Z. Wahrscheinlichkeitstheorie 2, 273—305 (1964)

Meßfehler und Information*

Von

V. Strassen

Einleitung

In dieser Arbeit wird untersucht, welche Folgen es für die Informationstheorie hat, wenn man Beobachtungsfehler berücksichtigt, deren statistische Natur man nicht genau kennt. Das Ergebnis ist eine Verallgemeinerung der Hauptsätze von Shannon auf den Fall, daß an die Stelle von Wahrscheinlichkeitsverteilungen totalmonotone bzw. totalalternierende Kapazitäten im Sinne der Potentialtheorie treten. Die Lektüre der Arbeit setzt keine Vorkenntnisse über Potentialtheorie voraus.

Wir beginnen mit einer heuristischen Betrachtung. Eine zufällige Größe (ein zufälliges Experiment) wird mathematisch durch eine WV (Wahrscheinlichkeitsverteilung) w beschrieben, die auf einem σ-Körper von Teilmengen einer Menge A definiert ist (Kolmogoroff [1]). In vielen Anwendungen ist A eine differenzierbare Mannigfaltigkeit. Die bei einer Messung der zufälligen Größe möglichen Meßwerte bilden dagegen stets eine endliche Menge X, die gewöhnlich aus Dezimalzahlen oder Vektoren von solchen besteht.

Den Übergang von A nach X, also die Beobachtung, beschreibt man häufig durch einen Markoffschen Kern P von A nach X: $P(E, a)$ ist die Wahrscheinlichkeit dafür, daß ein $x \in E$ beobachtet wird, falls a der wahre Wert der zufälligen Größe ist.

Unter der Annahme, daß P bekannt ist und sich bei unabhängiger Wiederholung des Experiments unabhängig reproduziert, werden die relativen Häufigkeiten der Meßwerte $x \in X$ ebenfalls durch eine WV, nämlich Pw geregelt $((Pw)(E) = \int P(E, a)\, w(da))$, wobei Pw bekannt ist, sobald man w kennt.

Die Voraussetzungen über P sind allerdings selten mit hinreichender Genauigkeit erfüllt. Die Fehlerquellen, die etwa beim Ablesen eines Voltmeters auftreten, verhalten sich meist entweder nicht invariant (z.B. tageszeitliche Temperaturschwankungen) oder nicht unabhängig (z.B. Parallaxe, gröbere mechanische Erschütterungen, Wünsche des Experimentators). Außerdem besteht gewöhnlich die einzige verwertbare Kenntnis, die man über P besitzt, darin, daß $P(\{x\}, a)$ nur für solche x positiv ist, die in unmittelbarer Nähe von a liegen (gegebenenfalls unter Vernachlässigung sehr kleiner Wahrscheinlichkeiten, vergleiche die übernächste Seite unten).

Man hat es also in Wahrheit nicht mit einem einzelnen P, sondern mit einer Klasse $\boldsymbol{K}$ von Kernen zu tun, deren Elemente sich bei unabhängiger Wiederholung des Experiments nicht unabhängig zu reproduzieren brauchen. Nach dem Voran-

---

* Die Arbeit entstand im Rahmen eines von der Deutschen Forschungsgemeinschaft finanzierten Forschungsvorhabens.

gehenden liegt es nahe, $\boldsymbol{K}$ folgendermaßen zu definieren: Jedem $a \in A$ wird eine nichtleere Teilmenge $T(a)$ von X zugeordnet mit der Bedeutung, daß ein $x \in T(a)$ beobachtet wird, falls a der wahre Wert der zufälligen Größe ist (ebensogut kann man jedem x die a zuordnen, die als wahre Versuchsergebnisse in Frage kommen, falls x abgelesen wurde: Dies deckt sich mit der empirischen Fehlerabschätzung des Physikers). Dann ist

$$\boldsymbol{K} = \{P : P \text{ Markoffscher Kern von } A \text{ nach } X, \quad P(T(a), a) = 1\}.$$

Bezüglich einer n-maligen Wiederholung des durch w beschriebenen Experiments wird man annehmen, daß die dem Ergebnis

$$a_n = (a^1, \ldots, a^n) \in \prod_1^n A$$

zugeordnete Menge $T_n(a_n)$ mit der Menge

$$T(a^1) \times \cdots \times T(a^n)$$

übereinstimmt und wird

$$\boldsymbol{K}_n = \{P_n : P_n \text{ Markoffscher Kern von } \prod_1^n A \text{ nach } \prod_1^n X,\ P_n(T_n(a_n), a_n) = 1\}$$

setzen.

$\boldsymbol{K}_n$ induziert in $\prod_1^n X$ eine Klasse von WVen, in der Sprechweise der Informationstheorie eine Klasse von Quellen, nämlich

$$\{P_n w_n : P_n \in \boldsymbol{K}_n\} \tag{1}$$

mit $w_n = \prod_1^n w$. Es scheint vernünftig, dieser Klasse mit einer Minimaxstrategie zu begegnen (vgl. ROOT [27]). Verschiedene bekannte Methoden, die Shannonschen Resultate auf Klassen von Quellen bzw. Kanälen im Sinne des Minimaxstandpunktes zu verallgemeinern (BLACKWELL-BREIMAN-THOMASIAN [2], [3], WOLFOWITZ [4], [5], KIEFER-WOLFOWITZ [6]), sind allerdings hier nicht anwendbar, da $\{P_n w_n : P_n \in \boldsymbol{K}_n\}$ nicht nur invariante oder unabhängige Quellen enthält.

Es erweist sich nun als zweckmäßig, den statistischen Gesetzmäßgkeiten, die in X oder $\prod_1^n X$ herrschen, eine andere äquivalente Formulierung zu geben. Dazu fragen wir direkt nach den relativen Häufigkeiten eines Ereignisses $E \subseteqq X$ bei unabhängigem Wiederholen des Experimentes w.

Zunächst ist klar, daß die Häufigkeit von E niemals die von

$$\{a : T(a) \subseteqq E\}$$

unterschreitet. Denn ist ein $b \in \{a : T(a) \subseteqq E\}$ der wahre Wert der zufälligen Größe, so wird sicher ein $x \in T(b) \subseteqq E$ beobachtet. Ebenso ist die Häufigkeit von E stets $\leqq$ der Häufigkeit von

$$\{a : T(a) \cap E \neq 0\}.$$

Sind nun die Mengen $\{a : T(a) \subseteqq E\}$ und $\{a : T(a) \cap E \neq 0\}$ meßbar (was man von der Abbildung T verlangen wird), so streben ihre relativen Häufigkeiten mit Wahrscheinlichkeit 1 gegen $w\{a : T(a) \subseteqq E\}$ und $w\{a : T(a) \cap E \neq 0\}$. Setzen wir also

$$u(E) = w\{a : T(a) \subseteqq E\},$$
$$v(E) = w\{a : T(a) \cap E \neq 0\},$$

so liegen alle limes-Punkte der Folge der relativen Häufigkeiten von E mit Wahrscheinlichkeit 1 zwischen $u(E)$ und $v(E)$. Mehr können wir andererseits über die Häufigkeit von E nicht allgemeingültig aussagen, so daß durch die beiden Mengenfunktionen u und v die gesuchten Gesetzmäßigkeiten für das Auftreten von Meßwerten beschrieben sind.

u ist eine totalmonotone, v eine totalalternierende Kapazität im Sinn von Choquet [7]. Das wird deutlich, wenn man den Raum Ω der nichtleeren Teilmengen von X einführt: T ist eine meßbare Abbildung von A nach Ω, transportiert also die WV w in eine WV μ in Ω, und nach Definition von μ gilt

$$\begin{aligned} u(E) &= \mu\{\omega : \omega \subseteqq E\}, \\ v(E) &= \mu\{\omega : \omega \cap E \neq 0\} = 1 - u(X - E). \end{aligned} \tag{2}$$

Sieht man von der Normierung $u(X) = v(X) = 1$ und einer additiven Konstanten ab, so läßt sich (2) als Definition totalmonotoner bzw. totalalternierender Kapazitäten verwenden (siehe Choquet [7], Satz 50.1).

Die hier für w und T durchgeführte Betrachtung kann man ebenso für w_n und T_n anstellen und erhält statt μ das Produktmaß $\mu_n = \prod_1^n \mu$ und statt u und v totalmonotone bzw. totalalternierende Kapazitäten u_n und v_n, die man die unabhängigen Produkte von n Exemplaren von u bzw. v nennen wird. Eine ganz analoge Interpretation haben die in Abschnitt 4 definierten Begriffe der Invarianz und Ergodizität totalmonotoner Kapazitäten.

Der Zusammenhang mit (1) wird durch die leicht zu beweisenden Formeln

$$\begin{aligned} u_n(E_n) &= \min\{(P_n w_n)(E_n) : P_n \in \boldsymbol{K}_n\} \\ v_n(E_n) &= \max\{(P_n w_n)(E_n) : P_n \in \boldsymbol{K}_n\} \end{aligned} \qquad \left(E_n \subseteqq \prod_1^n X\right)$$

hergestellt. Diese liefern zusammen mit der Minimax-Forderung (siehe oben) ein Rezept, Verallgemeinerungen von Sätzen der mathematischen Statistik (der Informationstheorie) auf totalmonotone bzw. totalalternierende Kapazitäten zu formulieren. So wird man Ungleichungen der Art $p(E) > 1 - \varepsilon$ (p eine WV) im allgemeinen durch $u(E) > 1 - \varepsilon$, Ungleichungen $p(E) < \varepsilon$ durch $v(E) < \varepsilon$ zu ersetzen haben.

Wir bemerken noch, daß auch die Beobachtung von stochastischen Prozessen mit kontinuierlicher Zeit auf totalmonotone Kapazitäten führt: Hier ist ein nichttrivialer Meßvorgang bezüglich der Zeit zu berücksichtigen.

Ferner kann man die Abbildung T von A nach Ω durch einen Markoffschen Kern von A nach Ω ersetzen (von dem man allerdings annehmen muß, daß er sich bei unabhängiger Wiederholung des Experiments unabhängig reproduziert). Ein Beispiel wäre der Kern S mit $S(\{T(a)\}, a) = 0{,}99$, $S(\{X\}, a) = 0{,}01$.

In Abschnitt 1 werden einige Hilfsmittel zusammengestellt.

In Abschnitt 2 werden das Neyman-Pearsonsche Lemma und eine Verschärfung des bekannten Resultats von SHANNON über das asymptotische Verhalten der Minimalmächtigkeit hochwahrscheinlicher Mengen für stationäre Quellen mit unabhängigen Zeichen auf totalmonotone Kapazitäten verallgemeinert (Satz 2.1, Satz 2.2). Die Methode ist repräsentativ in dem Sinne, daß im Fall des Neyman-Pearsonschen Lemmas zwei WVen, im Falle des Shannonschen Satzes eine Quelle konstruiert werden, die den jeweiligen totalmonotonen Kapazitäten im Hinblick auf das betrachtete Problem völlig äquivalent sind. Die Verallgemeinerung des Shannonschen Satzes auf nichtstationäre (z. B. fastperiodische, vgl. JACOBS [8], [28]) totalmonotone Kapazitäten mit unabhängigen Zeichen bietet keine Schwierigkeiten, die auf ergodische totalmonotone Kapazitäten wird in Abschnitt 5 behandelt.

In Abschnitt 3 wird ein Coding-Theorem mit schwacher Umkehrung für stationäre Unterkanäle ohne Gedächtnis bewiesen (Satz 3.1): Die Definition eines Unterkanals erhält man aus der des Kanals, indem man alle auftretenden WVen durch totalmonotone Kapazitäten ersetzt. Den Beweis des Coding-Theorems kann man leider nicht mit einem im obigen Sinne repräsentativen Verfahren führen (wie einfache Beispiele zeigen, vgl. auch den Ausdruck (3.13), (3.15) für die Kapazität). Die Verallgemeinerung auf Unterkanäle mit endlichem Gedächtnis ist leicht.

Satz 3.1 ist äquivalent zur folgenden Aussage: Sei $(P_n)_{n \geqq 1}$ ein stationärer Kanal ohne Gedächtnis von Y nach A (siehe Abschnitt 3; Y und A sind endlich), R eine reflexive und symmetrische Relation in A und $R_n = \prod_1^n R \subseteqq \prod_1^n A \times \prod_1^n A$ (bis auf triviale Isomorphie). Eine Menge

$$\{(y_n^{(i)}, E_n^{(i)}) : i = 1, \ldots, N\} \quad \text{mit} \quad y_n^{(i)} \in \prod_1^n Y, \quad E_n^{(i)} \subseteqq \prod_1^n A,$$
$$P_n(E_n^{(i)}, y_n^{(i)}) \geqq 1 - \varepsilon, \quad (E_n^{(i)} \times E_n^{(j)}) \cap R_n = 0 \qquad (i \neq j)$$

heiße ein ε-Code der Länge N für P_n bezüglich R_n. Das Maximum solcher N bezeichnen wir mit $N_R(n, \varepsilon)$.

Dann gibt es ein C mit

$$n C + o(n) < \log N_R(n, \varepsilon) < n C + \varepsilon\, 0(n)$$

(man kann nämlich A in die Menge Ω der nichtleeren Teilmengen einer geeigneten endlichen Menge X so einbetten, daß

$$R = \{(\omega, \omega') : \omega \cap \omega' \neq 0\} \cap (A \times A)$$

ist; hieraus folgt leicht die Äquivalenz mit Satz 3.1).

