_n}(n, \varepsilon)$ ist aber wieder gerechtfertigt).

An Stelle von (4.29) erhalten wir jetzt

$$\begin{aligned} 1 \geqq (P\bar{\alpha})_n\,(f^{-1}Y_n) = \sum_{\alpha\in A_n} (P\bar{\alpha})_n\,(f_\alpha^{-1}Y_n) \geqq \sum_{\alpha\in A_n} N_\alpha\,\beta_\alpha(n, \varepsilon) \geqq \\ \geqq N(n, \varepsilon)\min_{\alpha\in A_n} \beta_\alpha(n, \varepsilon)\,. \end{aligned} \tag{4.52}$$

Sei $\alpha \in A_n$ beliebig. Dann ist entweder

$$n\,\alpha(Z) \leqq \frac{140}{\delta^8} \tag{4.53}$$

oder

$$n\,\alpha(Z) > \frac{140}{\delta^8} \tag{4.54}$$

Im ersten Fall setzen wir

$$M = \max_{x,y:P(x,y)>0} \frac{P(x,y)}{(P\bar{\alpha})\{x\}}. \tag{4.55}$$

Wegen $\bar{\alpha} \in \bar{A}$ ist $M < \infty$ (siehe (3.31)). Aus (4.53), (4.55) und (4.47) folgt

$$\beta_\alpha(n,\varepsilon) \geqq \frac{1-\varepsilon}{\exp\{(n - n\,\alpha(Z))\,C\}\,M^{n\alpha(Z)}} \geqq \tag{4.56}$$

$$\geqq \frac{1-\varepsilon}{\exp\left\{nC + \frac{140}{\delta^8}\log M\right\}} = \exp\{-nC + O(1)\}$$

gleichmäßig für α mit (4.53).

Im zweiten Fall (4.54) wählen wir ein y_n mit $\alpha_{y_n} = \alpha$ so, daß die ersten $m = n\,\alpha(Z)$ y^k in Z liegen, die übrigen also in $Y - Z$. Aus (4.49) ergibt sich dann

$$p_n\left\{x_n \,\middle|\, \frac{dp_m}{dq_m}(x_m) > \mu_m, \frac{d(p^{m+1}\times\ldots\times p^n)}{d(q^{m+1}\times\ldots\times q^n)}(x^{m+1},\ldots,x^n) > 0\right\} < 1-\varepsilon \tag{4.57}$$

und nach Definition von Z gibt es γ_n mit

$$\left\{x_n \,\middle|\, \frac{dp_m}{dq_m}(x_m) > \mu_m, \frac{d(p^{m+1}\times\ldots\times p^n)}{d(q^{m+1}\times\ldots\times q^n)}(x^{m+1},\ldots,x^n) > 0\right\} =$$

$$= \left\{x_n \,\middle|\, \frac{dp_n}{dq_n}(x_n) > \gamma_n\right\}.$$

Hieraus, aus (4.57) und aus der Definition von $\beta_\alpha(n,\varepsilon)$ folgt

$$\log\beta_\alpha(n,\varepsilon) \geqq \log q_n\left\{x_n \,\middle|\, \frac{dp_m}{dq_m}(x_m) > \mu_m, \frac{d(p^{m+1}\times\ldots\times p^n)}{d(q^{m+1}\times\ldots\times q^n)}\cdot\right. \tag{4.58}$$

$$\left.\cdot(x^{m+1},\ldots,x^n) > 0\right\} = \log q_m\left\{x_m \,\middle|\, \frac{dp_m}{dq_m}(x_m) > \mu_m\right\} + \log(q^{m+1}\times\ldots\times q^n)\,.$$

$$\cdot\left\{(x^{m+1},\ldots,x^n) \,\middle|\, \frac{d(p^{m+1}\times\ldots\times p^n)}{d(q^{m+1}\times\ldots\times q^n)} > 0\right\} \geqq$$

