INFORMATIONSENTFERNUNGEN IM RAUM DER WAHRSCHEINLICHKEITSVERTEILUNGEN

von

I. CSISZÁR und J. FISCHER

Einleitung

Die relative Information (auch Informationsgewinn oder relative Entropie genannt) und die entsprechende J-Divergenz der Ordnung $\alpha > 0$ (s. Rényi [1], [2]) können als Masse der Verschiedenheit zweier Wahrscheinlichkeitsverteilungen angesehen werden. Mit Hilfe dieser Begriffe lassen sich »Umgebungen« einer Wahrscheinlichkeitsverteilung definieren.

Es ist eine natürliche Frage, ob diese Umgebungen eine Topologie erzeugen und ob sich im bejahenden Fall eine Funktion der Divergenz bzw. des Informationsgewinnes geben lässt, die eine dieselbe Topologie erzeugende Entfernung bzw. »Quasientfernung« darstellt.

In § 1 wird die erste Frage für $0 < \alpha < 1$ positiv beantwortet,[1] sowie die zweite ausführlich formuliert. (Dass die Informations- bzw. Divergenzumgebungen im Falle $\alpha \geqq 1$ keine Topologie erzeugen, folgt aus [3], § 3.)

§ 2 enthält einige allgemeine Beziehungen zwischen Quasientfernungen und Entfernungen, die in den späteren benützt werden.

In § 3 bzw. § 4 werden einige Funktionen der Information bzw. Divergenz der Ordnung α $(0 < \alpha < 1)$ aus dem Gesichtspunkt der Quasientfernungs-(Entfernungs)eigenschaft untersucht, wodurch die in § 1 gestellte Metrisationsfrage bejahend beantwortet wird. Das erste positive Resultat in dieser Richtung stammt von J. Czipszer [4] und wird im Rahmen des Satzes 3 angeführt. Die wesentlichsten Ergebnisse der weiteren Untersuchungen werden in den Sätzen 4, 5 und 6 zusammengefasst, in denen wir über einige Typen der Funktionen der Information bzw. der Divergenz entscheiden, in welchen Fällen sie Quasientfernungen bzw. Entfernungen darstellen. Dabei sind die bezüglich der Divergenz in § 4 erhaltenen Resultate denen über die Information in § 3 analog. In beiden Paragraphen werden auch einige offene Fragen erwähnt.

Es werden die folgenden Bezeichnungen benützt:

P, Q, R bezeichnen Masse auf einem messbaren Raum (X, S) mit

$$P(X) = Q(X) = R(X) = 1\,, \tag{1}$$

also Wahrscheinlichkeitsverteilungen auf (X, S) (falls wir (1) nicht erfordern. wird auf dies ausdrücklich angedeutet);

λ bezeichnet ein Mass, in bezug auf welchem die betrachteten Masse absolut stetig sind (zu einer höchstens abzählbar unendlichen Menge von Massen

[1] Das erhaltene Ergebnis folgt auch aus [3] § 3; hier wird jedoch ein alternativer Beweis, der in gewissem Sinne ein schärferes Resultat ergibt, angeführt.

ist ein solches dominierendes λ immer zu finden); p, q, r bedeuten die Radon-Nikodymschen Ableitungen $\frac{dP}{d\lambda}, \frac{dQ}{d\lambda}, \frac{dR}{d\lambda}$.

Die Grösse

$$\mathcal{J}_\alpha(P, Q) = \int_X p^\alpha q^{1-\alpha} d\lambda \tag{2}$$

wird das Informationsintegral der Ordnung α ($\alpha > 0$) der Wahrscheinlichkeitsverteilung P bezüglich Q genannt (im diskreten Fall $P = (p_1, p_2, \ldots)$, $Q = (q_1, q_2, \ldots)$ ist $\mathcal{J}_\alpha(P, Q) = \sum_k p_k^\alpha q_k^{1-\alpha}$). Offenbar hängt der Wert $\mathcal{J}_\alpha(P, Q)$ von der Wahl des dominierenden Masses λ nicht ab.

$$I_\alpha(P \| Q) = -\frac{1}{1-\alpha} \log \mathcal{J}_\alpha(P, Q) \tag{3}$$

(hierbei wird unter »log« der natürliche Logarithmus verstanden) ist die (relative) Information der Ordnung α von P in bezug auf Q (im Falle $\alpha = 1$ ergibt sich der Grenzwert $I_1(P \| Q) = \int_X p \log \frac{p}{q} d\lambda$).

$$J_\alpha(P, Q) = I_\alpha(P \| Q) + I_\alpha(Q \| P) \tag{4}$$

ist die J-Divergenz der Ordnung α der Wahrscheinlichkeitsmasse P und Q.

§ 1. Die Informations- (Divergenz)topologie und das Problem ihrer Metrisation

Die Informations- bzw. Divergenzumgebungen der Ordnung $\alpha > 0$ einer Wahrscheinlichkeitsverteilung Q auf einem messbaren Raum (X, S) seien durch

$$V_\alpha(Q, \varepsilon) = \{P : I_\alpha(P \| Q) < \varepsilon\} \tag{5}$$

bzw.

$$V_\alpha^*(Q, \varepsilon) = \{P : J_\alpha(P, Q) < \varepsilon\} \tag{6}$$

definiert.

Es wird gezeigt, dass diese Umgebungen bei $0 < \alpha < 1$ eine Topologie im Raum der Wahrscheinlichkeitsverteilungen auf (X, S) definieren, also dass die Mengen der Art (5) bzw. (6) als Basis der Umgebungen der Verteilungen Q auf (X, S) in einer Topologie aufgefasst werden können. (Im Falle $\alpha \geqq 1$ definieren sie aber keine Topologie, nur einen Konvergenzraum, s. [3], § 3.)

Beweisen wir zuerst den folgenden

Hilfssatz 1. *Für beliebige $0 \leqq \alpha \leqq 1$ besteht für die Totalvariation auf der Menge $A \in S$ der Differenz $P - Q$ zweier Wahrscheinlichkeitsmasse die folgende Gleichung*

$$|P - Q|(A) = \int_A |p^\alpha q^{1-\alpha} - p| \, d\lambda + \int_A |p^\alpha q^{1-\alpha} - q| \, d\lambda .$$

Beweis. Offenbar liegt $p^\alpha q^{1-\alpha}$ bei jedem nichtnegativen p, q und $0 \leqq \alpha \leqq 1$ zwischen p und q. Also gilt

$$|p - q| = |p^\alpha q^{1-\alpha} - p| + |p^\alpha q^{1-\alpha} - q| ,$$

woraus die Behauptung des Hilfssatzes folgt.

Bezeichnen wir die Totalvariation von $P - Q$ auf X mit $\varrho(P, Q)$. Dann gilt der folgende

Satz 1. *Zwischen der Grösse* $\varrho(P, Q)$ *und dem Informationsintegral* $\mathcal{J}_\alpha(P, Q)$ $(0 < \alpha < 1)$ *besteht die Beziehung*

$$1 - \mathcal{J}_\alpha(P, Q) \leqq \varrho(P, Q) . \tag{7}$$

Beweis. Es gilt

$$1 - \mathcal{J}_\alpha(P, Q) \leqq \int_X |p - p^\alpha q^{1-\alpha}| \, d\lambda \leqq \varrho(P, Q) ,$$

wobei die letztere Ungleichungen aus Hilfssatz 1 folgt, w. z. b. w.

Aus Satz 1 folgt nach dem Zusammenhang von $I_\alpha(P \,||\, Q)$ und $\mathcal{J}_\alpha(P, Q)$, dass zu gegebenem α $(0 < \alpha < 1)$ immer eine Konstante C_α^* derart gewählt werden kann, dass

$$I_\alpha(P \,||\, Q) \leqq C_\alpha^* \, \varrho(P, Q) , \tag{8}$$

falls eine der Grössen $I_\alpha(P \,||\, Q)$ und $\varrho(P, Q)$ genügend klein ist. Damit haben wir auf den Fall $0 < \alpha < 1$ ein schärferes Resultat als die aus [3] Satz 3 folgende Relation $I_\alpha(P \,||\, Q) \leqq C_\alpha^* \sqrt{\varrho(P, Q)}$ erhalten.

Aus der von [3] Satz 3 folgenden Beziehung $\varrho(P, Q) \leqq C_\alpha \sqrt[3]{I_\alpha(P \,||\, Q)}$ und aus (8) folgt, dass die Informationsumgebungen eine Topologie, und zwar eine der durch die Variationsentfernung $\varrho(P, Q)$ erzeugten identische definieren.

Offenbar gelten für die J-Divergenz ähnliche Behauptungen.

Da also die Informations- bzw. Divergenzumgebungen der Ordnung $0 < \alpha < 1$ eine metrisierbare Topologie erzeugen, ergibt sich die Frage, ob ihre Metrisierung auch durch eine Funktion der Information bzw. der Divergenz erzielt werden kann.

Die Information $I_\alpha(P \,||\, Q)$ ist keine symmetrische Funktion von P und Q $\left(\text{ausgenommen für } \alpha = \frac{1}{2}\right)$. Daher ist die Metrisierung durch irgendeine Funktion von $I_\alpha(P \,||\, Q)$ im gewöhnlichen Sinne nicht möglich. Diese Schwierigkeit lässt sich jedoch in einer ziemlich natürlichen Weise überwinden, und zwar durch die Benützung des Begriffes der Quasientfernung (s. [5] § 13 etwas allgemeiner als Quasiabstand — »quasi-écart« — vgl. hier S. 164).

Eine auf einer Menge M definierte Funktion $\varphi(a, b)$ wird eine Quasientfernung genannt, falls sie die folgenden Eigenschaften besitzt:

$$\text{(A)} \qquad \varphi(a, b) > \varphi(a, a) = 0 \qquad (a \neq b)$$

$$\text{(B)} \qquad \varphi(a, b) + \varphi(b, c) \geqq \varphi(a, c) .$$

Falls man hier auch die Symmetrie von $\varphi(a, b)$ erfordert, erhält man offenbar eine gewöhnliche Entfernung. (Das Problem der Metrisation einer Topologie durch eine (A) und (B) erfüllende nicht unbedingt symmetrische Funktion ist in [6] und [7] betrachtet.)

Bemerken wir, dass falls man an der rechten Seite von (B) $\varphi(a, c)$ durch $\varphi(c, a)$ ersetzt, sich eine gewöhnliche Entfernung — sogar wenn (A) durch die schwächere Forderung $\varphi(a, b) \neq \varphi(a, a) = 0$ ersetzt wird — ergibt, wie es sich einfach zeigen lässt.

Natürlicherweise erzeugt eine Quasientfernung ebenso eine Topologie, wie eine Entfernung.

Die natürliche Fragestellung bezüglich der Metrisation lautet also folgendermassen: man sucht eine Funktion der Information bzw. Divergenz der Ordnung $0 < \alpha < 1$, die eine — der durch die Informations- (Divergenz)umgebungen definierten identische Topologie erzeugende — Quasientfernung bzw. Entfernung über dem Raum der Wahrscheinlichkeitsverteilungen auf (X, S) darstellt.

