EINE INFORMATIONSTHEORETISCHE UNGLEICHUNG UND IHRE ANWENDUNG AUF DEN BEWEIS DER ERGODIZITÄT VON MARKOFFSCHEN KETTEN

von

Imre CSISZÁR

Einleitung

Setzen wir voraus, daß durch einen gestörten Kanal irgendwelche Signale übertragen werden, welche aus zwei verschiedenen Quellen stammen dürfen. Die Wahrscheinlichkeitsverteilungen der Signale der beiden Quellen seien verschieden. Nun ist es anschaulich klar, daß die Störungen während der Übertragung die Verschiedenheit der Verteilungen nur vermindern können, also die Wahrscheinlichkeitsverteilungen der Ausgangssignale, welche Eingangssignalen aus verschiedenen Quellen entsprechen, voneinander nur in kleinerem (genauer: nicht größerem) Maße abweichen können, als diejenigen der Eingangssignalen selbst. Satz 1 der vorliegenden Arbeit ist eine mathematische Wiedergabe dieser anschaulichen Aussage, indem er besagt, daß die f-Abweichung der Verteilungen der übertragenen Signale nicht zunehmen kann, wobei der Begriff der f-Abweichung — definiert durch (4) — die Information für Diskrimination von Kullback und Leibler [1] (relative Information erster Ordnung; auch relative Entropie genannt), die relative Information beliebiger Ordnung $\alpha > 0$ (vgl. [2]) sowie die gewöhnliche Variationsentfernung als Spezialfälle enthält.

Die Bedingung dafür, daß die Übertragung durch den Kanal keine Verminderung der f-Abweichung bedeutet, wird als Ergänzung des Satzes 1 angegeben. Im § 2 werden noch manche Folgerungen bzw. Spezialfälle des Satzes 1 dargestellt. Im § 3 betrachten wir den Fall der kleinen Abweichungsverminderung (Stabilitätsfrage der Ungleichung, Satz 2), und die Verallgemeinerung der Ungleichung für nicht normierbare Maße (Satz 3).

Als eine Anwendung wird ein rein informationstheoretischer Beweis der Ergodizität bestimmer Markoffschen Ketten dargestellt[1] (§ 4).

Herrn Professor Alfréd Rényi möchte ich für seine wertvollen Bemerkungen meinen Dank aussprechen.

§ 1. Bezeichnungen und Hilfssätze

In den folgenden wird $f(u)$ immer eine für $0 < u < +\infty$ definierte stetige und von unten konvexe, sonst aber beliebige Funktion bedeuten.

[1] Der erste informationstheoretische Beweis für Ergodizität Markoffscher Ketten stammt von A. Rényi [2]. Dort brauchte man aber auch matrizentheoretische Überlegungen, nämlich um die Existenz einer stationären Verteilung zu beweisen.

Es werden auch Ausdrücke der Form $f(0)$, $0 \cdot f\left(\frac{0}{0}\right)$ und $0 \cdot f\left(\frac{a}{0}\right)$ $(0 < a < +\infty)$ vorkommen; diese sind folgendermassen zu verstehen:

$$f(0) = \lim_{u \to 0} f(u) \tag{1}$$

$$0 \cdot f\left(\frac{0}{0}\right) = 0 \tag{2}$$

$$0 \cdot f\left(\frac{a}{0}\right) = \lim_{\varepsilon \to +0} \varepsilon f\left(\frac{a}{\varepsilon}\right) = a \lim_{u \to \infty} \frac{f(u)}{u} \qquad (0 < a < +\infty) \tag{3}$$

Die Größen $f(0)$ und $0 \cdot f\left(\frac{a}{0}\right)$ können also auch gleich $+\infty$ sein, $-\infty$ aber nicht, wegen der Konvexität von $f(u)$.

Mit den Buchstaben $\mathbf{P}$ und $\mathbf{Q}$ werden wir Wahrscheinlichkeitsmaße auf einem meßbaren Raum (X, S_X) bezeichnen, mit λ ein σ-endliches Maß, welches $\mathbf{P}$ und $\mathbf{Q}$ dominiert (also für welches $\mathbf{P} \ll \lambda$, $\mathbf{Q} \ll \lambda$ gilt; ein solches Maß ist z. B. $\mathbf{P} + \mathbf{Q}$). Wir werden uns mit Größen der Form

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \int_X q(x) f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) \tag{4}$$

beschäftigen, wobei $p(x) = \frac{\mathbf{P}(dx)}{\lambda(dx)}$, $q(x) = \frac{\mathbf{Q}(dx)}{\lambda(dx)}$ die λ-Dichten von $\mathbf{P}$ und $\mathbf{Q}$ bedeuten. (Die Radon-Nikodymschen Ableitungen sind so bestimmt, daß sie überall — und nicht nur fast überall — nichtnegativ und endlich sind.) Das Integral in (4) ist immer sinnvoll, da wegen der Konvexität von $f(u)$ für geeignete reelle Zahlen A und B $f(u) \geqq A + Bu$ gelten muß, also kann die Funktion $qf\left(\frac{p}{q}\right)$ von unten durch die λ-summierbare Funktion $Ap + Bq$ abgeschätzt werden. Ferner ist der Wert der Grösse $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$ von der Wahl des dominierenden Maßes λ offenbar unabhängig.

Die Größe $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$ wird der *f-Abweichung* der Wahrscheinlichkeitsverteilungen $\mathbf{P}$ und $\mathbf{Q}$ heissen, da sie als eine Verallgemeinerung verschiedener bekannten Abweichungsmaßzahlen anzusehen ist. Die relative Information[2] (erster Ordnung) zweier Verteilungen $\mathbf{P}$ und $\mathbf{Q}$ ist z. B. nichts anderes, als ihre $u \log u$-Abweichung. Es gilt nämlich für $f(u) = u \log u$

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \int_X q(x) \frac{p(x)}{q(x)} \log \frac{p(x)}{q(x)} \lambda(dx) = \int_X p(x) \log \frac{p(x)}{q(x)} \lambda(dx) = \tag{5}$$

$$= \left\{ \begin{array}{ll} \int_X \log \frac{\mathbf{P}(dx)}{\mathbf{Q}(dx)} \cdot \mathbf{P}(dx) & \text{falls } \mathbf{P} \ll \mathbf{Q} \\ +\infty & \text{sonst} \end{array} \right\} = I(\mathbf{P} \| \mathbf{Q}).$$

[2] Auch die Bezeichnungen »Information für Diskrimination«, »I-Divergenz«, »Informationsgewinn«, »verallgemeinerte Entropie« sind benutzt, vgl. z. B. [1], [2], [3], [4].

Auch für $\alpha \neq 1$ sind die Informationsgrössen $I_\alpha(\mathbf{P} \,||\, \mathbf{Q})$ (relative Information von Ordnung α, $\alpha > 0$, s. [2], vgl. auch [5]) aus den Grössen $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$ herleitbar. Setzt man nämlich

$$(6) \qquad f_\alpha(u) = \begin{cases} -u^\alpha & \text{für} \quad 0 < \alpha < 1 \\ u^\alpha & \text{für} \quad \alpha > 1 \end{cases},$$

so ergibt sich

$$(7) \qquad \mathcal{I}_{f\alpha}(\mathbf{P}, \mathbf{Q}) = \operatorname{sgn}(\alpha - 1) \int_X p^\alpha(x)\, q^{1-\alpha}(x)\, \lambda(dx) = \operatorname{sgn}(\alpha - 1)\, \mathcal{I}_\alpha(\mathbf{P}, \mathbf{Q})$$

wobei $\mathcal{I}_\alpha(\mathbf{P}, \mathbf{Q}) = \int_X p^\alpha(x) q^{1-\alpha}(x)\, \lambda(dx)$ das Informationsintegral von Ordnung α bedeutet (vgl. [5]), also gilt

$$(8) \qquad I_\alpha(\mathbf{P} \,||\, \mathbf{Q}) = \frac{1}{\alpha - 1} \log \mathcal{I}_\alpha(\mathbf{P}, \mathbf{Q}) = \frac{1}{\alpha - 1} \log |\mathcal{I}_{f\alpha}(\mathbf{P}, \mathbf{Q})|,$$

d. h. stellt $I_\alpha(\mathbf{P} \,||\, \mathbf{Q})$ eine monoton zunehmende Funktion von $\mathcal{I}_{f\alpha}(\mathbf{P}, \mathbf{Q})$ dar.

Die gewöhnliche Variationsentfernung[3] zweier Verteilungen $\mathbf{P}$ und $\mathbf{Q}$ ist auch ein Spezialfall der f-Abweichung; setzt man nämlich $f(u) = |u - 1|$, so gilt offenbar

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \int_X q(x) \left| \frac{p(x)}{q(x)} - 1 \right| \lambda(dx) = \int_X |p(x) - q(x)|\, \lambda(dx) = |\mathbf{P} - \mathbf{Q}|\,(X).$$

Es soll allerdings betont sein, daß die „f-Abweichung" zweier Verteilungen — für allgemeines f — auch negativ sein kann. Es gilt aber allgemein — s. (35) —

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq f(1) = \mathcal{I}_f(\mathbf{P}, \mathbf{P}),$$

so daß durch Addieren einer Konstante die f-Abweichung immer zu einer annehmbaren — obwohl im allgemeinen keiner symmetrischen — Verschiedenheitsmaßzahl gemacht werden kann.

Bemerkung. Die relative Information $I_\alpha(\mathbf{P} \,||\, \mathbf{Q})$ ist bekanntlich gleich dem Supremum der relativen Informationen bezüglich aller endlichen Unteralgebren von S_X, also gilt die Beziehung $I_\alpha(\mathbf{P} \,||\, \mathbf{Q}) = \sup_{A \subset S_X} I_\alpha(\mathbf{P}_A \,||\, \mathbf{Q}_A)$, wobei $\mathbf{P}_A$ bzw. $\mathbf{Q}_A$ die Einschränkung von $\mathbf{P}$ bzw. $\mathbf{Q}$ auf die endliche Algebra A bedeutet (für $\alpha = 1$ siehe [3], [4]; für beliebiges $\alpha > 0$ vgl. [5]). In derselben Weise kann gezeigt werden, daß diese Beziehung auch für $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$ bei beliebigem konvexem $f(u)$ gültig ist.

In den folgenden werden wir die Jensensche Ungleichung oft benützen und zwar in der folgenden Form:

Hilfssatz 1. *Ist μ ein beliebiges Wahrscheinlichkeitsmaß auf den Borelschen Mengen der Halbgerade $[0, +\infty)$ mit $\int_0^\infty u\, \mu(du) < +\infty$, so gilt für jedes konvexe $f(u)$*

$$(9) \qquad \int_0^\infty f(u)\, \mu(du) \geqq f\left(\int_0^\infty u\, \mu(du)\right).$$

[3] Unter der Variationsentfernung zweier Wahrscheinlichkeitsmaße wird die Totalvariation ihrer Differenz auf dem ganzen Definitionsbereich der beiden Maße, also — falls $\mathbf{P}$ und $\mathbf{Q}$ auf (X, S_X) definierte Wahrscheinlichkeitsmaße sind — die Grösse $|\mathbf{P} - \mathbf{Q}|\,(X) = \int_X |p(x) - q(x)|\, \lambda(dx)$ verstanden.

Ist die Funktion $f(u)$ streng konvex, enthält also ihr Graph keine gerade Strecke, so ist für die Gleichheit in (9) *notwendig und hinreichend, daß das Maß μ auf den einzigen Punkt $u_0 = \int\limits_0^\infty u\,\mu(du)$ konzentriert ist.*

Wir brauchen auch den folgenden Hilfssatz über die Stabilität der Jensenschen Ungleichung:

Hilfssatz 2. *Ist die Funktion $f(u)$ in der ε_1-Umgebung von $u_0 = \int\limits_0^\infty u\;\mu(du)$ zweimal differenzierbar, und gilt hier $f''(u) \geqq \tau > 0$, so folgt aus*

$$\int\limits_0^\infty f(u)\,\mu(du) - f\left(\int\limits_0^\infty u\,\mu(du)\right) \leqq \frac{\tau}{2}\,\varepsilon_1^2\,\varepsilon$$

die Ungleichung $\mu([u_0 - \varepsilon_1,\; u_0 + \varepsilon_1]) > 1 - \varepsilon_2$.