(nach (4.51) und (4.47))

$$\geqq -mC - \tfrac{1}{2}\log m - \frac{140}{\delta^8} - (n-m)\,C \geqq$$

$$\geqq -nC - \tfrac{1}{2}\log n - \frac{140}{\delta^8}$$

gleichmäßig für α mit (4.54). (4.56) und (4.58) liefern zusammen

$$\beta_\alpha(n, \varepsilon) > \exp\{-nC - \tfrac{1}{2}\log n + O(1)\}$$

gleichmäßig für alle $\alpha \in A_n$, so daß aus (4.52) jedenfalls

$$N(n, \varepsilon) < \exp\{nC - \sqrt{n}\,.\,\lambda T_1 + (|Y| - \tfrac{1}{2})\log n + O(1)\}$$

folgt. Diese Ungleichung ist damit allgemein (d. h. für $T_1 = 0$ und $T_1 > 0$) bewiesen.

Auf ähnliche Weise erhält man für $\varepsilon > \frac{1}{2}$

$$N(n, \varepsilon) < \exp\{nC - \sqrt{n}\,.\,\lambda T_{-1} + (|Y| - \tfrac{1}{2})\log n + O(1)\},$$

woraus der Satz folgt.

5. ANMERKUNGEN

(i) Im Beweis von Satz 3 wird nicht benutzt, daß die zugrundeliegenden Räume X^k endlich sind. Der Satz ist also für beliebige meßbare Räume X^k richtig. Das Gleiche gilt für die verallgemeinerte Fassung von Satz 1, wenn man die Endlichkeit der auftretenden Momente von h voraussetzt und die Summenschreibweise der Integrale vermeidet.

Ist h gitterförmig verteilt oder genügt die Verteilung der Bedingung C [10] (dies ist natürlich nur für unendliches X möglich), so kann man Satz 1 wahrscheinlich weiter verschärfen, indem man die bekannten asymptotischen Entwicklungen von Esseen [11] und Cramér [10] heranzieht.

Satz 1 und Satz 3 bleiben im wesentlichen richtig (siehe [14]), wenn man die WVen p bzw. p^k und die endlichen Maße q bzw. q^k durch totalmonotone Kapazitäten im Sinne von Choquet [15] ersetzt. Dies hat Anwendungen auf die Theorie der kontinuierlichen Informationsquellen bei Berücksichtigung von Beobachtungsfehlern, deren statistische Natur man nicht genau kennt (siehe [14]).

(ii) Beispiel eines Kanals mit $T_1 \neq T_{-1}$: Sei $X = \{0, 1, 2\}$, $Y = \{0, 1, 2, 3, 4, 5\}$,

(5.1) $$P(x, y) = \tfrac{1}{2}(1 - \delta_{xy}) \quad (x, y = 0, 1, 2)$$

und

(5.2) $$P(x, y) = a\{x + y \bmod 3\} \quad (x \in X,\ y = 3, 4, 5),$$

wobei $(x + y \bmod 3) \in \{0, 1, 2\}$ und a eine WV in X ist. Sei ferner

$$\alpha\{y\} = \begin{cases} \frac{1}{3} & (y = 0, 1, 2) \\ 0 & (\text{sonst}). \end{cases}$$

Es ist $(P\alpha)\{x\} = \frac{1}{3}$ $(x \in X)$ und

$$\sum_{x:P(x,y)>0} P(x, y) \log \frac{P(x, y)}{(P\alpha)\{x\}} = \log \tfrac{3}{2} \quad (y = 0, 1, 2).$$

Nach dem Zwischenwertsatz gibt es ein t mit $0 < t < 1$ so, daß für

$$a\{0\} = \frac{t}{4}, \quad a\{1\} = \tfrac{1}{4}, \quad a\{2\} = \frac{3 - t}{4}$$

$$\sum_x a\{x\} \log \frac{a\{x\}}{(P\alpha)\{x\}} = \log \tfrac{3}{2}$$

gilt. Dieses a verwenden wir in (5.2).

Aus Lemma 1 folgt nun $\alpha \in \bar{A}$, zusammen mit $G_\alpha = 0$ also $T_1 = 0$. Ist $\alpha'\{y\} = = \alpha\{5 - y\}$, so ist nach Lemma 1 auch $\alpha' \in \bar{A}$. Da $P\alpha'$ die Gleichverteilung in X ist, a überall positiv, aber nicht die Gleichverteilung ist, haben wir $G\alpha' > 0$, also $T_{-1} > 0$, also $T_1 \neq T_{-1}$.

