B. Forte (Ed.)

CIME Summer Schools

Functional Equations and Inequalities

54

La Mendola, Italy 1970

Springer

FONDAZIONE CIME ROBERTO CONTI

B. Forte (Ed.)

Functional Equations and Inequalities

Lectures given at a Summer School of the
Centro Internazionale Matematico Estivo (C.I.M.E.),
held in La Mendola (Trento), Italy,
August 20-28, 1970

Springer

FONDAZIONE
CIME
ROBERTO CONTI

C.I.M.E. Foundation
c/o Dipartimento di Matematica "U. Dini"
Viale Morgagni n. 67/a
50134 Firenze
Italy
cime@math.unifi.it

ISBN 978-3-642-11002-3 e-ISBN: 978-3-642-11004-7
DOI:10.1007/978-3-642-11004-7
Springer Heidelberg Dordrecht London New York

©Springer-Verlag Berlin Heidelberg 2010
Reprint of the 1st ed. C.I.M.E., Ed. Cremonese, Roma 1971
With kind permission of C.I.M.E.

Printed on acid-free paper

Springer.com

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

III Ciclo - La Mendola - dal 20 al 28 Agosto 1970

"FUNCTIONAL EQUATIONS AND INEQUALITIES"

Coordinatore: Prof. B. FORTE

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C. I. M. E.)

J. ACZÉL

SOME APPLICATIONS OF FUNCTIONAL EQUATIONS AND INEQUALITIES TO INFORMATION MEASURES

Corso tenuto a La Mendola (Trento) dal 20 al 28 agosto 1970

Some Applications of Functional Equations and Inequalities to Information Measures

by

J. Aczél, University of Waterloo, Waterloo, Ont., Canada

1. Let $\Gamma_N = \{(p_1,p_2,\ldots,p_N) \mid \sum_{k=1}^{N} p_k = 1,\ p_k \geq 0,\ k = 1,2,\ldots,N\}$

be the set of all complete finite discrete probability distributions (e.g. the probabilities of different outcomes of an experiment, contents of a communication, etc.) with N members (N = 2,3,...). C. E. Shannon (1948) has introduced the "Shannon entropy" (with the understanding 0 log 0 := 0)

$$(1) \quad H_N(p_1,p_2,\ldots,p_N) := -\sum_{k=1}^{N} p_k \log_2 p_k \text{ for all } (p_1,p_2,\ldots,p_N) \in \Gamma_N,\ n = 2,3,\ldots,$$

as measure of uncertainty (before the experiment was made, the message received etc.) or, equivalently, of information (received from the completed experiment, communication, etc.). What justifies the formula (1) and some further measures of uncertainty and information?

In this lecture we summarize some older and newer results in this direction with some proofs indicated. Detailed proofs (and more results) will be available in the book J. Aczél Z. Daróczy 1971.

2. One justification of (1) lies in coding theory (see e.g. A. Feinstein 1958). There we have a finite set X of messages (letters) to which finite sequences (codewords) of elements of a finite set A of symbols are bijectively associated. Such a mapping is a uniquely decodable code S(X, A). For every such code, the average length of a codeword is

$$(2) \qquad \sum_{k=1}^{N} p_k n_k \geq \frac{H_N(p_1,p_2,\ldots,p_N)}{\log_2 D}$$

where $N = |X|$, $D = |A|$, p_k is the probability of the message $x_k \in X$, and n_k is the length of the codeword associated to x_k $(k = 1,2,\ldots,N)$. On the other hand, there exists a code $S^*(X, A)$ for which the average length of a codeword is

$$\text{(3)} \qquad \sum_{k=1}^{N} p_k n_k^* < \frac{H_N(p_1,p_2,\ldots,p_N)}{\log_2 D} + 1.$$

If we permit L-tuples of messages (letters) to be encoded and take the

$$\frac{1}{L} \sum_{\underline{x}\in X^L} P(\underline{x}) n_{\underline{x}}$$

average length of codewords per messages, then (3) can be improved in the following way. For arbitrary small $\varepsilon > 0$ there exists a code $S^*(X,A)$, such that

$$\text{(4)} \qquad \frac{1}{L} \sum_{\underline{x}\in X^L} P(\underline{x}) n_{\underline{x}}^* < \frac{H_N(p_1,p_2,\ldots,p_N)}{\log D} + \varepsilon .$$