Abschnitt 4 enthält unter anderem das folgende Ergebnis (siehe Satz 4.3): Seien X und A metrisch kompakte Räume, R eine abgeschlossene Relation zwischen X und A und q und w je eine WV in X bzw. A. Notwendig und hinreichend für die Existenz einer WV in R mit q und w als Marginalverteilungen ist

$$q(E) \leqq w(\mathrm{proj}_A(E \times A) \cap R)$$

für jedes abgeschlossene $E \subseteqq X$. Ein Spezialfall dieser Aussage folgt auch aus dem Heiratssatz der Theorie der fastperiodischen Funktionen (siehe MAAK [29]).

1. Totalmonotone Kapazitäten auf endlichen Mengen

Sei $X = \{x, \ldots\}$ eine nichtleere endliche Menge, $\Omega = \Omega(X) = \{\omega, \ldots\}$ die Menge der nichtleeren Teilmengen von X (beliebige Teilmengen von X bezeichnen wir mit $E, F, \ldots$), sei $Q = \{q, p, \ldots\}$ die Menge der endlichen Maße in X (statt $q(\{x\})$ schreiben wir $q\{x\}$) und $M = \{\mu, \ldots\}$ die Menge der endlichen Maße in Ω.

Die durch

$$(\varphi\,\mu)(E) = \sum_{\omega \subseteqq E} \mu\{\omega\} = \mu\{\omega : \omega \subseteqq E\} \tag{1.1}$$

und

$$(\psi\,\mu)(E) = \sum_{\omega \cap E \neq 0} \mu\{\omega\} = \mu\{\omega : \omega \cap E \neq 0\} \qquad (E \subseteqq X) \tag{1.2}$$

definierten Abbildungen φ und ψ von M in die Menge der Mengenfunktionen in X sind eineindeutig (Shapley [9], Choquet [7], es gilt z. B.

$$\mu\{\omega\} = \sum_{E \subseteqq \omega} (-1)^{|\omega - E|} (\varphi\,\mu)(E),$$

wobei $|\omega - E|$ die Anzahl der x ist, die in ω, aber nicht in E liegen; siehe auch Lemma 4.2). Die $u \in \varphi M$ heißen totalmonotone Kapazitäten mit $u(0) = 0$ und die $v \in \psi M$ totalalternierende Kapazitäten mit $v(0) = 0$ (Choquet [7]; wir werden im folgenden „mit $u(0) = 0$“ und „mit $v(0) = 0$“ weglassen; wenn von u, v und μ die Rede ist, wird stets $u = \varphi\,\mu$, $v = \psi\,\mu$ angenommen). Es gilt

$$v(E) = u(X) - u(X - E) \qquad (E \subseteqq X). \tag{1.3}$$

Wir schreiben $q \geqq u$ für

$$q \in Q, \quad q(X) = u(X), \quad q(E) \geqq u(E) \qquad (E \subseteqq X)$$

und $q \leqq v$ für

$$q \in Q, \quad q(X) = v(X), \quad q(E) \leqq v(E) \qquad (E \subseteqq X),$$

$q \geqq u$ und $q \leqq v$ bedeuten also dasselbe.

Ist P ein Markoffscher Kern von Ω nach X, d. h. ist $P(E, \omega)$ für jedes $\omega \in \Omega$ als Funktion von $E \subseteqq X$ eine WV (= Wahrscheinlichkeitsverteilung) in X, so ist durch

$$(P\,\mu)(E) = \sum_{\omega \in \Omega} P(E, \omega)\,\mu\{\omega\}$$

ein $P\,\mu \in Q$ erklärt. Die Menge aller Markoffschen Kerne P von Ω nach X mit

$$P(\omega, \omega) = 1 \qquad (\omega \in \Omega)$$

bezeichnen wir mit $\boldsymbol{P}$.

Satz 1.1.

$$\{q : q \geqq u\} = \{P\,\mu : P \in \boldsymbol{P}\}.$$

Beweis.

$$\{P\,\mu : P \in \boldsymbol{P}\} \subseteqq \{q : q \geqq u\}$$

folgt aus

$$(P\,\mu)(E) \geqq \sum_{\omega \subseteqq E} P(E, \omega)\,\mu\{\omega\} = \sum_{\omega \subseteqq E} \mu\{\omega\} = u(E).$$

Sei umgekehrt $q \geqq u$,

$$V = \{E : q(E) = v(E)\}$$

und

$$V_\omega = \{E : q(E) = v(E), \quad \omega \cap E \neq 0\} \qquad (\omega \in \Omega).$$

Für $\omega \in \Omega$ mit $\mu\{\omega\} > 0$ ist V_ω ein Mengenverband. Denn aus

$$q(E) = v(E) = \sum_{\omega' \cap E \neq 0} \mu\{\omega'\},$$

$$q(F) = v(F) = \sum_{\omega' \cap F \neq 0} \mu\{\omega'\},$$

$$\omega \cap E \neq 0, \quad \omega \cap F \neq 0$$

folgt zunächst

$$\begin{aligned} q(E \cup F) &= q(E) + q(F) - q(E \cap F) \\ &\geqq \sum_{\omega' \cap E \neq 0} \mu\{\omega'\} + \sum_{\omega' \cap F \neq 0} \mu\{\omega'\} - \sum_{\omega' \cap E \cap F \neq 0} \mu\{\omega'\} \\ &\geqq \sum_{\omega' \cap (E \cup F) \neq 0} \mu\{\omega'\} = v(E \cup F) \geqq q(E \cup F). \end{aligned}$$

Da die Enden dieser Ungleichung übereinstimmen, ist jedes $\geqq$ ein $=$. Für die beiden äußeren $\geqq$ impliziert das

$$q(E \cap F) = v(E \cap F)$$

und

$$q(E \cup F) = v(E \cup F)$$

und für das mittlere

$$\omega \cap (E \cap F) \neq 0,$$

da sonst $\mu\{\omega\} > 0$ links des $\geqq$ zweimal, rechts davon aber nur einmal aufträte. Ebenso sieht man, daß V ein Mengenverband ist.

Wir nehmen nun an, q sei ein Extremalpunkt der konvexen kompakten Menge $\{q : q \geqq u\} = \{q : q \leqq v\}$. Dann trennt V Punkte. Sind nämlich $x, y \in X$, $x \neq y$ und gilt

$$q(E) < v(E)$$

für alle E, welche x und y trennen, so ist $q + \varepsilon\delta_x - \varepsilon\delta_y \leqq v$ für hinreichend kleines $|\varepsilon|$, was der Extremalität von q widerspricht (δ_x bedeutet die Einheitsmasse im Punkte x). Es gibt also ein $E \in V$, das x und y trennt. Hieraus und aus der Definition der V_ω folgt, daß V_ω die Punkte von ω trennt.

Für $\mu\{\omega\} > 0$ ist $\omega \cap \bigcap V_\omega$ nach dem Vorausgehenden eine einpunktige Menge, deren Element wir mit $f(\omega)$ bezeichnen. Für $\mu\{\omega\} = 0$ sei $f(\omega) \in \omega$ beliebig. Wir setzen

$$P(E, \omega) = \begin{cases} 1 & \text{wenn} \quad f(\omega) \in E \\ 0 & \text{sonst}. \end{cases}$$

Ersichtlich ist $P \in \boldsymbol{P}$. Nach der Definition von f gilt

$$(1.4) \qquad P(E, \omega) = 1 \qquad (E \in V_\omega, \mu\{\omega\} > 0).$$

Sei $E \in V$. Dann ist

$$\{\omega : \mu\{\omega\} > 0,\, \omega \cap E \neq 0\} = \{\omega : \mu\{\omega\} > 0,\, E \in V_\omega\}$$

und aus (1.4) folgt

$$\begin{aligned}(P\mu)(E) &= \sum_{\omega \in \Omega} P(E,\omega)\,\mu\{\omega\} \\ &= \sum_{\mu\{\omega\}>0,\,\omega \cap E \neq 0} P(E,\omega)\,\mu\{\omega\} \\ &= \sum_{\omega:\mu\{\omega\}>0,\,E \in V_\omega} P(E,\omega)\,\mu\{\omega\} \\ &= \sum_{\omega:\mu\{\omega\}>0,\,E \in V_\omega} \mu\{\omega\} = \sum_{\omega \cap E \neq 0} \mu\{\omega\} = v(E),\end{aligned}$$

d. h.

$$q(E) = (P\mu)(E).$$

Da V ein punktetrennender Verband ist, haben wir

$$q = P\mu.$$

Der Satz folgt jetzt aus dem Satz von Minkowski-Krein-Milman.

Anmerkung. Faßt man u als charakteristische Funktion eines kooperativen Spiels mit der Spielermenge X auf (siehe von Neumann-Morgenstern [*10*], Burger [*11*]), so charakterisiert Satz 1.1 diejenigen Imputationen, die von keiner anderen dominiert werden.

Setzen wir im Anschluß an Choquet [*7*]

$$\int f\,du = \int_0^\infty u\{f > t\}\,dt$$

$$\int f\,dv = \int_0^\infty v\{f > t\}\,dt$$

für nichtnegative Funktionen f auf X, so haben wir das

Corollar 1.2.

$$\begin{aligned}\int f\,du &= \min\{\textstyle\int f\,dq : q \geqq u\} \\ \int f\,dv &= \max\{\textstyle\int f\,dq : q \leqq v\}.\end{aligned}$$

Beweis. Aus

$$\int f\,dq = \int_0^\infty q\{f > t\}\,dt$$

folgt sofort

$$\int f\,du \leqq \min\{\textstyle\int f\,dq : q \geqq u\}.$$

Ist andererseits $P \in \boldsymbol{P}$ mit $P(\{x : x \in \omega,\ f(x) \leqq f(y)\ (y \in \omega)\}, \omega) = 1$ für alle $\omega \in \Omega$, so gilt

$$\int f\,du = \int f\,d(P\mu).$$

(das Corollar kann man auch direkt aus dem Satz von Hahn-Banach, angewandt auf die Halbnorm $\int |f|\,dv$ im linearen Raum der reellen Funktionen auf X gewinnen).

Seien $q,\ q' \in Q$ und $q' > 0$ (d.h. $q'\{x\} > 0$ $(x \in X)$). Dann heißt

$$H(q, q') = - \sum_{q\{x\} > 0} q\{x\} \log \frac{q\{x\}}{q'\{x\}} \tag{1.5}$$

die relative Entropie von q bezüglich q'. Sind $u,\ u' \in \varphi M$ und ist $u' > 0$ (d.h. $u'\{x\} > 0$ $(x \in X)$), so nennen wir

$$H(u, u') = \sup\{H(q, q') : q \geqq u,\ q' \geqq u'\} \tag{1.6}$$

die relative Entropie von u bezüglich u'.

Lemma 1.3. *Seien $u,\ u' \in \varphi M$ und $u' > 0$. Die Menge*

$$R = \{(q, q') : q \geqq u,\ q' \geqq u',\ H(u, u') = H(q, q')\}$$

ist konvex, kompakt und nichtleer. Die Dichte dq/dq' und deren Verteilung bezüglich q ist für alle $(q, q') \in R$ die gleiche. Ist $q \geqq u$, $q' \geqq u'$, so liegt (q, q') genau dann in R, wenn für alle $a \geqq 0$

$$\begin{aligned} q\left\{\frac{dq}{dq'} > a\right\} &= u\left\{\frac{dq}{dq'} > a\right\} \\ q'\left\{\frac{dq}{dq'} > a\right\} &= v'\left\{\frac{dq}{dq'} > a\right\} \end{aligned} \tag{1.7}$$

gilt.

Beweis. Aus der Stetigkeit von $H(q, q')$ auf $\{(q, q') : q \geqq u,\ q' \geqq u'\}$ folgt, daß R kompakt und nichtleer ist. Sei $(q, q') \in R$, $a \geqq 0$ und $E_a = \{dq/dq' > a\}$. Ist $u = \varphi\mu$, so gibt es nach Satz 1.1 ein $P \in \boldsymbol{P}$ mit $q = P\mu$. Angenommen, wir haben $q(E_a) > u(E_a)$, also

$$\sum_{\omega \in \Omega} P(E_a, \omega)\, \mu\{\omega\} > \sum_{\omega \in E_a} \mu\{\omega\}.$$

Für ein $\omega \nsubseteq E_a$ ist dann $P(E_a, \omega)\mu\{\omega\} > 0$. Sei $x \in E_a$, $P(\{x\}, \omega) > 0$ und $y \in \omega - E_a$. Definieren wir für $\omega' \in \Omega$ und $E \subseteqq X$

$$\widetilde{P}(E, \omega') = \begin{cases} P(E, \omega') & \text{wenn} \quad \omega' \neq \omega \\ P(E, \omega) + \varepsilon\,\delta_y - \varepsilon\,\delta_x & \text{wenn} \quad \omega' = \omega, \end{cases}$$

so ist $\widetilde{P} \in \boldsymbol{P}$ für hinreichend kleine $\varepsilon > 0$ und wegen

$$\frac{dq}{dq'}(x) > a \geqq \frac{dq}{dq'}(y)$$

gilt

$$H(q, q') < H(\widetilde{P}\mu, q')$$

(wie man leicht nachrechnet), im Widerspruch zu $(q, q') \in R$. Damit ist die erste der Gleichungen (1.7) bewiesen. Genau so folgt die zweite.