Aus dem Beweis von Satz 2 ergibt sich, daß es beim vorliegenden Kanal klug ist, für $\varepsilon < \frac{1}{2}$ nur die Symbole 0, 1, 2, für $\varepsilon > \frac{1}{2}$ nur die Symbole 3, 4, 5 zu senden (jedenfalls für große n). Dies hängt eng mit der Anwendbarkeit der Methode der Zufalls-Codes (siehe [1], [5], [16]) zusammen: Die $\alpha \in \bar{A}$ sind für die Konstruktion eines Zufalls-Codes nicht gleichwertig. Die Brauchbarkeit eines α hängt von der Irrtumswahrscheinlichkeit ab.

Die untere Abschätzung in Satz 2 kann man unter Berücksichtigung dieser Tatsache auch mit Hilfe der Methode der Zufalls-Codes erhalten.

(iii) Definiert man einen „ε-Code im Mittel für P_n“ als eine Abbildung g aus einem Teil von X_n nach Y_n mit

$$\text{(5.3)} \qquad \frac{1}{N'} \sum_{y_n \in g(Y_n)} P_n(g^{-1}\{y_n\}, y_n) \geqq 1 - \varepsilon,$$

wobei N' die Länge des Codes ist, und bezeichnet man mit $N'(n, \varepsilon)$ die maximale Länge von ε-Codes im Mittel für P_n, so gilt

$$\text{(5.4)} \qquad \log N'(n, \varepsilon) = nC - \sqrt{n} \,.\, \lambda T_{\operatorname{sign}\lambda} + O(\log n).$$

Denn da jeder ε-Code ein ε-Code im Mittel ist, hat man

$$\text{(5.5)} \qquad N(n, \varepsilon) \leqq N'(n, \varepsilon).$$

Ist umgekehrt g ein ε-Code im Mittel der Länge N' für P_n, so sei der $(\varepsilon + 1/\sqrt{n})$-Code f für P_n durch

$$\text{(5.6)} \qquad f = \left\{(x_n, y_n) \mid y_n = g(x_n),\, P_n(g^{-1}\{y_n\}, y_n) \geqq 1 - \varepsilon - \frac{1}{\sqrt{n}}\right\}$$

definiert. Sei N die Länge von f. Aus (5.3) und (5.6) folgt

$$1 - \varepsilon \leqq \frac{1}{N'} \sum_{y_n \in f(X_n)} P_n(g^{-1}\{y_n\}, y_n) + \frac{1}{N'} \sum_{y_n \in g(X_n) - f(X_n)} P_n(g^{-1}\{y_n\}, y_n) \leqq \frac{N}{N'} +$$

$$+ \frac{N' - N}{N'}\left(1 - \varepsilon - \frac{1}{\sqrt{n}}\right) = 1 - \varepsilon - \frac{1}{\sqrt{n}} + \frac{N}{N'}\left(\varepsilon + \frac{1}{\sqrt{n}}\right),$$

also

$$N' \leqq (1 + \sqrt{n} \,.\, \varepsilon)\, N$$

und deshalb

$$N'(n, \varepsilon) \leqq (1 + \sqrt{n} \,.\, \varepsilon)\, N\left(n, \varepsilon + \frac{1}{\sqrt{n}}\right). \tag{5.7}$$

Aus (5.5), (5.7) und Satz 2 ergibt sich (5.4).