The proof of the first statement (2) is based upon Kraft's inequality

$$\sum_{k=1}^{N} D^{-n_k} \le 1$$

and upon Shannon's inequality

$$\text{(5)} \quad H_N(p_1,p_2,\ldots,p_N) = -\sum_{k=1}^{N} p_k \log p_k \le -\sum_{k=1}^{N} p_k \log q_k$$

for all $(p_1,p_2,\ldots,p_N)\in\Gamma_n$, $(q_1,q_2,\ldots,q_N)\in\Gamma_N$, $q_k > 0$, $k = 1,2,\ldots,N$,

$$\Big(\sum_{k=1}^{N} p_k = \sum_{k=1}^{N} q_k = 1,\ p_k > 0,\ q_k > 0,\ k = 1,2,\ldots,N\Big).$$

The second statement (3) is proved by choosing n_k^* as the (only) integer in the interval

$$[-\frac{\log p_k}{\log D}, -\frac{\log p_k}{\log D} + 1) \quad (k = 1,2,\ldots,N),$$

and the third statement (4) follows from (3) and from an important property of the entropy, viz. additivity, to which we come back in a moment.

In other words, a combination of (2) and (4) means that the only solution of the inequality

$$\forall\varepsilon\ \exists\{n^*_{\underline{x}}\} : \frac{I(p_1,p_2,\ldots,p_N)}{\log D} \le \frac{1}{L}\sum_{\underline{x}\in X^L} P(\underline{x})n^*_{\underline{x}} < \frac{I(p_1,p_2,\ldots,p_N)}{\log D} + \varepsilon$$

is

$$I(p_1,p_2,\ldots,p_N) = H_N(p_1,p_2,\ldots,p_N).$$

The inequalities (2), (3) and (4), of course, give an excellent justification for the formula (1) (although they are based upon a particular way of averaging). So does the similar formula of thermodynamical entropy. These are the two origions of the probabilistic entropy concept.

However, the applications of entropy go farther than that. Also, in some applications, instead of (1), the entropies of order α

(6) $${}_\alpha H_N(p_1,p_2,\ldots,p_N) = \frac{1}{1-\alpha}\log\sum_{k=1}^{N} p_k^\alpha, \quad (\alpha \ne 1;\ (p_1,p_2,\ldots,p_N)\in\Gamma_N,\ N = 2,3,\ldots)$$

introduced by M. P. Schützenberger (1954), S. Kullback (1959) and A. Rényi (1960, 1965), are used instead of (1) (these too can be characterized by coding theorems, see e.g. L. L. Campbell 1966). Notice that $\lim_{\alpha\to 1} {}_\alpha H_N = H_N$, so the Shannon entropy can be considered as an entropy of order 1:

$${}_1 H_N(p_1,p_2,\ldots,p_N) := H_N(p_1,p_2,\ldots,p_N) \quad ((p_1,p_2,\ldots,p_N) \in \Gamma_N,\ N = 2,3,\ldots).$$

Thus it is important to investigate the properties of all these entropies (some of which make the above applications possible) and the question, which of these properties characterize them. In this way we can justify their use and also select in which applications which to use.

3. Some of the most important properties of the Shannon entropy (1) are the following (cf. J. Aczél 1968). We write here I_N instead of H_N, since later we want to find all expressions satisfying these properties.

1. Algebraic Properties

(7) N-Symmetry: I_N is a symmetric function of $p_1,p_2,\dots,p_N$,

or generally,

(8) Symmetry: All I_N $(N = 2,3,\dots)$ are symmetric functions.

(9) Normalization: $I_2(\frac{1}{2}, \frac{1}{2}) = 1$,

more generally,

(10) $I_N(\frac{1}{N},\dots,\frac{1}{N}) = \log N$.

(11) Expansibility: $I_{N+1}(p_1,p_2,\dots,p_N,0) = I_N(p_1,p_2,\dots,p_N)$ for all $(p_1,p_2,\dots,p_N) \in \Gamma_N$, $N = 2,3,\dots$, (this is satisfied for (1) if we agree upon $0 \log 0 := 0$ and for (6) also in the case $\alpha \le 0$, if $0^\alpha := 0$; or can serve as definition for the value of $H_N(p_1,p_2,\dots,p_N)$ and ${}_\alpha H_N(p_1,p_2,\dots,p_N)$ $(\alpha \le 0)$ if one - or more - of the p_k are 0).

(12) Decisivity: $I_2(1, 0) = 0$,

(cf. the remark after (11)).

(13) (M, N)-Additivity: $I_{MN}(p_1q_1,p_1q_2,\dots,p_1q_N,p_2q_1,p_2q_2,\dots,p_2q_N,\dots,p_Mq_1,p_Mq_2,\dots,p_Mq_N) = I_M(p_1,p_2,\dots,p_M) + I_N(q_1,q_2,\dots,q_N)$

$((p_1,p_2,\dots,p_M) \in \Gamma_M,\ (q_1,q_2,\dots,q_N) \in \Gamma_N)$,

or generally,

(14) Additivity: (13) holds for all M, N = 2,3,...

