PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 38, Number 2, April 1973

MEASURABLE SOLUTIONS OF FUNCTIONAL EQUATIONS RELATED TO INFORMATION THEORY[1]

PL. KANNAPPAN AND C. T. NG

ABSTRACT. Measurable solutions of functional equations connected with Shannon's measure of entropy, directed divergence or information gain and inaccuracy are found.

1. **Introduction.** While characterizing Shannon's entropy, one encounters the functional equation (cf. [1], [2], [3], [4], [7], [8], [10] and [11])

$$f(x)+(1-x)f(y/(1-x))=f(y)+(1-y)f(x/(1-y)), \tag{1}$$

and while characterizing directed divergence and inaccuracy one comes across the functional equation (cf. [6], [9])

$$\begin{aligned} &F(x,y)+(1-x)F(u/(1-x),v/(1-y)) \\ &\qquad = F(u,v)+(1-u)F(x/(1-u),y/(1-v)). \end{aligned} \tag{2}$$

As for (1), the general solutions as well as Lebesgue measurable, symmetric and nonsymmetric solutions are known (cf. [1], [8]).

Regarding (2), the general solutions having continuous first partial derivatives are given in [6] and [9].

In this paper, we solve first the functional equation

$$f(x)+(1-x)g(y/(1-x))=h(y)+(1-y)k(x/(1-y)) \tag{3}$$

in four unknown measurable functions f, g, h, k. This will yield, as a particular case, the most general measurable solutions of (1).

Next, we describe all solutions F of (2) which are measurable in each variable, through a reduction to the equation (3).

Notations. Let $I=[0,1]$, $I_1=[0,1[$, $I^\circ=]0,1[$ and R stand for the real numbers.

Received by the editors April 25, 1972 and, in revised form, July 14, 1972.

AMS (MOS) subject classifications (1970). Primary 39A15, 39A30, 28A20; Secondary 94A15.

Key words and phrases. Functional equations, measurable set, measurable, locally bounded, integrable functions.

[1] Work supported in part by the National Research Council of Canada grants.

2. **Measurable solutions of (3).** In this section we treat the functional equation (3) and obtain all the measurable solutions of (3) through a series of auxiliary results.

Let $g, k: I \to R$, and $f, h: I_1 \to R$ be functions such that (3) holds for all $x, y \in I_1$ with $x+y \in I$.

We will first reduce the functional equation (3) containing four unknown functions to a functional equation with one unknown function.

With $x=0$, (3) gives

$$h(y) = g(y) + a_1 y + b_1 \tag{4}$$

for $y \in I_1$, where a_1 and b_1 are constants.

Replacing y by $1-x$ in (3), (3) and (4) yield

$$f(x) = h(1-x) + a_2 x + b_2 = g(1-x) + a_3 x + b_3 \tag{5}$$

for $x \in I^\circ$, where a_2, b_2, a_3 and b_3 are constants.

From the equations (5) and (3) with $y=0$, we obtain

$$k(x) = f(x) + a_4 x + b_4 = g(1-x) + a_5 x + b_5 \tag{6}$$

for $x \in I^\circ$, where a_4, a_5, b_4 and b_5 are constants.

The equations (4), (5) and (6) enable us to write (3) as

$$\begin{aligned} g(1-x) &+ (1-x)g(y/(1-x)) \\ &= g(y) + (1-y)g((1-x-y)/(1-y)) + ax + by + c \end{aligned} \tag{7}$$

for $x \in I^\circ$, $y \in I_1$ with $x+y \in I^\circ$, where a, b, c are constants. It is to be noted that, contrary to (3), (7) is supposed valid only on $\{(x,y): x>0, y \geqq 0 \text{ with } 0<x+y<1\}$. Thus we have proved the following lemma.

LEMMA 1. *If the functions $g, k: I \to R$ and $f, h: I_1 \to R$ satisfy the functional equation* (3), *for all $x, y \in I_1$ with $x+y \in I$, then all the functions can be expressed as affine compositions of g given by* (4), (5) *and* (6) *such that g is a solution of* (7).

REMARK 1. It follows easily from (4), (5) and (6) that if any one of the four functions is Lebesgue measurable, then so are the other functions.