(iv) Gegeben eine stationäre Quelle $(p_n)_{n \geqq 1}$ mit unabhängigen Zeichen über dem Alphabet Z und ein stationärer Kanal $(P_n)_{n \geqq 1}$ ohne Gedächtnis von Y nach X. Wir schreiben $(p_n \xrightarrow{P_n})_\varepsilon$, wenn es für hinreichend großes n Abbildungen χ von Z_n nach Y_n und φ von X_n nach Z_n gibt derart, daß

$$\sum_{z_n \in Z_n} P_n(\varphi^{-1}\{z_n\}, \chi(z_n))\, p_n\{z_n\} \geqq 1 - \varepsilon \tag{5.8}$$

gilt, und $(p_n \xrightarrow{P_n}\!\!\mid)_\varepsilon$, wenn für hinreichend großes n keine solchen Abbildungen existieren. ($(p_n \xrightarrow{P_n})_\varepsilon$ z. B. bedeutet anschaulich, daß der Kanal fähig ist, die Nachrichten der Quelle auf die Dauer mit einer Irrtumswahrscheinlichkeit $\leqq \varepsilon$ zu übertragen).

Ist H die Entropie der Quelle und C die Kapazität des Kanals, so gelten bekanntlich ([1], [6], [17]) die Shannonschen Beziehungen

$$(p_n \xrightarrow{P_n})_\varepsilon \quad (0 < \varepsilon < 1)\,, \quad \text{falls } H < C$$

und

$$(p_n \xrightarrow{P_n}\!\!\mid)_\varepsilon \quad (0 < \varepsilon < 1)\,, \quad \text{falls } H > C\,.$$

Wir betrachten den Fall

$$H = C \tag{5.9}$$

und nehmen an, daß wenigstens eine der beiden Zahlen S und T_1 positiv ist. Sei $0 < \varepsilon < 1$, $(p_n \xrightarrow{P_n})_\varepsilon$ und n so groß, daß es Abbildungen χ und φ mit (5.8) gibt.

Sei $\delta > \varepsilon$ und < 1 und

$$E = \{z_n | z_n \in Z_n,\, P_n(\varphi^{-1}\{z_n\}, \chi(z_n)) > 1 - \delta\}\,.$$

Aus (5.8) folgt

$$1 - \varepsilon \leqq \sum_{z_n \in E} P_n(\varphi^{-1}\{z_n\}, \chi(z_n))\, p_n\{z_n\} +$$

$$+ \sum_{z_n \in Z_n - E} P_n(\varphi^{-1}\{z_n\}, \chi(z_n))\, p_n\{z_n\} \leqq p_n(E) + (1 - \delta)(1 - p_n(E))$$

und hieraus

$$p_n(E) \geqq 1 - \frac{\varepsilon}{\delta}.$$

Satz 1 und eine einfache Überlegung im Falle $S = 0$ liefern also

$$|E| \geqq \beta\left(n, \frac{\varepsilon}{\delta}\right) = \exp\left\{nH + \sqrt{n}\,.\,\lambda\left(\frac{\varepsilon}{\delta}\right)S + O(\log n)\right\}, \tag{5.10}$$

wobei $|E|$ die Mächtigkeit von E ist.

Sei die Abbildung f von $\varphi^{-1}E$ nach Y_n durch

$$f(x_n) = \chi(\varphi(x_n)) \quad (x_n \in \varphi^{-1}E)$$

definiert. Dann ist $f(X_n) = \chi(E)$ und nach Definition von E

$$P_n(f^{-1}\{y_n\}, y_n) = P_n\Big(\bigcup_{\substack{z_n \in E \\ \chi(z_n) = y_n}} \varphi^{-1}\{z_n\}, y_n\Big) = \sum_{\substack{z_n \in E \\ \chi(z_n) = y_n}} P_n(\varphi^{-1}\{z_n\}, \chi(z_n)) > \tag{5.11}$$
$$> (1 - \delta)\left|\chi^{-1}\{y_n\} \cap E\right| \geqq 1 - \delta \quad (y_n \in f(X_n)).$$

f ist also ein δ-Code für P_n der Länge $|\chi(E)|$.