(15) N-Recursivity: $I_N(p_1,p_2,p_3,\dots,p_N) = I_{N-1}(p_1+p_2,p_3,\dots,p_N) +$

$$+ (p_1+p_2)I_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right), \quad ((p_1,p_2,\dots,p_N) \in \Gamma_N,\ p_1 + p_2 > 0),$$

or generally,

(16) Recursivity: (15) holds for all $N = 3,4,\dots$,

Even more generally,

(17) Branching Property: There exist functions J_3, $J_4,\dots$ such that

$$I_N(p_1,p_2,p_3,\dots,p_N) - I_{N-1}(p_1+p_2,p_3,\dots,p_N) = J_N(p_1,p_2),$$

for all $(p_1,p_2,\dots,p_N) \in \Gamma_N$, $N = 3,4,\dots$

2. Representations

(18) Sum: There exists a function g, measurable in]0, 1), with $g(1) = 0$, such that $I_N(p_1,p_2,\dots,p_N) = \sum_{k=1}^{N} g(p_k)$,

$(p_1,p_2,\dots,p_N) \in \Gamma_N$; $(N = 2,3,\dots)$.

(19) Quasilinearity: There exists a continuous and strictly monotonic function ψ or R^+ such that $I_N(p_1,p_2,\dots,p_N) = \psi^{-1}[\sum_{k=1}^{N} p_k\psi(-\log_2 p_k)]$

$((p_1,p_2,\dots,p_N) \in \Gamma_N,\ p_k > 0;\ k = 1,2,\dots,N;\ N = 2,3,\dots)$.

3. Inequalities

(20) Nonnegativity: $I_2(1-q, q) \geq 0$ for all $q \in [0, 1]$.

(21) Maximality: $I_N(p_1,p_2,\dots,p_N) \leq I_N(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}) = \log_2 N$

for all $(p_1,p_2,\dots,p_N) \in \Gamma_N$, $N = 2,3,\dots$) (the last equality comes from (10)).

(22) Subadditivity: $I_{MN}(p_{11},p_{12},\dots,p_{1N},p_{21},p_{22},\dots,p_{2N},\dots,p_{M1},p_{M2},\dots,p_{MN}) \leq$

$$\leq I_M\left(\sum_{k=1}^{N} p_{1k}, \sum_{k=1}^{N} p_{2k}, \dots, \sum_{k=1}^{N} p_{Mk}\right) + I_N\left(\sum_{j=1}^{M} p_{j1}, \sum_{j=1}^{M} p_{j2}, \dots, \sum_{j=1}^{M} p_{jN}\right)$$

$((p_{11},\dots,p_{MN}) \in \Gamma_{MN};\ M,N = 2,3,\dots)$.

(23) Shannon-Inequality: $I_N(p_1,p_2,\ldots,p_N) \leq -\sum_{k=1}^{N} p_k \log q_k$

$(\sum_{k=1}^{N} p_k = \sum_{k=1}^{N} q_k = 1;\ p_k > 0,\ q_k > 0;\ k = 1,2,\ldots,N)$, (cf. (5)).

4. Regularity

(24) **Smallness** for small probabilities: $\lim_{q\to 0+} I_2(1-q,\ q) = 0$.

(25) Boundedness from above: There exists a $k \geq 1$, such that $I_2(1-q,\ q) \leq k$ for all $q \in [0,\ 1]$.

(26) Analyticity: $q \to I_N(p(1-q),\ pq,\ p_3,\ \ldots,\ p_N)$ are analytic in $(0,\ 1)$, $(N = 2,3,\ldots)$.

(27) Measurability: $f = (q \to I_2(1-q,\ q))$ is measurable in $(0,\ 1)$.

As hinted earlier also the entropies of order $\alpha \neq 1$ have most of these properties, - all but (22), (18), (17) [and (16), (15)]. However, (24) and (25) hold only if $\alpha > 0$.

4. We mention here the following characterization results, the first three based on the (16)-recursivity.

Generalizing theorems of A. I. Khinchin (1953) and D. K. Faddeev (1956; cf. A. Feinstein 1958), Z. Daróczy (1969; cf. R. Borges 1967) have proved the following result.