Sei nun $S = \{(q, q') : q \geqq u,\ q' \geqq u',\ (1.7)\}$. Wir zeigen, daß $H(q, q')$ auf S konstant ist. Für ein beliebiges Paar $(q, q') \in S$ besteht $\{\{dq/dq' > a\} : a \geqq 0\}$ aus endlich vielen ineinander geschachtelten Teilmengen von X, etwa

$$0 = F_0 \subset F_1 \subset \cdots \subset F_n.$$

Aus (1.7) folgt

$$(1.8)\qquad \begin{aligned} u(F_{i+1}) - u(F_i) &= q(F_{i+1} - F_i) \\ v'(F_{i+1}) - v'(F_i) &= q'(F_{i+1} - F_i) \qquad (0 \leqq i < n). \end{aligned}$$

Für beliebiges $E \supseteqq F_i$ gilt also nach Definition der F_k und wegen $(q, q') \in S$

$$\frac{u(F_{i+1}) - u(F_i)}{v'(F_{i+1}) - v'(F_i)} = \frac{q(F_{i+1} - F_i)}{q'(F_{i+1} - F_i)} \geqq \frac{q(E - F_i)}{q'(E - F_i)} = \frac{q(E) - q(F_i)}{q'(E) - q'(F_i)} \geqq \frac{u(E) - u(F_i)}{v'(E) - v'(F_i)}.$$

Das erste $\geqq$ ist ein $>$, sobald $E - F_{i+1} \neq 0$ ist. Daraus folgt

$$F_{i+1} = \bigcup\left\{F : F \supseteqq F_i, \frac{u(F) - u(F_i)}{v'(F) - v'(F_i)} \geqq \frac{u(E) - u(F_i)}{v'(E) - v'(F_i)} \text{ für alle } \quad E \supseteqq F_i\right\}.$$

$\{\{dq/dq' > a\} : a \geqq 0\}$ hängt also nicht von $(q, q') \in S$ ab. Wegen (1.8) und weil nach Definition der F_k die relative Dichte dq/dq' konstant auf $F_{i+1} - F_i$ ist, haben alle $(q, q') \in S$ dieselbe relative Dichte (also wegen $R \subseteqq S$ erst recht alle $(q, q') \in R$. Außerdem ergibt sich jetzt aus der Definition von S, daß S konvex ist). Der von dq/dq' induzierte Mengenkörper ist aus den $F_{i+1} - F_i$ als Atomen aufgebaut, wegen (1.8) hängt also auch die Verteilungsfunktion von dq/dq' bezüglich q und damit $H(q, q')$ nicht von $(q, q') \in S$ ab. Hieraus und aus $0 \neq R \subseteqq S$ folgen $R = S$ nach Definition von R und die restlichen Aussagen des Satzes (die Konvexität von R hätte man auch aus der wohlbekannten Konkavität von H ableiten können).

Das Lemma besitzt eine Verallgemeinerung für den Fall, daß statt der Voraussetzung $u' > 0$ nur die schwächere Totalstetigkeitsannahme

$$u(E) = 0 \qquad (E \subseteqq X, v'(E) = 0)$$

erfüllt ist. Die Formulierung macht aber etwas mehr Mühe.

2. Entropie

Seien $u, u' \in \varphi M$. Ein Test f ist eine Abbildung von X nach $\langle 0, 1\rangle$. Das Risiko von f bezüglich u, u' ist der Vektor

$$(2.1)\qquad r_{u,u'}(f) = \left(\int f\, dv, \int (1 - f)\, dv'\right).$$

Eine Klasse $\boldsymbol{F}$ von Tests heißt im wesentlichen vollständig für u, u', wenn es zu jedem Test g ein $f \in \boldsymbol{F}$ mit $r_{u,u'}(f) \leqq r_{u,u'}(g)$ gibt. Eine im wesentlichen vollständige Klasse heißt minimal im wesentlichen vollständig, wenn sie keine echte im wesentlichen vollständige Teilklasse enthält (siehe Lehmann [*12*], Wald [*13*], LeCam [*14*], Blackwell-Girshick [*15*]). Sind $q, q' \in Q$, $q' > 0$, so heißt

$$\boldsymbol{F}_{q,q'} = \{f_{a,b} : a \geqq 0,\ 0 \leqq b \leqq 1\}$$

mit

$$f_{a,b}(x) = \begin{cases} 1 & \text{wenn } \dfrac{q\{x\}}{q'\{x\}} < a \\[2ex] b & \text{wenn } \dfrac{q\{x\}}{q'\{x\}} = a \\[2ex] 0 & \text{wenn } \dfrac{q\{x\}}{q'\{x\}} > a \end{cases}$$

die Klasse der Neyman-Pearson-Tests zwischen q und q' (siehe etwa Lehmann [*12*]).

Satz 2.1. *Seien u, $u' \in \varphi M$, $u' > 0$, $q \geqq u$, $q' \geqq u'$ und $H(q, q') = H(u, u')$. Die Klasse $\boldsymbol{F}_{q,q'}$ der Neyman-Pearson Tests zwischen q und q' ist minimal im wesentlichen vollständig für u, u'. Darüber hinaus gilt*

$$r_{u,u'}(f) = r_{q,q'}(f) \qquad (f \in \boldsymbol{F}_{q,q'}). \tag{2.2}$$

Beweis. Da nach dem Lemma von Neyman und Pearson $\boldsymbol{F}_{q,q'}$ minimal im wesentlichen vollständig für q, q' ist und da für beliebige Tests f

$$r_{u,u'}(f) \geqq r_{q,q'}(f) \tag{2.3}$$

gilt (Corollar 1.2), genügt es, (2.2) zu beweisen. Aus (1.7) und der Definition von $\int (1-f)\,dv'$ folgt aber sofort

$$\int (1-f)\,dv' = \int (1-f)\,dq' \qquad (f \in \boldsymbol{F}_{q,q'}),$$

da $1 - f$ für $f \in \boldsymbol{F}_{q,q'}$ monoton von dq/dq' abhängt. Analog erhält man

$$\int f\,dv = \int f\,dq$$

indem man (1.3) heranzieht.

Ein Test zwischen u und u' mit dem Risiko (2.1) ist nach Corollar 1.2 nichts anderes als ein Test zwischen den zusammengesetzten Hypothesen $\{p : p \geqq u\}$ und $\{p' : p' \geqq u'\}$ mit dem üblichen supremum-Risiko. Der Satz behauptet die Existenz von q, q' aus den beiden Hypothesen mit der Eigenschaft, daß eine vernünftige Entscheidung zwischen q und q' zugleich auch eine vernünftige Entscheidung zwischen $\{q : p \geqq u\}$ und $\{p' : p' \geqq u'\}$ ist und daß das (nach q, q' berechnete) Risiko einer solchen Entscheidung mit dem supremum-Risiko übereinstimmt. Ein entsprechender Satz für beliebige konvexe kompakte Hypothesen ist falsch, selbst wenn die eine einfach ist (vgl. auch LEHMANN [*16*]): Setzt man

$$X = \{1, 2, 3\}, \quad H = \left\{\left(\frac{1}{2} + \frac{t}{6}\right)\delta_1 + \left(\frac{1}{2} - \frac{t}{3}\right)\delta_2 + \frac{t}{6}\delta_3 : 0 \leqq t \leqq 1\right\}$$

und $K = \{q'\}$ (wobei $q' = \sum_{x=1}^{3} \delta_x$ ist), so gibt es kein $q \in H$ mit

$$q\left\{\frac{dq}{dq'} > a\right\} = \min\left\{p\left\{\frac{dq}{dq'} > a\right\} : p \in H\right\}.$$

Wir gehen nun zu Anwendungen auf die Informationstheorie über. Ist

$$X_n = \{x_n = (x^1, \ldots, x^n) : x^k \in X\}, \quad \Omega_n = \{\omega_n = (\omega^1, \ldots, \omega^n) : \omega^k \in \Omega\},$$

so gilt bis auf triviale Isomorphie

$$\Omega_n \subseteqq \Omega(X_n).$$

Wir bilden die zu (1.1) und (1.2) analogen Abbildungen für die Menge X_n an Stelle von X, deren Einschränkungen auf Ω_n wir mit φ_n und ψ_n bezeichnen. Sei μ eine WV in Ω, μ_n das (unabhängige) Produkt von n Exemplaren von μ und

$$u_n = \varphi_n \mu_n$$

(statt u_1 schreiben wir auch u). Sei $\mu' \in M$,

$$u' = \varphi \mu' > 0, \tag{2.4}$$

μ_n' das Produkt von n Exemplaren von μ' und

$$u_n' = \varphi_n \mu_n', \quad v_n' = \psi_n \mu_n'$$

(es ist $u' = u_1'$; statt v_1' schreiben wir auch v'). Wir wollen das asymptotische Verhalten ($n \to \infty$) von

$$B(n, \varepsilon) = \min\{v_n'(E_n) : u_n(E_n) \geqq 1 - \varepsilon\} \tag{2.5}$$

und

$$\Gamma(n, \varepsilon) = \min\{\textstyle\int f_n\, dv_n' : \int f_n\, du_n \geqq 1 - \varepsilon\} \tag{2.6}$$

bei gegebenem ε mit $0 < \varepsilon < 1$ untersuchen (mit E_n werden Teilmengen von X_n bezeichnet, mit f_n Abbildungen von X_n in $\langle 0, 1\rangle$). Seien $q \geqq u$, $q' \geqq u'$ mit

$$H(q, q') = H(u, u').$$

q ist eine WV. Aus Lemma 1.3 folgt, daß die Verteilung der Funktion

$$h = -\log \frac{dq}{dq'}$$

bezüglich q von der Wahl von q und q' unabhängig ist. Wir bezeichnen den Erwartungswert von h (bezüglich q) mit H (es ist $H = H(u, u')$), die Streuung von h mit S und schließen den für das Folgende trivialen Fall $S = 0$ aus. Sei

$$Q(t) = \frac{\int (h-H)^3\, dq}{6 S^2} (t^2 - 1) \qquad (t \text{ reell}),$$

λ die durch

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-t^2/2} dt = 1 - \varepsilon$$

bestimmte reelle Zahl und sei für gegebenes $r > 0$

$$w(t) = -d(t) + \log\left(d(t) + \frac{r}{2} \frac{e^r + 1}{e^r - 1}\right) \qquad (t \text{ reell}),$$

wobei $d(t)$ der absolut kleinste Rest von $t \bmod r$ ist.

Satz 2.2. *Es gilt*

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

falls die Verteilung von h nicht gitterförmig ist und

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

falls die Verteilung von h gitterförmig ist mit der Gitterkonstanten r, wobei a eine reelle Zahl mit $p\{h - H = a\} > 0$ ist. Ferner

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

für nicht gitterförmig verteiltes h und

$$\log B(n, \varepsilon) = nH + \sqrt{n}\,\lambda S - \frac{1}{2}\log n + O(1)$$

allgemein.

Beweis. Wir bemerken zunächst, daß für beliebige $E^k \subseteqq X$

$$u_n(E^1 \times \cdots \times E^n) = u(E^1) \cdots u(E^n)$$

und

$$v_n'(E^1 \times \cdots \times E^n) = v'(E^1) \cdots v'(E^n)$$

ist (dies folgt leicht aus der Definition von u_n, v_n').

Sei q_n (bzw. q_n') das Produktmaß von n Exemplaren von q (bzw. q'). Dann gilt $q_n \geqq u_n$, $q_n' \leqq v_n'$. Für jedes $a \geqq 0$ ist die Menge

$$\left\{\frac{dq_n}{dq_n'} > a\right\}$$

eine Vereinigung von Durchschnitten von Mengen der Gestalt

$$\left\{x_n : \frac{q\{x^k\}}{q'\{x^k\}} > b\right\} \qquad (1 \leqq k \leqq n, b \geqq 0).$$

Es gilt nach Lemma 1.3 und der Definition von q, q'

$$q_n\left\{x_n : \frac{q\{x^k\}}{q'\{x^k\}} > b\right\} = q\left\{\frac{dq}{dq'} > b\right\} = u\left\{\frac{dq}{dq'} > b\right\} = u_n\left\{x_n : \frac{q\{x^k\}}{q'\{x^k\}} > b\right\}$$

und

$$q_n'\left\{x_n : \frac{q\{x^k\}}{q'\{x^k\}} > b\right\} = q'\left\{\frac{dq}{dq'} > b\right\} \cdot q'(X)^{n-1} = v'\left\{\frac{dq}{dq'} > b\right\} v'(X)^{n-1} =$$
$$= v_n'\left\{x_n : \frac{q\{x^k\}}{q'\{x^k\}} > b\right\}.$$

Wir wissen bereits, daß die Mengen $E_n \subseteqq X_n$ mit

$$q_n'(E_n) = v_n'(E_n)$$

einen Verband bilden (siehe Beweis zu Satz 1.1). Das gleiche gilt für die E_n mit

$$q_n(E_n) = u_n(E_n)$$

(wegen (1.3)). Daraus und aus dem Vorausgehenden folgt

$$q_n\left\{\frac{dq_n}{dq_n'} > a\right\} = u_n\left\{\frac{dq_n}{dq_n'} > a\right\}$$

und

$$q_n'\left\{\frac{dq_n}{dq_n'} > a\right\} = v_n'\left\{\frac{dq_n}{dq_n'} > a\right\},$$

d.h. nach Lemma 1.3

$$H(q_n, q_n') = H(u_n, u_n')$$

(wegen (2.4) und $u_n'\{x_n\} = u'\{x^1\} \ldots u'\{x^n\}$ ist $u_n' > 0$).