Hilfssatz 2 ist ein Spezialfall des Lemmas in [5], § 2, und ergibt sich daraus, indem man $X = [0, +\infty)$, $g(u) = u$, $A = u_0$, $\delta_1 = 0$, $\delta_2 = \int\limits_0^\infty f(u)\;\mu(du) - f(\int\limits_0^\infty u\;\mu(du))$, $a = \tau$ setzt.

Hilfssatz 3. *Es gilt für jedes $p_0 \geqq 0$, $p_1 \geqq 0$, $q \geqq 0$*

$$0 \cdot f\left(\frac{p_0}{0}\right) + qf\left(\frac{p_1}{q}\right) \geqq qf\left(\frac{p_0 + p_1}{q}\right), \tag{10}$$

mit Gleichheit nur im Falle $p_0 = 0$ oder $q = 0$. Ist ferner $f(u)$ in der ε_1-Umgebung von $u_0 = \dfrac{p_0 + p}{q}$ $(q > 0)$ zweimal differenzierbar und gilt hier $f''(u) \geqq \tau > 0$, so folgt aus

$$0 \cdot f\left(\frac{p_0}{0}\right) + qf\left(\frac{p_1}{q}\right) - qf\left(\frac{p_0 + p_1}{q}\right) < \frac{\tau}{2}\,\varepsilon_1^2\,\varepsilon_2\,q \qquad (\varepsilon_2 < 1) \tag{11}$$

die Ungleichung $\dfrac{p_0}{q} \leqq \varepsilon_1$.

Beweis. Für $q = 0$ ist (10) eine unmittelbare Folge von (2), (3). Für $q > 0$ folgt (10) aus (9), indem man das Maß μ durch $\mu\left(\left\{\dfrac{p_0}{\varepsilon}\right\}\right) = \dfrac{\varepsilon}{q + \varepsilon}$, $\mu\left(\left\{\dfrac{p_1}{q}\right\}\right) = \dfrac{q}{q + \varepsilon}$ definiert $(\varepsilon > 0)$, und dann ε gegen 0 streben läßt. Die zweite Behauptung folgt ähnlicherweise aus Hilfssatz 2, man soll bloß bemerken, daß aus (11) für beliebige Konstante K

$$\frac{\varepsilon}{q + \varepsilon} f\left(\frac{p_0}{\varepsilon}\right) + \frac{q}{q + \varepsilon} f\left(\frac{p_1}{q}\right) - f\left(\frac{p_0 + p_1}{q + \varepsilon}\right) \leqq \frac{\tau}{2}(\varepsilon_1 - K\varepsilon)^2\,\varepsilon_2$$

folgt, falls nur $\varepsilon > 0$ genügend klein ist, was nach Hilfssatz 2 die Ungleichung[4]

[4] Falls nur K so gewählt ist, daß die $(\varepsilon_1 - K\varepsilon)$-Umgebung von $\dfrac{p_0 + p_1}{q + \varepsilon}$ ein Teilintervall der ε_1-Umgebung von $\dfrac{p_0 + p_1}{q}$ darstellt, wofür z. B. $K = \dfrac{p_0 + p_1}{q^2}$ eine geeignete Wahl ist.

$\mu\left(\left[\frac{p_0+p_1}{q+\varepsilon}-(\varepsilon_1-K\varepsilon), \frac{p_0+p_1}{q+\varepsilon}+(\varepsilon_1-K\varepsilon)\right]\right) > 1-\varepsilon_2$ und daher, für genügend kleines $\varepsilon > 0$, $\frac{p_1}{q} \in \left[\frac{p_0+p_1}{q+\varepsilon}-(\varepsilon_1-K\varepsilon), \frac{p_0+p_1}{q+\varepsilon}+(\varepsilon_1-K\varepsilon)\right]$ mit sich bringt, woraus mit dem Grenzübergang $\varepsilon \to 0$ wirklich $\frac{p_0}{q} \leqq \varepsilon_1$ fogt.

§ 2. Das Abnehmen der f-Abweichung, wenn ein Übertragungskanal benützt wird

Fassen wir den meßbaren Raum (X, S_X) dem Raum der Eingangssignale eines Übertragungskanals auf. Der Ausgangsraum sei mit (Y, S_Y) bezeichnet, und die Übergangsfunktion mit $\nu(B \mid x)$. Also sei $\nu(B \mid x)$ für jedes feste $x \in X$ ein Wahrscheinlichkeitsmaß auf (Y, S_Y) und für jedes feste $B \in S_Y$ eine S_X-meßbare Funktion von x.

Sind $\mathbf{P}$ und $\mathbf{Q}$ zwei Wahrscheinlichkeitsmaße auf (X, S_X), die Wahrscheinlichkeitsmaße $\overline{\mathbf{P}}$ und $\overline{\mathbf{Q}}$ auf (Y, S_Y) werden durch

$$\overline{\mathbf{P}}(B) = \int_X \nu(B|x)\,\mathbf{P}(dx), \qquad \overline{\mathbf{Q}}(B) = \int_X \nu(B|x)\,\mathbf{Q}(dx) \tag{12}$$

definiert. $\overline{\mathbf{P}}$ bzw. $\overline{\mathbf{Q}}$ stellt offenbar die Wahrscheinlichkeitsverteilung der Ausgangssignale dar, wenn die Verteilung der Eingangssignale $\mathbf{P}$ bzw. $\mathbf{Q}$ ist.

Es ist anschaulich klar, daß die Störungen (»Geräusch«) während der Übertragung in Richtung der Verminderung der Verschiedenheit der Verteilungen von Signalen aus verschiedenen Quellen wirken, mit anderen Worten, daß das Geräusch einen Verlust an aus den Signalen für Bestimmung ihrer Quelle erhältlichen Information verursacht.

Man erwartet also die Gültigkeit der Ungleichung

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) \tag{13}$$

die wir in diesem Paragraphen beweisen wollen. Man bemerke, daß Ungleichung (13) im Falle $f(u) = u \log u$ auch in engerem Sinne einen Informationsverlust ausdrückt, da $\mathcal{I}_{u\log u}(\mathbf{P}, \mathbf{Q}) = I(\mathbf{P} \| \mathbf{Q})$ als die Maßzahl der für Unterscheidung der Verteilungen $\mathbf{P}$ und $\mathbf{Q}$ erhältlichen Information gilt [1].

Schicken wir noch manche Bezeichnungen voraus. Die Maße $\mathbf{P}$ und $\overline{\mathbf{P}}$ bzw. $\mathbf{Q}$ und $\overline{\mathbf{Q}}$ lassen sich als die Projektionen auf (X, S_X) und (Y, S_Y) von Maßen $\mathbf{P}^*$ bzw. $\mathbf{Q}^*$ auf dem Produktraum $(X \times Y,\ S_X \times S_Y)$ auffassen, wobei $\mathbf{P}^*$ bzw. $\mathbf{Q}^*$ für Mengen der Form $A \times B$ $(A \in S_X,\ B \in S_Y)$ durch

$$\mathbf{P}^*(A \times B) = \int_A \nu(B|x)\,\mathbf{P}(dx), \qquad \mathbf{Q}^*(A \times B) = \int_A \nu(B|x)\,\mathbf{Q}(dx) \tag{14}$$

und für das ganze $S_X \times S_Y$ in der gewöhnlichen Weise durch Erweiterung definiert sind.

Wählen wir ein σ-endliches Maß λ^* auf $(X \times Y,\ S_X \times S_Y)$ mit $\mathbf{P}^* \ll \lambda^*$, $\mathbf{Q} \ll \lambda^*$ derart, daß auch seine Projektionen auf (X, S_X) und (Y, S_Y) σ-endlich sind ($\lambda^* = \mathbf{P}^* + \mathbf{Q}^*$ ist z. B. geeignet); diese Projektionen dominieren offenbar

$\mathbf{P}$ und $\mathbf{Q}$ bzw. $\overline{\mathbf{P}}$ und $\overline{\mathbf{Q}}$, also dürfen sie als λ bzw. $\overline{\lambda}$ gewählt werden. In den folgenden werden $p(x)$ und $q(x)$ bzw. $\overline{p}(y)$ und $\overline{q}(y)$ die Dichten (also Radon-Nikodymsche Ableitungen, bestimmt als überall — und nicht nur fast überall — nichtnegative und endliche Funktionen) von $\mathbf{P}$ und $\mathbf{Q}$ bzw. $\overline{\mathbf{P}}$ und $\overline{\mathbf{Q}}$ bezüglich dieses λ bzw. $\overline{\lambda}$ bezeichnen.

Nun beweisen wir die Ungleichung (13) und zeigen, daß die Gleichheit (für streng konvexes $f(u)$ mit $\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < +\infty$) dann und nur dann besteht, wenn die Likelihood-Quotienen $\frac{p(x)}{q(x)}$ und $\frac{\overline{p}(y)}{\overline{q}(y)}$ mit $\mathbf{Q}^*$-Wahrscheinlichkeit 1 gleich sind.

Satz 1. *Es gilt für eine beliebige konvexe Funktion* $f(u)$ $(0 < u < +\infty)$ *und beliebige Wahrscheinlichkeitsmaße* $\mathbf{P}$ *und* $\mathbf{Q}$ *auf* (X, S_X)

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) . \tag{13}$$

Ergänzung. *Ist ferner die Funktion* $f(u)$ *streng konvex* (enthält also ihr Graph keine gerade Strecke), *so ist für die Gleichheit in der obigen Ungleichung notwendig und hinreichend, daß entweder* $\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) = +\infty$ *oder*

$$\frac{p(x)}{q(x)} = \frac{\overline{p}(y)}{\overline{q}(y)} \qquad [\mathbf{Q}^*] \tag{15}$$

d. h.

$$\nu\left(\left\{y : \frac{\overline{p}(y)}{\overline{q}(y)} = \frac{p(x)}{q(x)}\right\} | x\right) = 1 \qquad [\mathbf{Q}] \tag{16}$$

gilt.[5]

Beweis. Man darf $\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) < +\infty$ voraussetzen, da für $\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = +\infty$ nichts zu beweisen ist.

Um das Wesen des Beweises besser hervorzuheben, führen wir ihn zunächst im Spezialfall $\mathbf{P} \ll \mathbf{Q}$ durch. Aus $\mathbf{P} \ll \mathbf{Q}$ folgt offenbar auch $\mathbf{P}^* \ll \mathbf{Q}^*$ so daß man $\lambda^* = \mathbf{Q}^*$ wählen darf; dann gilt $p(x) = \frac{\mathbf{P}(dx)}{\mathbf{Q}(dx)}$, $\overline{p}(y) = \frac{\overline{\mathbf{P}}(dy)}{\overline{\mathbf{Q}}(dy)}$, $q(x) = 1$, $\overline{q}(y) = 1$.

Betrachte man nun die Wahrscheinlichkeitsalgebra $(X \times Y, S_X \times S_Y, \mathbf{Q}^*)$ und die auf ihr definierten Zufallsveränderlichen

$$\xi(x, y) = p(x)$$
$$\eta(x, y) = \overline{p}(y) .$$

Wir zeigen, daß $\eta = \overline{p}(y)$ gleich dem bedingten Erwartungswert von $\xi = p(x)$ bei gegebenem y, oder, genauer gesagt,

$$\eta = \mathbf{E}(\xi | S'_Y) \tag{17}$$

ist, wobei die σ-Algebra $S'_Y \subset S_X \times S_Y$ durch $S'_Y = \{X \times B : B \in S_Y\}$ definiert

[5] Bezeichnungen wie $[\mathbf{Q}^*]$, $[\overline{\mathbf{Q}}]$ u. s. w. bedeuten: fast überall bezüglich des Maßes $\mathbf{Q}^*$, $\overline{\mathbf{Q}}$, u. s. w.

ist. In der Tat ist η offenbar S'_Y-meßbar, ferner gilt für jede Menge aus S_Y d. h. für jede Menge der Form $X \times B$ $(B \in S_Y)$

$$\int_{X\times B} \xi\, \mathbf{Q}^*(dx, dy) = \int_{X\times B} p(x)\, \mathbf{Q}^*(dx, dy) = \int_X p(x)\, \nu(B|x)\, \mathbf{Q}(dx) =$$

$$= \int_X \nu(B|x)\, \mathbf{P}(dx) = \overline{\mathbf{P}}(B) = \int_B \bar{p}(y)\, \overline{\mathbf{Q}}(dy) = \int_{X\times B} \bar{p}(y)\, \mathbf{Q}^*(dx, dy) =$$

$$= \int_{X\times B} \eta\, \mathbf{Q}^*(dx, dy)\,,$$

wodurch (17) bewiesen ist.