The equation (7) with $y=0$ gives $(1-x)g(0)=g(0)+ax+c$ for all $x \in I^\circ$, so that $c=0$. Now, (7) reduces to

$$\begin{aligned} g(1-x) &+ (1-x)g(y/(1-x)) \\ &= g(y) + (1-y)g((1-x-y)/(1-y)) + ax + by, \end{aligned} \tag{8}$$

for $x \in I^\circ$, $y \in I_1$, with $x+y \in I^\circ$.

Making use of the transformation

$$u(x) = g(x) - (a+b)x + a, \tag{9}$$

for $x \in I_1$, the equation (8) takes the form

$$u(1-x) + (1-x)u(y/(1-x)) = u(y) + (1-y)u((1-x-y)/(1-y)), \tag{10}$$

for $x \in I^\circ$, $y \in I_1$, with $x+y \in I^\circ$.

First we will seek functions u satisfying (10) for $x, y \in I^\circ$ with $x+y \in I^\circ$ and then we will determine all the solutions u of (10) for all $x \in I^\circ$ and $y \in I_1$, with $x+y \in I^\circ$.

We give the following lemma without proof. For details refer to [**1**] and [**5**, p. 143].

LEMMA 2. *If A, B, C are Lebesgue measurable subsets of R with finite measure, then $x \to \mu(A \cap (1-x)B \cap (1-xC))$ is continuous.*

LEMMA 3. *If u is measurable in I° and satisfies* (10) *for all $x, y \in I^\circ$ with $x+y \in I^\circ$, then u is locally bounded in I° and hence locally integrable.*

PROOF. We will make the following observations regarding measurable functions which will make clear the proof of this lemma.

If a function u is measurable on I°, then there exists a measurable subset A of I° on which u is bounded, such that $\mu(A)$ can be made as close to 1 as desired; and in this case for every y in a certain neighbourhood of a given $y_0 \in]0, 1[$, there is an $x \in]0, 1[$ so that $1-x$, $y/(1-x)$, $(1-x-y)/(1-y)$ are all in A and thus u is locally bounded at y_0.

In fact, let $y_0 \in I^\circ$ be arbitrary but fixed. Let $A_n = u^{-1}([-n, n])$, for all $n \geqq 1$. Then (A_n) is a sequence of measurable subsets of I° increasing to the whole interval I°. Let ε be an arbitrary fixed positive number. Then there is an A_N such that $\mu(I^\circ \backslash A_N) \leqq \varepsilon$. Let $z_0 = \min(y_0, 1-y_0)$. It follows then

(i) $\mu(]0, z_0[\backslash(1-A_N)) \leqq \mu(]0, 1[\backslash(1-A_N)) = \mu(]0, 1[\backslash A_N) \leqq \varepsilon;$

(ii)
$$\begin{aligned}\mu(]0, z_0[\backslash(1-y_0A_N^{-1})) &= \mu(]1-z_0, 1[\backslash y_0A_N^{-1}) \\ &\leqq \mu\left(\left]\frac{1-z_0}{y_0}, \frac{1}{y_0}\right[\backslash A_N^{-1}\right) \leqq \frac{y_0^2}{(1-z_0)^2}\mu\left(\left]y_0, \frac{y_0}{1-z_0}\right[\backslash A_N\right) \\ &\leqq \mu(]y_0, y_0/(1-z_0)[\backslash A_N) \leqq \mu(]0, 1[\backslash A_N) \leqq \varepsilon;\end{aligned}$$

(iii)
$$\begin{aligned}\mu(]0, z_0[\backslash(1-y_0)(1-A_N)) &\leqq \mu(]0, z_0/(1-y_0)[\backslash(1-A_N)) \\ &\leqq \mu(]0, 1[\backslash(1-A_N)) \leqq \varepsilon.\end{aligned}$$

If we choose $\varepsilon = \frac{1}{6}\mu(]0, z_0[)$, then (i), (ii) and (iii) lead to

$$\begin{aligned}\mu((1-A_N) \cap (1-y_0A_N^{-1}) \cap (1-y_0)(1-A_N)) &\geqq \mu(]0, z_0[) - 3\varepsilon \\ &= \tfrac{1}{2}\mu(]0, z_0[) > 0.\end{aligned}$$