Theorem 1. *Iff $\{I_N\}$ is (7) 3-symmetric, (9) normalized, (16) recursive and (24) small for small probabilities, then*

(28) $I_N(p_1,p_2,\ldots,p_N) \equiv H_N(p_1,p_2,\ldots,p_N)$ *for all* $(p_1,p_2,\ldots,p_N)\in\Gamma_N$, $N = 2,3,\ldots$

The *proof* first shows that (7) 3-symmetry, (9) normalization and (15) 3-recursivity imply that

(29) $$f = (q \to I_2(1-q,\ q))$$

satisfies the functional equations and boundary conditions

$$(30)\begin{cases} f(x) + (1-x)f(\frac{y}{1-x}) = f(y) + (1-y)f(\frac{x}{1-y}) \text{ whenever } x,\ y < [0,1),\ x+y \le 1 \\ f(1-x) = f(x) \quad (x \in [0,\ 1]),\quad f(1) = 0,\quad f(\frac{1}{2}) = 1. \end{cases}$$

Functions satisfying (30) are called information functions. With aid of the information function (29) one gets from the (16) recursivity

(31) $$I_N(p_1,p_2,\ldots,p_N) = \sum_{k=2}^{N} (p_1 + p_2 + \ldots + p_k)f(\frac{p_k}{p_1+p_2+\ldots+p_k}).$$

Then one proves that the arithmetical function

(32) $$\phi = (N \to \frac{1}{N}\sum_{k=1}^{N} kf(\frac{1}{k});\quad N = 1,2,\ldots) \text{ i.e.}\begin{cases}\phi(1) = 0, \\ \phi(N) = I_N(\frac{1}{N},\frac{1}{N},\ \ldots,\ \frac{1}{N}),\ (N=2,3,\ldots\end{cases}$$

is completely additive:

(33) $$\phi(MN) = \phi(M) + \phi(N) \text{ for all } M,\ N = 1,2,\ldots\ .$$

Then, (24), (30), and a particular case of the Lemma 1 below gives

(34) $$\phi(N) = \log_2 N \qquad (n = 1,2,\ldots).$$

from which one deduces, by applying (30) and (31), that

(35) $$f(r) = -r\ \log_2 r - (1-r)\log_2(1-r)$$

for all rational $r \in [0,\ 1]$. Then from (30) and (24) one deduces the continuity of f, thus extending (35) to all reals in $[0,\ 1]$. But this means

$$I_2(p_1,p_2) \equiv H_2(p_1,p_2)$$

and then the (16) recursivity finishes the proof.

This last step and the steps (29)-(33) are used also in the proofs of the next two theorems.

The following result is due to P. M. Lee (1964) and generalizes results of H. Tverberg (1958) and D. G. Kendall (1963, cf. R. Borges 1967).

Theorem 2. Iff $\{I_N\}$ is (7) 3-symmetric, (9) normalized, (16) recursive and (27) measurable, then (28) holds.

In the proof of this theorem, beyond the steps common with the proof of Theorem 1, one has to prove (which is not quite easy) that there exists a nonvanishing interval on which f (as defined by (29)) is bounded and then that this holds on every closed subinterval of (0, 1) so f is Lebesgue integrable on these intervals. Then, integrating (30) with respect to x over [a, b] $(0 < y < y+a < y+b < 1)$, we get

$$(b-a)f(y) = \int_a^b f(x)dx + y^2 \int_{y/(1-a)}^{y/(1-b)} x^{-3} f(x)dx - (1-y)^2 \int_{a/(1-y)}^{b/(1-y)} f(x)dx.$$

Thus f is continuous and the proof can be finished without difficulty.

Finally, the following result is due to Z. Daróczy and I. Kátai (1969) (on basis of Theorem 5 we have dropped one of their assumptions).

Theorem 3. Iff $\{I_N\}$ is (7) 3-symmetric, (9) normalized, (16) recursive, and (20) nonnegative, then (28) holds.

The proof proceeds to (33), as for Theorem 1, then makes use of the following lemma (P. Erdös 1958, I. Kátai 1967).

Lemma 1. Iff a (13) completely additive number theoretical function satisfies

$$\lim\inf[\phi(N+1) - \phi(N)] \geq 0,$$

then there exists a constant c *such that*

$$\phi(N) = c \log_2 N \qquad (N = 1,2,3,\ldots).$$

Again, (30) implies c = 1 and (35). By another nontrivial application of (30), f can be proved continuous, and from there on the proof goes again as that of Theorem 1.

Recently, B. Jessen, J. Karpf and A. Thorup (1968) have proved the following theorem.

Theorem 4. *The general solution of the system*

$$F(a, b) = F(b, a),$$

$$F(a, b) + F(a + b, c) = F(b, c) + F(a, b + c)$$

$$F(ac, bc) = cF(a, b)$$

$(a, b, c \in R^+)$ *of functional equations is of the form*

$$F(x, y) = \phi(x + y) - \phi(x) - \phi(y),$$

where ϕ *satisfies the functional equation*

$$\phi(xy) = y\phi(x) + x\phi(y).$$

If we define f and F by

$$f(x) = F(1-x, x)$$

and

$$F(x, y) = (x + y)f\left(\frac{y}{x + y}\right),$$

respectively, we see that the system of functional equations in Theorem 4 is equivalent to

$$f(1-x) = f(x) \qquad \text{and} \qquad f(x) + (1-x)f\left(\frac{y}{1-x}\right) = f(y) + f\left(\frac{x}{1-y}\right)$$

from (30) and with $L = (x \to \phi(x)/x)$; $(x > 0)$ we get the following general theorem.

Theorem 5. All solutions of (30) are of the form

$$f(x) = -xL(x) - (1-x)L(1-x) \quad (\text{convention: } 0L(0) = 0, \text{ if necessary}),$$

where L satisfies

$$L(xy) = L(x) + L(y), \qquad L(2) = 1.$$

Theorems 1-3 are consequences (essentially) of Theorem 5, but the tools for the proof of Theorem 4 are rather intricate.

5. With respect to the branching property (17), the following holds (generalizing results of B. Forte - Z. Daróczy 1968, Z. Daróczy 1970).

Theorem 6. Iff $\{I_N\}$ is (7) 4-symmetric, (9) normalized, (11) expansible, (13) (2,3)-additive, (17) branching, and (20) nonnegative, then (28) holds.

The proof of this theorem first derives the (2,2)-additivity from the (2,3)-additivity and from the (11) expansibility. Then from these and from (20)

$$J_N(p_1,p_2) = (p_1 + p_2)I_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_1}{p_1 + p_2}\right) \quad (N = 3,4,\ldots; \ 0 < p_1+p_2 \leq 1)$$

follows, that is, $\{I_N\}$ is (16) recursive and thus the proof of Theorem 6 is reduced to that of Theorem 3.

The (17) branching property can also be considered as a representation, in particular the (18) sum representation implies that (17) is satisfied with $J_N(p_1,p_2) = g(p_1) + g(p_2) - g(p_1 + p_2)$.

As to the (18) sum representation, generalizing previous results by T. W. Chaundy - J. B. McLeod (1960) and J. Aczél - Z. Daróczy (1963A), Z. Daróczy (1970) has proved the following.

<u>Theorem 7.</u> <u>Iff</u> $\{I_N\}$ <u>has the (18) sum-representation, is (13) (2,3)-additive, and (9) normalized, then (28) holds.</u>

The <u>proof</u> uses similar tools as that of Theorem 6.

If we restrict ourselves to (0, 1), then we can take in (18) $g(p) = ph(p)$, and we have $I_N(p_1,p_2,\ldots,p_N) = \sum_{k=1}^{N} p_k h(p_k)$, in particular on the lefthand side of the (23) Shannon-inequality. If we write also the righthand side similarly, we can ask the question, which are the functions h for which

$$\sum_{k=1}^{N} p_k h(p_k) \leq \sum_{k=1}^{N} p_k h(q_k) \quad ((p_1,p_2,\ldots,p_N) \in \Gamma_N,\ (q_1,q_2,\ldots,q_N) \in \Gamma_N;\ p_k > 0,\ q_k > 0;\ k = 1,2,\ldots,N). \tag{36}$$

The following theorem is due to P. Fischer (1970; cf. also J. McCarthy 1956, where a slightly misstated similar result is mentioned), who has generalized a result of J. Aczél - J. Pfanzagl (1966).

<u>Theorem 8.</u> <u>The inequality (36) holds for a fixed</u> $N \geq 3$ <u>iff</u>

$$h(p) = c \log p + b \quad \text{for all} \quad p \in (0, 1) \tag{37}$$

<u>where</u> $c \leq 0$, b <u>are constants, that is, iff up to an additive and a multiplicative constant, the lefthand side of (36) is the Shannon entropy</u> $H_N(p_1,p_2,\ldots,p_N)$

(For N = 2 the theorem is not true, see J. Marschak 1959, J. Aczel - J. Pfanzagl 1966).