Wir haben nun

$$\text{(2.7)} \qquad \begin{aligned} \Gamma(n, \varepsilon) &= \min\{\textstyle\int f_n\, dv_n' : \int f_n\, du_n \geqq 1 - \varepsilon\} \\ &= \min\{\textstyle\int f_n\, dv_n' : \int (1 - f_n)\, dv_n \leqq \varepsilon\} \end{aligned}$$

(benutze Corollar 1.2)

$$= \min\{\textstyle\int f_n\, dv_n' : 1 - f_n \in \boldsymbol{F}_{q_n, q_n'}, \int (1 - f_n)\, dv_n \leqq \varepsilon\}$$

(Satz 2.1)

$$= \min\{\textstyle\int f_n\, dq_n' : 1 - f_n \in \boldsymbol{F}_{q_n, q_n'}, \int (1 - f_n)\, dq_n \leqq \varepsilon\}$$

(Satz 2.1)

$$= \min\{\textstyle\int f_n\, dq_n' : \int (1 - f_n)\, dq_n \leqq \varepsilon\}$$

(Neyman-Pearsonsches Lemma)

$$= \min\{\textstyle\int f_n\, dq_n' : \int f_n\, dq_n \geqq 1 - \varepsilon\}$$

$$= \gamma(n, \varepsilon) \quad \text{(etwa)}.$$

Setzen wir

$$\beta(n, \varepsilon) = \min\{q_n'(E_n) : q_n(E_n) \geqq 1 - \varepsilon\},$$

so folgt aus dem Beweis von Satz 1 in [*17*] (oder auch direkt)

$$\log \gamma(n, \varepsilon) = \log \beta(n, \varepsilon) + o(1),$$

also nach (2.7)

$$\log \Gamma(n, \varepsilon) = \log \beta(n, \varepsilon) + o(1). \tag{2.8}$$

Anwendung von Satz 1 (allgemeinere Fassung) von [*17*] liefert die Behauptung über $\Gamma(n, \varepsilon)$.

Aus dem Beweis von Satz 1 in [*17*] (siehe (2.7), (2.10), (2.11), (2.17) und (2.19)) ergibt sich für nicht gitterförmig verteiltes h die Existenz von a_n mit

$$q_n\left\{\frac{dq_n}{dq_n'} > a_n\right\} \geqq 1 - \varepsilon$$

und

$$\log q_n'\left\{\frac{dq_n}{dq_n'} > a_n\right\} = \log \beta(n, \varepsilon) + o(1).$$

Nach Definition von $B(n, \varepsilon)$ und Lemma 1.3 hat man deshalb

$$\log B(n, \varepsilon) \leqq \log \beta(n, \varepsilon) + o(1),$$

also wegen $\Gamma(n, \varepsilon) \leqq B(n, \varepsilon)$ und (2.8)

$$\log B(n, \varepsilon) = \log \beta(n, \varepsilon) + o(1),$$

woraus nach Satz 1 von [*17*] die erste Hälfte der Behauptung über $B(n, \varepsilon)$ folgt. Die zweite ergibt sich analog.

Ist v_n' speziell das Anzahlmaß *in* X_n, so liefert Satz 2.2 eine Aussage über Klassen von Quellen, der wir das folgende Resultat (Satz 2.3) gegenüberstellen:

Sei T eine beliebige Menge von WVen in X, T^* die konvexe abgeschlossene Hülle von T,

$$T_n = \{q_n : q_n = q^1 \times \cdots \times q^n \quad \text{mit} \quad q^k \in T\}$$

und

$$b(n, \varepsilon) = \min\{p_n(E_n) : q_n(E_n) \geqq 1 - \varepsilon \quad \text{für alle} \quad q_n \in T_n\},$$

wobei $0 < \varepsilon < 1$ und $p_n(E_n)$ die Anzahl der Punkte von E_n ist. Sei ferner

$$H = \sup\{-\sum_{q\{x\} > 0} q\{x\} \log q\{x\} : q \in T^*\}.$$

Satz 2.3.* *Es ist*

$$\log b(n, \varepsilon) = nH + o(n).$$

Beweis. Wir können ohne Beschränkung der Allgemeinheit annehmen, daß für kein $x \in X$ alle $q\{x\}$ verschwinden ($q \in T^*$). Sei $q^* \in T^*$ mit

$$-\sum_{q^*\{x\}>0} q^*\{x\} \log q^*\{x\} = H.$$

Aus der eben gemachten Annahme folgt leicht, daß $q^* > 0$, also $h(x) = \log q^*\{x\}$ endlich ist. Es gilt

$$\int h\, dq^* \leqq \int h\, dq \qquad (q \in T^*), \tag{2.9}$$

wie weiter unten bewiesen wird. Sei $\delta > 0$ und

$$E_n = \left\{x_n : \sum_{k=1}^{n} h(x^k) > n \int h\, dq^* - n\delta\right\}.$$

Dann folgt aus (2.9) für hinreichend große n

$$\begin{aligned} q_n(E_n) &\geqq q_n\left\{\sum_1^n h(x^k) > n \int h\, d\left(\frac{1}{n}\sum_{i=1}^n q^i\right) - n\delta\right\} \\ &= q_n\left\{\frac{1}{n}\sum_1^n \left(h(x^k) - \int h\, dq^k\right) > -\delta\right\} \geqq 1 - \varepsilon \end{aligned}$$

gleichmäßig in $q_n \in T_n$. Andererseits ist aber

$$\begin{aligned} p_n(E_n) &= p_n\left\{\exp\left(\sum_1^n h(x^k)\right) > \exp\left(n\int h\, dq^* - n\delta\right)\right\} \\ &\leqq \exp\left(-n\int h\, dq^* + n\delta + n \log \int e^h\, dp_1\right) = e^{n(H+\delta)} \end{aligned}$$

nach der Tschebyscheffschen Ungleichung und der Definition von h und H. Also haben wir

$$\log b(n, \varepsilon) \leqq \log p_n(E_n) \leqq nH + n\delta$$

für hinreichend große n, und da $\delta > 0$ beliebig ist,

$$\log b(n, \varepsilon) \leqq nH + o(n).$$

Ist q_n^* das unabhängige Produkt von n Exemplaren von q^*, so liegt q_n^* in der konvexen abgeschlossenen Hülle von T_n und es gilt

$$\log b(n, \varepsilon) \geqq \log \min\{p_n(E_n) : q_n^*(E_n) \geqq 1 - \varepsilon\} = nH + o(n)$$

(bekanntlich).

Wir haben noch (2.9) zu beweisen.

Dazu sei T_1 eine beliebige konvexe kompakte Teilmenge von T^* so, daß $q^* \in T_1$ und $q > 0$ für alle $q \in T_1$ ist. Das Zweipersonennullsummenspiel mit den Strategienmengen T_1, T_1 und der Auszahlungsfunktion

$$\sum_x q'\{x\} \log q\{x\}$$

* Es ist nicht anzunehmen, daß dieser Satz neu ist.

(wobei q eine Strategie des ersten, q' eine Strategie des zweiten Spielers bedeutet) hat den Gleichgewichtspunkt (q^*, q^*), wie sich leicht aus der bekannten Ungleichung

$$\sum_x q'\{x\} \log q\{x\} < \sum_x q'\{x\} \log q'\{x\} \qquad (q \neq q')$$

ergibt. Also gilt

$$\sum_x q^*\{x\} \log q^*\{x\} \leqq \sum_x q'\{x\} \log q^*\{x\} \qquad (q' \in T_1),$$

woraus (2.9) folgt.

Beschränkt man sich in Satz 2.2 auf eine Genauigkeit von $o(n)$ und ist v_n' das Anzahl-Maß in X_n, so erhält man keinen Spezialfall von Satz 2.3, da im allgemeinen u_n nicht das Minimum von Produktmaßen q_n mit $q_n \geqq u_n$ ist.

3. Das Coding Theorem

Sei $Y = \{y, \ldots\}$ eine zweite endliche Menge, $Y_n = \{y_n = (y^1, \ldots, y^n) : y^k \in Y\}$ und $(P_n)_{n \geqq 1}$ ein stationärer Kanal ohne Gedächtnis von Y nach Ω (siehe etwa [*17*]), d. h. sei

$$P_n(\omega_n, y_n) = P(\omega^1, y^1) \ldots P(\omega^n, y^n),$$

wobei P ein Markoffscher Kern von Y nach Ω ist ($P(\omega, y)$ ist eine Abkürzung für $P(\{\omega\}, y)$). Wir setzen

$$U_n(\quad, y_n) = \varphi_n P_n(\quad, y_n) \qquad (y_n \in Y_n),$$

also

$$U_n(E_n, y_n) = P_n(\{\omega_n : \omega_n \subseteqq E_n\}, y_n) \qquad (E_n \subseteqq X_n).$$

Die Folge $(U_n)_{n \geqq 1}$ heißt ein stationärer Unterkanal ohne Gedächtnis von Y nach X (statt U_1 schreiben wir auch U). Ein Code für U_n ist eine Abbildung f_n einer Teilmenge von X_n nach Y_n. Gilt

$$U_n(f_n^{-1}\{y_n\}, y_n) \geqq 1 - \varepsilon \qquad (y_n \in f_n(X_n)),$$

so heißt f_n ein ε-Code für U_n. Die Anzahl der Punkte von $f_n(X_n)$ heißt die Länge des Codes. Wir wollen das Verhalten für große n der maximalen Länge $N(n, \varepsilon)$ von ε-Codes für U_n untersuchen.

Sei μ eine WV in Y und h eine reelle Funktion auf $X \times Y$. Wir führen den Raum

$$\tilde{\Omega} = \Omega \times Y = \{\tilde{\omega} = (\omega, y) : \omega \in \Omega, y \in Y\}$$

ein und setzen

$$(\min_\omega h)(y) = (\min h)(\tilde{\omega}) = \min\{h(x, y) : x \in \omega\}$$

$$(\max_\omega h)(y) = \max\{h(x, y) : x \in \omega\}.$$

Seien die WVen $\tilde{\mu}$ in $\tilde{\Omega}$, μ' in Ω und μ_ω in Y durch

$$\begin{aligned} \tilde{\mu}\{\tilde{\omega}\} &= P(\omega, y)\,\mu\{y\}, \\ \mu'\{\omega\} &= \tilde{\mu}(\{\omega\} \times Y) = \int P(\omega, y)\,\mu(dy) \\ \mu_\omega\{y\} &= \tilde{\mu}\{\tilde{\omega}\}/\mu'\{\omega\} \qquad (\omega \in \Omega) \end{aligned}$$

erklärt und sei

$$H = \int \min_{\omega} h \, d\tilde{\mu} = \int \mu'(d\omega) \int \min_{\omega} h \, d\mu_\omega, \tag{3.1}$$

$$G = \int \mu'(d\omega) \log \int \exp(\max_{\omega} h) \, d\mu. \tag{3.2}$$

Ist N eine nichtnegative ganze Zahl, so bezeichnen wir mit μ_N das unabhängige Produkt von N Exemplaren von μ, μ_N ist also eine WV in

$$Y_N = \{y_N = (y^1, \ldots, y^N) : y^i \in Y\}.$$

Jedem $y_N \in Y_N$ ordnen wir den Code f_{y_N} für U mit

$$f_{y_N}(x) = y^i \quad \text{falls} \quad h(x, y^i) > H - \delta \geqq h(x, y^j) \qquad (i \neq j)$$

zu, wobei $\delta > 0$ ist und f_{y_N} nicht definiert wird, falls es kein i mit $1 \leqq i \leqq N$ gibt, das die rechtsstehende Bedingung erfüllt. Gelingt es uns, für ein ε mit $0 < \varepsilon < 1$

$$\int U(f_{y_N}^{-1}\{y^i\}, y^i) \, \mu_N(dy_N) > 1 - \frac{\varepsilon}{2} \qquad (1 \leqq i \leqq N) \tag{3.3}$$

zu beweisen, so ist auch die Existenz eines ε-Codes für U einer Länge $\geqq N/2$ gesichert. Denn aus (3.3) folgt zunächst

$$\int \frac{1}{N} \sum_{i=1}^{N} U(f_{y_N}^{-1}\{y^i\}, y^i) \, \mu_N(dy_N) > 1 - \frac{\varepsilon}{2},$$

also für mindestens ein y_N

$$\frac{1}{N} \sum_{i} U(f_{y_N}^{-1}\{y^i\}, y^i) > 1 - \frac{\varepsilon}{2}.$$

Das ist aber nur möglich, wenn für mindestens die Hälfte der i

$$U(f_{y_N}^{-1}\{y^i\}, y^i) > 1 - \varepsilon$$

ist.