By Lemma 2, because of the continuity of

$$y \to \mu((1-A_N) \cap (1-yA_N^{-1}) \cap (1-y)(1-A_N))$$

at y_0 there exists a neighbourhood $N(y_0)$ of y_0 such that

$$\mu((1-A_N) \cap (1-yA_N^{-1}) \cap (1-y)(1-A_N)) > 0,$$

for every $y \in N(y_0)$. Thus, in particular

$$(1-A_N) \cap (1-yA_N^{-1}) \cap (1-y)(1-A_N) \neq \varnothing,$$

for all $y \in N(y_0)$. Hence, for each $y \in N(y_0)$, there is an x (depending on y) in $(1-A_N)\cap(1-yA_N^{-1})\cap(1-y)(1-A_N)$; which is equivalent to $1-x$, $y/(1-x)$, $(1-x-y)/(1-y) \in A_N$. Hence from (10) it follows that

$$|u(y)| = |u(1-x) + (1-x)u(y/(1-x)) - (1-y)u((1-x-y)/(1-y))| \leqq 3N,$$

for all $y \in N(y_0)$. Thus we have proved that u is locally bounded at y_0. As y_0 is arbitrary in I°, u is locally bounded in I°. Hence u is locally integrable in I°. This completes the proof of Lemma 3.

Next we determine the measurable solutions of (10), for $x, y \in I^\circ$ with $x+y \in I^\circ$.

LEMMA 4. *The general measurable solution of* (10), *for* $x, y \in I^\circ$ *with* $x+y \in I^\circ$, *is given by*

$$u(x) = AS(x), \tag{11}$$

where A is an arbitrary constant and S is the Shannon function given by

$$S(x) = -x\log x - (1-x)\log(1-x). \tag{12}$$

PROOF. First we will show that u is differentiable infinitely in I°. Indeed, for arbitrary but fixed $y_0 \in I^\circ$, it is possible to choose s, t $(s<t) \in I^\circ$ such that $(1-y-s)/(1-y)$, $(1-y-t)/(1-y) \in I^\circ$, for y in a certain neighbourhood of y_0. On integrating (10) with respect to x from s to t, we get

$$\begin{aligned}(t-s)u(y) = \int_{1-t}^{1-s} u(x)\,dx + y\int_{y/(1-s)}^{y/(1-t)} \frac{u(x)}{x^3}\,dx \\ + (1-y)^2\int_{(1-y-s)/(1-y)}^{(1-y-t)/(1-y)} u(x)\,dx.\end{aligned} \tag{13}$$

The continuity of the right side of (13) as a function of y at y_0 implies the continuity of u at y_0. Thus u is continuous on I°. Now, the continuity of u in the right side of (13) shows that the right side of (13) is differentiable at

y_0 and hence the u in the left side of (13) is differentiable at y_0 and so everywhere in I°. Repetition of the above argument yields the differentiability of u of all orders in I°.

Now differentiating (10), first with respect to x and then the resultant with respect to y and making the substitutions $y/(1-x)=t$ and $x/(1-y)=1-s$ in the latter, we obtain, after cancelling out $1-t+ts$, that is $(1-x-y)/(1-x)(1-y)$ which is not zero, that

$$t(1-t)u''(t) = s(1-s)u''(s) = -A \text{ (say)}, \tag{14}$$

for $s, t \in I^\circ$. By successive integration, we have

$$u(x) = -A[x \log x + (1-x)\log(1-x)] + a_6 x + b_6, \tag{15}$$

for $x \in I^\circ$, where a_6 and b_6 are constants.

The function u given by (15) satisfies (10), provided $a_6=b_6=0$; that is, when u has the form given by (11).

We are now ready to describe all the measurable solutions of (3).

THEOREM 1. *The most general measurable solutions of* (3) *have the form,*

$$\begin{aligned} f(x) &= AS(x) + B_1 x + D, \\ g(y) &= AS(y) + B_2 y + B_1 - B_4, \\ h(x) &= AS(x) + B_3 x + B_1 + B_2 - B_3 - B_4 + D, \\ k(y) &= AS(y) + B_4 y + B_3 - B_2, \end{aligned} \tag{16}$$

for $x \in I_1$ and $y \in I$, where S is the Shannon function and A, B_1, B_2, B_3, B_4 and D are arbitrary constants.