Aus (5.11) folgt

$$\left|\chi^{-1}\{y_n\} \cap E\right| \leqq \frac{1}{1 - \delta} \quad (y_n \in \chi(E)),$$

und wir erhalten

$$|E| \leqq \frac{1}{1-\delta}|\chi(E)| \leqq \frac{1}{1-\delta} N(n, \delta) =$$
$$= \exp\{nC - \sqrt{n}\,.\,\lambda(\delta)\,T_{\operatorname{sign}\lambda(\delta)} + O(\log n)\}.$$

Dies ergibt zusammen mit (5.10) und (5.9)

$$\sqrt{n}\,.\,\lambda\left(\frac{\varepsilon}{\delta}\right)S + O(\log n) \leqq -\sqrt{n}\,.\,\lambda(\delta)\,T_{\operatorname{sign}\lambda(\delta)} + O(\log n), \tag{5.12}$$

oder

$$\lambda\left(\frac{\varepsilon}{\delta}\right)S + \lambda(\delta)\,T_{\operatorname{sign}\lambda(\delta)} \leqq 0 \tag{5.13}$$

für alle δ mit $\varepsilon < \delta < 1$ als notwendige Bedingung für $(p_n \xrightarrow{P_n})_\varepsilon$. Tatsächlich haben wir etwas mehr bewiesen. Falls nämlich für ein δ mit $\varepsilon < \delta < 1$

$$\lambda\left(\frac{\varepsilon}{\delta}\right)S + \lambda(\delta)\,T_{\operatorname{sign}\lambda(\delta)} > 0 \tag{5.14}$$

ist, gilt sogar $(p_n \xrightarrow{P_n}\!\!|)_\varepsilon$, da (5.13) schon folgt, wenn (5.12) für unendlich viele n richtig ist. Hieraus erhält man durch Einsetzen von $\delta = 2\varepsilon$ bzw. $\delta = \frac{1}{2}$ z. B.

$$(p_n \xrightarrow{P_n}\!\!|)_\varepsilon \quad (\varepsilon < \tfrac{1}{4}). \tag{5.15}$$

Ist entweder $S = 0$ oder $T_{-1} = 0$, so ergibt sich durch Einsetzen von $\delta = \varepsilon$ (für $S = 0$) oder $\delta > 2\varepsilon$ (für $T_{-1} = 0$) etwas schärfer

$$\left(p_n \overset{P_n}{\neg}\right)_\varepsilon \quad \left(\varepsilon < \tfrac{1}{2}\right).$$

Seien nun $0 < \gamma < 1$, $0 < \delta < 1$ so, daß

$$\lambda(\gamma)\, S + \lambda(\delta)\, T_{\operatorname{sign}\lambda(\delta)} < 0 \tag{5.16}$$

ist. Nach Satz 1 (bzw. einer einfachen überlegung im Falle $S = 0$) und Satz 2 folgt hieraus

$$\beta(n, \gamma) \leqq N(n, \delta)$$

für hinreichend große n. Für solche n gibt es also eine Menge $E_{n,\gamma}$ mit

$$p_n(E_{n,\gamma}) \geqq 1 - \gamma$$

und einen δ-Code f für P_n so, daß

$$|E_{n,\gamma}| \leqq |f(X_n)|\,.$$

Sei χ_0 eine eineindeutige Abbildung von $E_{n,\gamma}$ in $f(X_n)$ und seien χ und φ durch

$$\chi(z_n) = \begin{cases} \chi_0(z_n) & (z_n \in E_{n,\gamma}) \\ y_n^0 & (z_n \in Z_n - E_{n,\gamma}) \end{cases}$$

und

$$\varphi(x_n) = \begin{cases} \chi_0^{-1}(f(x_n)) & (x_n \in f^{-1}\,\chi_0(E_{n,\gamma})) \\ z_n^0 & (x_n \in X_n - f^{-1}\,\chi_0(E_{n,\gamma})) \end{cases}$$

erklärt, wobei y_n^0 und z_n^0 beliebige fest gewählte Elemente von Y_n bzw. Z_n sind.