Die Existenz solcher Funktionen ist gar nicht trivial. Falls nämlich auf einer Menge E die durch eine Funktion $\varphi(x, y)$ $(x, y \in E)$ definierte Umgebungen

$$V(x, \varepsilon) = \{y : \varphi(x, y) < \varepsilon\}$$

eine metrisierbare Topologie definieren, ist damit die Existenz einer Funktion f, für die $f(\varphi(x, y))$ eine Quasientfernung darstellt, noch nicht gesichert. Für die Existenz einer solchen lässt sich eine allgemeine Bedingung geben, die aber in unserem Fall die Konstruktion nicht erleichtert, aus welchem Grunde auf diese Frage hier nicht eingegangen wird.

§ 2. Eine allgemeine Methode zur Herleitung weiterer Quasientfernungen aus gegebenen

In diesem Paragraphen werden allgemeine Beziehungen zwischen Quasientfernungen betrachtet und einige in den folgenden benötigten Zusammenhänge hergeleitet.

Schicken wir vor allem einen einfachen Hilfssatz voraus.

Hilfssatz 2. *Es seien H_1 und H_2 beliebige halbgeordnete Mengen, auf denen je eine ordungstreue Operation definiert ist (der Einfachheit halber wird auf beiden Mengen die Ordnungsrelation mit $\prec$ und die Operation mit $\circ$ bezeichnet), also*

$$u \circ v \prec u' \circ v' \quad \textit{falls} \quad u \prec u', \quad v \prec v' \qquad (u, v, u', v' \in H_i; i = 1, 2).$$

Ferner sei ψ eine Abbildung einer Teilmenge $K \subset H_1$ in H_2, die die folgende Eigenschaft besitzt:

$$\psi(w) \prec \psi(u) \circ \psi(v) \quad \textit{falls} \quad w \prec u \circ v \qquad (u, v, w \in K). \tag{9}$$

Ist dann M eine beliebige Menge und $t(a, b)$ eine Abbildung von $M \otimes M$ in K, die der abstrakten »Dreiecksungleichung«

$$t(a, c) \prec t(a, b) \circ t(b, c) \qquad (a, b, c \in M) \tag{10}$$

genügt, so genügt auch $\psi(t(a, b))$ der abstrakten Dreiecksungleichung, also

$$\psi(t(a, c)) \prec \psi(t(a, b)) \circ \psi(t(b, c)). \tag{11}$$

Beweis. (11) folgt unmittelbar aus (9) und (10).

Falls man in Hilfssatz 2 für H_1 eine Funktionenmenge, für H_2 die reelle Achse wählt und die Abbildung $t(a, b)$ mit Hilfe von gegebenen, die Dreiecksungleichung (B) befriedigenden Funktionen $\varphi_x(a, b)$ definiert, lässt sich aus Hilfssatz 2 eine allgemeine Methode zur Herstellung neuer Quasientfernungen

(Entfernungen) aus gegebenen ableiten. Dabei genügt es statt (9) etwas weniger vorauszusetzen, was für die folgenden auch von Bedeutung sein wird.

Es gilt nämlich der folgende

Satz 2. *Es sei X eine beliebige Menge und K eine Klasse der auf X definierten nichtnegativen Funktionen, für die mit $u(x)$, $v(x)$, $w(x) \in K$ auch $\frac{u(x)}{u(x)+v(x)} w(x) \in K$ gilt, falls die letztere Funktion für jedes $x \in X$ definiert ist*[2] *und $\psi(u(x))$ ein auf K definiertes Funktional, das die folgenden Eigenschaften besitzt:*

$$\text{(C)} \qquad \psi(u(x)) \leqq \psi(v(x)) \quad \textit{falls} \quad u(x) \leqq v(x) \ \textit{für jedes} \quad x \in X$$

(Monotonität)

$$\text{(D)} \qquad \psi(u(x)+v(x)) \leqq \psi(u(x)) + \psi(v(x)) \quad \text{falls} \quad u(x)+v(x) \in K$$

(Subadditivität).

Sind dann $\varphi_x(a, b)$ $(x \in X)$ auf einer beliebigen Menge M definierte, die Bedingung (B) *erfüllende nichtnegative Funktionen so dass $\varphi_x(a, b)$ als eine Funktion von x für jedes $a, b \in M$ zur Klasse K gehört, erfüllt auch die Funktion*

$$\varphi^*(a, b) = \psi(\varphi_x(a, b))$$

die Bedingung (B).

Beweis. Aus (C) und (D) folgt, dass im Falle $w(x) \leqq u(x) + v(x)$ $(u(x), v(x), w(x) \in K)$ die Ungleichung

$$\text{(12)} \qquad \psi(w(x)) \leqq \psi(u(x)) + \psi(v(x))$$

besteht (auch wenn $u(x) + v(x) \notin K$).

Setzen wir nämlich

$$u^*(x) = \frac{u(x)}{u(x)+v(x)} w(x), \qquad v^*(x) = \frac{v(x)}{u(x)+v(x)} w(x),$$

dann sind $u^*(x)$ und $v^*(x)$ für jedes $x \in X$ definiert (da im Falle $u(x) = v(x) = 0$ nach $0 \leqq w(x) \leqq u(x) + v(x)$ auch $w(x) = 0$ gelten muss) und gehören nach Voraussetzung zu K, also gilt wegen $w(x) = u^*(x) + v^*(x)$ nach (D) und (C)

$$\text{(13)} \qquad \psi(w(x)) \leqq \psi(u^*(x)) + \psi(v^*(x)) \leqq \psi(u(x)) + \psi(v(x)).$$

Aus (13) folgt unmittelbar mit der Wahl $u(x) = \varphi_x(a, b)$, $v(x) = \varphi_x(b, c)$, $w(x) = \varphi_x(a, c)$ die Behauptung des Satzes 2.

Folgerung 1. Wählen wir in Satz 2 das Funktional ψ derart, dass

$$\text{(E)} \qquad \psi(0) = 0, \quad \psi(u(x)) > 0 \quad \text{falls} \quad u(x) > 0 \quad \text{für jedes} \quad x \in X$$

---

[2] Falls für ein $x \in X$ $u(x) = v(x) = w(x) = 0$ ist, setzt man $\frac{u(x)}{u(x)+v(x)} w(x) = 0$.

gelte. Sind dann die Funktionen $\varphi_x(a, b)$ Quasientfernungen auf M, so ist auch $\varphi^*(a, b)$ eine Quasientfernung auf M; gilt hier $\varphi^*(a, b) = \varphi^*(b, a)$, so ist $\varphi^*(a, b)$ natürlich eine Entfernung; das ist z. B. der Fall, wenn schon $\varphi_x(a, b) = \varphi_x(b, a)$ gilt, also falls die Funktionen $\varphi_x(a, b)$ selbst Entfernungen sind.

Bemerkung. Eine nichtnegative Funktion $\varphi(a, b)$ heisst einen Quasiabstand auf M, falls sie die Bedingung (B) erfüllt, statt (A) jedoch nur $\varphi(a, a) = 0$ verlangt wird. Ein symmetrischer Quasiabstand wird einen Abstand genannt. Offenbar folgen aus Satz 2 für Quasiabstände (bzw. Abstände) $\varphi_x(a, b)$ zur Folgerung 1 analogische Behauptungen: die Funktion $\varphi^*(a, b)$ ist ein Quasiabstand (bzw. Abstand) — auch wenn statt (E) nur $\psi(0) = 0$ verlangt wird. Falls es dabei für jedes $a, b \in M$, $a \neq b$ wenigstens ein $x \in X$ mit $\varphi_x(a, b) \neq 0$ gibt (dann heisst die Funktionenmenge $\{\varphi_x(a, b) : x \in X\}$ eine Quasimetrik auf M) und $\psi(u(x)) = 0$ nur für $u(x) \equiv 0$ gilt, ist die Funktion $\varphi^*(a, b)$ eine Quasientfernung (bzw. Entfernung) auf M.

Folgerung 2. Es seien $\varphi_1(a, b), \ldots, \varphi_n(a, b)$ Quasientfernungen auf M und $\psi(u_1, \ldots, u_n)$ eine im n-dimensionalen Intervall $0 \leqq u_i \leqq \sup_{a,b \in M} \varphi_i(a, b) = m_i$ $(i = 1, \ldots, n)$ definierte Funktion mit den Eigenschaften

$$\text{(C}'\text{)} \quad \psi(u_1, \ldots, u_n) \leqq \psi(v_1, \ldots, v_n) \quad \text{für} \quad 0 \leqq u_i \leqq v_i \leqq m_i \quad (i = 1, \ldots, n)$$

(Monotonität)

$$\text{(D}'\text{)} \quad \psi(u_1 + v_1, \ldots, u_n + v_n) \leqq \psi(u_1, \ldots, u_n) + \psi(v_1, \ldots, v_n)$$

$$(0 \leqq \min(u_i, v_i) \leqq u_i + v_i \leqq m_i\,;\; i = 1, \ldots, n)$$

(Subadditivität)

$$\text{(E}'\text{)} \quad \psi(0, \ldots, 0) = 0, \; \psi(u_1, \ldots, u_n) > 0 \text{ falls } u_i > 0 \; (i = 1, \ldots, n)\,.$$

Dann ist auch $\varphi^*(a, b) = \psi(\varphi_1(a, b), \ldots, \varphi_n(a, b))$ eine Quasientfernung auf M. Das folgt aus Folgerung 1 mit der Wahl

$$X = \{1, \ldots, n\}, \qquad K = \{(u_1, \ldots, u_n) : 0 \leqq u_i \leqq m_i\}\,.$$

Was die Symmetrie dieser Quasientfernung bzw. den Fall der Quasiabstände anbetrifft, sind natürlich die in Folgerung 1 bzw. in ihrer Bemerkung gesagten gültig.

Bemerkung. Aus Folgerung 2 ergibt sich als Spezialfall $(n = 1)$, dass falls $\varphi(a, b)$ eine Quasientfernung auf M und $\psi(u)$ eine monoton zunehmende, subadditive Funktion mit $\psi(0) = 0$ ist, so ist auch $\psi(\varphi(a, b))$ eine Quasientfernung auf M.

Es sei bemerkt, dass für die Subadditivität einer monoton zunehmenden Funktion $\psi(u)$ mit $\psi(0) = 0$ ihre Konkavität eine hinreichende Bedingung ist. Im folgenden wird häufig vorkommen, dass die Funktion $\psi(u)$ in der Form $\psi(u) = \psi_2(\psi_1^{-1}(u))$ gegeben wird, wobei $\psi_1(u)$ und $\psi_2(u)$ stetige und stückweise differenzierbare Funktionen mit $\psi_1(0) = \psi_2(0) = 0$ und $\psi_1'(u) > 0$, $\psi_2'(u) \geqq 0$ sind. Dann gilt

$$\psi'(u) = \frac{\psi_2'(\psi_1^{-1}(u))}{\psi_1'(\psi_1^{-1}(u))}$$

(wo die rechte Seite nur einen Sinn hat) und daher ist für die Konkavität von $\psi(u)$ notwendig und hinreichend, dass $\frac{\psi_2'(u)}{\psi_1'(u)}$ (wo es nur existiert) monoton abnimmt. Die letztere Bedingung ist also für die Subadditivität von $\psi(u)$ hinreichend.