PROOF. From Lemmas 1 and 4, and the equation (9), we deduce that f, g, h and k must have the form

$$f(x) = AS(x) + d_1 x + c_1, \qquad g(x) = AS(x) + d_2 x + c_2,$$
$$h(x) = AS(x) + d_3 x + c_3, \qquad k(x) = AS(x) + d_4 x + c_4,$$

for $x \in I^\circ$, where d_i, c_i ($i=1, 2, 3, 4$) are constants. A direct substitution of these f, g, h, k into (3) leads to the form (16) on the interval I°. An examination at the boundary points 0 and 1 reveals that f, g, h, k have the form (16) on the respective domains.

COROLLARY 1. *When all f, g, h, k are the same in* (3), *that is, when f satisfies* (1) *and is measurable, it is easy to see from* (16) *that*

$$f(x) = AS(x) + Bx \tag{17}$$

for some constants A and B.

REMARK 2. Let f be any measurable solution of (1). Then $\tilde{f}(x)=f(x)-f(1)x$ also satisfies (1) and further $\tilde{f}(x)=\tilde{f}(1-x)$. Thus by the previous papers quoted in the introduction, $\tilde{f}(x)=AS(x)$ and hence f has the form (17) (cf. [**1**]).

3. **Measurable solutions of** (2). Let $F: I\times I^\circ\to R$ satisfy (2) for $x, u\in I_1$, $y, v\in I^\circ$ with $x+u\leqq 1$ and $y+v<1$.

For each specified $y, v\in I^\circ$ with $y+v<1$, (2) is of the form (3) in the variables x and u. So, by Theorem 1, there exist constants $A(y, v)$, $B_i(y, v)$, $i=1, 2, 3, 4$, and $D(y, v)$ such that

$$\begin{aligned} F(x, y) &= A(y, v)S(x) + B_1(y, v)x + D(y, v),\\ F(x, v/(1-y)) &= A(y, v)S(x) + B_2(y, v)x + B_1(y, v) - B_4(y, v),\\ F(x, v) &= A(y, v)S(x) + B_3(y, v)x + B_1(y, v) + B_2(y, v)\\ &\quad - B_3(y, v) - B_4(y, v) + D(y, v),\\ F(x, y/(1-v)) &= A(y, v)S(x) + B_4(y, v)x + B_3(y, v) - B_2(y, v). \end{aligned}$$

The functions A, B_i and D give F consistently if and only if $A(y, v)\equiv$ constant$=A$ (say), $B_1(y, v)\equiv$a function of $y=B(y)$ (say), $D(y, v)\equiv$a function of $y=C(y)$ (say) and B and C which are evidently measurable satisfy the equations

$$B(y) - B(y/(1-v)) = C(v/(1-y)), \tag{18}$$

and

$$C(v/(1-y)) - C(y/(1-v)) + C(y) = C(v), \tag{19}$$

for $y, v\in I^\circ$ with $y+v<1$, so that F is of the form

$$F(x, y) = AS(x) + B(y)x + C(y), \tag{20}$$

for $x\in I_1$ and $y\in I^\circ$.

LEMMA 5. *If C defined on I° is a measurable function satisfying* (19) *for $y, v, y+v\in I^\circ$ then, and only then, there exist two arbitrary constants d and e such that*

$$C(x) = d\log(1-x) + e. \tag{21}$$

PROOF. Similar to the proof of Lemma 3, it can be shown that, for every fixed $y_0\in I^\circ$, there exists a neighbourhood $N(y_0)$ of y_0 and a measurable subset A_N of I° on which C is bounded by N so that, for each $y\in N(y_0)$, $\mu(A_N\cap(1-yA_N^{-1})\cap(1-y)A_N)>0$, which in turn implies that C is bounded by $3N$ on $N(y_0)$, so, C is locally bounded and hence locally integrable.

Integrating (19) with regard to y from μ to λ, we get

$$(\lambda-\mu)C(v)=\int_{\mu}^{\lambda}C(t)\,dt+v\int_{v/(1-\mu)}^{v/(1-\lambda)}\frac{C(t)}{t^2}\,dt-(1-v)\int_{\mu/(1-v)}^{\lambda/(1-v)}C(t)\,dt,$$

which implies the differentiability of C of all orders.