Es folgt

$$\sum_{z_n \in Z_n} P_n(\varphi^{-1}\{z_n\}, \chi(z_n))\, p_n\{z_n\} \geqq \sum_{z_n \in E_{n,\gamma}} P_n(f^{-1}\{\chi(z_n)\}, \chi(z_n))\, p_n\{z_n\} \geqq$$
$$\geqq (1 - \delta)\, p_n(E_{n,\gamma}) \geqq (1 - \delta)(1 - \gamma) = 1 - (\delta + \gamma - \delta\gamma)\,,$$

so daß

$$\left(p_n \xrightarrow{P_n}\right)_{\delta + \gamma - \delta\gamma}.$$

Für $\delta = \gamma > \frac{1}{2}$ ist (5.16) stets erfüllt. Also gilt

$$\left(p_n \xrightarrow{P_n}\right)_\varepsilon \quad \left(\varepsilon > \tfrac{3}{4}\right).$$

Ist entweder $S = 0$ oder $T_1 = 0$, so ergibt sich ähnlich wie oben

$$\left(p_n \xrightarrow{P_n}\right)_\varepsilon \quad \left(\varepsilon > \tfrac{1}{2}\right).$$

Ich danke Herrn K. Jacobs für seine Unterstützung während der Entstehung dieser Arbeit und ihm sowie Herrn A. Renyi und Herrn R. Ahlswede für verschiedene interessante Bemerkungen.

LITERATUR

[1] C. E. Shannon: *A mathematical Theory of Communication.* Bell System Tech. J., Vol. 27 (1948).

[2] A. Feinstein: *Foundations of Information Theory.* New York, McGraw-Hill, 1958.

[3] Juschkewitsch: *Über einen Grenzwertsatz, der mit dem Begriff der Entropie einer Markoffschen Kette zusammenhängt.* Uspechi Mat. Nauk VIII, 5 (1953), (russisch).

[4] B. V. Gnedenko, A. N. Kolmogoroff: *Grenzverteilungen für Summen von unabhängigen Zufallsvariablen.* Berlin 1959.

[5] J. Wolfowitz: *Coding Theorems of Information Theory.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Berlin-Göttingen-Heidelberg 1961.

[6] R. L. Dobruschin: *Mathematical Problems in the Shannon Theory of Optimal Coding of Information.* Fourth Berkeley Symposium, 1961.

[7] L. Weiss: *On the Strong Converse of the Coding Theorem for Symmetric Channels Without Memory.* Quart. Appl. Math. 18, 3 (1960).

[8] K. Jacobs: *Almost Periodic Channels.* Colloquium on Combinatorial Methods in Probability Theory, Aarhus 1962.

[9] E. L. Lehmann: *Testing Statistical Hypotheses,* New York 1959.

[10] H. Cramér: *Random Variables and Probability Distributions.* Cambridge Tracts in Mathematics and Mathematical Physics, No. 36, London 1937.

[11] C. G. Esseen: *Fourier Analysis of Distribution Functions.* Acta Mathematica, Vol. 37.

[12] C. E. Shannon: *Some Geometric Results in Channel Capacity.* Nachr. Tech. Fachber., Vol. 6 (1956).

[13] F. und R. Nevanlinna: *Absolute Analysis, Die Grundlehren der Mathematischen Wissenschaften,* Band 102, Berlin-Göttingen-Heidelberg 1959.

[14] V. Strassen: *Messfehler und Information,* Zeitschrift für Wahrscheinlichkeitstheorie, 1964.

[15] G. Choquet: *Theory of Capacities,* Ann. Inst. Fourier, 5 (1953).

[16] K. Jacobs: *Bericht über Informationstheorie,* Jahresbericht der Deutschen Mathematiker Vereinigung, 1963.

[17] K. Jacobs: *Einführung in die Informationstheorie,* Göttingen 1959.

[18] H. Cramér: *Sur un nouveau théorème-limite de la théorie des probabilités.* Colloque consacré à la théorie probabilités, Paris 1938.

[19] W. Feller: *Generalization of a Probability Limit Theorem of Cramér.* Trans. Am. Math. Soc., 54 (1943), 3.

[20] A. Feinstein: *A New Basic Theorem of Information Theory.* I. R. E. Trans. I. T. – 4,2 (1954)

[21] P. Elias: *Coding for Two Noisy Channels,* Information Theory (C. Cherry, Editor), London 1956.

[22] R. M. Fano: *Transmission of Information.* New York, London 1961.

UNIVERSITÄT ZU GÖTTINGEN

MATHEMATISCHES INSTITUT

UNIVERSITY OF CALIFORNIA BERKELEY

DEPARTMENT OF STATISTICS