Folgerung 3. Es sei (X, S) ein messbarer Raum, M^* eine Menge von σ-endlichen Massen auf (X, S), die durch ein σ-endliches Mass λ dominiert sind (also $\mu \ll \lambda$ für $\mu \in M^*$). Ist dann $\varphi(u, v)$ eine auf der Halbgeraden $[0, +\infty)$ definierte Quasientfernung, für welche

$$\varphi^*(\mu, \nu) = \left[\int_X \varphi^{\omega} \left(\frac{d\mu}{d\lambda}, \frac{d\nu}{d\lambda} \right) d\lambda \right]^{\frac{1}{\omega}} \qquad (\omega \geqq 1)$$

bei jedem $\mu \in M^*$, $\nu \in M^*$ existiert und endlich ist, so stellt $\varphi^*(\mu, \nu)$ eine Quasientfernung auf M^* dar.

Um das zu erhalten nehmen wir in Satz 2 K die Menge der endlichen nichtnegativen Funktionen $u(x)$ auf X, für welche $u^{\omega}(x)$ λ-summierbar ist, und für $u(x) \in K$

$$\psi(u(x)) = \left[\int_X u^{\omega}(x)\, d\lambda(x) \right]^{\frac{1}{\omega}}.$$

Dieses Funktional $\psi(u(x))$ besitzt offenbar (mit Rücksicht auf die Minkowskische Ungleichung) die Eigenschaften (C), (D), (E). M sei die Menge der Dichtefunktionen[3] $a(x) = \frac{d\mu}{d\lambda}$ ($\mu \in M^*$) und $\varphi_X(a, b) = \varphi(a(x), b(x))$ $(a(x), b(x) \in M)$. Dann ist nach Folgerung 2

$$\varphi^*(a(x), b(x)) = \left[\int_X \varphi^{\omega}(a(x), b(x))\, d\lambda(x) \right]^{\frac{1}{\omega}}$$

eine Quasientfernung auf M. Daraus folgt die Behauptung indem man $a(x) = \frac{d\mu}{d\lambda}$, $b(x) = \frac{d\nu}{d\lambda}$ setzt.

Folgerung 3 kann auch unmittelbar aus Hilfssatz 2 hergeleitet werden, indem man für $H_1 = K$ die Menge der nichtnegativen Funktionen $u(x)$ $(x \in X)$ mit λ-summierbarem $u^{\omega}(x)$, für H_2 die nichtnegative Halbgerade mit der natürlichen Anordnung wählt, für die Operation $\circ$ die Addition und für die Abbildung ψ

$$\psi(u) = \left[\int_X u^{\omega}(x)\, d\lambda \right]^{\frac{1}{\omega}} \tag{14}$$

setzt. Hierbei soll $M = M^*$ und $t(\mu, \nu) = \varphi \left(\frac{d\mu}{d\lambda}, \frac{d\nu}{d\lambda} \right)$ für $\mu, \nu \in M$ gewählt werden.

Da in den obigen eine Quasientfernung über der nichtnegativen Halbgeraden als bekannt vorausgesetzt wurde, scheint es von Interesse zu sein,

[3] Diese sind nur fast überall bezüglich λ eindeutig bestimmt, nun denken wir aber ihre Werte irgendwelcher Weise überall festgelegt und zwar so, dass für jedes $x \in X$ $0 \leq a(x) < \infty$ ist.

eine hinreichende Bedingung, dass eine Funktion eine Quasientfernung über einem Intervall sei, anzuführen.

Hilfssatz 3. *Es sei E ein offenes Intervall an der Zahlengeraden und $\varphi(u, v)$ eine auf dem Quadrat $E \otimes E$ definierte stetige Funktion, für die φ_{uv} bei $u \in E$, $v \in E$, $u \neq v$ existiert. Falls dann für $u \in E$, $v \in E$ die Bedingungen*

$$1) \qquad \varphi(u, u) = 0$$

$$2) \qquad \varphi_u(u, v) < 0 < \varphi_v(u, v) \qquad u < v$$

$$\varphi_v(u, v) < 0 < \varphi_u(u, v) \qquad u > v$$

$$3) \qquad \varphi_{uv}(u, v) \geqq 0 \qquad u \neq v$$

erfüllt sind, so ist $\varphi(u, v)$ eine Quasientfernung über E. Falls ferner E^* die Vereinigung von E mit einem oder beiden seiner Endpunkte bezeichnet und $\varphi(u, v)$ auch auf E^* stetig ist, so stellt sie auch auf E^* eine Quasientfernung dar. (Um einen Quasiabstand zu erhalten, genügt es in 2) statt $<$ nur $\leqq$ zu erfordern.)

Beweis. Offenbar genügt wegen der Stetigkeit von $\varphi(u, v)$ den Hilfssatz für E zu beweisen.

Es seien $u, v, w \in E$. Aus 1) und 2) folgt, dass die stetige Funktion $\varphi(u, v)$ die Bedingung (A) der Quasientfernung erfüllt. Ferner folgt aus 2), dass falls u zum festen v bzw. v zum festen u angenähert wird, die Funktion $\varphi(u, v)$ abnimmt. Das bedeutet aber, dass falls v die kleinste oder die grösste der Zahlen u, v, w ist, dann ist ein Glied an der linken Seite der Dreiecksungleichung (B) schon an sich grösser als die rechte Seite. Es ist also noch übrig, die Gültigkeit von (B) in den Fällen $u \leqq v \leqq w$ bzw. $w \leqq v \leqq u$ zu beweisen. Aus 3) folgt aber, dass φ_u bei festem u als eine Funkion von v nicht abnimmt, also ist $\varphi_u(u, w) - \varphi_u(u, v)$ nichtnegativ oder nichtpositiv je nachdem ob $u < v \leqq w$ oder $w \leqq v < u$ besteht. Daher ist $\varphi(u, w) - \varphi(u, v)$ für $u \leqq v \leqq w$ eine nicht abnehmende und für $w \leqq v \leqq u$ eine nicht zunehmende Funktion von u ($u = v$ kann hier wegen der Stetigkeit von $\varphi(u, v)$ zugelassen werden). Daraus folgt mit Rücksicht auf 1)

$$\varphi(u, w) - \varphi(u, v) \leqq \varphi(v, w),$$

also die Gültigkeit von (B) auch für die beiden Fälle $u \leqq v \leqq w$ bzw. $w \leqq v \leqq u$ w. z. b. w.

Bemerkung. Falls man in 3) die Gleichheit nicht zulässt, besteht die Gleichheit in der bewiesenen Dreiecksungleichung offenbar nur in den Fällen $u = v$ bzw. $v = w$.

§ 3. Informationsquasientfernungen

In § 1 wurde das Problem gestellt, ob sich eine Funktion f der Information der Ordnung α $(0 < \alpha < 1)$ finden lässt, für die $f(I_\alpha(P \| Q))$ eine Quasientfernung in Raum der Verteilungen auf (X, S) ist. Dabei möchte man solche Quasientfernungen herstellen, welche auch die durch die Informationsumgebungen $V_\alpha(Q, \varepsilon)$ definierte Topologie erzeugen. Daher wollen wir in den folgenden solche Funktionen f mit der obigen Eigenschaft angeben, die nichtabnehmend mit $f(u) > f(0) = 0$ für $u > 0$ und im Ursprung stetig sind.

Es ist leicht zu sehen, dass $I_\alpha(P \,||\, Q)$ selbst keine Quasientfernung darstellen kann. Falls nämlich die drei Verteilungen P, Q, R so gewählt werden, dass $P \pm Q$, $Q \pm R$, $P \perp R$ ist, dann gilt $I_\alpha(P \,||\, Q) < +\infty$, $I_\alpha(Q \,||\, R) < +\infty$, $I_\alpha(P \,||\, R) = +\infty$, also ist die Dreiecksungleichung für P, Q, R nicht erfüllt. Ähnlich lässt sich zeigen, dass auch keine Funktion $f\big(I_\alpha(P \,||\, Q)\big)$ mit $f(u) < f(+\infty) = +\infty$ eine Quasientfernung darstellen kann. Es kann also keine Potenz $I_\alpha^\beta(P \,||\, Q)$ der Information eine Quasientfernung sein (da für $\beta > 0$ die Funktion $f(u) = u^\beta$ der obigen Bedingung genügt, $\beta \leqq 0$ jedoch gar nicht in Betracht kommt, nämlich ist dann $I_\alpha^\beta(P \,||\, P) = 0$ nicht erfüllt.)

Man kann jedoch noch vermuten, dass eine Funktion der Form

$$f(u) = \begin{cases} u^\beta & \text{für} \quad u \leqq c \\ c^\beta & \text{für} \quad u \geqq c \end{cases} \tag{15}$$

geeignet sein mag, um eine Quasientfernung zu erzeugen — da für solche $f(+\infty) < +\infty$ ist.

Zur Entscheidung dieser Frage genügt zu untersuchen, ob die Ungleichung

$$I_\alpha^\beta(P \,||\, Q) + I_\alpha^\beta(Q \,||\, R) \geqq \min\big(I_\alpha^\beta(P \,||\, R), c^\beta\big) \tag{16}$$

im Falle $I_\alpha(P \,||\, Q) \leqq c$, $I_\alpha(Q \,||\, R) \leqq c$ mit genügend kleinem $c > 0$ besteht.

Für $\beta = 1$ lässt sich leicht zeigen, dass dies nicht der Fall sein kann. Setzen wir nämlich

$$P = (0, 1), \qquad Q_\delta = (\delta, 1-\delta), \qquad R_\varepsilon = (\varepsilon, 1-\varepsilon) \tag{17}$$

mit so kleinen $\delta > 0$ und $\varepsilon > 0$, dass schon $I_\alpha(P \,||\, Q_\delta) \leqq c$, $I_\alpha(Q_\delta \,||\, R_\varepsilon) \leqq c$, $I_\alpha(P \,||\, R_\varepsilon) \leqq c$ gilt. Dann ist die Ungleichung (16) für (17) mit

$$(1-\delta)^{1-\alpha}\left[\delta^\alpha \varepsilon^{1-\alpha} + (1-\delta)^\alpha (1-\varepsilon)^{1-\alpha}\right] \leqq (1-\varepsilon)^{1-\alpha} \tag{18}$$

äquivalent. Eine einfache Rechnung zeigt, dass (18) dann und nur dann besteht, falls $\delta \geqq \varepsilon$ ist. Für $\beta > 1$ kann (16) mit Rücksicht auf die Bemerkung zur Folgerung 2 des Satzes 2 noch weniger gelten, da dann $I_\alpha(P \,||\, Q)$ eine konkave Funktion von $I_\alpha^\beta(P \,||\, Q)$ ist.

Für $0 < \beta < 1$ ist die Frage komplizierter, da in diesem Fall die Ungleichung (16) nicht auf Produktform gebracht werden kann. Da es sich jedoch hier um kleine Informationswerte handelt, scheint die Ersetzung der Informationen $I_\alpha(P \,||\, Q)$ durch die annähernden Grössen

$$\frac{1}{1-\alpha}\left(1 - \mathcal{J}_\alpha(P, Q)\right)$$

zweckmässig zu sein.