Da aus der Konvexität von $f(u)$ bekanntlicherweise

$$\mathbf{E}(f(\xi)|S'_Y) \geqq f(\mathbf{E}(\xi|S'_Y)) \qquad [\mathbf{Q}^*] \tag{18}$$

folgt, (vgl. (29)), erhält man auf Grund von (17)

$$\mathcal{J}_f(\mathbf{P}, \mathbf{Q}) = \mathbf{E}(f(\xi)) = \mathbf{E}(\mathbf{E}(f(\xi)|S'_Y)) \geqq$$

$$\geqq \mathbf{E}(f(\mathbf{E}(\xi|S'_Y))) = \mathbf{E}(f(\eta)) = \mathcal{J}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})\,,$$

also (13). Für die Gleichheit ist notwendig und hinreichend, daß in (18) $\mathbf{Q}^*$-f. ü. die Gleichheit besteht, was aber für eine streng konvexe Funktion $f(u)$ dann und nur dann zutrifft, wenn — vgl. (32) —

$$\xi = \mathbf{E}(\xi|S'_Y) = \eta \qquad [\mathbf{Q}^*]$$

also (15) gilt, w. z. b. w.

Es sei bemerkt, daß für Funktionen $f(u)$ mit $\lim\limits_{u\to\infty} \dfrac{f(u)}{u} = +\infty$ die Voraussetzung $\mathbf{P} \ll \mathbf{Q}$ keine Beschränkung der Allgemeinheit bedeutet, da sonst $\mathcal{J}_f(\mathbf{P}, \mathbf{Q}) = +\infty$ wäre. Im allgemeinen Falle kann jedoch die f-Abweichung zweier Wahrscheinlichkeitsverteilungen auch dann endlich sein, wenn $\mathbf{P} \ll \mathbf{Q}$ nicht gilt. In diesem Falle können wir folgendermaßen vorgehen: man setze

$$X_0 = \{x : q(x) = 0\}\,, \qquad Y_0 = \{y : \bar{q}(y) = 0\} \tag{19}$$

und zerlege das Maß $\overline{\mathbf{P}}$ als $\overline{\mathbf{P}} = \overline{\mathbf{P}}_0 + \overline{\mathbf{P}}_1$,

$$\begin{aligned} \overline{\mathbf{P}}_0(B) &= \mathbf{P}^*(X_0 \times B) = \int_{X_0} \nu(B|x)\, \mathbf{P}(dx) \\ \overline{\mathbf{P}}_1(B) &= \mathbf{P}^*((X - X_0) \times B) = \int_{X-X_0} \nu(B|x)\, \mathbf{P}(dx)\,. \end{aligned} \tag{20}$$

Dementsprechend setzen wir

$$\bar{p}_0(y) = \frac{\overline{\mathbf{P}}_0(dy)}{\bar{\lambda}(dy)}\,, \qquad \bar{p}_1(y) = \frac{\overline{\mathbf{P}}_1(dy)}{\bar{\lambda}(dy)} \qquad (\bar{p}_0(y) + \bar{p}_1(y) = \bar{p}(y))\,. \tag{21}$$

Betrachte man wieder die Wahrscheinlichkeitsalgebra $(X\times Y,\ S_X\times S_Y,\ \mathbf{Q}^*)$ und definiere die Zufallsveränderliche $\xi=\xi(x,y)$ durch

$$\xi(x,y)=\begin{cases}\dfrac{p(x)}{q(x)} & \text{für} \quad x\notin X_0\\[2mm] 0 & \text{für} \quad x\in X_0 .\end{cases} \tag{22}$$

Es wird zweckmäßig sein, die bedingte Verteilung (im schwachen Sinne) von ξ bei gegebenem y, d. h. bezüglich der σ-Algebra $S'_Y=\{X\times B : B\in S_Y\}$ einzuführen. Diese bedingte Verteilung ist — vgl. [6], S. 29 — eine Schar von Wahrscheinlichkeitsmaßen $\mathbf{Q}_0(\,\cdot\mid x,y)$ $((x,y)\in X\times Y)$ auf den Borelschen Mengen der Zahlengerade, wobei $\mathbf{Q}_0(M\mid x,y)$ für jede Borelsche Menge M eine S'_Y-meßbare Funktion von (x,y) darstellt — man kann also $\mathbf{Q}_0(\,\cdot\mid y)$ statt $\mathbf{Q}_0(\,\cdot\mid x,y)$ schreiben — so daß für jede Borelsche Menge M

$$\mathbf{Q}^*\{\xi\in M\,|\,S'_Y\}=\mathbf{Q}_0(M\,|\,y) \tag{23}$$

und für jede Bairesche Funktion $f(u)$ mit $\mathbf{E}(|f(\xi)|)<+\infty$

$$\mathbf{E}(f(\xi)\,|\,S'_Y)=\int_0^\infty f(u)\,\mathbf{Q}_0(du\,|\,y) \tag{24}$$

gilt[6] (mit Wahrscheinlichkeit 1, d. h. $\mathbf{Q}^*$-fast überall; man kann statt dessen auch $\overline{\mathbf{Q}}$-f. ü. sagen, da es sich um bloß von y abhängenden (x,y)-Funktionen handelt).

Nun beweisen wir die Relation

$$\mathbf{E}(\xi\,|\,S'_Y)=\int_0^\infty u\,\mathbf{Q}_0(du\,|\,y)=\frac{\bar p_1(y)}{\bar q(y)}\qquad[\overline{\mathbf{Q}}]. \tag{25}$$

Die Zufallsveränderliche

$$\eta(x,y)=\begin{cases}\dfrac{\bar p_1(y)}{\bar q(y)} & \text{für} \quad y\notin Y_0\\[2mm] 0 & \text{für} \quad y\in Y_0\end{cases}$$

ist offenbar S'_Y-meßbar, ferner gilt wegen (22), (14) und (20) für jede Menge aus S'_Y, d. h. für jede Menge der Form $X\times B$ $(B\in S_Y)$

$$\int_{X\times B}\xi(x,y)\,\mathbf{Q}^*(dx,dy)=\int_{(X-X_0)\times B}\frac{p(x)}{q(x)}\,\mathbf{Q}^*(dx,dy)=\int_{X-X_0}\frac{p(x)}{q(x)}\,\nu(B\,|\,x)\,\mathbf{Q}(dx)=$$

$$=\int_{X-X_0}\nu(B\,|\,x)\,\mathbf{P}(dx)=\overline{\mathbf{P}}_1(B)=\overline{\mathbf{P}}_1(B-Y_0)=\int_{B-Y_0}\frac{\bar p_1(y)}{\bar q(y)}\,\overline{\mathbf{Q}}(dy)= \tag{26}$$

$$=\int_{X\times(B-Y_0)}\frac{\bar p_1(y)}{\bar q(y)}\,\mathbf{Q}^*(dx,dy)=\int_{X\times B}\eta(x,y)\,\mathbf{Q}^*(dx,dy).$$

---

[6] In (24) wurde darum nur $\int_0^\infty$ statt $\int_{-\infty}^{+\infty}$ geschrieben, weil die Funktion ξ nach ihrer Definition nichtnegativ ist.

Hier folgt die Gleichheit $\overline{\mathbf{P}}_1(B) = \overline{\mathbf{P}}_1(B - Y_0)$ daraus, daß wegen (19)

$$\int_{X-X_0} \nu(Y_0|x)\, q(x)\, \lambda(dx) = \int_X \nu(Y_0|x)\, \mathbf{Q}(dx) = \overline{\mathbf{Q}}(Y_0) = 0\,,$$

also $\nu(Y_0 \mid x) = 0$ λ-f. ü. auf $X - X_0$ gilt, woraus man

$$\overline{\mathbf{P}}_1(Y) = \int_{X-X_0} \nu(Y_0|x)\, \mathbf{P}(dx) = 0 \tag{27}$$

erhält. Da aus (26) $\mathbf{E}(\xi \mid S'_Y) = \eta(x, y)$ ($\mathbf{Q}^*$-f. ü., also $\overline{\mathbf{Q}}$-f. ü.) folgt, ist (25) bewiesen.

Nun gilt

$$\begin{aligned} \mathscr{I}_f(\mathbf{P}, \mathbf{Q}) &= \int_{X-X_0} q(x)\, f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) + \int_{X_0} q(x)\, f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) = \\ &= \mathbf{E}(f(\xi)) + \mathbf{P}(X_0) \lim_{u\to\infty} \frac{f(u)}{u}\,, \end{aligned} \tag{28}$$

wobei, wegen (24), (9) und (25)

$$\begin{aligned} \mathbf{E}(f(\xi)) &= \mathbf{E}(\mathbf{E}(f(\xi) \mid S'_Y)) = \mathbf{E}\left(\int_0^\infty f(u)\, \mathbf{Q}_0(du|y)\right) \geqq \mathbf{E}\left(f\left(\int_0^\infty u\, \mathbf{Q}_0(du|y)\right)\right) = \\ &= \mathbf{E}\left(f\left(\frac{\bar{p}_1(y)}{\bar{q}(y)}\right)\right) = \int_{Y-Y_0} f\left(\frac{\bar{p}_1(y)}{\bar{q}(y)}\right) \overline{\mathbf{Q}}(dy) = \int_Y \bar{q}(y)\, f\left(\frac{\bar{p}_1(y)}{\bar{q}(y)}\right) \bar{\lambda}(dy) \end{aligned} \tag{29}$$

gilt. (Die letzte Gleichheit folgt daraus, daß auf Y_0 wegen (27) auch $\bar{p}_1(y)$ verschwindet $[\bar{\lambda}]$.)

Ferner gilt nach (14), (20) und (3)

$$\begin{aligned} \mathbf{P}(X_0) \lim_{u\to\infty} \frac{f(u)}{u} &= \mathbf{P}^*(X_0 \times Y) \lim_{u\to\infty} \frac{f(u)}{u} = \\ &= \overline{\mathbf{P}}_0(Y) \lim_{u\to\infty} \frac{f(u)}{u} = \int_Y 0 \cdot f\left(\frac{\bar{p}_0(y)}{0}\right) \bar{\lambda}(dy)\,. \end{aligned} \tag{30}$$

Aus (28), (29) und (30) folgt — Gebrauch machend auch von Hilfssatz 3 und (21) —

$$\begin{aligned} \mathscr{I}_f(\mathbf{P}, \mathbf{Q}) &\geqq \int_Y \bar{q}(y)\, f\left(\frac{\bar{p}_1(y)}{\bar{q}(y)}\right) \bar{\lambda}(dy) + \int_Y 0 \cdot f\left(\frac{\bar{p}_0(y)}{0}\right) \bar{\lambda}(dy) \geqq \\ &\geqq \int_Y \bar{q}(y)\, f\left(\frac{\bar{p}_0(y) + \bar{p}_1(y)}{\bar{q}(y)}\right) \bar{\lambda}(dy) = \mathscr{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})\,, \end{aligned} \tag{31}$$

wodurch Satz 1 bewiesen ist.

Um auch die Ergänzung zu verifizieren, bemerken wir, daß man in (31) nach dem obigen Beweis dann und nur dann an beiden Stellen die Gleichheit statthat (vorausgesetzt $\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < +\infty$), wenn

$$\int_0^\infty f(u)\,\mathbf{Q}_0(du\,|\,y) = f\Big(\int_0^\infty u\,\mathbf{Q}_0(du\,|\,y)\Big) \quad [\overline{\mathbf{Q}}] \tag{32}$$

und

$$0 \cdot f\left(\frac{\bar{p}_0(y)}{0}\right) + \bar{q}(y)\,f\left(\frac{\bar{p}_1(y)}{\bar{q}(y)}\right) = \bar{q}(y)\,f\left(\frac{\bar{p}_0(y)+\bar{p}_1(y)}{\bar{q}(y)}\right) \qquad [\lambda] \tag{33}$$

gilt. Für ein streng konvexes $f(u)$ besteht (32) — nach Hilfssatz 1 und (25) — genau im Falle

$$\mathbf{Q}_0\left(\left\{\frac{\bar{p}_1(y)}{\bar{q}(y)}\right\}\Big|\,y\right) = 1 \qquad [\overline{\mathbf{Q}}]\,,$$

was nach (23) mit

$$\frac{p(x)}{q(x)} = \frac{\bar{p}_1(y)}{\bar{q}(y)} \qquad [\mathbf{Q}^*] \tag{15'}$$

gleichbedeutend ist.