Sei nun $i \in \{1, \ldots, N\}$ fest gewählt. Nach Definition von f_{y_N} gilt

$$\begin{aligned} \{\omega : \omega \subseteqq f_{y_N}^{-1}\{y^i\}\} &= \{\omega : \omega \subseteqq (\bigcap_{j \neq i} \{x : h(x, y^j) \leqq H - \delta\} - \{x : h(x, y^i) \leqq H - \delta\})\} \\ &= \bigcap_{j \neq i} \{\omega : (\max_{\omega} h)(y^j) \leqq H - \delta\} - \{\omega : (\min_{\omega} h)(y^i) \leqq H - \delta\}, \end{aligned}$$

nach Definition von U also

$$\begin{aligned} &\int U(f_{y_N}^{-1}\{y^i\}, y^i) \, \mu_N(dy_N) = \int P(\{\omega : \omega \subseteqq f_{y_N}^{-1}\{y^i\}\}, y^i) \, \mu_N(dy_N) \geqq \\ &\geqq \int P(\bigcap_{j \neq i} \{\omega : (\max_{\omega} h)(y^i) \leqq H - \delta\}, y^i) \, \mu_N(dy_N) \\ &- \int P(\{\omega : (\min_{\omega} h)(y^i) \leqq H - \delta\}, y^i) \, \mu_N(dy_N). \end{aligned} \tag{3.4}$$

Hier ist das zweite Integral

$$\begin{aligned} &= \int P(\{\omega : (\min_{\omega} h)(y) \leqq H - \delta\}, y) \, \mu(dy) \\ &= \tilde{\mu}\{\min h \leqq H - \delta\}, \end{aligned}$$

das erste ist aus Symmetriegründen

$$= \int P\Big(\bigcap_{j=1}^{N-1}\{\omega : (\max_\omega h)(y^j) \leqq H - \delta\}, y^N\Big)\, \mu_N(dy_N)$$

$$= \int \mu'\Big(\bigcap_{1}^{N-1}\{\omega : (\max_\omega h)(y^j) \leqq H - \delta\}\Big)\, \mu_{N-1}(dy_{N-1})$$

$$= \int \mu_{N-1}\Big(\bigcap_{1}^{N-1}\{y_{N-1} : (\max_\omega h)(y^j) \leqq H - \delta\}\Big)\, \mu'(d\omega)$$

$$= \int (\mu\{\max_\omega h \leqq H - \delta\})^{N-1}\, \mu'(d\omega)\,.$$

Aus (3.4) folgt also

$$\begin{aligned} \text{(3.5)} \qquad & \int U(f_{y_N}^{-1}\{y^i\}, y^i)\, \mu_N(dy_N) \geqq \\ & \geqq \int (\mu\{\max_\omega h) \leqq H - \delta\})^N \mu'(d\omega) - \tilde{\mu}\{\min h \leqq H - \delta\}\,. \end{aligned}$$

Nach (3.3) gibt es einen ε-Code für U einer Länge $\geqq N/2$, sobald die rechte Seite von (3.5) $> 1 - \varepsilon/2$ ist. Wir sind allerdings nicht so sehr an einem Code für U interessiert, als vielmehr an einem für U_n (n groß). Deshalb denken wir uns die vorangehende Betrachtung an U_n statt U durchgeführt, indem wir an Stelle von μ und h jetzt μ_n (unabhängiges Produkt aus n Exemplaren von μ) und h_n mit $h_n(x_n, y_n) = \frac{1}{n}\sum_{k=1}^{n} h(x^k, y^k)$ verwenden (aus $\tilde{\mu}$, μ' werden ebenfalls Produktmaße $\tilde{\mu}_n$, μ_n'). Wir erhalten dann

$$\int (\mu_n\{\max_{\omega_n} h_n \leqq H - \delta\})^N \mu_n'(d\omega_n) - \tilde{\mu}_n\{\min h_n \leqq H - \delta\} > 1 - \frac{\varepsilon}{2}$$

als hinreichende Bedingung für

$$\text{(3.6)} \qquad N(n, \varepsilon) \geqq \frac{N}{2}\,.$$

Nun ist

$$\begin{aligned} \text{(3.7)} \qquad & \int (\mu_n\{\max_{\omega_n} h_n \leqq H - \delta\})^N \mu_n'(d\omega_n) \\ & = \int \left(\mu_n\left\{y_n : \sum_{k=1}^{n} (\max_{\omega^k} h)(y^k) \leqq n(H - \delta)\right\}\right)^N \mu_n'(d\omega_n) \\ & \geqq \int_{\Theta_n} (1 - N\,\mu_n\{y_n : \sum_k (\max_{\omega^k} h)(y^k) > n(H - \delta)\})\, \mu_n'(d\omega_n) \end{aligned}$$

nach der Bernoullischen Ungleichung, wobei

$$\Theta_n = \left\{\omega_n : \left|\frac{1}{n}\sum_{k=1}^{n} \log \int \exp\,(\max_{\omega^k} h)\, d\mu - G\right| < \delta\right\}$$

sei. Für $\omega_n \in \Theta_n$ haben wir

$$(3.8)\qquad \mu_n\{y_n : \sum_k (\max_{\omega^k} h)(y^k) > n(H-\delta)\}$$

$$= \mu_n\left\{y_n : \exp\left(\sum_{k=1}^{n} (\max_{\omega^k} h)(y^k) > \exp(nH - n\delta)\right)\right\}$$

$$\leqq \exp\left(-n(H-\delta) + \sum_{k=1}^{n} \log \int \exp(\max_{\omega^k} h)\, d\mu\right)$$

(nach der Tschebyscheffschen Ungleichung)

$$\leqq e^{-n(H-G-2\delta)}.$$

Wir wählen jetzt n so groß, daß aus dem schwachen Gesetz der großen Zahl

$$(3.9)\qquad \tilde{\mu}_n\{\min h_n \leqq H-\delta\}$$

$$= \tilde{\mu}_n\left\{\tilde{\omega}_n : \frac{1}{n}\sum_{k=1}^{n} (\min h)(\tilde{\omega}^k) \leqq H-\delta\right\} < \frac{\varepsilon}{6}$$

und

$$(3.10)\qquad \mu_n'(\Theta_n) > 1 - \frac{\varepsilon}{6}$$

folgt (siehe (3.1) und (3.2)), und außerdem

$$(3.11)\qquad e^{-n\delta} < \frac{\varepsilon}{6}$$

gilt. Ist

$$N \leqq e^{n(H-G-3\delta)},$$

so haben wir wegen (3.7) bis (3.11)

$$\int_{\omega_n} (\mu_n\{\max h_n \leqq H-\delta\})^N \mu_n'(d\omega_n) - \tilde{\mu}_n\{\min h_n \leqq H-\delta\}$$

$$> (1 - N e^{-n(H-G-2\delta)})\, \mu_n'(\Theta_n) - \frac{\varepsilon}{6}$$

$$> \left(1-\frac{\varepsilon}{6}\right)^2 - \frac{\varepsilon}{6} > 1 - \frac{\varepsilon}{2}.$$

Aus (3.6) folgt nun

$$(3.12)\qquad N(n,\varepsilon) > e^{n(H-G-4\delta)} \qquad (n \text{ hinreichend groß}).$$

Wir setzen

$$(3.13)\qquad C(U) = \sup_{\mu,h}(H-G) = \sup_{\mu,h} \int \mu'(d\omega) \log \frac{\exp\,(\int_\omega \min h\, d\mu_\omega)}{\int_\omega \exp(\max h)\, d\mu},$$

wobei das supremum über alle WVen μ in Y und alle Funktionen h auf $X \times Y$ zu erstrecken ist. Aus (3.12) folgt, daß $C(U)$ endlich ist, $C(U) \geqq 0$ ist trivial ($h = 0$). Ebenso wie $C(U)$ können wir $C(U_n)$ bilden. Durch Einsetzen von μ und h in Produkt- bzw. Summengestalt erkennt man

$$(3.14)\qquad C(U_{n+m}) \geqq C(U_n) + C(U_m),$$

so daß

$$C = \lim_{n\to\infty} \frac{1}{n} C(U_n) = \sup_n \frac{1}{n} C(U_n) \tag{3.15}$$

existiert (vgl. Chintschin [*18*]). Aus (3.12), (3.13) und (3.15) folgt leicht

$$\log N(n, \varepsilon) > n\,C + o(n).$$

Ist andererseits f ein ε-Code für U der Länge N, so setze man

$$\mu = \frac{1}{N} \sum_{y \in f(X)} \delta_y,$$

und

$$h(x, y) = \begin{cases} \log N & \text{falls} \quad y = f(x) \\ 0 & \text{sonst}. \end{cases}$$

Eine einfache Rechnung liefert

$$H - G \geqq (1 - 2\varepsilon) \log N - (1 - \varepsilon) \log (2 - 1/N)$$

(Fanos Lemma). Das entsprechende Ergebnis erhalten wir bei einem ε-Code für U_n der Länge N. Es folgt also

$$\log N(n, \varepsilon) < n C + \varepsilon\, 0(n).$$

Damit haben wir den

Satz 3.1. *(Coding Theorem und schwache Umkehrung.) Ist* $(U_n)_{n \geqq 1}$ *ein stationärer Unterkanal ohne Gedächtnis,* $0 < \varepsilon < 1$ *und* C *durch* (3.13), (3.15) *definiert, so gilt*

$$n C + o(n) < \log N(n, \varepsilon) < n C + \varepsilon\, 0(n).$$

C ist im allgemeinen verschieden von $C(U)$, wie das folgende aus Shannon [*19*] stammende Beispiel zeigt: $X = Y = \{1, 2, 3, 4, 5\}$, $P(\ \ , y) = \delta_{\{y, y+1\}}$ (Addition mod 5). Man hat

$$C(U) = \sup_{\mu, h} \sum_{y=1}^{5} \mu\{y\} \log \frac{\exp(\min\{h(y, y), h(y+1, y)\})}{\sum\limits_{z=1}^{5} \exp(\max\{h(y, z), h(y+1, z)\})\, \mu\{z\}}$$

$$\leqq \sup_{\mu, h} \sum_{y=1}^{5} \mu\{y\} \log \frac{e^{h(y,y)}}{e^{h(y-1, y-1)}\, \mu\{y-1\} + e^{h(y,y)}\, \mu\{y\} + e^{h(y+1, y+1)}\, \mu\{y+1\}}$$

(da man sich, wie man leicht sieht, auf Funktionen h mit $h(y, y) = h(y, y+1)$ $(y \in Y)$ beschränken kann)

$$\leqq \sup_{\mu, h} \log \sum_{y=1}^{5} \frac{e^{h(y,y)}\, \mu\{y\}}{e^{h(y-1, y-1)}\, \mu\{y-1\} + e^{h(y,y)}\, \mu\{y\} + e^{h(y+1, y+1)}\, \mu\{y+1\}}$$

$$\leqq \sup_g \log \sum_{y=1}^{5} \frac{g(y)}{g(y-1) + g(y) + g(y+1)},$$

wobei das supremum über alle nichtnegativen Funktionen g auf Y zu erstrecken ist. Durch eine einfache Rechnung folgt (vgl. Mordell [*30*])

$$C(U) \leqq \log 2.$$

Es ist aber $C \geqq \frac{1}{2} \log 5$ (siehe Shannon [*19*], S. 9).

Der Ausdruck (3.13), (3.15) für die Kapazität C eines stationären Unterkanals ohne Gedächtnis ist ziemlich unhandlich. Wir haben eben gesehen, daß sich der Grenzübergang in (3.15) nicht ohne weiteres vermeiden läßt. Dagegen kann man die Konkurrenzmenge des supremums über h in der Definition von $C(U_n)$ (siehe (3.13)) wesentlich verkleinern, ohne daß man für C einen kleineren Wert erhält: Sei etwa $n = 1$, also $U_n = U$. Bei gegebenem μ definiere man die totalmonotonen Kapazitäten u und u' in $X \times Y$ durch

$$\begin{aligned} u(\widetilde{E}) &= \widetilde{\mu}\{\widetilde{\omega} : \widetilde{\omega} \subseteqq \widetilde{E}\} \\ u'(\widetilde{E}) &= (\mu' \times \mu)\{\widetilde{\omega} : \widetilde{\omega} \subseteqq \widetilde{E}\} \end{aligned} \qquad (\widetilde{E} \subseteqq X \times Y)$$

und bilde dq/dq' wie in Lemma 1.3. Setzt man

$$C'(U) = \sup_{\mu} \sup_{h \in \{t \log(dq/dq') : t \geqq 0\}} (H - G) \tag{3.16}$$

und

$$C' = \lim_{n \to \infty} \frac{1}{n} C'(U_n),$$

so ist

$$C' = C.$$

Wegen $C' \leqq C$ genügt es nämlich, ein Fanosches Lemma für $C'(U_n)$ zu beweisen. Der Leser kann die Rechnung mit Hilfe von Lemma 1.3 leicht selbst durchführen, indem er

$$t = \frac{\log N}{2 \log \frac{1-\varepsilon}{\varepsilon}}$$

setzt (siehe (3.16)).

Ob die starke Umkehrung für beliebige stationäre Unterkanäle ohne Gedächtnis gilt, ist nicht bekannt. Im übrigen ist eine ähnliche Bemerkung am Platze wie im Anschluß an Satz 2.3 (vgl. aber BLACKWELL-BREIMAN-THOMASIAN [*3*] und KIEFER-WOLFOWITZ [*6*]).

4. Totalmonotone Kapazitäten in metrisch kompakten Räumen

Sei X ein metrisch kompakter Raum (mit der Metrik $|\,,|$), Ω die Menge der nichtleeren kompakten Teilmengen von X. Mit der Hausdorff-Metrik

$$|\omega, \tau| = \max\{\sup_{x \in \omega} \inf_{y \in \tau} |x, y|, \sup_{y \in \tau} \inf_{x \in \omega} |x, y|\} \qquad (\omega, \tau \in \Omega)$$

ist bekanntlich auch Ω metrisch kompakt. Sei $C(X)$ (bzw. $C(\Omega)$) die Banachalgebra der stetigen Funktionen auf X (bzw. Ω) und D (bzw. B) der σ-Körper der Borelschen Mengen in X (bzw. Ω). Für beliebige beschränkte reellwertige Funktionen f auf X definieren wir Λf durch

$$(\Lambda f)(\omega) = \min_{x \in \omega} f(x) \qquad (\omega \in \Omega).$$

Lemma 4.1. *Die lineare Hülle L von $\Lambda C(X)$ liegt normdicht in $C(\Omega)$.*

Beweis. Offenbar ist $\Lambda C(X) \subseteqq C(\Omega)$. Da L Punkte trennt und die 1 enthält, können wir den Satz von Stone-Weierstraß anwenden, sobald wir wissen, daß etwa die abgeschlossene Hülle von L mit h, h' auch $h \cdot h'$ enthält. Dazu müssen wir zeigen:

Ist $\varepsilon > 0$ und $f_k \in C(X)$ $(k = 1, 2)$, so gibt es a_m reell, $g_m \in C(X)$ mit

$$||\Lambda f_1 \Lambda f_2 - \sum_m a_m \Lambda g_m|| < \varepsilon. \tag{4.1}$$

Wir können $f_k \geqq 0$ annehmen (Translation). Aus

$$\sum_{i \geqq 1} \delta \chi_{\{f_k \geqq \delta i\}} \leqq f_k \leqq \sum_{i \geqq 1} \delta \chi_{\{f_k > \delta(i-1)\}} \qquad (\delta > 0)$$

(χ_E ist der Indikator der Menge E) folgt

$$\sum_{i \leqq 1} \delta \Lambda \chi_{\{f_k \geqq \delta i\}} \leqq \Lambda f_k \leqq \sum_{i \geqq 1} \delta \Lambda \chi_{\{f_k > \delta(i-1)\}},$$

also

$$\sum_{i,j \geqq 1} \delta^2 \Lambda \chi_{Ai,j} \leqq \Lambda f_1 \Lambda f_2 \leqq \sum_{i,j \geqq 1} \delta^2 \Lambda \chi_{Ui,j} \tag{4.2}$$

mit $A_{i,j} = \{f_1 \geqq \delta i, f_2 \geqq \delta j\}$ und $U_{i,j} = \{f_1 > \delta(i-1), f_2 > \delta(j-1)\}$.