Also werden wir uns mit der Frage beschäftigen, ob die Grössen $\left(1 - \mathcal{J}_\alpha(P, Q)\right)^\beta$ die Dreiecksungleichung bei geeigneter Wahl von $\beta > 0$ (wenigstens für kleine Werte der $I_\alpha(P \,||\, Q)$) erfüllen.

Betrachten wir zuerst wieder die diskreten Verteilungen (17), d. h.

$$P = (0, 1), \qquad Q_\delta = (\delta, 1-\delta), \qquad R_\varepsilon = (\varepsilon, 1-\varepsilon)\,,$$

also untersuchen wir ob die Ungleichung

$$\left(1-(1-\delta)^{1-\alpha}\right)^\beta + \left(1 - \left(\delta^\alpha \varepsilon^{1-\alpha} + (1-\delta)^\alpha (1-\varepsilon)^{1-\alpha}\right)\right)^\beta \geqq \left(1-(1-\varepsilon)^{1-\alpha}\right)^\beta \tag{19}$$

gültig ist. Offenbar besteht sie für $\delta = 0$ mit der Gleichheit. Die partielle Ableitung der linken Seite nach δ ist

$$\beta(1-\alpha)\bigl(1-(1-\delta)^{1-\alpha}\bigr)^{\beta-1}(1-\delta)^{-\alpha}+$$
$$+\beta\bigl(1-(\delta^{\alpha}\varepsilon^{1-\alpha}+(1-\delta)^{\alpha}(1-\varepsilon)^{1-\alpha})\bigr)^{\beta-1}\bigl(-\alpha\delta^{\alpha-1}\varepsilon^{1-\alpha}+\alpha(1-\delta)^{\alpha-1}(1-\varepsilon)^{1-\alpha}\bigr)^{\beta}.$$

Diese strebt für $\delta \to 0$ im Falle $\beta > \alpha$ bei jedem festen $\varepsilon > 0$ gegen $-\infty$ (da der erste Glied für $\delta \to 0$ asymptotisch gleich $K_1\,\delta^{\beta-1}$, der zweite aber $-K_2\,\delta^{\alpha-1}$ ist, vobei K_1 und K_2 positive — von α, β und ε abhängende — Konstanten sind). Also kann zu jedem $\varepsilon > 0$ die Zahl $\delta > 0$ derart gewählt werden, dass die Ungleichung (19) nicht gilt. Das bedeutet, dass die Dreiecksungleichung für die Grössen der Form $\bigl(1-\mathcal{J}_\alpha(P,Q)\bigr)^{\beta}$ mit $\beta > \alpha$ nicht einmal bei genügend kleinen Werten von $I_\alpha(P\,||\,Q)$, $I_\alpha(Q\,||\,R)$ und $I_\alpha(P\,||\,R)$ allgemein gültig sein kann.

Ähnliches kann für $\beta > 1-\alpha$ mit der Wahl $P = (1, 0)$, $Q = (1-\delta, \delta)$, $R = (1-\varepsilon, \varepsilon)$ in derselben Weise gezeigt werden.

Also gilt

Hilfssatz 4. *Die Funktion* $\bigl(1-\mathcal{J}_\alpha(P,Q)\bigr)^{\beta}$ *kann im Fall* $\beta > \alpha' = \min(\alpha, 1-\alpha)$ *die Dreiecksungleichung selbst für beliebig kleine Werte der Ausdrücke der Form* $1-\mathcal{J}_\alpha(P,Q)$ *(also auch der entsprechenden Informationen) nicht allgemein erfüllen.*

Daraus folgt schon dieselbe Behauptung auch für $I_\alpha(P\,||\,Q)$. Falls nämlich die Dreiecksungleichung bezüglich I_α^{β} bei genügend kleinen I_α-Werten gültig wäre, so müsste sie nach der Bemerkung zur Folgerung 2 des Satzes 2 auch für $(1-\mathcal{J}_\alpha)^{\beta}$ (bei genügend kleinem I_α) gelten, was aber für $\beta > \alpha' = \min(\alpha, 1-\alpha)$ in Widerspruch mit den bewiesenen ist. Hier haben wir benützt, dass die Grössen $I_\alpha^{\beta}(P\,||\,Q)$ und $\bigl(1-\mathcal{J}_\alpha(P,Q)\bigr)^{\beta}$ aus $(1-\alpha)\,I_\alpha(P\,||\,Q) = -\log \mathcal{J}_\alpha(P,Q)$ durch die Funktionen $\psi_1(u) = \left(\dfrac{u}{1-\alpha}\right)^{\beta}$ bzw. $\psi_2(u) = (1-e^{-u})^{\beta}$ entstehen, wobei

$$\frac{\psi_2'(u)}{\psi_1'(u)} = (1-\alpha)^{\beta}\left(\frac{u}{e^{\frac{u}{1-\beta}}(1-e^{-u})}\right)^{1-\beta} = (1-\alpha)^{\beta}(1-\beta)^{1-\beta}\left(\frac{\dfrac{u}{1-\beta}}{e^{\frac{u}{1-\beta}}-e^{\beta\frac{u}{1-\beta}}}\right)^{1-\beta} \tag{20}$$

wegen des Zunehmens der Funktion

$$\frac{e^{u}-e^{\beta u}}{u} = e^{\beta u}\,\frac{e^{(1-\beta)u}-1}{u} = e^{\beta u}\sum_{k=1}^{\infty}\frac{(1-\beta)^k}{k!}\,u^{k-1} \tag{21}$$

monoton abnimmt.

Was nun den Fall $\beta \leqq \alpha'$ anbetrifft, beweisen wir zuerst den folgenden grundlegenden Satz:

Satz 3. *Die Grösse*

$$\Delta_\alpha(P,Q) = \bigl(1-\mathcal{J}_\alpha(P,Q)\bigr)^{\alpha'} \qquad \bigl(\alpha' = \min(\alpha, 1-\alpha),\ 0 < \alpha < 1\bigr) \tag{22}$$

stellt eine Quaisentfernung im Raum der Wahrscheinlichkeitsverteilungen auf (X, S) *dar. Also lässt sich die durch die Informationsumgebungen* $V_\alpha^*(Q, \varepsilon)$ *erzeugte Topologie in dem in § 1 angedeuteten Sinne metrisiert werden.*

Allgemeiner ist die — von der Wahl des dominierenden Masses unabhängige — Grösse

(23) $$\Delta_\alpha(P,Q) = \left(\alpha P(X) + (1-\alpha)\, Q(X) - \int_X p^\alpha q^{1-\alpha}\, d\lambda\right)^{\alpha'}$$

eine Quasientfernung im Raum der endlichen Masse auf (X, S).

(Die Tatsache, dass der Ausdruck $(1 - \int_X p^\alpha q^{1-\alpha}\, d\lambda)^{\alpha'}$ und allgemeiner die für beliebige nichtnegative λ-summierbare Funktionen $u(x)$, $v(x)$ definierte Funktional

$$\left(\alpha \int_X u(x)\, d\lambda(x) + (1-\alpha) \int_X v(x)\, d\lambda(x) - \int_X u^\alpha(x)\, v^{1-\alpha}(x)\, d\lambda(x)\right)^{\alpha'}$$

der Dreiecksungleichung genügt, wurde von Herrn J. Czipszer [4] erkannt. Für die liebenswürdige Erlaubnis der Publikation möchten ihm die Verfasser ihre Dankbarkeit aussprechen. Hier wird im wesentlichen der Beweis von Herrn J. Czipszer angeführt; unsere Modifikation besteht nur darin, dass der Beweis in die allgemeine Theorie eingebaut wurde.)

Bemerkung. Im Spezialfall $\alpha = \dfrac{1}{2}$ wurde die Entfernungseigenschaft von (22) schon von Bhattacharyya [8] erkannt. In diesem Falle ist die oben eingeführte Grösse $\Delta_\alpha(P, Q)$ im wesentlichen die Entfernung der Masse P und Q in einem Hilbertschen Raum der endlichen signierten Masse auf (X, S). Um diesen zu erhalten, führen wir die Operationen folgendermassen ein: Zu zwei endlichen signierten Massen P und Q wählen wir ein dominierendes Mass λ und setzen $p = \dfrac{dP}{d\lambda}$, $q = \dfrac{dQ}{d\lambda}$. Die Operationen des Hilbertschen Raumes definieren wir als die für $\operatorname{sg} p \cdot \sqrt{|p|}$ und $\operatorname{sg} q \cdot \sqrt{|q|}$ im Raum $L^2(X, S, \lambda)$, also die Addition durch

$$(P \oplus Q)(A) = \int_A \operatorname{sg}\left(\operatorname{sg} p \cdot \sqrt{|p|} + \operatorname{sg} q \cdot \sqrt{|q|}\right)\left(\operatorname{sg} p \cdot \sqrt{|p|} + \operatorname{sg} q \cdot \sqrt{|q|}\right)^2 d\lambda \quad (A \in S),$$

die Multiplikation mit einem Skalar m durch

$$(m \odot P)(A) = \int_A \operatorname{sg}(mp)\left(m \operatorname{sg} p \cdot \sqrt{|p|}\right)^2 d\lambda \quad (A \in S), \text{ also } m \odot P = m\,|m|\,P,$$

das Skalarprodukt durch

$$\langle P, Q\rangle = \int_X \operatorname{sg} p \cdot \sqrt{|p|} \cdot \operatorname{sg} q \cdot \sqrt{|q|}\, d\lambda = \int_X \operatorname{sg}(pq) \cdot \sqrt{|pq|}\, d\lambda$$

und daher die Norm durch $\| P \| = \sqrt{\int_X |p|\, d\lambda} = \sqrt{|P|(X)}$.

Diese Operationen sind von der Wahl des dominierenden Masses λ offenbar unabhängig und befriedigen die Axiomen des Hilbertschen Raumes. Man sieht leicht, dass der dadurch definierte Hilbetsche Raum bei beliebigem (X, S) vollständig ist; für abzählbares X ist er sogar separabel und auch einem L^2-Raum isomorph.

Für gewöhnliche Masse P und Q gilt nun in diesem Raum

$$\|P \ominus Q\| = \sqrt{\int_X (\sqrt{p} - \sqrt{q})^2\, d\lambda} = \sqrt{2}\, \Delta_{\frac{1}{2}}(P, Q).$$

Wir möchten noch die Bemerkung machen, dass im obigen Raum zwei gewöhnliche Masse P und Q genau dann orthogonal sind, falls $P \perp Q$ im Sinne der Masstheorie gilt. Ferner ist ein gewöhnliches Mass genau dann eine Wahrscheinlichkeitsverteilung, falls es am Einheitskugel liegt.

Beweis des Satzes 3. In der Dreiecksungleichung bezüglich (23) für die Masse P, Q und R ist die Ersetzung von α durch $1-\alpha$ mit der Vertauschung von P und R äquivalent. Daher genügt die Quasientfernungseigenschaft von (23) für $0 < \alpha \leqq \frac{1}{2}$ zu verifizieren.

Um die Ungleichung

$$\Big(\alpha \int_X p d\lambda + (1-\alpha) \int_X q d\lambda - \int_X p^\alpha q^{1-\alpha} d\lambda\Big)^\alpha + $$
$$+ \Big(\alpha \int_X q d\lambda + (1-\alpha) \int_X r d\lambda - \int_X q^\alpha r^{1-\alpha} d\lambda\Big)^\alpha \geqq$$
$$\geqq \Big(\alpha \int_X p d\lambda + (1-\alpha) \int_X r d\lambda - \int_X p^\alpha r^{1-\alpha} d\lambda\Big)^\alpha$$

zu beweisen (wobei P, Q und R beliebige, durch λ dominierte Masse bedeuten und $0 < \alpha \leqq \frac{1}{2}$ ist), reicht mit Rücksicht auf Folgerung 3 des Satzes 2 aus, die Quasientfernungseigenschaft der Funktion

$$\varphi(u, v) = \big(\alpha u + (1-\alpha) v - u^\alpha v^{1-\alpha}\big)^\alpha \qquad \left(0 < \alpha \leqq \frac{1}{2}\right) \tag{24}$$

über der nichtnegativen Halbgeraden einzusehen.

Dass $\varphi(u, v)$ für $u \geqq 0, v \geqq 0$ stets definiert ist, folgt aus der Ungleichung zwischen den arithmetischen und geometrischen Mitteln. Ferner ist $\varphi(u, v)$ für $u \geqq 0$, $v \geqq 0$ offenbar stetig und seine zweite gemischte Ableitung für $u > 0$, $v > 0$, $u \neq v$ existiert. Also genügt zu zeigen, dass die Funktion (24) die Bedingungen 1) bis 3) des Hilfssatzes 3 erfüllt $\big($mit $E = (0, +\infty)\big)$.

Die Gültigkeit von 1) ist trivial. Ferner gilt

$$\varphi_u = \alpha^2\big(\alpha u + (1-\alpha) v - u^\alpha v^{1-\alpha}\big)^{\alpha-1} (1 - u^{\alpha-1} v^{1-\alpha}),$$
$$\varphi_v = \alpha(1-\alpha)\big(\alpha u + (1-\alpha) v - u^\alpha v^{1-\alpha}\big)^{\alpha-1} (1 - u^\alpha v^{-\alpha}),$$

woraus 2) unmittelbar folgt.

Was 3) anbetrifft, ist

$$\varphi_{uv} = \alpha^2(1-\alpha)\big(\alpha u + (1-\alpha) v - u^\alpha v^{1-\alpha}\big)^{\alpha-2} \times$$
$$\times \left[\alpha u^{2\alpha-1} v^{1-2\alpha} + (1-2\alpha) u^\alpha v^{-\alpha} - (1-\alpha)\right], \tag{25}$$

was sich durch einfache Rechnung ergibt.

Setzen wir $u^\alpha v^{-\alpha} = x$. Dann ist der letzte Faktor der rechten Seite von (25) gleich

$$\alpha x^{2-\frac{1}{\alpha}} + (1-2\alpha) x - (1-\alpha).$$

Dieser Ausdruck ist aber für $0 < \alpha \leqq \frac{1}{2}$ nichtnegativ, da die Gerade $y = -(1-2\alpha)x + (1-\alpha)$ eine Tangente der konvexen Kurve $y = \alpha x^{2-\frac{1}{\alpha}}$

ist (und sie im Punkte $(1, \alpha)$ berührt). Daraus sieht man gleich, dass φ_{uv} im Falle $0 < \alpha \leqq \frac{1}{2}$ nichtnegativ (für $0 < \alpha < \frac{1}{2}$ sogar streng positiv) ist.

Damit ist die Quasientfernungseigenschaft von (24) und dadurch auch Satz 3 vollständig bewiesen.

Für $0 < \alpha < \frac{1}{2}$ ersieht man auch, mit Rücksicht auf die Bemerkung zu Hilfssatz 3, dass die Gleichheit in der Dreiecksungleichung bezüglich $\Delta_\alpha(P, Q)$ für P, Q und R nur in den Fällen $P = Q$ bzw. $Q = R$ besteht. Dass diese Behauptung auch für $\alpha = \frac{1}{2}$ gültig ist, folgt am einfachsten aus der Bemerkung dieses Satzes.

Folgerung. Die Grösse $(1 - \mathcal{J}_\alpha^\gamma(P, Q))^\beta$ ist für jede $0 < \beta \leqq \alpha' = \min(\alpha, 1 - \alpha)$ und $\gamma \geqq 1$ eine Quasientfernung im Raume der Wahrscheinlichkeitsverteilungen auf (X, S).

Beweis. Für $\gamma = 1$ ist die Folgerung mit Rücksicht auf die Bemerkung zur Folgerung 2 des Satzes 2 trivial, da $(1 - \mathcal{J}_\alpha(P, Q))^\beta$ die $\frac{\beta}{\alpha'}$-te Potenz, also für $0 < \beta < \alpha'$ eine konkave Funktion von $\Delta_\alpha(P, Q)$ ist. Was den Fall $\gamma > 1$ anbetrifft, genügt nach der erwähnten Bemerkung zu zeigen, dass — mit der Bezeichnung $\psi_1(u) = u^\beta$, $\psi_2(u) = (1 - (1 - u)^\gamma)^\beta$ — der Quotient $\frac{\psi_2'(u)}{\psi_1'(u)}$ im Intervalle $(0, 1)$ monoton abnimmt. (Die Bemerkung wird auf die Funktionen $\psi_1(u)$ und $\psi_2(u)$ mit der Substitution $u = 1 - \mathcal{J}_\alpha(P, Q)$ angewendet). Da

$$\frac{\psi_2'(u)}{\psi_1'(u)} = \gamma \frac{(1 - (1 - u)^\gamma)^{\beta-1}(1 - u)^{\gamma-1}}{u^{\beta-1}} = \gamma \left[\frac{u(1 - u)^{\frac{\gamma-1}{1-\beta}}}{1 - (1 - u)^\gamma}\right]^{1-\beta}$$

ist, soll man dazu das Abnehmen der Funktion

$$f(u) = \frac{u(1 - u)^\vartheta}{1 - (1 - u)^\gamma} \qquad \left(\vartheta = \frac{\gamma - 1}{1 - \beta} > 0\right)$$

verifizieren. Eine einfache Rechnung ergibt

$$f'(u) = \frac{(1 - u)^{\vartheta-1}}{[1 - (1 - u)^\gamma]^2}\left(1 - (1 + \vartheta)\, u + (\vartheta - \gamma)\, u(1 - u)^\gamma - (1 - u)^{\gamma+1}\right),$$

also ist das Vorzeichen von $f'(u)$ $(0 < u < 1)$ gleich dem Vorzeichen der Funktion $g(u) = 1 - (1 + \vartheta)u + (\vartheta - \gamma)u(1 - u)^\gamma - (1 - u)^{\gamma+1}$. Hier ist $g(0) = g'(0) = 0$, $g(1) = -\vartheta < 0$, ferner gilt wegen

$$g''(u) = \gamma(1 - u)^{\gamma-2}\left[(1 + \gamma\vartheta + \vartheta - \gamma^2)\, u - (2\vartheta - \gamma + 1)\right]$$

$g''(0) = -\gamma(2\vartheta - \gamma + 1) < 0$ (da $\vartheta = \frac{\gamma - 1}{1 - \beta} > \gamma - 1 > 0$ ist) und die Funktion $g(u)$ kann im Intervall $(0, 1)$ höchstens eine Inflexion besitzen. Daraus folgt schon, dass die Kurve $g(u)$ im Intervall $(0, 1)$ unter der Abszissenachse liegt.

Damit ist das Abnehmen der Funktion $f(u)$ und dadurch auch die Folgerung vollständig bewiesen.

Für $0 < \gamma < 1$ $(0 < \beta \leqq \alpha')$ ist das Problem, ob und bei welchen Werten von γ $\left(1-\mathcal{I}_\alpha^\gamma(P, Q)\right)^\beta$ eine Quasientfernung darstellt, noch nicht entschieden. Hier sei nur behauptet, dass man zu jedem $\beta > 0$ ein so kleines $\gamma > 0$ wählen kann, dass der obige Ausdruck schon keine Quasientfernung ist.

Betrachte man nämlich die Verteilungen $P=(1, 0)$, $Q=(a, b)$ $(a + b = 1)$, $R = (0, 1)$. Falls $\left(1 - \mathcal{I}_\alpha^\gamma(P, Q)\right)^\beta$ eine Quasientfernung ist, soll für diese

$$(26)\qquad (1 - a^{\gamma(1-\alpha)})^\beta + (1 - b^{\gamma\alpha})^\beta \geqq 1$$

gelten. Für genügend kleines $\gamma > 0$ kann dies jedoch nicht mehr zutreffen, da die linke Seite von (26) für $\gamma \to 0$ gegen 0 strebt.

Bemerken wir jedoch, dass falls die Grösse $\left(1 - \mathcal{I}_\alpha^\gamma(P, Q)\right)^\beta$ mit einem $\gamma > 0$ eine Quasientfernung darstellt, so gilt dies auch für jedes $\gamma' > \gamma$. Das kann mit Hilfe der im Beweis der Folgerung des Satzes 3 benützten Funktionen $\psi_1(u)$ und $\psi_2(u)$ eingesehen werden, bloss soll man γ durch $\frac{\gamma'}{\gamma}$ und die Substitution $u = 1 - \mathcal{I}_\alpha(P, Q)$ durch $u = 1 - \mathcal{I}_\alpha^\gamma(P, Q)$ ersetzen.

Die bisherigen Betrachtungen ermöglichen eine ziemlich ausführliche Diskussion der am Anfang dieses Paragraphen gestellten Frage über die Funktionen der Form (15) von $I_\alpha(P \| Q)$. Es gilt nämlich

Satz 4. *Im Falle $0 < \beta < \alpha' = \min(\alpha, 1 - \alpha)$ $(0 < \alpha < 1)$ lässt sich immer eine (von α und β abhängende) Konstante $c > 0$ derart angeben, dass*

$$(27)\qquad \min\left(I_\alpha^\beta(P \| Q), c^\beta\right)$$

eine Quasientfernung im Raume der Wahrscheinlichkeitsverteilungen ist.

Im Falle $\beta > \alpha'$ jedoch ist dies nicht möglich.