Ferner ist (33) nach Hilfssatz 3 mit $\bar{p}_0(y)\,\bar{q}(y) = 0$ $[\lambda]$ also mit

$$\bar{p}_0(y) = 0 \qquad [\overline{\mathbf{Q}}] \tag{34}$$

äquivalent.

Aus (15') und (34) folgt bereits (15). Gilt umgekehrt (15), so ist $\xi(x, y)$ bei fast jedem y mit (bedingter) Wahrscheinlichkeit 1 gleich $\dfrac{\bar{p}(y)}{\bar{q}(y)}$, also sind fast alle Maße $\mathbf{Q}_0(\,\cdot\,|\,y)$ auf die Punkte $\dfrac{\bar{p}(y)}{\bar{q}(y)}$ konzentriert. Daraus folgt aber (32) und auch $\dfrac{\bar{p}(y)}{\bar{q}(y)} = \dfrac{\bar{p}_1(y)}{\bar{q}(y)}$ $[\overline{\mathbf{Q}}]$, also auch $\bar{p}(y) = \bar{p}_1(y)$ $[\overline{\mathbf{Q}}]$. Letztere Gleichung ist aber mit (34) gleichbedeutend, woraus (33) folgt.

Damit ist auch die Ergänzung vollständig bewiesen, da (15) und (16) offenbar äquivalent sind.

Wählt man alle Maße $\nu(\,\cdot\,|\,x)$ miteinander gleich, so ergibt sich aus dem bewiesenen Satz die Ungleichung

$$\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) \geqq f(1) \tag{35}$$

mit der Gleichheit — für streng konvexes f — genau im Falle $\mathbf{P} = \mathbf{Q}$. Dies läßt sich selbstverständlich auch direkt leicht verifizieren.

Es seien noch einige interessantere und weniger triviale Folgerungen des Satzes 1 erwähnt. Die anschauliche Deutung der folgenden Aussagen ist klar.

Folgerung 1. *Haben die Maße $\nu(\,\cdot\,|\,x)$ die Eigenschaft, daß $\nu(\,\cdot\,|\,x_1)$ und $\nu(\,\cdot\,|\,x_2)$ für keine $x_1 \in X$ und $x_2 \in X$ zueinander orthogonal sind* (diese Bedingung bedeutet, daß für den Kanal $(X, Y, \nu(\,\cdot\,|\,x))$ (kein Code — wohlauch der Länge 2 — mit Fehlerwahrscheinlichkeit 0 existiert), *so gilt die Gleichheit*

$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$ *für eine streng konvexe Funktion* $f(u)$ *mit* $\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < < + \infty$ *nur im Falle* $\mathbf{P} = \mathbf{Q}$.

In der Tat hat im Falle der Gleichheit die Menge

$$X' = \left\{x : \nu\left(\left\{y : \frac{\bar{p}(y)}{\bar{q}(y)} = \frac{p(x)}{q(x)}\right\} \middle| x\right) = 1\right\} \tag{36}$$

nach (16) das $\mathbf{Q}$-Maß 1. Wegen der vorausgesetzten Nichtorthogonalität der Maße $\nu(\,\cdot\,|\,x)$ gilt für jedes $x_0 \in X'$, $x \in X$

$$\nu\left(\left\{y : \frac{\bar{p}(y)}{\bar{q}(y)} = \frac{p(x_0)}{q(y_0)}\right\} \middle| x\right) > 0\,,$$

woraus und aus (36) für $x \in X'$ $\dfrac{p(x)}{q(x)} = \dfrac{p(x_0)}{q(x_0)}$ folgt. Dies bedeutet wegen $\mathbf{Q}(X') = 1$ und (15), daß

$$\frac{p(x)}{q(x)} = \frac{\bar{p}(y)}{\bar{q}(y)} = C \qquad [\mathbf{Q}^*]$$

d. h.

$$p(x) = Cq(x) \quad [\mathbf{Q}]\,, \quad \bar{p}(y) = C\bar{q}(y) \quad [\overline{\mathbf{Q}}] \tag{37}$$

gilt, woraus durch Integrieren

$$\mathbf{P}(X - X_0) = C\mathbf{Q}(X - X_0) = C, \quad \overline{\mathbf{P}}(Y - Y_0) = C\overline{\mathbf{Q}}(Y - Y_0) = C$$

folgt. Hieraus erhält man wegen (27) und (20)

$$\overline{\mathbf{P}}_1(Y - Y_0) = \overline{\mathbf{P}}_1(Y) = \mathbf{P}(X - X_0) = C\,,$$

also

$$C = \overline{\mathbf{P}}(Y - Y_0) = \overline{\mathbf{P}}_0(Y - Y_0) + \overline{\mathbf{P}}_1(Y - Y_0) = \int_{X_0} \nu(Y - Y_0\,|\,x)\,\mathbf{P}(dx) + C\,,$$

d. h.

$$\int_{X_0} \nu(Y - Y_0\,|\,x)\,\mathbf{P}(dx) = 0\,. \tag{38}$$

Da aber wegen $\int_X \nu(Y_0\,|\,x)\,\mathbf{Q}(dx) = \overline{\mathbf{Q}}(Y_0) = 0$ $\mathbf{Q}$-f. ü. $\nu(Y_0\,|\,x) =$ gilt, kann die Größe $\nu(Y - Y_0\,|\,x)$ — wegen der vorausgesetzten Nichtorthogonalität der Maße $\nu(\,\cdot\,|\,x)$ — für kein $x \in X$ verschwinden, so daß (38) mit $\mathbf{P}(X_0) = 0$ äquivalent ist. Dies bedeutet, daß die erste Gleichung in (37) auch λ-f. ü. besteht, ferner daß $C = \mathbf{P}(X - X_0) = 1$ ist. Mit dem erhaltenen Relation

$$p(x) = q(x) \quad [\lambda]$$

ist Folgerung 1 bewiesen.

Folgerung 2. *Läßt sich* Y *in der Form*

$$Y = \bigcup_{x \in X} B_x \quad (B_x \in S_Y, \quad B_{x_1} \cap B_{x_2} = 0 \quad \text{für} \quad x_1 \neq x_2)$$

mit $\nu(B_x\,|\,x) = 1$ *darstellen* (diese Bedingung bedeutet, daß $(X, Y, \nu(\,\cdot\,|\,x))$ einen Kanal ohne Verlust darstellt, also daß die Eingangssignale auf Grund der empfangenen Signale eindeutig bestimmbar sind), *so gilt* $\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$ *für je zwei Wahrscheinlichkeitsmaße* $\mathbf{P}$ *und* $\mathbf{Q}$.

Wähle man nämlich $\lambda^* = \mathbf{P}^* + \mathbf{Q}^*$; dann ist $\lambda = \mathbf{P} + \mathbf{Q}$, $\overline{\lambda} = \overline{\mathbf{P}} + \overline{\mathbf{Q}}$, also gilt für beliebiges $B \in S_Y$

$$\bar{\lambda}(B) = \int_X \nu(B \mid x)\, \lambda(dx) = \int_X \nu(B \cap B_x \mid x)\, \lambda(dx)\,.$$

Man verifiziert dann leicht, daß man für $\bar{p}(y)$ und $\bar{q}(y)$ $\bar{p}(y) = p(r(y))$, $\bar{q}(y) = q(r(y))$ wählen kann, wobei $r(y)$ durch $r(y) = x$ für $y \in B_x$ definiert ist. Daraus folgt für $q(x) > 0$

$$\nu\left(\left\{y : \frac{\bar{p}(y)}{\bar{q}(y)} = \frac{p(x)}{q(x)}\right\} \Big| \, x\right) \geqq \nu(B_x \mid x) = 1$$

w. z. b. w.

Folgerung 3. *Sind alle Maße* $\nu(\,\cdot \mid x)$ *auf je einen Punkt* Tx *konzentriert* —haben wir also einen Kanal ohne Geräusch, oder, was dasselbe ist, *eine meßbare Transformation* T *von* (X, S_X) *in* (Y, S_Y) —*so ergibt Satz* 1 *die Ungleichung* $\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\mathbf{P}T^{-1}, \mathbf{Q}T^{-1})$, *wobei für die Gleichheit notwendig und hinreichend ist* (bei streng konvexem $f(u)$ und endlichem $\mathcal{I}_f(\mathbf{P}T^{-1}, \mathbf{Q}T^{-1})$), daß die Likelihood-Quotienten $\dfrac{p(x)}{q(x)}$ und $\dfrac{\bar{p}(y)}{\bar{q}(y)}$ $\mathbf{Q}$-f. ü. übereinstimmen. Letztere Bedingung ist äquivalent mit der Existenz einer S_X-meßbaren Funktion $r(x)$ mit

$$p(x) = \bar{p}(Tx)\, r(x)\,, \qquad q(x) = \bar{q}(Tx)\, r(x) \qquad [\lambda],$$

was bedeutet, *daß* T *eine erschöpfende Schätzfunktion bezüglich der Maße* $\mathbf{P}$ *und* $\mathbf{Q}$ *darstellt*. Im Spezialfall $f(u) = u \log u$ erhalten wir also den wohlbekannten Satz von KULLBACK und LEIBLER [1] (vgl. auch [3]).

Folgerung 4. *Es sei* $F \subset S_X$ *eine* σ*-Algebra von Teilmengen von* X; betrachte man die Abbildung T von (X, S_X) auf (X, F) definiert durch $Tx = x$. Dann ist offenbar $\mathbf{P}T^{-1} = \mathbf{P}_F$, $\mathbf{Q}T^{-1} = \mathbf{Q}_F$, wobei $\mathbf{P}_F$ bzw. $\mathbf{Q}_F$ die Einschränkung von $\mathbf{P}$ bzw. $\mathbf{Q}$ auf F bezeichnet. Folgerung 3 ergibt also *die Ungleichung*

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\mathbf{P}_F, \mathbf{Q}_F) \tag{13'}$$

wobei die Gleichheit (bei streng konvexem f und endlichem $\mathcal{I}_f(\mathbf{P}_F, \mathbf{Q}_F)$) *dann und nur dann besteht, wenn* F *eine erschöpfende* σ*-Algebra bezüglich der beiden Wahrscheinlichkeitsmaßen* $\mathbf{P}$ *und* $\mathbf{Q}$ *darstellt.*[7]

Es sei noch bemerkt, daß auch umgekehrt, die Folgerung 4 den Satz 1 selbst zu Folge hat

Nämlich verifiziert man leicht, daß die f-Abweichung von $\mathbf{P}^*$ und $\mathbf{Q}^*$ immer gleich der f-Abweichung von $\mathbf{P}$ und $\mathbf{Q}$ ist, ferner daß $\overline{\mathbf{P}}$ und $\overline{\mathbf{Q}}$ als die Einschränkungen von $\mathbf{P}^*$ und $\mathbf{Q}^*$ auf S'_Y anzusehen sind (vgl. den Beweis von Satz 1). Wendet man also (13') anstatt X, S_X, $\mathbf{P}$, $\mathbf{Q}$ und F auf $X \times Y$, $S_X \times S_Y$, $\mathbf{P}^*$, $\mathbf{Q}^*$ und S'_Y an, so ergibt sich $\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \mathcal{I}_f(\mathbf{P}^*, \mathbf{Q}^*) \geqq \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$, also (13).

---

[7] Die erhaltene Behauptung ist für $f(u) = u \log u$ wohlbekannt, und wird gewöhnlich in einer dem Beweis des Satzes 1 im wesentlichen ähnlichen Weise bewiesen (vgl. z. B. [3] oder auch [7]), wobei eine Vereinfachung bedeutet, daß man sich auf den Fall $\mathbf{P} \ll \mathbf{Q}$ beschränken darf.