Da A_{ij} abgeschlossen, U_{ij} offen und $A_{ij} \subseteqq U_{ij}$ ist, gibt es $g_{ij} \in C(X)$ mit

$$\chi_{Aij} \leqq g_{ij} \leqq \chi_{Uij},$$

also

$$\sum_{ij} \delta^2 \Lambda \chi_{Aij} \leqq \sum_{i,j} \delta^2 \Lambda g_{ij} \leqq \sum_{i,j} \delta^2 \Lambda \chi_{Uij}.$$

Hieraus und aus (4.2) folgt

$$\begin{aligned}||\Lambda f_1 \Lambda f_2 - \sum_{i,j} \delta^2 \Lambda g_{ij}|| &\leqq ||\sum_{i,j} \delta^2 \Lambda \chi_{Uij} - \sum_{i,j} \delta^2 \Lambda \chi_{Aij}|| \\ &= ||\Lambda \sum_i \delta \chi_{\{f_1 > \delta(i-1)\}} \Lambda \sum_j \delta \chi_{\{f_2 > \delta(j-1)\}} - \Lambda \sum_i \delta \chi_{\{f_1 \geqq \delta i\}} \Lambda \sum_j \delta \chi_{\{f_2 \geqq \delta j\}}|| \\ &\leqq \delta(||f_2|| + \delta) + ||f_1||\, \delta,\end{aligned}$$

und daraus (4.1).

Sei $Q = \{q, p, \ldots\}$ die Menge der endlichen Maße auf D, $M = \{\mu, \nu, \ldots\}$ die der endlichen Maße auf B.

Lemma 4.2. (Choquet [7]). *Die durch*

$$(\varphi \mu)(E) = \mu\{\omega : \omega \subseteqq E\} \qquad (\mu \in M,\ E \subseteqq X \textit{ abgeschlossen})$$

definierte Abbildung φ ist eineindeutig.

Beweis. Sei $u = \varphi \mu$. Dann gilt

$$\int_0^\infty u\{f \geqq t\}\, dt = \int_0^\infty \mu\{\omega : (\Lambda f)(\omega) \geqq t\}\, dt = \int \Lambda f\, d\mu.$$

Die linke Seite ist durch u festgelegt, die rechte Seite bestimmt μ nach Lemma 4.1 und dem Satz von Riesz.

Die $u \in \varphi M$ heißen endliche totalmonotone Kapazitäten mit $u(0) = 0$, kurz totalmonotone Kapazitäten. Wenn im Zusammenhang von u und μ die Rede ist, wird $u = \varphi\mu$ angenommen. Wir schreiben $q \geqq u$ für

$$q \in Q, \quad q(X) = u(X), \quad q(E) \geqq u(E), \qquad (E \subseteqq X \text{ abgeschlossen}).$$

Ist P ein Markoffscher Kern von Ω nach X (bezüglich D und B; siehe JACOBS [20]) und ist $\mu \in M$, so sei $P\mu \in Q$ durch

$$(P\mu)(E) = \int P(E, \omega)\, \mu(d\omega) \qquad (E \in D)$$

definiert. Für jedes $\mu \in M$ ist die Menge der Markoffschen Kerne von Ω nach X in der von den Funktionalen

$$P \to \int f\, d(P\nu) \qquad (f \in C(X), \nu \in M, \nu \ll \mu) \tag{4.3}$$

erzeugten Topologie ein kompakter (i. a. nicht Hausdorffscher) Raum (siehe LECAM [14], DUNFORD-SCHWARTZ [21]). Sei

$$\boldsymbol{P} = \{P : P \text{ Markoffscher Kern von } \Omega \text{ nach } X,\ P(\omega, \omega) = 1 \qquad (\omega \in \Omega)\}.$$

Satz 4.3. *Die Menge $\{q : q \geqq u\}$ ist konvex und w^*-kompakt, d. h. in der von den Funktionalen $f \to \int f\, dq$ $(f \in C(X))$ erzeugten Topologie kompakt. Es gilt*

$$\{q : q \geqq u\} = \{P\mu : P \in \boldsymbol{P}\} \tag{4.4}$$

und

$$u(E) = \min\{q(E) : q \geqq u\} \qquad (E \text{ abgeschlossen}). \tag{4.5}$$

Beweis. Die Konvexität von $\{q : q \geqq u\}$ ist trivial, die w^*-Kompaktheit folgt aus der bedingten w^*-Kompaktheit normbeschränkter Mengen in Q (siehe etwa JACOBS [20], LOOMIS [22]) zusammen mit der w^*-Oberhalbstetigkeit der Funktionale $E \to q(E)$ (E abgeschlossen). Klar ist ferner $\supseteqq$ in (4.4). Sei umgekehrt $q \geqq u$ und seien $A^1, A^2, \ldots$ die Komplemente einer abzählbaren Basis der Topologie von X, D^i die von $\{A^1, \ldots, A^i\}$ erzeugten Mengenkörper. Es gilt

$$q(E) \geqq \mu\{\omega : \omega \subseteqq E\} \qquad (E \in D^i), \tag{4.6}$$

denn zu jedem $E \in D^i$ gibt es eine aufsteigende Folge $(E_n)_{n \geqq 1}$ von abgeschlossenen Mengen mit

$$E = \bigcup_n E_n,$$

also

$$\{\omega : \omega \subseteqq E\} = \bigcup_n \{\omega : \omega \subseteqq E_n\},$$

also (4.6)

Wir wählen in jedem Atom von D^i einen Punkt, bezeichnen mit X^i die Menge dieser Punkte und mit β^i die Abbildung, die jedem $x \in X$ das $x^i \in X^i$ zuordnet, das im gleichen Atom von D^i liegt. Sind

$$q^i(E^i) = q\{x : \beta^i x \in E^i\} \qquad (E^i \subseteqq X^i)$$

und

$$\mu^i(\Theta^i) = \mu\{\omega : \beta^i \omega \in \Theta^i\} \qquad (\Theta^i \subseteqq \Omega(X^i)),$$

so folgt aus (4.6)

$$q^i(E^i) \geqq \mu^i\{\omega^i : \omega^i \in \Omega(X^i), \omega^i \subseteqq E^i\}.$$

Nach Satz 1.1 gibt es einen Markoffschen Kern P^i von $\Omega(X^i)$ nach X^i mit

$$P^i(\omega^i, \omega^i) = 1 \qquad (\omega^i \in \Omega(X^i))$$

und

$$q^i = P^i \mu^i .$$

Sei

$$P_i(E, \omega) = P^i(E \cap X^i, \beta^i \omega) \qquad (E \subseteqq X,\ \omega \in \Omega).$$

P_i ist ein Markoffscher Kern von Ω nach X, der (wegen $A^k \cap X^i = \beta^i A^k$ $(k \leqq i)$)

$$(4.7) \qquad P_i(A^k, \omega) = P^i(\beta^i A^k, \beta^i \omega) \geqq P^i(\beta^i \omega, \beta^i \omega) = 1 \qquad (k \leqq i,\ \omega \subseteqq A^k)$$

und

$$(4.8) \qquad \begin{aligned} q(A^k) &= q^i(\beta^i A^k) = \int P^i(\beta^i A^k, \omega^i)\, \mu^i(d\omega^i) \\ &= \int P^i(\beta^i A^k, \beta^i \omega)\, \mu(d\omega) = \int P_i(A^k, \omega)\, \mu(d\omega) \\ &= (P_i \mu)(A^k) \qquad (k \leqq i) \end{aligned}$$

erfüllt. Die Funktionale

$$P \to (P\nu)(A^k) \qquad (\nu \in M,\ \nu \ll \mu,\ k \geqq 1)$$

sind oberhalbstetig in der durch (4.3) definierten Topologie. Ist also P' ein Häufungspunkt der Folge P_i, so gilt wegen (4.8)

$$q(A^k) \leqq (P'\mu)(A^k) \qquad (k \geqq 1),$$

d. h. wegen $q(X) = (P'\mu)(X)$

$$(4.9) \qquad q = P'\mu .$$

Aus (4.7) folgt

$$(P_i \nu)(A^k) \geqq \nu\{\omega : \omega \subseteqq A^k\} \qquad (k \leqq i,\ \nu \in M),$$

daraus

$$(P'\nu)(A^k) \geqq \nu\{\omega : \omega \subseteqq A^k\} \qquad (k \geqq 1,\ \nu \in M,\ \nu \ll \mu)$$

und hieraus

$$P'(A^k, \omega) = 1 \qquad (\mu\text{-fastalle } \omega \subseteqq A^k).$$

Es existiert also eine μ-Nullmenge $\Theta \in B$ mit

$$P'(A^k, \omega) = 1 \qquad (k \geqq 1,\ \omega \subseteqq A^k,\ \omega \in \Omega - \Theta).$$

Da jedes $\omega \in \Omega$ Durchschnitt gewisser A^k ist, ergibt sich

$$(4.10) \qquad P'(\omega, \omega) = 1 \qquad (\omega \in \Omega - \Theta).$$

Um die Nullmenge Θ zu eliminieren, zeigen wir, daß es meßbare Abbildungen g von Ω nach X mit

$$g(\omega) \in \omega \qquad (\omega \in \Omega)$$

gibt. Die folgende Konstruktion stammt von K. Jacobs: Sei

$$f_1(\omega) = \omega,$$

$$f_{k+1}(\omega) = \begin{cases} f_k(\omega) \cap A^{k+1} & \text{falls } f_k(\omega) \cap A^{k+1} \neq 0 \\ f_k(\omega) & \text{sonst}, \end{cases}$$

$$(4.11) \qquad f(\omega) = \bigcap_{k \geqq 1} f_k(\omega) \neq 0 .$$

Die f_k sind offenbar B-meßbare Selbstabbildungen von Ω. Für $\omega \in \Omega$ und $k \geqq 1$ gilt entweder $f(\omega) \subseteqq A^k$ oder $f(\omega) \cap A^k = 0$, $f(\omega)$ ist also einpunktig (die A^k trennen die Punkte von X). Sei g die Abbildung von Ω nach X mit

$$f(\omega) = \{g(\omega)\}.$$

Wegen

$$g^{-1}(A^k) = \{\omega : f_k(\omega) \subseteqq A^k\} \qquad (k \geqq 1)$$

ist g meßbar,

$$g(\omega) \in \omega$$

ist trivial.

Setzt man nun

$$P(E,\omega) = \begin{cases} P'(E,\omega) & \text{falls} \quad \omega \in \Omega - \Theta \\ 1 & \text{falls} \quad \omega \in \Theta, \quad g(\omega) \in E \\ 0 & \text{sonst,} \end{cases}$$

so ist $P \in \boldsymbol{P}$ und $q = P\mu$ (wegen (4.10) und (4.9)). Damit ist (4.4) bewiesen.

Um (4.5) zu zeigen, nehmen wir an, E sei ein abgeschlossener Teil von X. Ist $(E_k)_{k \geqq 1}$ eine aufsteigende Folge von abgeschlossenen Mengen mit

$$X - E = \bigcup_k E_k,$$

so definieren wir die Selbstabbildung h von Ω ebenso wie f (siehe (4.11)), nur unter Verwendung der E_k an Stelle der A^k. h ist dann meßbar und es gilt

$$0 \neq h(\omega) \subseteqq \omega$$

und

(4.12) $$h(\omega) \cap E = 0 \qquad (\omega - E \neq 0).$$

Setzt man

$$P(\ ,\omega) = \delta_{g(h(\omega))} \qquad (\omega \in \Omega),$$

so ist offenbar $P \in \boldsymbol{P}$ und (wegen (4.12))

$$(P\mu)(E) = u(E).$$

Daraus folgt (4.5) und der Satz ist bewiesen.

Sei T ein Homeomorphismus von X. T kann auch als Homeomorphismus von Ω aufgefaßt werden. Eine totalmonotone Kapazität u heißt invariant (bezüglich T), wenn μ (bezüglich T) invariant ist, u heißt ergodisch, wenn μ eine ergodische WV ist (siehe JACOBS [*20*]). Wir bezeichnen die Menge aller invarianten Maße aus $\{q : q \geqq u\}$ mit $J(u)$.

Lemma 4.4. *Ist u invariant, so ist $J(u)$ konvex, w^*-kompakt und nichtleer. Ein $q \in Q$ liegt genau dann in $J(u)$, wenn es einen Markoffschen Kern P von Ω nach X mit*

(4.13) $$q = P\mu,$$

(4.14) $$P(\omega,\omega) = 1 \qquad (\mu\text{-fastalle } \omega)$$

(4.15) $$P(TE, T\omega) = P(E,\omega) \qquad (E \in D,\ \mu\text{-fastalle } \omega)$$

gibt.