Beweis. Die negative Behauptung ist bereits bewiesen. Die positive Behauptung lässt sich aus Satz 3 hergeleitet werden. Dazu genügt es mit Rücksicht auf die Bemerkung zur Folgerung 2 des Satzes zu zeigen, dass für die Funktionen

$$\psi_1(u) = (1 - e^{-u})^{\delta_1}, \quad \psi_2(u) = \left(\frac{u}{1-\alpha}\right)^{\delta_2} \qquad (0 < \delta_2 < \delta_1 < 1)$$

der Quotient $\frac{\psi_2'(u)}{\psi_1'(u)}$ im Intervall $(0, \eta)$ monoton abnimmt, falls $\eta > 0$ genügend klein ist. (Nämlich ist bei $\delta_1 = \alpha'$, $\delta_2 = \beta$ mit $u = (1 - \alpha) I_\alpha(P \| Q)$

$$\psi_1(u) = \left(1 - \mathcal{I}_\alpha(P, Q)\right)^{\alpha'}, \quad \psi_2(u) = I_\alpha^\beta(P \| Q)\,.)$$

Durch Differenzieren der beiden Funktionen $\psi_1(u)$ und $\psi_2(u)$ erhält man

$$\frac{\psi_2'(u)}{\psi_1'(u)} = \frac{\delta_2}{(1-\alpha)^{\delta_2}\,\delta_1(1-\delta_1)^{1-\delta_2}} \left(\frac{e^{\frac{u}{1-\delta_1}} - e^{\delta_1 \frac{u}{1-\delta_1}}}{\left(\frac{u}{1-\delta_1}\right)^\tau} \right)^{1-\delta_1} \qquad \left(\tau = \frac{1-\delta_2}{1-\delta_1} > 1\right).$$

Da mit der Bezeichnung $g(u) = \dfrac{e^u - e^{\delta_1 u}}{u^\tau}$

$$g'(u) = \frac{e^u - \delta_1 e^{\delta_1 u} - \tau u^{-1}(e^u - e^{\delta_1 u})}{u^\tau} \tag{28}$$

gilt, erhält man $\lim\limits_{u\to+0} u^\tau g'(u) = (1-\delta_1)(1-\tau) < 0$. Das bedeutet aber, dass die Funktion $g(u)$ und damit auch $\dfrac{\psi_2'(u)}{\psi_1'(u)}$ in einem genügend kleinen Intervall $(0, \eta)$ monoton abnimmt, wodurch Satz 4 bewiesen ist.

Es sei bemerkt, dass die Frage, ob ein Ausdruck der Form $\min(I_\alpha^\beta(P \,||\, Q), c^\beta)$ auch für $\beta = \alpha' = \min(\alpha, 1-\alpha)$ eine Quasientfernung darstellen kann, noch offen ist (mit der Ausnahme des Falles $\alpha = 1/2$, s. am Ende dieses Paragraphen). Mit der obigen Methode kann diese Frage darum nicht entschieden werden, weil im Falle $\delta_1 = \delta_2 = \beta$, also $\tau = 1$ der Quotient $\dfrac{\psi_2'(u)}{\psi_1'(u)}$ selbst für kleine Werte von u zunimmt (vgl. (20) — mit verkehrter Bezeichnung).

Durch den Bewiesenen kann die Frage, ob vielleicht die Grösse $(1 - \mathcal{J}_\alpha^\gamma(P, Q))^\beta$ bei irgendwelchem γ auch für ein $\beta > \alpha'$ eine Quasientfernung darstellt, bereits negativ beantwortet werden. Falls man nämlich in den im Beweis des Satzes 4 auftretenden Funktionen $\psi_1(u)$ bzw. $\psi_2(u)$ mit $\delta_1 = \beta$ und $\delta_2 = \delta < \beta$ $u = \gamma(1-\alpha) I_\alpha(P \,||\, Q)$ setzt, ergibt sich

$$\psi_1(u) = (1 - \mathcal{J}_\alpha^\gamma(P, Q))^\beta, \quad \psi_2(u) = \gamma^\delta I_\alpha^\delta(P \,||\, Q).$$

Also folgt aus dem Abnehmen von $\dfrac{\psi_2'(u)}{\psi_1'(u)}$ für kleine Werte[4] von u nach der Bemerkung zur Folgerung 2 des Satzes 2, dass falls $(1 - \mathcal{J}_\alpha^\gamma(P, Q))^\beta$ für genügend kleine Informationswerte die Dreiecksungleichung befriedigt, so gilt dasselbe auch für $I_\alpha^\delta(P \,||\, Q)$ $(0 < \delta < \beta)$. Falls aber $\beta > \alpha'$ ist, erhält man daraus z. B. mit der Wahl $\delta = \dfrac{\alpha' + \beta}{2}$ einen Widerspruch. (Es sei erwähnt, dass die eben bewiesene negative Behauptung auch mit Hilfe des an der Seite 167 angeführten Gegenbeispiels einfach zu verifizieren ist).

Auf Grund derselben Methode kann gezeigt werden, dass bei $\beta < \alpha'$ zu beliebigem $\gamma > 0$ stets eine positive Zahl c_γ (die auch von β und α' abhängt) derart gewählt werden kann, dass

$$\min\left[(1 - \mathcal{J}_\alpha^\gamma(P, Q))^\beta, c_\gamma^\beta\right] \tag{29}$$

eine Quasientfernung darstellt (wobei für genügend kleines $\gamma > 0$ bestimmt $c_\gamma < 1$ sein muss, vgl. die über (26) gesagten). Dies folgt nämlich nach der Bemerkung zur Folgerung 2 des Satzes 2 aus der Behauptung des Satzes 4, nach welcher (27) eine Quasientfernung ist, und aus (20) — mit der Substitution $u = (1-\alpha)\gamma I_\alpha(P \,||\, Q)$ in den dortigen Funktionen $\psi_1(u)$ und $\psi_2(u)$.

---

[4] Oben wurde das Abnehmen von $\dfrac{\psi_2'(u)}{\psi_1'(u)}$ $(0 < \delta_2 < \delta_1)$ bei kleinen u-Werten nur für $\delta_1 < 1$ gezeigt. Ebenso leicht verifiziert man es aber auch für $\delta_1 \geq 1$.

Aus den obigen sieht man auch, dass für die Konstan te c_γ die Zah

(30) $$c_\gamma = 1 - e_{-(1-\alpha)\gamma c}\,,$$

also für $\gamma \to 0$ $c_\gamma \sim (1-\alpha)\,\gamma\,c$ gewählt werden kann, wobei c die Konstante in (27) bedeutet.

Umgekehrt folgt daraus, dass (29) mit $c_\gamma = (1-\alpha)\,\gamma\,c$ (wobei $c > 0$ beliebig ist) wenigstens für genügend kleine γ-Werte eine Quasientfernung darstellt, die Quasientfernungseigenschaft von (27) mit diesem c. Nämlich ist $\lim\limits_{\gamma\to 0}\dfrac{1-x^\gamma}{\gamma} = -\log x$, also kann die Dreiecksungleichung für $\min\left(I_\alpha^\beta(P\,||\,Q), c^\beta\right)$ aus derselben für $\min\left((1-\mathcal{I}_\alpha^\gamma(P,Q))^\beta; (1-\alpha)^\beta\gamma^\beta c^\beta\right)$ durch den Grenzübergang $\gamma\to 0$ hergeleitet werden. Falls wir also mit $\bar c$ bzw. $\bar c_\gamma$ die grösstmöglichen Konstanten c bzw. c_γ in (27) bzw. (29) bezeichnen, so gilt offenbar

(31) $$\bar c = \varliminf_{\gamma\to 0}\frac{\bar c_\gamma}{\gamma}\,.$$

Bemerken wir noch, dass die Frage, ob eine Grösse der Form $\min\left((1-\mathcal{I}_\alpha^\gamma(P,Q))^{\alpha'}, c_\gamma^\alpha\right)$ mit $0<\gamma<1$ der Dreiecksungleichung genügen kann, ebenso wie das analoge Problem für $I_\alpha(P\,||\,Q)$ selbst, im allgemeinen noch nicht entschieden ist.

Die bisher betrachteten Informationsentfernungen wurden aus Satz 3 auf Grund des Satzes 2 hergeleitet. Man könnte leicht daran denken, dass dies mit allen möglichen Quasientfernungen der Fall ist. Zum Abschluss dieses Paragraphen möchten wir jedoch ein spezielles Gegenbeispiel anführen, und zwar für $\alpha = \dfrac{1}{2}$. Die Grösse

(32) $$\operatorname{arc\,cos}\mathcal{I}_\alpha(P,Q)$$

genügt nämlich im Spezialfall $\alpha = \dfrac{1}{2}$ der Dreiecksungleichung, da sie dann offenbar der Winkel im Hilbertschen Raum der endlichen signierten Masse (s. Bemerkung des Satzes 3) zwischen den — am Einheitskugel liegenden — Verteilungen P und Q ist.

Diese Quasientfernung (die sogar eine Entfernung ist) kann aber aus Satz 3 mit unseren üblichen Methoden nicht abgeleitet werden, da sie aus $(1-\mathcal{I}_{\frac{1}{2}}(P,Q))^{\frac{1}{2}}$ durch die streng konvexe — also keineswegs subadditive — Funktion $\psi(u) = 2\operatorname{arc\,sin}\dfrac{u}{\sqrt{2}}$ erzeugt werden kann. Die Grösse (32) — mit $\alpha = \dfrac{1}{2}$ — ist sogar eine in gewissem Sinne möglichst schärfste Entfernung, indem sie keine streng subadditive Funktion anderer Entfernungen sein kann (wobei die strenge Subadditivität bedeutet, dass die Gleichheit in (D′) nur dann besteht, wenn entweder alle u_i oder alle v_i gleich Null sind). Diese Eigenschaft folgt unmittelbar daraus, dass in der Dreiecksungleichung bezüglich (32) für $\sqrt{q} = \varkappa_1\sqrt{p} + \varkappa_2\sqrt{r}$ ($\varkappa_1 \geqq 0, \varkappa_2 \geqq 0$) die Gleichheit gilt (wenn also der zugrunde liegende Raum X nur aus zwei Punkten besteht, gilt für beliebige drei Verteilungen auf X bei ihrer geeigneten Reihenfolge stets die Gleichheit). Bei den anderen in dieser Arbeit angeführten Quasientfernungen (Entfernungen) kann jedoch die Gleichheit in der Dreiecksungleichung nur in den Fällen $P = Q$ bzw. $Q = R$ bestehen, was eine unmittelbare Folge der Tatsache ist, dass sie aus $\Delta_\alpha(P, Q)$ mittels subadditiver Funktionen hergeleitet werden können.

Als eine Folge der gesagten sei erwähnt, dass im Falle $\alpha = \frac{1}{2}$ genau dann eine Entfernung der Form $\min\left((1-\mathcal{J}^{\gamma}_{\frac{1}{2}}(P, Q))^{\frac{1}{\gamma}}, c^{\frac{1}{\gamma}}\right)$ existiert, falls $\gamma > \frac{2}{3}$ ist[5] — nämlich entsteht $(1-\mathcal{J}^{\gamma}_{\frac{1}{2}}(P, Q))^{\frac{1}{\gamma}}$ aus (32) durch die Funktion $\psi(u) = (1 - \cos^{\gamma} u)^{\frac{1}{\gamma}}$, die für genügend kleine Werte von u bei $0 < \gamma \leqq \frac{2}{3}$ streng konvex (also die inverse Funktion streng konkav und um so mehr streng subadditiv), bei $\gamma > \frac{2}{3}$ dagegen konkav ist. Um so weniger kann eine Grösse der Form $\min\left(I^{\frac{1}{\gamma}}_{\frac{1}{2}}(P \| Q), c^{\frac{1}{\gamma}}\right)$ eine Entfernung sein. Dadurch haben wir die Lösung der im Vorigen gestellten Probleme im Spezialfall $\alpha = \frac{1}{2}$ angegeben, und zwar in negativem Sinne. Die Verfasser vermuten jedoch, dass für $0 < \alpha < \frac{1}{2}$ die eben erwähnten Fragen positiv zu beantworten sind.