§ 3. Die Stabilitätsfrage der Ungleichung $\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$ und die Verallgemeinerung der Ungleichung für nicht normierbare Maße

Im vorigen Paragraphen wurde die Ungleichung

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) \tag{13}$$

bewiesen, sowie die Bedingung der Gleichheit betrachtet. Nun wenden wir uns dem Problem zu, ob was sich über die Maße $\mathbf{P}$ und $\mathbf{Q}$ behaupten läßt, wenn in (13) nur annäherungsweise die Gleichheit besteht, also — mit anschaulichem Ausdruck — wenn die Übertragung die Abweichung der Verteilungen der Signale der beiden Quellen nur wenig vermindert. Diese Frage ist nicht nur an sich interessant, sondern auch für die Anwendungen (vgl. § 4) wichtig.

Satz 2. *Ist die konvexe Funktion $f(u)$ $(0 < u < +\infty)$ zweimal differenzierbar und gilt*

$$\inf_{0<u<b} f''(u) > 0 \qquad \textit{für jedes} \quad 0 < b < +\infty \tag{39}$$

so kann zu jedem $\varkappa$ mit $0 < \varkappa < 1$ eine nur von f und $\varkappa$ abhängende positive Zahl δ derart angegeben werden, daß für beliebiges $0 < \varepsilon < \frac{1}{\varkappa}$ und für zwei beliebige Wahrscheinlichkeitsmaße $\mathbf{P}$, $\mathbf{Q}$ auf (X, S_X) mit endlichem $\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$ und

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) - \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < \delta\varepsilon^2 \tag{40}$$

die Abschätzung

$$\mathbf{Q}^*\left(\left\{(x, y) : \left|\frac{p(x)}{q(x)} - \frac{\overline{p}(y)}{\overline{q}(y)}\right| \leqq \varepsilon\right\}\right) \geqq 1 - \varkappa \tag{41}$$

gilt.

Bemerkung. Satz 2 bedeutet, daß sich die Abweichung der Verteilungen nur in demfalle in kleinem Maße vermindert, wenn die Likelihood-Quotienten $\frac{p(x)}{q(x)}$ und $\frac{\overline{p}(y)}{\overline{q}(y)}$ mit großer $\mathbf{Q}^*$-Wahrscheinlichkeit fast übereinstimmen.

Beweis. Wir benützen dieselbe Bezeichnungen wie im Beweis des Satzes 1. Man setze $\tau = \tau(\varkappa) = \inf\limits_{0<u<\frac{5}{\varkappa}} f''(u)$; dann ist nach Voraussetzung $\tau > 0$.

Bezeichnen wir mit N_0, N_1, N_2 die folgenden Mengen aus S_Y:

$$N_0 = \left\{y : \int_0^\infty u\,\mathbf{Q}_0(du \mid y) \neq \frac{\overline{p}_1(y)}{\overline{q}(y)}\right\} \tag{42}$$

$$N_1 = \left\{ y : \begin{array}{l} \displaystyle\int_0^\infty f(u)\,\mathbf{Q}_0(du \mid y) - f\left(\int_0^\infty u\,\mathbf{Q}_0(du \mid y)\right) \geqq \frac{\tau\varkappa}{16}\varepsilon^2 \quad \text{oder} \\ \displaystyle 0 \cdot f\left(\frac{p_0(y)}{0}\right) + \overline{q}(y)\, f\left(\frac{\overline{p}_1(y)}{\overline{q}(y)}\right) - \overline{q}(y)\, f\left(\frac{\overline{p}(y)}{\overline{q}(y)}\right) \geqq \frac{\tau\varkappa}{16}\varepsilon^2 \overline{q}(y) \end{array} \right\} \tag{43}$$

$$N_2 = \left\{y : \frac{\overline{p}(y)}{\overline{q}(y)} \geqq \frac{4}{\varkappa}\right\}. \tag{44}$$

Dann gilt nach (25)

$$\overline{\mathbf{Q}}(N_0) = 0\,; \tag{45}$$

nach dem Beweis von Satz 1 — vgl. (29) und (31) — folgt aus (40)

$$\overline{\mathbf{Q}}(N_1) \cdot \frac{\tau\varkappa}{16}\,\varepsilon^2 \leqq \mathcal{I}_f(\mathbf{P}, \mathbf{Q}) - \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < \delta\varepsilon^2$$

also

$$\overline{\mathbf{Q}}(N_1) < \frac{16\,\delta}{\tau\varkappa}, \tag{46}$$

ferner folgt aus (44)

$$\mathbf{Q}(N_2) = \int\limits_{N_2} \bar{q}(y)\,\bar{\lambda}(dy) \leqq \frac{\varkappa}{4} \int\limits_{N_2} \bar{p}_1(y)\,\bar{\lambda}(dy) = \frac{\varkappa}{4}\,\overline{\mathbf{P}}_1(N_2) \leqq \frac{\varkappa}{4}. \tag{47}$$

Nun besteht für $y \notin N_0 \cup N_1 \cup N_2$ erstens

$$\int\limits_0^\infty f(u)\,\mathbf{Q}_0(du \mid y) - f\left(\int\limits_0^\infty u\,\mathbf{Q}_0(du \mid y)\right) < \frac{\tau}{2}\left(\frac{\varepsilon}{2}\right)^2 \cdot \frac{\varkappa}{2},$$

wobei in der $\frac{\varepsilon}{2}$-Umgebung von $\int\limits_0^\infty u\,\mathbf{Q}_0(du \mid y) = \frac{\bar{p}_1(y)}{\bar{q}(y)}$ $f''(u) > \tau$ gilt, woraus man nach Hilfssatz 2

$$\mathbf{Q}_0\left(\left[\frac{\bar{p}_1(y)}{\bar{q}(y)} - \frac{\varepsilon}{2},\ \frac{\bar{p}_1(y)}{\bar{q}(y)} + \frac{\varepsilon}{2}\right] \Big| \, y\right) \geqq 1 - \frac{\varkappa}{2} \qquad (y \notin N_0 \cup N_1 \cup N_2) \tag{48}$$

erhält, und zweitens

$$0 \cdot f\left(\frac{\bar{p}_0(y)}{0}\right) + \bar{q}(y)\,f\left(\frac{\bar{p}_1(y)}{\bar{q}(y)}\right) - \bar{q}(y)\,f\left(\frac{\bar{p}_0(y) + \bar{p}_1(y)}{\bar{q}(y)}\right) < \frac{\tau}{2}\left(\frac{\varepsilon}{2}\right)^2 \cdot \frac{\varkappa}{2}\,\bar{q}(y),$$

wobei in der $\frac{\varepsilon}{2}$-Umgebung von $\frac{\bar{p}_0(y) + \bar{p}_1(y)}{\bar{q}(y)} = \frac{\bar{p}(y)}{\bar{q}(y)}$ $f''(u) > \tau$ gilt, woraus sich nach Hilfssatz 3

$$\frac{\bar{p}_0(y)}{\bar{q}(y)} \leqq \frac{\varepsilon}{2} \qquad (y \notin N_0 \cup N_1 \cup N_2) \tag{49}$$

ergibt. Aus (48) und (49) folgt — vgl. (23) —, daß für $y \notin N_0 \cup N_1 \cup N_2$ mit Wahrscheinlichkeit 1 (also $\mathbf{Q}$-f. ü.)

$$\mathbf{Q}^*\left(\left|\frac{p(x)}{q(x)} - \frac{\bar{p}(y)}{\bar{q}(y)}\right| \leqq \varepsilon \,\Big|\, S'_Y\right) \geqq 1 - \frac{\varkappa}{2}$$

gilt.

Daraus erhält man, mit Rücksicht auf (45), (46) und (47)

$$\mathbf{Q}^*\left(\left\{(x,y)\ :\ \left|\frac{p(x)}{q(x)}-\frac{\bar p(y)}{\bar q(y)}\right|\leqq\varepsilon\right\}\right)\geqq\left(1-\frac{\varkappa}{2}\right)\left(1-\frac{16\,\delta}{\tau\varkappa}-\frac{\varkappa}{4}\right). \tag{50}$$

Wählt man nun z. B. $\delta=\dfrac{\tau\varkappa^2}{64}$, so folgt aus (50) die Abschätzung (41) wodurch Satz 2 bewiesen ist.

Bemerkung. Dem obigen Beweis liegt die folgende Behauptung zugrunde: *Für jede Funktion f mit der Eigenschaft* (39) *kann man zu jedem $0<\varkappa<1$ eine nur von f und $\varkappa$ abhängende positive Zahl δ derart angeben, daß für jede Zufallsvariable ξ (mit endlichem Erwartungswert) und jede σ-Algebra $\mathscr{F}_1\subset\mathscr{F}$* (wobei $(\Omega,\mathscr{F},\mathbf{P})$ diejenige Wahrscheinlichkeitsalgebra bezeichnet, auf welcher ξ definiert ist) *aus*

$$\mathbf{E}(f(\xi))-\mathbf{E}\big(f(\mathbf{E}(\xi\,|\,\mathscr{F}_1))\big)<\delta\varepsilon^2 \tag{40'}$$

die Relation

$$\mathbf{P}(|\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1)|\leqq\varepsilon)\geqq 1-\varkappa \tag{41'}$$

folgt.

Der Beweis dieser Aussage ist im wesentlichen in demselben des Satzes 2 enthalten. Auf diese Tatsache hat mich Herr Professor A. RÉNYI aufmerksam gemacht, und auch darauf, daß sich die obige Behauptung im Spezialfall $f''(u)\geqq\tau>0$ $(0<u<+\infty)$ auch folgendermaßen beweisen läßt:

Nach der Taylorschen Formel gilt

$$\begin{aligned} f(\xi)-f(\mathbf{E}(\xi\,|\,\mathscr{F}_1)) &= f'(\mathbf{E}(\xi\,|\,\mathscr{F}_1))\,(\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))+\\ &+\frac{1}{2}f''\big(\mathbf{E}(\xi\,|\,\mathscr{F}_1)+\vartheta(\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))\big)\,(\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))^2\geqq\\ &\geqq f'(\mathbf{E}(\xi\,|\,\mathscr{F}_1))\,(\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))+\frac{\tau}{2}\,(\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))^2, \end{aligned}$$

woraus

$$\mathbf{E}(f(\xi)\,|\,\mathscr{F}_1)-f(\mathbf{E}(\xi\,|\,\mathscr{F}_1))\geqq\frac{\tau}{2}\,\mathbf{E}\big((\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))^2\,|\,\mathscr{F}_1\big).$$

und daraus

$$\mathbf{E}(f(\xi))-\mathbf{E}\big(f(\mathbf{E}(\xi\,|\,\mathscr{F}_1))\big)\geqq\frac{\tau}{2}\,\mathbf{E}\big((\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))^2\big)$$

folgt. Dies bedeutet, daß (40') die Ungleichung

$$\mathbf{E}\big((\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1))^2\big)\leqq\frac{2\,\delta\varepsilon^2}{\tau}$$

zu Folge hat, woraus sich nach der Tschebyscheffschen Ungleichung

$$\mathbf{P}(|\xi-\mathbf{E}(\xi\,|\,\mathscr{F}_1)|\geqq\varepsilon)\leqq\frac{2\,\delta}{\tau},$$

also (41') — wenn man $\delta=\dfrac{\tau}{2}\varkappa$ wählt — ergibt.

Es sei bemerkt, daß wenn statt (39) schwächer nur

$$(39') \qquad \inf_{a<u<b} f''(u) > 0 \quad \text{für jede} \quad 0 < a < b < +\infty$$

vorausgesetzt wird, so braucht die Behauptung des Satzes 2 nicht mehr gültig zu bleiben. Dies ist bereits der Fall für die Funktionen $f_\alpha(u) = u^\alpha$ $(\alpha > 2)$ — die nur (39') erfüllen, (39) aber nicht — wie es der folgende Beispiel zeigt.