The <u>proof</u> (essentially due to A. Rényi, unpublished) first derives from (36)

$$p_1[h(q_1) - h(p_1)] \geq p_2[h(p_2) - h(q_2)] \tag{38}$$

for all p_1, p_2, q_1, q_2 for which $p_1 > 0$, $p_2 > 0$, $q_1 > 0$, $q_2 > 0$, $p_1 + p_2 = q_1 + q_2 < 1$. From (38) it follows that h is non-increasing, thus differentiable almost everywhere. The inequality (38) implies also that

(39) $$ph'(p) = c \text{ (constant)}$$

in the points where h is known to be differentiable and in the remaining points for the Dini numbers the inequalities

$$c \leq pD_-h(p) \leq pD^-h(p) \leq c \leq pD_+h(p) \leq pD^+h(p) \leq c$$

hold, thus h is differentiable and (39) holds <u>everywhere</u>. This gives (37) and concludes the proof.

<u>6</u>. The <u>entropies of order</u> $\alpha > 0$ are characterized by the following theorem of Z. Daróczy (1964) which generalizes a previous result of J. Aczél and Z. Daróczy (1963A) and is stated here in a form given by J. Aczél and Z. Daróczy (1963B).

<u>Theorem 9.</u> <u>Iff</u> $\{I_N\}$ <u>is (14) additive, (19) quasilinear and (24) small for small probabilities, then there exists an</u> $\alpha > 0$ <u>such that</u>

$$I_N(p_1,p_2,\ldots,p_N) = {}_\alpha H_N(p_1,p_2,\ldots,p_N) \text{ for all } (p_1,p_2,\ldots,p_n) \in \Gamma_N,\ N = 2,3,\ldots$$

<u>(including, for</u> $\alpha = 1$, <u>the Shannon entropy)</u>.

The <u>proof</u> is based on two lemmas.

<u>Lemma 2.</u> <u>The equality</u>

(40) $$\phi^{-1}\left[\sum_{k=1}^{N} q_k\phi(-\log_2 q_k)\right] = \psi^{-1}\left[\sum_{k=1}^{N} q_k\psi(-\log_2 q_k)\right],$$

<u>where</u> ϕ, ψ <u>are continuous, strictly monotonic on</u> R^+ <u>and</u> $o(2^x)$ <u>as</u> $x \to \infty$, <u>holds for all</u> $(q_1,q_2,\ldots,q_N) \in \Gamma_N$, $q_k > 0$; $k = 1,2,\ldots,N$; $n = 2,3,\ldots$, <u>iff</u>

(41) $$\phi = A\psi + B \quad \text{on } R^+ = [0, \infty)$$

<u>(</u>$A \geq 0$, <u>B are constants)</u>.

If we put (19) and $p_j = \frac{1}{M}$ $(j = 1,2,\dots,M)$ into (14) [cf. (13)], then we get an equality of the form (40) with $\phi(x) = \psi(x - \log M)$, so, by (41)

$$(42) \qquad \psi(x - \log_2 M) = A(M)\psi(x) + B(M) \qquad (x \in [0, \infty), \ \ M = 1,2,\dots)$$

and we apply the following lemma in order to conclude the proof of Theorem 7.

Lemma 3. All continuous, strictly monotonic and $o(2^x)$ (for $x \to \infty$) solutions of (42) are for all $x \in R^+$ either of the form

$$\psi(x) = ax + b$$

or of the form

$$\psi(x) = a2^{(1-\alpha)x} + b \qquad (\alpha > 0, \ \ \alpha \neq 1)$$

$(a \neq 0,\ b$ constants).

7. The characterization of the Shannon entropy and even more so that of the Rényi entropies is simpler if also incomplete finite discrete probability distributions $(p_1,p_2,\dots,p_N)$, for which $0 < \sum_{k=1}^{N} p_k < 1$ $(p_k \geq 0,\ k = 1,2,\dots,N;\ N = 1,2,3,\dots)$, are admitted (cf. A. Rényi 1960, Z. Daróczy 1963, J. Aczél 1964). Then

$$_\alpha H_N(p_1,p_2,\dots,p_N) = \frac{1}{1-\alpha} \log_2\left(\sum_{k=1}^{N} p_k^\alpha \Big/ \sum_{k=1}^{N} p_k\right) \qquad (\alpha \neq 1),$$

$$_1 H_N(p_1,p_2,\dots,p_N) = H_N(p_1,p_2,\dots,p_N) = -\sum_{k=1}^{N} p_k \log_2 p_k \Big/ \sum_{k=1}^{N} p_k .$$

Both these contain the entropy of a single event $H_1(p) = {}_\alpha H_1(p) = -\log p$ $(p \in [0, 1])$ (cf. A. Rényi 1960).

The Shannon inequality (5) can also be written as

$$\sum_{k=1}^{N} p_k \log \frac{p_k}{q_k} \geq 0.$$

The left-hand side (cf. S. Kullback 1959) and the more general expression

$$\sum_{k=1}^{N} p_k \log \frac{q_k}{r_k}$$

for <u>three</u> probability distributions $(p_1,p_2,\ldots,p_N)$, $(q_1,q_2,\ldots,q_N)$, $(r_1,r_2,\ldots,r_N)$ are sometimes called <u>directed divergences</u>. These and their analogues of order α and for incomplete distributions were characterized e.g. by P. Nath (1970) and J. Aczél - P. Nath (1971).

Another direction of research is that of nonprobabilistic information measures (e.g. B. Forte - N. Pintacuda 1968), but this will be the subject of talks by J. Kampé de Feriet and B. Forte.

A summary of this lecture was given at the Meeting on Information Measures, Kitchener-Waterloo, Ont., Canada, on April 10, 1970.

References

1948 C.E. Shannon, A Mathematical Theory of Communication. Bell System Tech. J. 27 (1948), 379-423, 623-656.

1953 A.I. Khinchin, The Concept of Entropy in the Theory of Probability (Russian). Uspehi Mat. Nauk 8 (1953), no. 3 (55), 3-20. English translation in "Mathematical Foundations of Information Theory", Dover, New York, 1957, 1-30.

1954 M.P. Schützenberger, Contribution aux applications statistiques de la théorie de l'information. Publ. Inst. Statist. Univ., Paris, 3 (1954), 3-117.

1956 D.K. Faddeev, On the Concept of Entropy of a Finite Probabilistic Scheme (Russian), Uspehi Mat. Nauk 11 (1956), No. 1 (67), 227-231.

J. McCarthy, Measures of the Value of Information. Proc. Nat. Acad. Sci., U.S.A., 42 (1956), 654-655.

1958 P. Erdös, On the Distribution of Additive Arithmetical Functions. Rend. Sem. Mat. Fis. Milano 27 (1958), 3-7.

A. Feinstein, Foundations of Information Theory. McGraw Hill, New York - Toronto, 1958.

H. Tverberg, A New Derivation of the Information Function. Math. Scand. 6 (1958), 297-298.

1959 S. Kullback, Information Theory and Statistics. Wiley, New York, 1959.

J. Marschak, Remarks on the Economy of Information. Contributions to Scientific Research in Management, Univ. of Calif., Los Angeles 1959, Univ. of Calif. Press, Los Angeles, Calif. 1960, 79-98.

1960 T.W. Chaundy - J.B. McLeod, On a Functional Equation. Proc. Edinburgh Math. Soc. Edinburgh Math. Notes 43 (1960), 7-8.

A. Rényi, On Measures of Entropy and Information. Proc. 4th Berkeley Symp. Math. Statist. and Probability 1960, Univ. of Calif. Press, Berkeley, Calif., 1961, Vol. I, 547-561.

1963 J. Aczél - Z. Daróczy, A, Charakterisierung der Entropien positiver Ordnung und der Shannonschen Entropie. Acta Math. Acad. Sci. Hungar. 14 (1963), 95-121.

B. Sur la caractérisation axiomatique des entropies d'ordre positif, y compris l'entropie de Shannon. C.R. Acad. Sci. Paris 257 (1963), 1581-1584.

Z. Daróczy, Über die gemeinsame Charakterisierung der zu den nicht vollständigen Verteilungen gehörigen Entropien von Shannon und Rényi. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 1 (1963), 381-388.

D.G. Kendall, Functional Equations in Information Theory. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2 (1963), 225-229.

1964 J. Aczél, Zur gemeinsamen Charakterisierung der Entropien α-ter Ordnung und der Shannonschen Entropie nicht unbedingt vollständiger Verteilungen. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 3 (1964), 177-183.

Z. Daróczy, Über Mittelwerte und Entropien vollständiger Wahrscheinlichkeitsverteilungen. Acta Math. Acad. Sci. Hungar. 15 (1964), 203-210.

P.M. Lee, On the Axioms of Information Theory. Ann. Math. Statist. 35 (1964), 414-441.

J. Aczél

1965 A. Rényi, On the Foundations of Information Theory. Rev. Inst. Internat. Statist. 33 (1965), 1-14.

1966 J. Aczél - J. Pfanzagl, Remarks on the Measurement of Subjective Probability and Information. Metrika 11 (1966), 91-105.

L.L. Campbell, Definition of Entropy by Means of a Coding Problem. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 6 (1966), 113-118.