Beweis. Ist $q \in J(u)$, so gibt es nach Satz 4.3 ein $P_0 \in \boldsymbol{P}$ mit $q = P_0 \mu$. Durch

$$(\tilde{T} P)(E, \omega) = P(TE, T\omega) \qquad (E \in D,\ \omega \in \Omega) \tag{4.16}$$

ist eine stetige Abbildung $\tilde{T}$ im kompakten Raum der Markoffschen Kerne von Ω nach X definiert (siehe (4.3)). Die Folge

$$P_n = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{T}^k P_0$$

hat einen Häufungspunkt P. Ein bekannter Schluß (siehe JACOBS [20], S. 9) liefert

$$\int f\, d(P\nu) = \int f\, d((\tilde{T} P)\nu) \qquad (f \in C(X),\ \nu \in M,\ \nu \ll \mu),$$

d.h. (4.15). (4.14) ist äquivalent zu

$$(P\nu)(A^k) \geqq \nu\{\omega : \omega \subseteqq A^k\} \qquad (k \geqq 1,\ \nu \in M,\ \nu \ll \mu) \tag{4.17}$$

(siehe Beweis von Satz 4.3), folgt also aus der Gültigkeit von (4.17) für die $P_n \in \boldsymbol{P}$. (4.13) und der Rest des Lemmas sind klar.

Lemma 4.5. *Ist u ergodisch, so sind alle Extremalpunkte von $J(u)$ ergodisch.*

Beweis. Sei q ein Extremalpunkt von $J(u)$, P ein Markoffscher Kern mit (4.13) bis (4.15). Für beliebiges invariantes $E \in D$ ist $P(E, \quad)$ eine μ-fastinvariante B-meßbare Funktion, also μ-fastkonstant, etwa $= a$ mit $0 \leqq a \leqq 1$. Ist $0 < a < 1$, so definiere man P_1, P_2 durch

$$P_1(F, \omega) = \begin{cases} \dfrac{1}{a} P(F \cap E, \omega) & \text{wenn } P(E, \omega) = a \\ P(F, \omega) & \text{sonst} \end{cases}$$

und

$$P_2(F, \omega) = \begin{cases} \dfrac{1}{1-a} P(F - E, \omega) & \text{wenn } P(E, \omega) = a \\ P(F, \omega) & \text{sonst}. \end{cases}$$

Man hat $P_1\mu$, $P_2\mu \in J(u)$, $P_1\mu \neq P_2\mu$ und

$$q = a P_1 \mu + (1 - a) P_2 \mu,$$

im Widerspruch zur Extremalität von q. Also ist $a \in \{0,1\}$ und damit

$$q(E) = (P\mu)(E) \in \{0,1\},$$

woraus die Ergodizität von q folgt.

Die Umkehrung des Lemmas ist falsch. Das liegt daran, daß verschiedene u dasselbe $J(u)$ besitzen können.

5. Ergodische Unterquellen

Sei X^0 eine endliche Menge,

$$X = \{x = (\ldots x^{-1}, x^0, x^1, \ldots) : x^i \in X^0 \quad (i \text{ ganz})\} \tag{5.1}$$

und

$$\Theta = \{\vartheta = (\ldots \omega^{-1}, \omega^0, \omega^1, \ldots) : \omega^i \in \Omega(X^0) \qquad (i \text{ ganz})\}.$$

Dann sind X und Θ metrisch kompakt und Θ ist in natürlicher Weise homeomorph ($\vartheta \to \cdots \times \omega^{-1} \times \omega^0 \times \omega^1 \times \cdots$) einem abgeschlossenen Teil von $\Omega = \Omega(X)$. Wir identifizieren Θ mit diesem Teil.

Eine totalmonotone Kapazität in X heißt eine Unterquelle, wenn

$$\mu(\Theta) = 1, \quad \mu(\Omega - \Theta) = 0$$

gilt.

Sei u eine Unterquelle,

$$X_n = \{x_n = (x^1, \ldots, x^n) : x^i \in X^0 \quad (1 \leqq i \leqq n)\}$$

und

$$u_n(E_n) = \mu\{x : (x^1, \ldots, x^n) \in E_n\} \qquad (E_n \subseteqq X_n).$$

Man sieht leicht, daß u_n eine totalmonotone Kapazität ist. u_n heißt die Projektion von u auf X_n. Ist π irgendeine WV in X_n, so bezeichnen wir mit $H(\pi)$ die Entropie von π, also

$$H(\pi) = -\sum_{\pi\{x_n\} > 0} \pi\{x_n\} \log \pi\{x_n\}.$$

Wir setzen

$$H(u_n) = \sup\{H(\pi) : \pi \geqq u_n\} \tag{5.2}$$

(es gilt $H(u_n) = H(u_n, q_n')$ (siehe (1.6)), wobei q_n' das Anzahlmaß in X_n ist) und nennen

$$H = H(u) = \limsup_{n \to \infty} \frac{1}{n} H(u_n) \tag{5.3}$$

die Entropie von u.

Satz 5.1. *Angenommen, es gibt ein $p \geqq u$ mit*

$$\lim_n \int \left| \frac{1}{n} \log p_n\{x_n\} + H \right| p(dx) = 0 \tag{5.4}$$

(p_n ist die Projektion von p auf X_n), dann gilt für jedes $\varepsilon > 0$

$$\min\{q_n'(E_n) : E_n \subseteqq X_n, u_n(E_n) \geqq 1 - \varepsilon\} = e^{nH + o(n)}. \tag{5.5}$$

Beweis. Aus (5.4) folgt bekanntlich

$$\min\{q_n'(E_n) : p_n(E_n) \geqq 1 - \varepsilon\} = e^{nH + o(n)},$$

also $\geqq$ in (5.5) (wegen $p_n \geqq u_n$). Um $\leqq$ in (5.5) zu zeigen, genügt es, für beliebige $\varepsilon, \delta > 0$ etwa

$$q_n\{x_n : q_n\{x_n\} > e^{-n(H+3\delta)}\} > 1 - 5\varepsilon \qquad (n \text{ hinreichend groß}) \tag{5.6}$$

nachzuweisen, wobei $q_n \geqq u_n$ und

$$H(q_n) = H(u_n) \tag{5.7}$$

ist (q_n ist nach Lemma 1.3 eindeutig bestimmt). Denn die Menge

$$E_n = \{x_n : q_n\{x_n\} > e^{-n(H+3\delta)}\}$$

hat sicher nicht mehr als $e^{n(H+3\delta)}$ Punkte, andererseits folgt aus Lemma 1.3 und (5.6) sofort

$$u_n(E_n) > 1 - 5\varepsilon.$$

Leider werden die q_n im allgemeinen nicht durch eine $W\,V\,q$ in X induziert, d.h. es gilt nicht notwendig

$$q_n\{y_n\} = q_{n+1}\{x_{n+1} : x_n = y_n\} \qquad (y_n \in X_n).$$

Nun ergibt sich aber aus (5.4) unmittelbar

$$(5.8) \qquad p_n\{x_n : p_n\{x_n\} > e^{-n(H+\delta)}\} > 1 - \varepsilon \qquad (n \text{ hinreichend groß})$$

und

$$\lim_n \frac{1}{n} H(p_n) = H,$$

also wegen (5.7) und (5.3)

$$\limsup_{n\to\infty} \frac{1}{n} (H(q_n) - H(p_n)) = 0$$

oder

$$(5.9) \qquad H(q_n) - H(p_n) < n\,\varepsilon\,\delta \qquad (n \text{ hinreichend groß}).$$

Ferner haben wir nach (5.7) und Lemma 1.3

$$(5.10) \qquad q_n\{x_n : q_n\{x_n\} > a\} \leqq p_n\{x_n : q_n\{x_n\} > a\} \qquad (a \geqq 0).$$

In Lemma 5.2 und Lemma 5.3 wird nun gezeigt, daß (5.6) aus (5.8) bis (5.10) folgt (man ersetze in Lemma 5.3 a durch $e^{-n(H+\delta)}$ und b durch $e^{-n\delta}$). Daraus ergibt sich der Satz.

Wir vereinfachen die Bezeichnungen. Sei X eine endliche Menge, und seien q und p endliche Maße in X. $q \swarrow p$ bedeute, daß es endliche Maße $p_j\,(j = 1, \ldots, k)$ $a_j > 0$, $x_j, y_j \in X\,(j = 1, \ldots, k-1)$ mit

$$\begin{aligned} &p_1 = p, \quad p_k = q \\ &p_{j+1} = p_j + a_j\,\delta_{y_j} - a_j\,\delta_{x_j} \qquad (j = 1, \ldots, k-1) \\ &p_{j+1}(y_j) \leqq p_{j+1}(x_j) \\ &\{x_j : j = 1, \ldots, k-1\} \cap \{y_j : j = 1, \ldots, k-1\} = 0 \end{aligned}$$

gibt. Die Folge $(p_j)_{j=1,\ldots,k}$ heißt eine Abtragung von p nach q.

Lemma 5.2. *Sind p,q $W\,V$en und gilt*

$$(5.11) \qquad q\{x : q\{x\} > a\} \leqq p\{x : q\{x\} > a\} \qquad (a \geqq 0),$$

so ist $q \swarrow p$.

Beweis. Um Induktion nach der Mächtigkeit von X durchführen zu können, beweisen wir etwas allgemeiner:

Sind q,p endliche Maße in X, ist

$$E \subseteqq \{x : q\{x\} \leqq p\{x\}\},$$

$$X - E \subseteqq \{x : q\{x\} \geqq p\{x\}\},$$

und ist (5.11) erfüllt, so gibt es ein p' mit

$$\begin{aligned} &p' \swarrow p\,, \\ &p'\{x\} \geqq q\{x\} && (x \in E)\,, \\ &p'\{x\} = q\{x\} && (x \in X - E) \end{aligned}$$

und

$$x_j \in E, \quad y_j \in X - E \qquad (j = 1, \ldots, k-1)\,.$$

Für einpunktiges X ist dies trivial. Allgemein sei

(5.12) $$Z = \{z : q\{z\} = \min\{q\{x\} : x \in X\}\}\,.$$

Nach Induktionsvoraussetzung gibt es ein p'' in $X - Z$ mit folgenden Eigenschaften:

$$\begin{aligned} &p'' \swarrow p \\ &p''\{x\} \geqq q\{x\} && (x \in X - Z) \\ &p''\{x\} = q\{x\} && (x \in X - (E \cup Z)) \\ &\tilde{x}_i \in E - Z, \quad \tilde{y}_i \in X - (E \cup Z) && (i = 1, \ldots, l-1) \end{aligned}$$

($\tilde{x}_i, \tilde{y}_i$ seien die bei einer Abtragung von p nach p'' auftretenden Punkte.) Ist q_1 die Einschränkung von q auf $E \cup Z$ und p_1 in $E \cup Z$ so definiert, daß p_1 in Z mit p und in $E - Z$ mit p'' übereinstimmt, so genügt es, die Induktionsaussage für q_1, p_1, E in der Menge $E \cup Z$ (an Stelle von q, p, E in X) zu beweisen, da ein p' mit den gesuchten Eigenschaften dann in naheliegender Weise durch Zusammensetzen der beiden gefundenen Abtragungen gewonnen werden kann. Wir wissen, daß $q_1\{x\}$ als Funktion von x auf $Z - E$ (sogar auf Z) konstant ist und daß

$$q_1\{x\} \leqq p_1\{x\} \qquad (x \in E)$$

und

$$q_1(E \cup Z) \leqq p_1(E \cup Z)$$

gelten. Der Beweis für die Existenz einer Abtragung der gesuchten Art ist unter diesen Bedingungen leicht zu führen und wird dem Leser überlassen.

Lemma 5.3. *Sind q, p WVen in X, $a > 0$, $0 < b < \varepsilon < 1$ und gilt $q \swarrow p$,*

(5.13) $$H(q) - H(p) < -\varepsilon \log b$$

und

(5.14) $$p\{x : p\{x\} \leqq a\} < \varepsilon\,,$$

so hat man

(5.15) $$q\{x : q\{x\} \leqq a b^2\} < 5\,\varepsilon\,.$$

Beweis. Wir wählen eine Abtragung $(p_j)_{1 \leqq j \leqq k}$ von p nach q so, daß für kein j

(5.16) $$p_j\{x_j\} > a\,b > p_{j+1}\{x_j\}$$

ist (das ist durch Aufspaltung der einzelnen Schritte (und der a_j) immer möglich) und zerlegen $F = \{x : q\{x\} \leqq a b^2\}$ in die Menge F_1 der Punkte, die als ein y_j

vorkommen und in den Rest F_2. Diesen zerlegen wir weiter in

$$F_{21} = \{x : x \in F_2, p\{x\} > a\} \quad \text{und} \quad F_{22} = F_2 - F_{21}.$$

Seien

$$\beta_1 = \{j : 1 \leqq j \leqq k-1, y_j \in F_1, p\{x_j\} \leqq a\},$$
$$\beta_2 = \{j : 1 \leqq j \leqq k-1, y_j \in F_1, p\{x_j\} > a, p_{j+1}\{x_j\} < a\,b\},$$
$$\beta_3 = \{j : 1 \leqq j \leqq k-1, y_j \in F_1, p\{x_j\} > a, p_{j+1}\{x_j\} \geqq a\,b\}.$$

Beim j-ten Schritt der Abtragung wird die Wahrscheinlichkeit von F_1 um $a_j \chi_{F_1}(y_j)$ erhöht, während die jeder Teilmenge von F_2 höchstens erniedrigt wird. Wir haben also

$$q(F) = q(F_1) + q(F_{21}) + q(F_{22})$$
$$\leqq p(F_1) + \sum_{j \in \beta_1} a_j + \sum_{j \in \beta_2} a_j + \sum_{j \in \beta_3} a_j + q(F_{21}) + p(F_{22}).$$

Aus

$$p\{x\} < q\{x\} \leqq a\,b^2 < a \qquad (x \in F_1),$$

$F \cap F_{22} = 0$ und (5.14) folgt

$$p(F_1) + p(F_{22}) < \varepsilon.$$

Wegen

$$p\{x\} > a \qquad (x \in F_{21})$$

ist die Mächtigkeit von $F_{21} < 1/a$, also

$$q(F_{21}) < \frac{1}{a}\,a\,b^2 < \varepsilon.$$

$$\sum_{j \in \beta_1} a_j < \varepsilon$$

folgt aus (5.14) ($\sum_{j \in \beta_1} a_j$ ist eine aus $\{x : p\{x\} \leqq a\}$ abgetragene Masse).