§ 4. Informations- und Divergenzentfernungen

Wie bereits erwähnt wurde, kann wegen der Asymmetrie der Informationsgrösse $I_\alpha(P \| Q)$ $\left(\alpha \neq \frac{1}{2}\right)$ keine ihrer Funktionen eine Entfernung darstellen. Es besteht jedoch die Möglichkeit, durch symmetrische Funktionen von $I_\alpha(P \| Q)$ und $I_\alpha(Q \| P)$ Entfernungen im Raum der Wahrscheinlichkeitsverteilungen zu erzeugen. Falls z. B. $f(I_\alpha(P \| Q))$ eine Quasientfernung ist, so sind die Grössen

$$f(I_\alpha(P \| Q)) + f(I_\alpha(Q \| P))$$

bzw.

$$\max\left[f(I_\alpha(P \| Q)), f(I_\alpha(Q \| P))\right]$$

offenbar Entfernungen. Diese sind die beiden Grenzfälle $\omega = 1$ bzw. $\omega = +\infty$ der Informationsentfernungen der Form

$$[f^{\omega}(I_\alpha(P \| Q)) + f^{\omega}(I_\alpha(Q \| P))]^{\frac{1}{\omega}} \tag{33}$$

---

[5] Für $0 < \gamma \leqq \frac{2}{3}$ gilt die Dreiecksungleichung nicht einmal für die Verteilungen der Form (17), falls nur $\delta > 0$ genügend klein ist, wie es den an der Seite 168 gesagten ähnlich leicht zu zeigen ist.

(wobei $f(I_\alpha(P \,\|\, Q))$ eine beliebige Informationsquasientfernung bedeutet), deren Entfernungseigenschaft aus Folgerung 2 des Satzes 2 mit der Wahl $\psi(u_1, u_2) = (u_1^\omega + u_2^\omega)^{\frac{1}{\omega}}$ folgt, berücksichtigend, dass dieses $\psi(u_1, u_2)$ nach der Minkowskischen Ungleichung subadditiv ist.

Falls man in (33) speziell

$$f(I_\alpha(P \,\|\, Q)) = (1 - \mathcal{J}_\alpha(P, Q))^{\alpha'} = \Delta_\alpha(P, Q)$$

(vgl. Satz 3) und $\omega = \dfrac{1}{\alpha'}$ setzt, ergibt sich, dass die Grösse

$$(2 - \mathcal{J}_\alpha(P, Q) - \mathcal{J}_\alpha(Q, P))^{\alpha'} \text{ und damit auch } \left(1 - \frac{\mathcal{J}_\alpha(P, Q) + \mathcal{J}_\alpha(Q, P)}{2}\right)^{\alpha'}$$

eine Entfernung darstellt.

Obzwar wir einige Informationsentfernungen bereits angeben konnten, wäre zur positiven Beantwortung der in § 1 gestellten Frage die Herleitung einer Informationsentfernung, die eine Funktion der Divergenz ist, nötig. In dieser Hinsicht können wir folgendes behaupten:

Satz 5. *Die durch die Divergenzumgebungen $V_\alpha^*(Q, \varepsilon)$ $(0 < \alpha < 1)$ erzeugte Topologie kann durch eine Funktion der Divergenz der Ordnung α metrisiert werden; eine solche Funktion ist*

$$\Delta_\alpha^*(P, Q) = (1 - \mathcal{J}_\alpha(P, Q)\,\mathcal{J}_\alpha(Q, P))^{\alpha'} = (1 - e^{-(1-\alpha)J_\alpha(P,Q)})^{\alpha'} \tag{34}$$

$$(\alpha' = \min(\alpha, 1 - \alpha)).$$

Beweis. Die Entfernunseigenschaft der Grösse (34) wird aus Satz 3 hergeleitet. Dazu genügt wegen

$$\Delta_\alpha^*(P, Q) = \left[\Delta_\alpha^{\frac{1}{\alpha'}}(P, Q) + \Delta_\alpha^{\frac{1}{\alpha'}}(Q, P) - \Delta_\alpha^{\frac{1}{\alpha'}}(P, Q)\,\Delta_\alpha^{\frac{1}{\alpha'}}(Q, P)\right]^{\alpha'}$$

— mit Rücksicht auf Folgerung 2 des Satzes 2 — zu zeigen, dass die Funktion

$$\psi(u_1, u_2) = (u_1^\omega + u_2^\omega - u_1^\omega u_2^\omega)^{\frac{1}{\omega}} \qquad (\omega \geqq 1) \tag{35}$$

die Bedingungen (C′), (D′), (E′) (mit $n = 2$, $m_1 = m_2 = 1$) befriedigt. Das Erfülltsein von (C′) und (E′) ist trivial. Das von (D′), also die Subadditivität von $\psi(u_1, u_2)$ kann folgendermassen bewiesen werden:

Die Funktion $f(x) = (A + x^\omega)^{\frac{1}{\omega}}$ ist bei $A \geqq 0$, $\omega \geqq 1$, $x \geqq 0$ konvex, da

$$f'(x) = (A + x^\omega)^{\frac{1}{\omega} - 1} x^{\omega - 1} = (Ax^{-\omega} + 1)^{\frac{1}{\omega} - 1}$$

monoton zunimmt. Also gilt nach der Jensenschen Ungleichung bei beliebigen nichtnegativen a_1, a_2, x_1, x_2 mit $a_1 + a_2 > 0$

$$a_1 f(x_1) + a_2 f(x_2) \geqq (a_1 + a_2)\, f\left(\frac{a_1 x_1 + a_2 x_2}{a_1 + a_2}\right).$$

Setzt man hier $a_1 = u_1$, $a_2 = v_1$, $x_1 = \dfrac{u_2}{u_1}$, $x_2 = \dfrac{v_2}{v_1}$ und $A = 1 - (u_2 + v_2)^\omega$, so ergibt sich

$$u_1\left[1+\left(\frac{u_2}{u_1}\right)^{\omega}-(u_2+v_2)^{\omega}\right]^{\frac{1}{\omega}}+v_1\left[1+\left(\frac{v_2}{v_1}\right)^{\omega}-(u_2+v_2)^{\omega}\right]^{\frac{1}{\omega}}\geqq$$

$$\geqq(u_1+v_1)\left[1+\left(\frac{u_2+v_2}{u_1+v_1}\right)^{\omega}-(u_2+v_2)^{\omega}\right]^{\frac{1}{\omega}}\quad(u_1,v_1>0\,;\,u_2,v_2\geqq 0\,;\,u_2+v_2\leqq 1),$$

woraus die Ungleichung

$$(36)\qquad \begin{aligned}&(u_1^{\omega}+u_2^{\omega}-u_1^{\omega}u_2^{\omega})^{\frac{1}{\omega}}+(v_1^{\omega}+v_2^{\omega}-v_1^{\omega}v_2^{\omega})^{\frac{1}{\omega}}\geqq\\ &\geqq[(u_1+v_1)^{\omega}+(u_2+v_2)^{\omega}-(u_1+v_1)^{\omega}(u_2+v_2)^{\omega}]^{\frac{1}{\omega}}\end{aligned}$$

(bei den obigen Einschränkungen) unmittelbar folgt. Wegen der Stetigkeit gilt (36) offenbar auch dann, falls man auch $u_1=0$ bzw. $v_1=0$ zulässt. Das bedeutet aber, dass die Funktion (35) auch der Bedingung (D') genügt,[6] wodurch Satz 5 vollständig bewiesen ist.

Bemerkung. In § 3 wurde gezeigt, dass bei $\beta<\alpha'$ zu jedem $\gamma>0$ eine positive Zahl c_γ derart gegeben werden kann, dass $\min\left((1-\mathcal{I}_\alpha^\gamma(P,Q))^\beta, c_\gamma^\beta\right)$ (s. (29)) eine Quasientfernung ist. Mit Hilfe des obigen Gedankens (durch die Anwendung derselben Funktion $\psi(u_1,u_2)$) sieht man, dass dann auch

$$(37)\qquad \min\left((1-\mathcal{I}_\alpha^\gamma(P,Q)\,\mathcal{I}_\alpha^\gamma(Q,P))^\beta, c_\gamma^\beta\right)$$

eine Entfernung darstellt.

Was den Fall $\beta>\alpha'$ anbetrifft, gilt

Hilfssatz 5. *Die Dreiecksungleichung für* $(1-\mathcal{I}_\alpha(P,Q)\,\mathcal{I}_\alpha(Q,P))^\beta$ *ist selbst für beliebig kleine Werte der in ihr auftretenden Grössen (also auch der entsprechenden Divergenzen) nicht allgemein gültig.*

Beweis. Man betrachte wie in § 3 die Verteilungen der Form

$$P=(0,1)\,,\quad Q_\delta=(\delta,1-\delta)\,,\quad R_\varepsilon=(\varepsilon,1-\varepsilon)\,.$$

Um unsere Behauptung zu beweisen genügt wegen

$$\begin{aligned}&(1-\mathcal{I}_\alpha(P,Q_0)\,\mathcal{I}_\alpha(Q_0,P))^\beta+(1-\mathcal{I}_\alpha(Q_0,R_\varepsilon)\,\mathcal{I}_\alpha(R_\varepsilon,Q_0))^\beta=\\ &=(1-\mathcal{I}_\alpha(P,R_\varepsilon)\,\mathcal{I}_\alpha(R_\varepsilon,P))^\beta\end{aligned}$$

zu zeigen, dass

$$(38)\quad h(\delta)=\frac{d}{d\delta}\left[(1-\mathcal{I}_\alpha(P,Q_\delta)\,\mathcal{I}_\alpha(Q_\delta,P))^\beta+(1-\mathcal{I}_\alpha(Q_\delta,R_\varepsilon)\,\mathcal{I}_\alpha(R_\varepsilon,Q_\delta))^\beta\right]$$

[6] Aus der bewiesenen Subadditivität der Funktion (35) kann das allgemeine Resultat hergeleitet werden, dass die Funktion $\psi(u_1,\ldots,u_n)=F^{-1}\left(\sum_{k=1}^{n}F(u_k)\right)$ mit der Wahl $F(u)=-\log(1-u^{\omega})$ subadditiv ist (also die Eigenschaft (D') hat). Dieses Ergebnis enthält auch die Minkowskische Ungleichung für Vektorräume.