1967 R. Borges, Zur Herleitung der Shannonschen Information. Math. Z. 96 (1967), 282-287.

I. Kátai, A Remark on Additive Arithmetical Functions. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 12 (1967), 81-83.

1968 J. Aczél, On Different Characterizations of Entropies. Proc. Internat. Symp. Probability and Information Theory, McMaster Univ., Canada, April, 1968, Lecture Notes in Math. Vol. 89, Springer, Berlin-Heidelberg-New York, 1969, 1-11.

B. Forte - Z. Daróczy, A Characterization of Shannon's Entropy. Boll. Un. Mat. Ital. (4) 1 (1968), 631-635.

B. Forte - N. Pintacuda, Sull'informazione associata alle esperienze incomplete. Ann. Mat. Pura Appl. (4) 80 (1968), 215-224.

B. Jessen - J. Karpf - A. Thorup, Some Functional Equations in Groups and Rings. Math. Scand. 22 (1968), 257-265.

1969 Z. Daróczy, On the Shannon Measure of Information (Hungarian). Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. 19 (1969), 9-24.

Z. Daróczy - I. Kátai, Additive zahlentheoretische Funktionen und das Mass der Information. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 14-15 (1969-70), in print.

1970 Z. Daróczy, On the Measurable Solution of a Functional Equation. Acta Math. Acad. Sci. Hungar. 21 (1970), in print.

P. Fischer, On the Inequality $\sum_{i=1}^{n} p_i f(p_i) \geq \sum_{i=1}^{n} p_i f(q_i)$. Metrika 17 (1970), in print.

P. Nath, An Axiomatic Characterization of Inaccuracy in Information Theory. Metrika 17 (1970), in print.

1971 J. Aczél - Z. Daróczy, Measures of Information and Their Characterizations. Academic Press, New York, 1971, in preparation.

J. Aczél - P. Nath, Axiomatic Characterizations of Some Measures of Divergence in Information, in print.

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C. I. M. E.)

J. A. BAKER

FUNCTIONAL EQUATIONS IN VECTOR SPACES, PART II

Corso tenuto a La Mendola (Trento) dal 20 al 28 agosto 1970

FUNCTIONAL EQUATIONS IN VECTOR SPACES, PART II

J O H N   A .   B A K E R

UNIVERSITY OF WATERLOO

In these lectures we are concerned with regularity theorems for functional equations and with some specific examples. For simplicity we consider functions defined in the real p-dimensional Euclidean space R^p with values in a Banach space X although most of the results are known in a more general setting.

If $x = (x_1,\ldots,x_p) \in R^p$ then $|x| = (x_1^2+\ldots+x_p^2)^{\frac{1}{2}}$. If $A, B \subset R^p$ then $A + B = \{x+y : x \in A, y \in B\}$, $-A = \{-x : x \in A\}$ and, if $a \in R^p$, $A + a = \{x + a : x \in A\}$. We let m denote the p-dimensional Lebesgue measure on R^p. In what follows, X is a Banach space with the norm of $x \in X$ denoted by $\|x\|$.

J. A. Baker

REGULARITY THEOREMS

Let us begin by considering the Cauchy functional equation

$$f(x+y) = f(x) + f(y).$$

A real valued function of a real variable satisfying this equation for all real x and y will be called an additive function. It is well known that if an additive function f is continuous at a point, or bounded on an open set, then it is continuous everywhere and there is a real constant a such that $f(x) = ax$ for all real x. (See Aczél [1] page 34). M. Fréchet [5] was among the first to prove that a Lebesgue measurable additive function is continous. More generally Ostrowski [16] has shown that if an additive function is bounded on one side on a set of positive Lebesgue measure then it is continuous.

To prove this last statement, suppose f is an additive function that is bounded above by M on a set A of positive Lebesgue measure. Then f is bounded above on A + A by 2M. It can be shown (e.g. [7] page 296) that A + A contains and open set, say U. If $a \in U$ then f is bounded above on $V = -a + U$ by $f(-a) + 2M$. Since V is a neighborhood of 0 and $f(-x) = -f(x)$ for all real x, f is bounded below on -V. Thus f is bounded on the open set $V \cap -V$ and is thus continuous.

Now, if a real function is Lebesgue measurable on a set of positive Lebesgue measure, it is certainly bounded on some set of positive Lebesgue measure. Hence, the theorem of Ostrowski immediately implies that an additive function which is measurable on a set of positive Lebesgue measure is continuous.

Regularity theorems of the type "boundedness implies continuity" have been proved by Kurepa [12] for the quadratic functional equation

$$f(x+y) + f(x-y) = 2f(x) + 2f(y)$$

and for more general equations by Kemperman [11].