Beispiel. Betrachten wir einen Übertragungskanal mit je drei Eingangs- und Ausgangssignale und mit dem Übergangswahrscheinlichkeit-Matrix

$$\begin{Vmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{Vmatrix}.$$

Mit anderen Worten, X bzw. Y bestehe aus je drei Punkten, die mit 0, 1, 2 bezeichnet werden können. S_X und S_Y bestehen aus allen Teilmengen von X bzw. Y und $\nu(\,\cdot\,|\,x)$ sei durch

$$(51) \qquad \begin{gathered} \nu(0|0) = 1, \quad \nu(1|0) = \nu(2|0) = 0, \quad \nu(0|1) = \nu(0|2) - 0, \\ \nu(1|1) = \nu(2|1) = \nu(1|2) = \nu(2|2) = \frac{1}{2} \end{gathered}$$

definiert. Die Wahrscheinlichkeitsmaße $\mathbf{P}$ und $\mathbf{Q}$ auf (X, S_X) seien durch

$$(52) \qquad \begin{aligned} &\mathbf{P}(0) = 1 - 2\varepsilon_1, \quad &&\mathbf{P}(1) = \varepsilon_1 + \varepsilon_2, \quad \mathbf{P}(2) = \varepsilon_1 - \varepsilon_2, \\ &\mathbf{Q}(0) = 2\varepsilon_1, &&\mathbf{Q}(1) = \mathbf{Q}(2) = \frac{1}{2} - \varepsilon_1 \end{aligned}$$

angegeben, wobei $\varepsilon_2 \ll \varepsilon_1 \ll 1$ vorausgesetzt wird. Dann gilt nach (51) und (52)

$$(53) \quad \overline{\mathbf{P}}(0) = 1 - 2\varepsilon_1, \ \ \overline{\mathbf{P}}(1) = \overline{\mathbf{P}}(2) = \varepsilon_1, \ \ \overline{\mathbf{Q}}(0) = 2\varepsilon_1, \ \ \overline{\mathbf{Q}}(1) = \overline{\mathbf{Q}}(2) = \frac{1}{2} - \varepsilon_1$$

also

$$\mathcal{I}_{f\alpha}(\mathbf{P}, \mathbf{Q}) = \mathcal{I}_\alpha(\mathbf{P}, \mathbf{Q}) = (2\varepsilon_1)^{1-\alpha}(1 - 2\varepsilon_1)^\alpha + \left(\frac{1}{2} - \varepsilon_1\right)^{1-\alpha}(\varepsilon_1 + \varepsilon_2)^\alpha + \\ + \left(\frac{1}{2} - \varepsilon_1\right)^{1-\alpha}(\varepsilon_1 - \varepsilon_2)^\alpha$$

$$\mathcal{I}_{f\alpha}(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) = \mathcal{I}_\alpha(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) = (2\varepsilon_1)^{1-\alpha}(1 - 2\varepsilon_1)^\alpha + 2\left(\frac{1}{2} - \varepsilon_1\right)^{1-\alpha}\varepsilon_2^\alpha$$

das heißt

$$\mathcal{I}_{f\alpha}(\mathbf{P}, \mathbf{Q}) - \mathcal{I}_{f\alpha}(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) = \left(\frac{1}{2} - \varepsilon_1\right)^{1-\alpha}[(\varepsilon_1 + \varepsilon_2)^\alpha + (\varepsilon_1 - \varepsilon_2)^\alpha - 2\varepsilon_1^\alpha] \sim \\ \sim 2^{\alpha-1}\varepsilon_1^\alpha \alpha(\alpha - 1)\left(\frac{\varepsilon_2}{\varepsilon_1}\right)^2 = 2^{\alpha-1}\alpha(\alpha - 1)\varepsilon_1^{\alpha-2}\varepsilon_2^2 .$$

Ist nun $\alpha > 2$ und ist δ eine beliebig kleine positive Zahl, so folgt aus (54) — indem man die positive Zahl ε_1 so wählt, daß $2^{\alpha-1}\alpha(\alpha-1)\varepsilon_1^{\alpha-2} \ll \delta$ gelte, und $\varepsilon = \varepsilon_2$ setzt — daß zu beliebig kleinem $\delta > 0$ ein beliebig kleines $\varepsilon > 0$ und zwei Maße $\overline{\mathbf{P}}$ und $\overline{\mathbf{Q}}$ mit $\mathcal{I}_{f\alpha}(\mathbf{P}, \mathbf{Q}) - \mathcal{I}_{f\alpha}(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < \delta\varepsilon^2$ derart angegeben werden können, daß die Menge $\left\{(x, y) : \left|\frac{p(x)}{q(x)} - \frac{\bar{p}(y)}{\bar{q}(y)}\right| \leqq \varepsilon\right\}$ nicht nur keine Menge von $\mathbf{Q}^*$-Maß nahe zu Eins, sondern — da sie nach (52) und (53) aus dem einzigen Punkt (0, 0) besteht — gerade eine Menge von beliebig kleinem $\mathbf{Q}^*$-Maß darstellt.

Obwohl — wie eben gezeigt wurde — aus der Voraussetzung (39') die Behauptung des Satzes 2 nicht folgt, eine schwächere Behauptung kann auch auf Grund von (39') bewiesen werden. Es gilt nämlich der folgende

Satz 2'. *Ist die konvexe Funktion $f(u)$ $(0 < u < +\infty)$ zweimal differenzierbar und gilt*

$$(39) \qquad \inf_{a<u<b} f''(u) > 0 \textit{ für jede } 0 < a < b < +\infty,$$

so kann zu jedem $\varkappa$ mit $0 < \varkappa < 1$ und $\varepsilon' > 0$ eine von f, $\varkappa$ und ε' abhängende positive Zahl δ' derart angegeben werden, daß aus dem Erfülltsein von

$$(55) \qquad \mathcal{I}_f(\mathbf{P}, \mathbf{Q}) - \mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < \delta' \qquad (\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) < +\infty)$$

die Ungleichung

$$(56) \qquad \mathbf{Q}^*\left(\left\{(x, y) : \left|\frac{p(x)}{q(x)} - \frac{\bar{p}(y)}{\bar{q}(y)}\right| \leqq \varepsilon'\right\}\right) \geqq 1 - \varkappa$$

folgt.

Beweis. Man kann das im Beweis von Satz 2 benützte Verfahren anwenden. Es seien δ und ε (vorläufig) beliebige positive Zahlen und setzen wir voraus, daß (40) gültig ist. Wir setzen

$$\tau = \inf_{\frac{\varepsilon}{2} < u < \frac{5}{\varkappa}} f''(u),$$

sonst werden die Bezeichnungen des Beweises von Satz 2 unverändert benützt.

Für $y \in Y - (N_0 \cup N_1 \cup N_2)$ unterscheiden wir die zwei Fälle $\frac{\bar{p}_1(y)}{\bar{q}(y)} > \varepsilon$ und $\frac{\bar{p}_1(y)}{\bar{q}(y)} \leqq \varepsilon$. Im ersten Falle gilt in der $\frac{\varepsilon}{2}$- Umgebung von $\int_0^\infty u\, \mathbf{Q}_0(du|y) = \frac{\bar{p}_1(y)}{\bar{q}(y)}$ nach der Definition von τ $f''(u) \geqq \tau$, so daß Hilfssatz 2 anwendbar ist, wodurch sich die Ungleichung (48) ergibt. Ähnlicherweise erhält man — benützend Hilfssatz 3 — auch die Ungleichung (49).

Im zweiten Falle ist Hilfssatz 2 nicht anwendbar. Doch folgt aus $\int_0^\infty u\, \mathbf{Q}_0(du|y) = \frac{\bar{p}_1(y)}{\bar{q}(y)} \leqq \varepsilon$ die Ungleichung

$$\mathbf{Q}_0\left(\left[0, \frac{\varepsilon'}{2}\right] \Big| y\right) \geqq 1 - \frac{2\varepsilon}{\varepsilon'}$$

und für $\varepsilon < \frac{\varepsilon'}{2}$ um so mehr

$$\mathbf{Q}_0\left(\left[\frac{\bar{p}_1(y)}{\bar{q}(y)} - \frac{\varepsilon'}{2},\ \frac{\bar{p}_1(y)}{\bar{q}(y)} + \frac{\varepsilon'}{2}\right] \middle|\ y\right) \geqq 1 - \frac{2\varepsilon}{\varepsilon'}. \tag{57}$$

Was nun die Ungleichung (49) anbetrifft, im Falle $\frac{\bar{p}(y)}{\bar{q}(y)} > \varepsilon$ folgt ihr Erfülltsein wieder aus Hilfssatz 3; im Falle $\frac{\bar{p}(y)}{\bar{q}(y)} \leqq \varepsilon$ wird nur das triviale

$$\frac{\bar{p}_0(y)}{\bar{q}(y)} \leqq \varepsilon \tag{58}$$

benützt.

Wählt man nun $\varepsilon < \frac{\varkappa}{4}\varepsilon'$, so gelten nach den obigen (vgl. (48), (49), (57), (58)) für $y \in Y - (N_0 \cup N_1 \cup N_2)$ jedenfalls die Ungleichungen

$$\mathbf{Q}_0\left(\left[\frac{\bar{p}_1(y)}{\bar{q}(y)} - \frac{\varepsilon'}{2}, \frac{\bar{p}_1(y)}{\bar{q}(y)} + \frac{\varepsilon'}{2}\right] \middle| y\right) \geqq 1 - \frac{\varkappa}{2}, \tag{48'}$$

$$\frac{\bar{p}_0(y)}{\bar{q}(y)} \leqq \frac{\varepsilon'}{2}. \tag{49'}$$

Diese sind mit (48) und (49) vollständig analogisch und haben — bei geeigneter Wahl von δ, z. B. bei $\delta = \frac{\tau\varkappa^2}{64}$ — die Beziehung (56) zu Folge, ebenso wie (48) und (49) die Beziehung (50), also auch (41) zu Folge hatte. Dadurch ist Satz 2' bereits bewiesen, bloß soll man $\delta' = \delta\varepsilon^2$ setzen.

Bemerkung zum Satz 2'. Aus den obigen ist ersichtlich, daß bei der Voraussetzung (39') die Behauptung des Satzes 2 im Allgemeinen darum nicht gilt, weil für diejenige $y \in Y$, für welche $\frac{\bar{p}_1(y)}{\bar{q}(y)}$ klein ist, der Hilfssatz 2 nicht angewendet werden kann.

Diese Schwierigkeit kommt aber nicht vor, wenn für $\bar{\mathbf{Q}}$-fast alle $y \in Y$ etwa

$$\frac{\bar{p}(y)}{\bar{q}(y)} \geqq \varepsilon_0 > 0 \tag{59}$$

gilt. Dann folgt nämlich aus (49) für $y \in N_0 \cup N_1 \cup N_2$ auch $\frac{\bar{p}_1(y)}{\bar{q}(y)} \geqq \frac{2}{3}\varepsilon_0$ — angenommen $\varepsilon \leqq \frac{2}{3}\varepsilon_0$ — also läßt sich der Beweis von Satz 2 für $\varepsilon \leqq \frac{2}{3}\varepsilon_0$ — mit $\tau = \inf\limits_{\frac{\varepsilon_0}{3} < u < \frac{5}{\varkappa}} f''(u)$ — wörtlich durchführen.

Eine für das Bestehen von (59) hinreichende Bedingung ist z. B.

$$\frac{\nu(B|x_1)}{\nu(B|x_2)} \geqq \varepsilon_0 \quad \text{für jede} \quad x_1 \in X,\ x_2 \in X \quad \text{und} \quad B \in S_Y. \tag{60}$$

Die Bedingung (60) läßt sich auch so formulieren, daß die Maße $\nu(\cdot \mid x)$ miteinander äquivalent sind und für ihre Dichtefunktionen $h(y \mid x)$ — bezüglich eines dominierenden Maßes auf (Y, S_Y) — bei jedem $x_1 \in X$, $x_2 \in X$ fast überall $\frac{h(y|x_1)}{h(y|x_2)} \geqq \varepsilon_0$ gilt.

Zum Schluß des vorliegenden Paragraphen möchte ich die Frage der Verallgemeinerung des Satzes 1 für nicht normierbare Maße betrachten.

Erstens bemerke man, daß die Mehrzahl der in den Paragraphen 1 und 2 angeführten Bezeichnungen auch dann sinnvoll sind, wenn $\mathbf{P}$ und $\mathbf{Q}$ statt Wahrscheinlichkeitsmaße beliebige σ-endliche Maße bedeuten (auf die möglichen Ausnahmen wird unten hingewiesen); daher werden wir sie auch in diesem allgemeinen Falle benützen.