bei jedem festen $\varepsilon > 0$ für genügend kleines $\delta > 0$ negativ ist. Da

$$(1 - \mathcal{J}_\alpha(P, Q_\delta)\, \mathcal{J}_\alpha(Q_\delta, P))^\beta + (1 - \mathcal{J}_\alpha(Q_\delta, R_\varepsilon)\, \mathcal{J}_\alpha(R_\varepsilon, Q_\delta))^\beta =$$

$$= \delta^\beta + \left[1 - (\delta^\alpha \varepsilon^{1-\alpha} + (1-\delta)^\alpha (1-\varepsilon)^{1-\alpha})\,(\delta^{1-\alpha}\varepsilon^\alpha + (1-\delta)^{1-\alpha}(1-\varepsilon)^\alpha\right]^\beta$$

gilt, erhält man durch Differenzieren

$$h(\delta) = \beta\delta^{\beta-1} + \beta\left[1 - (\delta^\alpha \varepsilon^{1-\alpha} + (1-\delta)^\alpha (1-\varepsilon)^{1-\alpha})\,(\delta^{1-\alpha}\varepsilon^\alpha + (1-\delta)^{1-\alpha}(1-\varepsilon)^\alpha)\right]^{\beta-1} \cdot$$

$$\cdot \left[(-\alpha\delta^{\alpha-1}\varepsilon^{1-\alpha} + \alpha(1-\delta)^{\alpha-1}(1-\varepsilon)^{1-\alpha})\,(\delta^{1-\alpha}\varepsilon^\alpha + (1-\delta)^{1-\alpha}(1-\varepsilon)^\alpha) + \right.$$

$$\left. + (\delta^\alpha\varepsilon^{1-\alpha} + (1-\delta)^\alpha(1-\varepsilon)^{1-\alpha})\,(-(1-\alpha)\,\delta^{-\alpha}\varepsilon^\alpha + (1-\alpha)\,(1-\delta)^{-\alpha}(1-\varepsilon)^\alpha\right] \sim$$

$$\sim \beta\delta^{\beta-1} - K_1(\varepsilon)\,\delta^{-\alpha} - K_2(\varepsilon)\,\delta^{\alpha-1},$$

wobei $K_1(\varepsilon)$ und $K_2(\varepsilon)$ für $\varepsilon > 0$ positiv sind. Also gilt für (38) im Falle $\beta > \alpha' = \min(\alpha, 1-\alpha)$

$$\lim_{\delta \to +0} h(\delta) = -\infty,$$

wodurch aber bewiesen ist, dass die Dreiecksungleichung für

$$(1 - \mathcal{J}_\alpha(P, Q)\, \mathcal{J}_\alpha(Q, P))^\beta$$

im Falle $\beta > \alpha'$ selbst bei beliebig kleinen Divergenzen ungültig sein kann.

Satz 5 und Hilfssatz 5 sind dem Satz 3 bzw. Hilfssatz 4 vollständig analogisch, es stehen bloss an Stelle der Grössen der Form $\mathcal{J}_\alpha(P, Q)$ diejenige der Form $\mathcal{J}_\alpha(P, Q)\, \mathcal{J}_\alpha(Q, P)$. Das bedeutet aber, dass sich aus den Behauptungen, die in § 3 aus Satz 3 und Hilfssatz 4 mittels subadditiver Funktionen hergeleitet wurden, wieder gültige Behauptungen ergeben, falls man überall $\mathcal{J}_\alpha(P, Q)\, \mathcal{J}_\alpha(Q, P)$ statt $\mathcal{J}_\alpha(P, Q)$ setzt.

Möchte man das Analogon des Satzes 4 erhalten, soll man den obigen gemäss $J_\alpha(P, Q)$ anstatt $I_\alpha(P \,\|\, Q)$ schreiben. Demnach gilt also

Satz 6. *Im Falle* $\beta < \alpha' = \min(\alpha, 1-\alpha)$ $(0 < \alpha < 1)$ *lässt sich stets eine (von* α *und* β *abhängende) Konstante* $c^* > 0$ *angeben, bei welcher*

$$\min\left(J_\alpha^\beta(P, Q), c^{*\beta}\right)$$

eine Entfernung im Raum der Wahrscheinlichkeitsverteilungen darstellt. Für $\beta > \alpha'$ *ist das dagegen unmöglich.*

Bemerkung. Die positive Behauptung dieses Satzes kann auch unmittelbar aus Satz 4 hergeleitet werden, indem man auf $I_\alpha^\beta(P \,\|\, Q)$ und $I_\alpha^\beta(Q \,\|\, P)$ die — nach der Minkowskischen Ungleichung subadditive — Funktion $\psi(u_1, u_2) = (u_1^{\frac{1}{\beta}} + u_2^{\frac{1}{\beta}})^\beta$ anwendet. Daher ist für c^* die in Satz 4 auftretende Konstante c bestimmt geeignet.

Die Frage, ob sich Satz 6 auf $\beta = \alpha'$ in positivem oder negativem Sinne ausdehnen lässt, ist für $\alpha \neq \frac{1}{2}$ — ebenso wie das analogische Problem in § 3 — offen. ($J_{\frac{1}{2}}^{\frac{1}{2}}(P, Q)$ ist eine Vielfache von $I_{\frac{1}{2}}^{\frac{1}{2}}(P \,\|\, Q)$, für welche wir die obige Frage in § 3 bereits negativ beantwortet haben.)

Beschäftigen wir uns noch ein wenig mit den Analoga der in § 3 über die Ausdrücke $(1 - \mathcal{J}_\alpha^\gamma(P, Q))^\beta$ angeführten Behauptungen. Also ist die Grösse

$$(1 - \mathcal{J}_\alpha^\gamma(P, Q)\, \mathcal{J}_\alpha^\gamma(Q, P))^\beta$$

für $\beta \leqq \alpha'$ und $\gamma \geqq 1$ stets eine Entfernung, für $\beta > \alpha'$ kann sie jedoch selbst für kleine Werte der Divergenzen bei keinem $\gamma > 0$ der Dreiecksungleichung genügen. Falls $\beta < \alpha'$, $0 < \gamma < 1$ ist, so können wir nur behaupten, das

$$\min\left(\left(1 - \mathcal{I}_\alpha^\gamma(P, Q)\, \mathcal{I}_\alpha^\gamma(Q, P)\right)^\beta, c_\gamma^{*\beta}\right) \tag{40}$$

bei geeignetem $c_\gamma^* > 0$ eine Entfernung darstellt[7] und dass (39) selbst keine Entfernung sein kann, falls $\gamma > 0$ genügend klein gewählt wird.

Es ist unentschieden, für welche $0 < \gamma < 1$ (39) bei gegebenem $\beta \leqq \alpha'$ noch eine Entfernung bleibt. Für $\beta = \alpha'$ ist diese Frage auch für (40) offen (mit der Ausnahme von $\alpha = \frac{1}{2}$, wobei sich das Problem wegen $\mathcal{I}_{\frac{1}{2}}(P, Q) = \mathcal{I}_{\frac{1}{2}}(Q, P)$ auf dasjenige in § 3 reduziert und daher sich genau für $\gamma > \frac{1}{3}$ positiv entscheiden lässt). Falls jedoch mit einem $\gamma > 0$ eine dieser beiden Fragen positiv zu beantworten ist, so gilt dies auch für jedes $\gamma' > \gamma$, ebenso wie es auch bei den analogischen Behauptungen bezüglich $\mathcal{I}_\alpha(P, Q)$ der Fall war.

Es liegt die Vermutung nahe, dass (39) mit $\gamma = \frac{1}{2}$ noch selbst bei $\beta = \alpha'$ eine Entfernung darstellen kann. Es scheint nämlich, dass dies ein in »Schärfe« der Quasientfernung $\left(1 - \mathcal{I}_\alpha(P, Q)\right)^{\alpha'}$ entsprechender symmetrischer Ausdruck ist (im Falle $\alpha = \frac{1}{2}$ sind diese beiden Grössen sogar identisch). Mit den in der vorliegenden Arbeit benützten Methode kann diese Hypothese auf Grund des Satzes 3 allerdings nicht verifiziert werden.

Die Verfasser haben die Absicht, auf die hier offen gebliebenen Fragen noch zurückzukehren.

(Eingegangen: 5. Februar, 1962.)

LITERATURVERZEICHNIS

[1] RÉNYI, A.: „Az információelmélet néhány alapvető kérdése”. *A Magyar Tudományos Akadémia III. (Matematikai és Fizikiai) Osztályának Közleményei* **10** (1960) 251—282.

[2] RÉNYI, A.: „On measures of entropy and information”. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, Berkeley and Los Angeles, 1960, 547—561.

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

[4] CZIPSZER, J., persönliche Mitteilung.

[5] CSÁSZÁR, Á.: *Fondements de la Topologie Générale.* Akadémiai Kiadó, Budapest, 1960.

[6] RIBEIRO, H.: „Sur les espaces à métrique faible”. *Portugaliae Math.* **4** (1943) 21—40 und 65—68.

[7] BALANZAT, M.: „On the metrisation of quasi-metric spaces”. *Gaz. Mat., Lisboa* **12** (1951) No. 50, 91—94.

[8] BHATTACHARYYA, A.: „On some analogues of the amount of information and their use in statistical estimation”. *Sankhya* **8** (1946) 1—14.

[7] Dabei ist $c_\gamma^* = 1 - e^{-(1-\alpha)\gamma c^*}$ mit dem c^* in Satz 6 eine geeignete Wahl (dies folgt ebenso wie (30)), ferner kann nach der Bemerkung zu Satz 5 auch $c_\gamma^* = c_\gamma$ gesetzt werden.

ИНФОРМАЦИОННЫЕ РАССТОЯНИЯ В ПРОСТРАНСТВЕ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ

I. CSISZÁR — J. FISCHER

Резюме

В качестве теоретико-информационной меры отличий распределения вероятностей мы можем рассматривать относительную информацию (энтропию) порядка α ($\alpha>0$), соотв. J-дивергенцию. Соответственно этому мы можем говорить об информационных или дивергенционных окрестностях некоторого распределения вероятностей которые в случае $0<\alpha<1$ (и только в этом случае) осуществляют топологию в пространстве распределений, а именно, ту же самую, которую осуществляет вариационное расстояние (§ 1).

Цель статьи задать такие функции от относительной информации порядка α $(0<\alpha<1)$, соответственно дивергенции, которые осуществляют метрику для вышеприведенной топологии, где из-за несимметричности информации мы не требуем симметрии расстояний, иначе говоря, где мы допускаем также т. н. квазирасстояния.

В § 2 авторы приводят несколько общих результатов относящихся к соотношениям между расстояниями и квазирасстояниями. При их помощи в § 3, соотв. § 4 они решают вопрос являются ли некоторые функции определенного типа от информации порядка α $(0<\alpha<1)$, соотв. дивергенции квазирасстояниями соотв. расстояниями.

Они доказывают, что если обозначить относительную информацию порядка α через $I_\alpha(P\,||\,Q)$, J-дивергенцию через $J_\alpha(P, Q)$ и $\min(\alpha, 1-\alpha)$ через α', тогда в случае $0<\beta<\alpha'$ можно задать такое число $c>0$ (зависящее от α и от β), что $\min\left(I_\alpha^\beta(P\,||\,Q), c^\beta\right)$ является квазирасстоянием (теорема 4) а $\min\left(J_\alpha^\beta(P, Q), c^\beta\right)$ является расстоянием (теорема 6) в пространстве распределений вероятностей; однако в случае $\beta>\alpha$ такого $c>0$ уже не существует. Эти результаты авторы выводят из того факта, что выражения вида $[1-\exp\{-(1-\alpha)\,I_\alpha(P\,||\,Q)\}]^\beta$ (теорема 3, по существу от J. Czipszer), соответсвенно $[1-\exp(-(1-\alpha)\,J_\alpha(P, Q))]^\beta$ (теорема 5) удовлетворяют неравенству треугольника точно в случае $0<\beta\leqq\alpha'$.

Они занимаются еще некоторыми другими функциями от информации, соотв. дивергенции и ставят также несколько нерешенных проблем.