Die Mächtigkeit der Menge

$$E = \{x : x = x_j \text{ für geeignetes } j \in \beta_2\}$$

ist $< 1/a$, setzt man also

$$j(x) = \min\{i : i \in \beta_2, x_i = x\} \qquad (x \in E),$$

so folgt aus dem Verbot von (5.16) und der Definition von β_2

$$\sum_{\substack{j \in \beta_2 \\ x_j = x}} a_j < p_{j(x)}\,\{x\} \leqq a\,b \qquad (x \in E),$$

also

$$\sum_{j \in \beta_2} a_j = \sum_{x \in E} \sum_{\substack{j \in \beta_2 \\ x_j = x}} a_j < \frac{1}{a}\,a\,b < \varepsilon.$$

Sei nun $j \in \beta_3$,

$$p^t = p_j + t\,a_j\,\delta_{y_j} - t\,a_j\,\delta_{x_j} \qquad (0 \leqq t \leqq 1).$$

Dann ist

$$\frac{d}{dt} H(p^t) = a_j \log \frac{p^t\{x_j\}}{p^t\{y_j\}} \geqq a_j \log \frac{p_{j+1}\{x_j\}}{p_{j+1}\{y_i\}}$$
$$\geqq a_j \log \frac{p_{j+1}\{x_j\}}{q\{y_j\}} \geqq a_j \log \frac{a\,b}{a\,b^2} = -\,a_j \log \mathrm{b},$$

also auch

$$H(p_{j+1}) - H(p_j) \geqq -\,a_j \log b.$$

Da sich bei jedem Schritt der Abtragung von p nach q die Entropie vergrößert, folgt hieraus und aus (5.13)

$$-\log b \sum_{j\in\beta_3} a_j \leqq \sum_{j\in\beta_3} (H(p_{j+1}) - H(p_j)) \leqq H(q) - H(p) < -\,\varepsilon \log b,$$

also

$$\sum_{j\in\beta_3} a_j < \varepsilon.$$

Fassen wir unsere Ergebnisse zusammen, so erhalten wir (5.15).

Sei nun wieder X der Produktraum (5.1) und sei T die Shift-Transformation:

$$(T\,x)^i = x^{i+1} \qquad (i \text{ ganz},\ x \in X).$$

Satz 5.4. *Ist u eine ergodische Unterquelle in X, $H = H(u)$ die Entropie von u und q_n' das Anzahlmaß in X_n, so gilt für jedes $\varepsilon > 0$*

$$\min\{q_n'(E_n) : u_n(E_n) \geqq 1 - \varepsilon\} = e^{nH + o(n)}.$$

Beweis. Wir setzen

$${}_iX_k = \{{}_ix_k = (x^i, \ldots, x^k) : \quad x^j \in X^0 \quad (j = i, \ldots, k)\}$$

und

$${}_i\Theta_k = \{{}_i\vartheta_k = (\omega^i, \ldots, \omega^k) : \omega^j \in \Omega(X^0) \quad (j = i, \ldots, k)\} \qquad (i \leqq k \text{ ganz}).$$

Seien ${}_iu_k$ bzw. ${}_i\mu_k$ die Projektionen von u bzw. μ auf ${}_iX_k$ bzw. ${}_i\Theta_k$, ferner ${}_i\varphi_k$ und ${}_i\boldsymbol{P}_k$ für ${}_iX_k$ wie φ und $\boldsymbol{P}$ für X definiert. Fassen wir ${}_i\Theta_k$ als Teilmenge von $\Omega({}_iX_k)$ auf, so ist ${}_i\mu_k$ eine $W\,V$ in $\Omega({}_iX_k)$ und es gilt

$${}_iu_k = {}_i\varphi_k({}_i\mu_k).$$

Ist ${}_iq_k \geqq {}_iu_k$ mit

$$H({}_iq_k) = H({}_iu_k),$$

so gibt es nach Satz 1.1 ein ${}_iP_k \in {}_i\boldsymbol{P}_k$ mit

$${}_iq_k = {}_iP_k({}_i\mu_k).$$

Wir sehen ${}_i\mu_k$ wieder als $W\,V$ in ${}_i\Theta_k$ und ${}_iP_k$ (nach Einschränkung des Definitionsbereichs) als Markoffschen Kern von ${}_i\Theta_k$ nach ${}_iX_k$ an. Sei $k \geqq 1$ und

$$\overset{k}{P}(\ ,\vartheta) = \prod_{i \text{ ganz}}{}^{\times}\ {}_{ik+1}P_{(i+1)k}(\ \ , {}_{ik+1}\vartheta_{(i+1)\,k}).$$

Wir haben

$$\overset{k}{P} \in \mathbf{P} \tag{5.17}$$

und für $i \geqq 1$ folgt aus den drei letzten Gleichungen

$$\left| \frac{1}{ik} H({}_1(\overset{k}{P}\mu)_{ik}) - \frac{1}{k} H({}_1u_k) \right| = \left| \frac{1}{ik} H({}_1(\overset{k}{P}\mu)_{ik}) - \frac{1}{k} H({}_1(\overset{k}{P}\mu)_k) \right|$$

$({}_j(\overset{k}{P}\mu)_l$ ist die Projektion von $P\mu$ auf ${}_jX_l)$

$$= \frac{1}{ik} \left| H({}_1(\overset{k}{P}\mu)_{ik} - \sum_{j=0}^{i-1} H({}_{jk+1}(\overset{k}{P}\mu)_{(j+1)k}) \right|$$

$$= \frac{1}{ik} \left| H({}_1\overset{k}{P}_{ik}({}_1\mu_{ik})) - H\left({}_1\overset{k}{P}_{ik}\left(\prod_{j=0}^{i-1}{}^{\times}\, {}_{jk+1}\,\mu_{(j+1)k}\right)\right) \right|$$

$$= -\frac{1}{ik} H\left({}_1\overset{k}{P}_{ik}({}_1\mu_{ik}),\quad {}_1\overset{k}{P}_{ik}\left(\prod_{j=0}^{i-1}{}^{\times}\, {}_{kj+1}\,\mu_{(j+1)k}\right)\right)$$

(siehe (1.5))

$$\leqq -\frac{1}{ik} H\left({}_1\mu_{ik},\quad \prod_{j=0}^{i-1}{}^{\times}\, {}_{jk+1}\mu_{(j+1)k}\right)$$

(siehe etwa KULLBACK-LEIBLER [*23*], DOBRUSCHIN [*24*])

$$= \frac{1}{ik} \left| H({}_1\mu_{ik}) - H\left(\prod_{j=0}^{i-1}{}^{\times}\, {}_{jk+1}\mu_{(j+1)k}\right) \right| = \left| \frac{1}{ik} H({}_1\mu_{ik}) - \frac{1}{k} H({}_1\mu_k) \right|.$$

Ist $\varepsilon > 0$ und k hinreichend groß, so haben wir

$$\left| \frac{1}{k} H({}_1u_k) - H \right| < \frac{\varepsilon}{2}$$

und

$$\left| \frac{1}{ik} H({}_1\mu_{ik}) - \frac{1}{k} H({}_1\mu_k) \right| < \frac{\varepsilon}{2} \qquad (i \geqq 1)$$

(siehe CHINTSCHIN [*18*]), also auch

$$\left| \frac{1}{ik} H({}_1(\overset{k}{P}\mu)_{ik} - H \right| < \varepsilon \qquad (i \geqq 1). \tag{5.18}$$

Setzen wir

$$P = \frac{1}{k} \sum_{j=0}^{k-1} \widetilde{T}^j \overset{k}{P}$$

(siehe (4.16)), so ist $P\mu$ invariant (da P invariant ist), $P\mu \geqq u$ (wegen $P \in \mathbf{P}$, siehe (5.17)) und

$$H(P\mu) \geqq H(u) - \varepsilon$$

(wie man leicht aus (5.18) und der Konkavität von H folgert), also

$$H = \sup\{H(p) : p \in J(u)\}.$$

Hieraus folgt die Existenz eines Extremalpunktes p von $J(u)$ mit

$$H = H(p)$$

(siehe BREIMAN [25]). Nach Lemma 4.5 ist p ergodisch, nach dem Satz von MACMILLAN erfüllt P die Voraussetzung von Satz 5.1. Damit ist Satz 5.4 bewiesen.

Herrn KONRAD JACOBS gilt mein dauernder Dank für seine langjährige Förderung, sein Interesse an dieser Arbeit und zahlreiche anregende Bemerkungen. Einen wichtigen Literaturhinweis verdanke ich auch Herrn HEINZ BAUER.

Literatur

[1] KOLMOGOROFF, A.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik **2**, No. 3, Berlin 1933.
[2] BLACKWELL, D., L. BREIMAN and A. J. THOMASIAN: The capacity of a class of channels. Ann. math. Statistics **30**, 1229 (1959).
[3] — — — The capacities of certain channel classes under random coding. Ann. mat. Statistics **31**, 558 (1960).
[4] WOLFOWITZ, J.: Simultaneous channels. Arch. Rational Mech. Anal. **4**, No. 4, 371 (1960).
[5] — Coding Theorems of Information Theory. Ergebnisse der Mathematik, Heft 31, Berlin 1961.
[6] KIEFER, J., and J. WOLFOWITZ: Channels with arbitrary varying channel probability functions. Information and Control. Vol. 5, No. 1 (1962).
[7] CHOQUET, G.: Theory of Capacities. Ann. Inst. Fourier. **5**, 131 (1953/54).
[8] JACOBS, K.: Die Übertragung diskreter Informationen durch periodische und fastperiodische Kanäle. Math. Ann. **137**, 125 (1959).
[9] SHAPLEY, L. S.: A value for n-Person Games. Contributions to the Theory of Games II (1953).
[10] NEUMANN, J. VON, und O. MORGENSTERN: Spieltheorie und wirtschaftliches Verhalten. Würzburg 1961.
[11] BURGER, E.: Einführung in die Theorie der Spiele. Berlin 1959.
[12] LEHMANN, E. L.: Testing Statistical Hypotheses. New York 1961.
[13] WALD, A.: Statistical Decision Functions. New York 1950.
[14] LE CAM, L.: A Generalization of the Theory of A. Wald. Ann. math. Statistics **26**, 69 (1955).
[15] BLACKWELL, D., and M. A. GIRSHICK: Theory of Games and Statistical Decisions. New York 1954.
[16] LEHMANN, E. L.: On the existence of least favorable distributions. Ann. math. Statistics **23**, 408 (1952).
[17] STRASSEN, V.: Asymptotische Abschätzungen in Shannons Informationstheorie. Transactions of the third Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (zum Druck eingereicht).
[18] CHINTSCHIN, A. J.: Über grundlegende Sätze der Informationstheorie. Math. Forschungsberichte, Berlin 1957.
[19] SHANNON, C. E.: The zero capacity of a noisy channel. IRE Trans., Inform. Theory, IT-2 (1956).
[20] JACOBS, K.: Neuere Methoden und Ergebnisse der Ergodentheorie. Ergebnisse der Mathematik, Heft 29, Berlin 1960.
[21] DUNFORD, N., and J. T. SCHWARTZ: Linear Operators. London 1958.
[22] LOOMIS, L. H.: An Introduction to Abstract Harmonic Analysis. New York 1953.
[23] KULLBACK, S., and R. A. LEIBLER: Information and Statistics. New York 1959.
[24] DOBRUSCHIN, R. L.: A general formulation of the fundamental Shannon theorem in information theory. Uspechi mat. Nauk. **14**, 3 (1959).
[25] BREIMAN, L.: On achieving channel capacity in finite-memory channels. Illinois J. Math. **4**, 246 (1960).

[26] PEREZ, A.: Sur la théorie de l'information dans le cas d'un alphabet abstrait. Trans. first Prague Conference of Information Theory, Statistical Decision Functions and Random Processes (1957).
[27] ROOT, W. L.: Communications through Unspecified Additive Noise. Information and Control **4**, Nr. 1, 15 (1961).
[28] JACOBS, K.: Almost Periodic Channels. Colloquium on Combinatorial Methods in Probability Theory, Aarhus 1962.
[29] MAAK, W.: Fastperiodische Funktionen. Berlin 1950.
[30] MORDELL, L. J.: On the inequality $\sum_{r=1}^{n} \frac{x_r}{x_{r+1}+x_{r+2}} \geqq \frac{n}{2}$ and some others. Abh. math. Sem. Univ. Hamburg **22**, (1958).

Institut für Mathematische
Statistik der Universität
34 Göttingen

Statistics Department
University of California
Berkeley, California

(Eingegangen am 15. März 1963)