Es ist klar, daß Satz 1 und seine Ergänzung auch dann gültig bleiben, wenn $\mathbf{P}$ kein Wahrscheinlichkeitsmaß, sondern ein beliebiges endliches Maß auf (X, S_X) darstellt, nämlich wurde im Bewies kein Gebrauch von $\mathbf{P}(X) = 1$ gemacht. Ebenso könnte auch $\mathbf{Q}$ ein beliebiges endliches Maß bedeuten, bloß sollte man in diesem Falle im Beweis die Maße $\frac{1}{\mathbf{Q}(X)}\mathbf{Q}$, $\frac{1}{\mathbf{P}(X)}\mathbf{P}$ benützen, wobei $\frac{1}{\mathbf{Q}(X)}\mathbf{Q}$ bereits ein Wahrscheinlichkeitsmaß darstellt. Will man Satz 1 auch für nicht normierbares (allerdings σ-endliches) $\mathbf{P}$ verallgemeinern, so betrifft man die Schwierigkeit, daß das Definitionsintegral (4) von $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$ nicht mehr unbedingt sinnvoll, sowie $\overline{\mathbf{P}}$ nicht unbedingt σ-endlich zu sein braucht. Ist aber mit $\mathbf{P}$ auch $\overline{\mathbf{P}}$ σ-endlich und sind $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$ und $\mathcal{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$ beide sinnvoll, so läßt sich der Beweis von Satz 1 und seiner Ergänzung auch für nicht normierbares $\mathbf{P}$ wörtlich durchführen (falls $\mathbf{Q}(X) \neq 1$ aber doch endlich ist, so soll man $\mathbf{P}$ und $\mathbf{Q}$ erstens durch $\frac{1}{\mathbf{Q}(X)}\mathbf{Q}$ und $\frac{1}{\mathbf{Q}(X)}\mathbf{P}$ ersetzen). Ist schließlich auch $\mathbf{Q}$ nur σ-endlich, so kann man folgendermaßen vorgehen (vorausgesetzt, daß auch $\overline{\mathbf{P}}$ und $\overline{\mathbf{Q}}$ σ-endlich sind):

Man stelle die Menge X in der Form $X = \bigcup_{i=0}^{\infty} X_i$ ($X_i \in S_X$, $X_i \cap X_j = 0$ für $i \neq j$) mit $\mathbf{Q}(X_i) < +\infty$ dar,[8] und setze

$$\mathbf{P}_i(A) = \mathbf{P}(A \cap X_i), \quad \mathbf{Q}_i(A) = \mathbf{Q}(A \cap X_i) \quad (A \in S_X\,;\ i = 0, 1, \ldots) \tag{61}$$

$$\begin{aligned} \overline{\mathbf{P}}_i(B) &= \int_X \nu(B|x)\,\mathbf{P}_i(dx) = \int_{X_i} \nu(B|x)\,\mathbf{P}(dx) \\ \overline{\mathbf{Q}}_i(B) &= \int_X \nu(B|x)\,\mathbf{Q}_i(dx) = \int_{X_i} \nu(B|x)\,\mathbf{Q}(dx) \end{aligned} \quad (B \in S_Y\,;\ i = 0, 1, \ldots) \tag{62}$$

$$\bar{p}_i(y) = \frac{\overline{\mathbf{P}}_i(dx)}{\bar{\lambda}(dx)}, \quad \bar{q}_i(y) = \frac{\overline{\mathbf{Q}}_i(dx)}{\bar{\lambda}(dx)} \qquad (i = 0, 1, \ldots). \tag{63}$$

[8] Wobei X_0 die durch (19) definierte Menge bedeutet (also gilt $\mathbf{Q}(X_0) = 0$).

Dann erhält man durch Anwendung der im Beweis des Satzes 1 benützten Methode — s. (28) und (29) — auf die Maße $\frac{1}{\mathbf{Q}(X_i)}\,\mathbf{Q}_i$ und $\frac{1}{\mathbf{P}(X_i)}\,\mathbf{P}_i$

$$\int\limits_{X_i} q(x)\, f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) \geqq \int\limits_{Y} \bar{q}_i(y)\, f\left(\frac{\bar{p}_i(y)}{\bar{q}_i(y)}\right) \bar{\lambda}(dy) \qquad (i = 1, 2, \ldots), \tag{64}$$

ferner gilt

$$\int\limits_{X_0} q(x)\, f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) = \mathbf{P}(X_0) \lim_{u\to\infty} \frac{f(u)}{u} = \bar{\mathbf{P}}(Y_0) \lim_{u\to\infty} \frac{f(u)}{u} = \int\limits_{Y} 0 \cdot f\left(\frac{\bar{p}_0(y)}{0}\right) \bar{\lambda}(dy)\,, \tag{65}$$

also besteht (64) auch für $i = 0$ (mit der Gleichheit).

Nun gilt nach Hilfssatz 1, angewandt für das diskrete Wahrscheinlichkeitsmaß definiert durch

$$\mu(\{u\}) = \sum_{i\,:\,\frac{\bar{p}_i(y)}{\bar{q}(y)} = u} \frac{\bar{q}_i(y)}{\bar{q}(y)}\,,$$

die Ungleichung

$$\sum_{i=1}^{\infty} \bar{q}(y)\, f\left(\frac{\bar{p}_i(y)}{\bar{q}_i(y)}\right) \geqq \bar{q}(y)\, f\left(\frac{\bar{p}(y) - \bar{p}_0(y)}{\bar{q}(y)}\right) \tag{66}$$

(es wurden auch die aus (61), (62), (63) folgenden Relationen $\sum_{i=0}^{\infty} \bar{p}_i(y) = \bar{p}(y)$, $\sum_{i=0}^{\infty} \bar{q}_i(y) = \bar{q}(y)$, $\bar{q}_0(y) = 0$ benutzt), ferner gilt nach Hilfssatz 2

$$0 \cdot f\left(\frac{\bar{p}_0(y)}{0}\right) + \bar{q}(y)\, f\left(\frac{\bar{p}(y) - \bar{p}_0(y)}{\bar{q}(y)}\right) \geqq \bar{q}(y)\, f\left(\frac{\bar{p}(y)}{\bar{q}(y)}\right). \tag{67}$$

Aus (64), (65), (66), (67) ergibt sich sofort

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) = \int\limits_{X} q(x)\, f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) = \sum_{i=0}^{\infty} \int\limits_{X_i} q(x)\, f\left(\frac{p(x)}{q(x)}\right) \lambda(dx) \geqq$$

$$\geqq \int\limits_{Y} \bar{q}(y)\, f\left(\frac{\bar{p}(y)}{\bar{q}(y)}\right) \bar{\lambda}(dy) = \mathcal{I}_f(\bar{\mathbf{P}}, \bar{\mathbf{Q}})\,,$$

vorausgesetzt, daß beide Integrale sinnvoll sind. Man erhält in dieser Weise die folgende Verallgemeinerung des Satzes 1:

Satz 3. *Sind* $\mathbf{P}$ *und* $\mathbf{Q}$ *beliebige σ-endliche Maße auf* (X, S_X) *mit endlichem* $\mathcal{I}_f(\mathbf{P}, \mathbf{Q})$, *sind ferner auch* $\bar{\mathbf{P}}$ *und* $\bar{\mathbf{Q}}$ *σ-endlich, so ist auch* $\mathcal{I}_f(\bar{\mathbf{P}}, \bar{\mathbf{Q}})$ *sinnvoll, und es gilt*

$$\mathcal{I}_f(\mathbf{P}, \mathbf{Q}) \geqq \mathcal{I}_f(\bar{\mathbf{P}}, \bar{\mathbf{Q}})\,,$$

wobei die Gleichheit für ein streng konvexes $f(u)$ *dann und nur dann gilt, wenn* (15) *(bzw.* (16)*) besteht.*

§ 4. Eine Anwendung

Betrachten wir eine homogene Markoffsche Kette mit den m-stufigen Übergangswahrscheinlichkeiten $p_{ij}^{(m)}$.

Ist die Wahrscheinlichkeitsverteilung der Anfangszustände $\mathbf{P}^{(1)} = (p_1^{(1)}, p_2^{(1)}, \ldots)$ bzw. $\mathbf{Q}^{(1)} = (q_1^{(1)}, q_2^{(1)}, \ldots)$ so bezeichne $p_k^{(n)}$ bzw. $q_k^{(n)}$ die Wahrscheinlichkeit des k-ten Zustandes im Zeitpunkt n, und man setze $\mathbf{P}^{(n)} = (p_1^{(n)}, p_2^{(n)}, \ldots)$, $\mathbf{Q}^{(n)} = (q_1^{(n)}, q_2^{(n)}, \ldots)$. Dann bestehen offenbar die Gleichungen

$$p_k^{(n+m)} = \sum_i p_i^{(n)} p_{ik}^{(m)}, \qquad q_k^{(n+m)} = \sum_i q_i^{(n)} p_{ik}^{(m)}, \tag{68}$$

die bedeuten, daß $\mathbf{P}^{(n+m)}$ bzw. $\mathbf{Q}^{(n+m)}$ (für beliebige natürliche Zahlen n und m) auffaßbar sind als die Verteilungen bestimmter, durch einen Kanal mit den Übergangswahrscheinlichkeiten $p_{ik}^{(m)}$ übertragener Signale, deren ursprüngliche Verteilung $\mathbf{P}$ bzw. $\mathbf{Q}$ war.

Es gilt also nach Satz 1 für eine beliebige konvexe Funktion $f(u)$

$$\mathcal{I}_f(\mathbf{P}^{(n)}, \mathbf{Q}^{(n)}) \geqq \mathcal{I}_f(\mathbf{P}^{(n+m)}, \mathbf{Q}^{(n+m)}). \tag{69}$$

Ist $\mathcal{I}_f(\mathbf{P}^{(1)}, \mathbf{Q}^{(1)}) < +\infty$, so folgt aus (69) (mit der Wahl $m = 1$) und aus (35), daß die Größen $\mathcal{I}_f(\mathbf{P}^{(n)}, \mathbf{Q}^{(n)})$ für $n \to \infty$ einen endlichen Grenzwert haben, und infolge dessen

$$\mathcal{I}_f(\mathbf{P}^{(n)}, \mathbf{Q}^{(n)}) - \mathcal{I}_f(\mathbf{P}^{(n+m)}, \mathbf{Q}^{(n+m)}) \to 0 \qquad \text{für} \quad n \to \infty \tag{70}$$

— gleichmäßig in m — gilt.

Aus (70) folgt nach Satz 2' — falls die Funktion $f(u)$ die Voraussetzungen dieses Satzes erfüllt — die Beziehung

$$\lim_{n\to\infty} \sum_{(i,j)\in N_{n,m}^{\varepsilon}} q_i^{(n)} p_{ij}^{(m)} = 1, \qquad N_{n,m}^{\varepsilon} = \left\{(i,j) : \left| \frac{p_i^{(n)}}{q_i^{(n)}} - \frac{p_j^{(n+m)}}{q_j^{(n+m)}} \right| \leqq \varepsilon \right\} \tag{71}$$

für jedes $\varepsilon > 0$.

Als Anwendung der Ergebnisse der vorliegenden Arbeit wollen wir nun auf Grund von (71) den Markoffschen Satz beweisen, daß jede homogene Markoffsche Kette mit endlich vielen möglichen Zuständen N, für die es ein m_0 und j_0 mit $p_{ij_0}^{(m_0)} > 0$ $(i = 1, 2, \ldots, N)$ gibt, ergodisch ist.

Wählen wir also eine beliebige in $(0, +\infty)$ definierte zweimal differenzierbare Funktion $f(u)$ mit

$$\inf_{a<u<b} f''(u) > 0 \quad \text{für jede} \quad 0 < a < b < +\infty \tag{39'}$$

und $f(0) = \lim_{u\to+0} f(u) < +\infty$. Wählen wir ferner etwa $\mathbf{Q}^{(1)} = \left(\frac{1}{N}, \frac{1}{N}, \ldots \frac{1}{N}\right)$. Dann ist die Endlichkeit von $\mathcal{I}_f(\mathbf{P}^{(1)}, \mathbf{Q}^{(1)})$ für beliebige Anfangsverteilung $\mathbf{P}^{(1)}$ gesichert, also gilt — nach den oben Gesagten — die Beziehung (71).

Setzen wir $K_n(\varepsilon) = \left\{ i : q_i^{(n)} \geqq \frac{\varepsilon}{N} \right\}$ und

$$d = \min_{1\leqq i\leqq N} p_{ij_0}^{(m_0)} > 0. \tag{72}$$

Dann muß nach (71) für jedes $i \in K_n(\varepsilon)$ und genügend großes n $(i, j_0) \in N^{\varepsilon}_{n,m_0}$, d. h.

$$|p_i^{(n)} q_{j_0}^{(n+m_0)} - q_i^{(n)} p_{j_0}^{(n+m_0)}| \leqq \varepsilon q_i^{(n)} q_{j_0}^{(n+m_0)} \qquad \text{für} \quad i \in K_n(\varepsilon) \tag{73}$$

gelten, woraus durch Summieren

$$\left|q_{j_0}^{(n+m_0)} \sum_{i \in K_n(\varepsilon)} p_i^{(n)} - p_{j_0}^{(n+m_0)} \sum_{i \in K_n(\varepsilon)} q_i^{(n)}\right| \leqq \varepsilon q_{j_0}^{(n+m_0)} \sum_{i \in K_n(\varepsilon)} q_i^{(n)} \leqq \varepsilon q_{j_0}^{(n+m_0)} \tag{74}$$

folgt. Hier besteht nach der Definition von $K_n(\varepsilon)$

$$\sum_{i \in K_n(\varepsilon)} q_i^{(n)} > 1 - \varepsilon\,; \tag{75}$$

um auch $\sum_{i \in K_n(\varepsilon)} p_i^{(n)}$ abschätzen zu können, multiplizieren wir (73) durch $p_{ij_0}^{(m_0)}$ und summieren nach i; dadurch ergibt sich

$$\left|q_{j_0}^{(n+m_0)} \sum_{i \in K_n(\varepsilon)} p_i^{(n)} p_{ij_0}^{(m_0)} - p_{j_0}^{(n+m_0)} \sum_{i \in K_n(\varepsilon)} q_i^{(n)} p_{ij_0}^{(m_0)}\right| \leqq$$
$$\leqq \varepsilon q_{j_0}^{(n+m_0)} \sum_{i \in K_n(\varepsilon)} q_i^{(n)} p_{ij_0}^{(m_0)} \leqq \varepsilon (q_{j_0}^{(n+m_0)})^2$$

und hieraus — benützend (68) —

$$\left|q_{j_0}^{(n+m_0)} \sum_{i \notin K_n(\varepsilon)} p_i^{(n)} p_{ij_0}^{(m_0)} - p_{j_0}^{(n+m_0)} \sum_{i \notin K_n(\varepsilon)} q_i^{(n)} p_{ij_0}^{(n+m_0)}\right| \leqq \varepsilon (q_{j_0}^{(n+m_0)})^2. \tag{76}$$

Aus (76) folgt nach der Definition von $K_n(\varepsilon)$ und nach (62)

$$\sum_{i \notin K_n(\varepsilon)} p_i^{(n)} \leqq \frac{\sum_{i \notin K_n(\varepsilon)} p_i^{(n)} p_{ij_0}^{(m_0)}}{d} \leqq \left(\frac{p_{j_0}^{(n+m_0)}}{q_{j_0}^{(n+m_0)}} + q_{j_0}^{(n+m_0)}\right) \frac{\varepsilon}{d}\,. \tag{77}$$

Berücksichtigt man (75) und (77), so erhält man aus (74)

$$|q_{j_0}^{(n+m_0)} - p_{j_0}^{(n+m_0)}| \leqq \varepsilon q_{j_0}^{(n+m_0)} + [p_{j_0}^{(n+m_0)} + (q_{j_0}^{(n+m_0)})^2] \frac{\varepsilon}{d} + \tag{78}$$
$$+ \varepsilon p_{j_0}^{(n+m_0)} \leqq 2\left(1 + \frac{1}{d}\right) \varepsilon\,.$$

Auf Grund von (78) und (73) sieht man gleich, indem man auch die aus (68) und (72) folgende Beziehung $q_{j_0}^{(n)} \geqq d$ berücksichtigt, daß $|p_i^{(n)} - q_i^{(n)}|$ bei genügend großem n für jedes $i \in K_n(\varepsilon)$ „klein" sein muß. Dasselbe gilt aber offenbar — vgl. (77) — auch für $i \notin K_n(\varepsilon)$. Dadurch ist die Limesrelation

$$\lim_{n \to \infty} |p_i^{(n)} - q_i^{(n)}| = 0 \qquad (i = 1, 2, \ldots, N) \tag{79}$$

für $\mathbf{Q}^{(1)} = \left(\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}\right)$ und beliebiges $\mathbf{P}^{(1)}$ bewiesen, woraus aber ihre Gültigkeit für beliebige $\mathbf{P}^{(1)}$ und $\mathbf{Q}^{(1)}$ folgt.

Es sei nun $\mathbf{P}^{(1)}$ eine beliebige Anfangsverteilung und setzen wir $\mathbf{Q}^{(1)} = \mathbf{P}^{(2)}$; dann gilt $\mathbf{Q}^{(n)} = \mathbf{P}^{(n+1)}$ und (79) ergibt

$$\lim_{n \to \infty} |p_i^{(n)} - p_i^{(n+1)}| = 0 \qquad (i = 1, 2, \ldots, N)\,. \tag{80}$$

Wählen wir also eine Folge $n_1, n_2, \ldots$ von natürlichen Zahlen derart, daß jede der Folgen $\{p_i^{n_k}\}_{k=1}^{\infty}$ $(i = 1, 2, \ldots, N)$ konvergent ist: $p_i^{(n_k)} \to p$ $(i = 1, 2, \ldots, N)$, so muß $\mathbf{P} = (p_1, p_2, \ldots, p_N)$ nach (80) eine stationäre Verteilung der betrachteten Markoffschen Kette sein, und aus (79) folgt, indem man nun $\mathbf{P}^{(1)} = \mathbf{P}$ setzt, daß die Verteilung $\mathbf{Q}^{(n)}$ für $n \to \infty$ bei beliebiger Anfangsverteilung $\mathbf{Q}^{(1)}$ gegen diese stationäre Verteilung $\mathbf{P}$ strebt.

Setzt man insbesondere $q_j^{(1)} = \begin{cases} 1 & \text{für } j = i \\ 0 & \text{sonst} \end{cases}$

so ergibt sich

$$\lim_{n \to \infty} p_{ik}^{(n)} = p_k \qquad \text{für jede } i \text{ und } k, \tag{81}$$

also die Ergodizität der betrachteten Markoffschen Kette.

Bemerkung. Die Homogenität der Markoffschen Kette wurde erst beim Beweis der Existenz einer stationären Verteilung benutzt. *Also ist die Beziehung* (79) *auch für jede inhomogene Markoffsche Kette bewiesen, für welche die Voraussetzung* (72) *gültig ist* (wobei nun m_0 und j_0 sowie die Übergangswahrscheinlichkeiten $p_{ij}^{(m)} = p_{ij}^{(m)}(n)$ selbst von dem Ausgangszeitpunkt n abhängen dürfen, die Zahlen $d(n) = \inf_{1 \leq i \leq N} p_{ij_0}^{(m_n)}(n)$ aber eine positive untere Grenze haben sollen.)

Führt man den obigen Beweis mit der Wahl $f(u) = u \log u$ durch, so kann sie als ein rein informationstheoretischer Beweis angesehen werden, indem dann im Beweis nur Eigenschaften der relativen Information (von Ordnung 1) $I(\mathbf{P} \,\|\, \mathbf{Q})$ benützt werden.

Die oben benützte Methode scheint nicht nur für den Beweis der Ergodizität von Markoffschen Ketten, sondern auch für den Beweis von anderen Grenzverteilungssätzen geeignet zu sein. Da die Ungleichung (13) nicht nur für Wahrscheinlichkeitsmaße, sondern auch für bestimmte σ-endliche Maße gültig ist (vgl. Satz 3) kann man erwarten, daß mit Hilfe der in diesem Paragraphen angewandten Methode — nach entsprechender Verallgemeinerung von Satz 2 bzw. 2' — auch manche Sätze bezüglich nicht normierbaren bedingten Verteilungen (vgl. [8]) beweisbar werden.

(Eingegangen: 16. Januar, 1963.)

LITERATURVERZEICHNIS

[1] KULLBACK, S.—LEIBLER, R. A.: „On information and sufficiency". *Annals of Mathematical Statistics* **22** (1951) 79—86.

[2] RÉNYI, A.: „On measures of entropy and information." *Proceedings of the 4th Berkeley Symposium on Probability and Statistics, I*, Berkeley, 1960, 547—561.

[3] PÉREZ, A.: „Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales". *Transactions of the First Prague Conference*, Prague, 1957, 183—208.

[4] ПИНСКЕР, М. С.: *Информация и информационная устойчивость случайных величин и процессов.* Издательство Академии Наук СССР, Москва, 1960.

[5] CSISZÁR, I.: „Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen". *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **7** (1962) 137—158.

[6] DOOB, J. L.: *Stochastic Processes.* Wiley, 1953.

[7] РОЗАНОВ, Ю. А.: „О плотности одной гауссовской меры относительно другой". *Теория Вероятностей и ее Применения* **7** (1962) 84—89.

[8] RÉNYI, A.: „On a new axiomatic theory of probability". *Acta Mathematica Academiae Scientiarum Hungaricae* **6** (1955) 285—355.

ОДНО ТЕОРЕТИКО-ИНФОРМАЦИОННОЕ НЕРАВЕНСТВО И ЕГО ПРИМЕНЕНИЕ К ДОКАЗАТЕЛЬСТВУ ЭРГОДИЧНОСТИ ЦЕПЕЙ МАРКОВА

I. CSISZÁR

Резюме

Для некоторой выпуклой функции $f(u)$ и для вероятностных мер $\mathbf{P}$ и $\mathbf{Q}$ на некотором измеримом пространстве (X, S_X) мы обозначим через $\mathscr{I}_f(\mathbf{P}, \mathbf{Q})$ величину $\int\limits_X qf\left(\frac{p}{q}\right) \lambda(dx)$ где p и q — плотности мер $\mathbf{P}$ и $\mathbf{Q}$ относящиеся к σ-конечной мере λ на (X, S_X). $\mathscr{I}_f(\mathbf{P}, \mathbf{Q})$ является обобщением понятия относительной информации (энтропии), так как в случае $f(u) = u\log u$ мы имеем $\mathscr{I}_f(\mathbf{P}, \mathbf{Q}) = I(\mathbf{P} \| \mathbf{Q}) = H_Q(\mathbf{P})$.

Если мы толкуем $\mathbf{P}$ и $\mathbf{Q}$ как распределения сигналов при входе канала $(X, Y, \nu(\cdot \mid x))$, происходящих из двух различных источников, тогда распределения $\overline{\mathbf{P}}$ и $\overline{\mathbf{Q}}$ выходных сигналов получаются формулами $\overline{\mathbf{P}}(E) = \int\limits_X \nu(E \mid x)\, \mathbf{P}(dx)$, $\overline{\mathbf{Q}}(E) = \int\limits_X \nu(E \mid x)\, \mathbf{Q}(dx)$.

В настоящей работе доказывается, что $\mathscr{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) \leqq \mathscr{I}_f(\mathbf{P}, \mathbf{Q})$ (теорема 1), кроме того даны условие равенства $\mathscr{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}}) = \mathscr{I}_f(\mathbf{P}, \mathbf{Q})$ и его применения к некоторым частным случаям а также одна оценка в случае малости разности $\mathscr{I}_f(\mathbf{P}, \mathbf{Q}) - \mathscr{I}_f(\overline{\mathbf{P}}, \overline{\mathbf{Q}})$ (дополнение к теореме 1, теоремы 2 и 2'). Теорема 1 обобщается и для σ-конечных мер (теорема 3).

В качестве применения результатов статьи в § 3 автор дает теоретико-информационное доказательство эргодичности некоторых цепей Маркова; при этом он не пользуется соображениями теории матриц. Доказательство такого характера дается здесь впервые.