Differentiating (19) first with respect to y and then the resultant by v and making the substitutions $s=v/(1-y)$ and $t=y/(1-v)$ in the latter we get

$$(1-s)^2[C'(s)+sC''(s)]=(1-t)^2[C'(t)+tC''(t)]=\text{constant},$$

from which, by successive integration, we have

$$C(t)=k\log t+d\log(1-t)+e,$$

for $t\in I^\circ$, where k, d and e are arbitrary constants. This C satisfies (19) if and only if $k=0$, so that C has the form given by (21). This proves Lemma 5.

From Lemma 5 and (18), we see that B satisfies the equation

$$B(y)-B(y/(1-v))=d\log(1-v/(1-y))+e \tag{22}$$

for y, v, $y+v\in I^\circ$.

LEMMA 6. *B satisfying (22) for y, v, $y+v$ in I° has the form*

$$B(t)=d\log t-d\log(1-t)+q \tag{23}$$

for $t\in I^\circ$, where q is an arbitrary constant. Consequently $e=0$.

PROOF. For $y\in\,]0,\tfrac12[$, putting $1-v=2y$ in (22), we get

$$B(y)-B(\tfrac12)=d\log(1-(1-2y)/(1-y))+e;$$

that is,

$$B(y)=d\log(y/(1-y))+e_1 \tag{24}$$

for $y\in\,]0,\tfrac12[$, where e_1 is a constant.

Again for $v=\tfrac12$ with $y\in\,]0,\tfrac12[$, (22) gives

$$B(y)-B(2y)=d\log(1-1/2(1-y))+e. \tag{25}$$

From (24) and (25), we see that $B(2y)=d\log[2y/(1-2y)]+q$, for $y\in\,]0,\tfrac12[$ where q is a constant, showing thereby that B indeed has the form (23).

THEOREM 2. *The general functions F on $I\times I^\circ$, measurable in each variable, satisfying (2) for $x,u\in I_1$, $y,v\in I^\circ$ with $x+u\leqq 1$, $y+v<1$, are given* *by*

(26) $$F(x, y) = AS(x) + dx \log y + d(1 - x)\log(1 - y) + qx,$$

where A, q and d are arbitrary constants.

PROOF. By (20), Lemma 5 and Lemma 6, it follows that F must be given by (26) on $I_1 \times I^\circ$. Further examination of equation (2) on the remaining boundary with the help of (26) on $I_1 \times I^\circ$ shows that F has the required form (26) on the whole domain.

We thank the referee for the kind suggestions which greatly improved the presentation of the paper.

REFERENCES

1. J. Aczél and Z. Daróczy, *Measures of information and their characterizations*, Academic Press, New York (to appear).

2. T. Borges, *Zur Herleitung der Shannonschen Information*, Math. Z. **96** (1967), 282–287.

3. Z. Daróczy, *On the Shannon measure of information*, Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. **19** (1970), 9–24. (Hungarian) MR **40** #8497.

4. Z. Daróczy and I. Kátai, *Additive zahlentheoretische Funktionen und das Mass der Information*, Ann. Univ. Sci. Budapest Sect. Math. **13** (1970), 83–88.

5. E. Hewitt and K. Stromberg, *Real and abstract analysis*, Springer-Verlag, Berlin and New York, 1965.

6. Pl. Kannappan and P. N. Rathie, *On a characterization of directed divergence*, Information and Control (to appear).

7. D. G. Kendall, *Functional equations in information theory*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **2** (1964), 225–229. MR **29** #3270.

8. P. M. Lee, *On the axioms of information theory*, Ann. Math. Statist. **35** (1964), 415–418. MR **28** #1990.

9. P. N. Rathie and Pl. Kannappan, *On a characterization of inaccuracy* (submitted).

10. ———, *On a functional equation connected with Shannon's entropy*, Funkcial. Ekvac. **14** (1971), 153–159.

11. H. Tverberg, *A new derivation of the information function*, Math. Scand. **6** (1958), 297–298. MR **21** #3691.

FACULTY OF MATHEMATICS, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA