ON THE AXIOMS OF INFORMATION THEORY

By P. M. Lee

Churchill College, University of Cambridge

1. Introduction. The uniqueness of Shannon's measure of information is here proved under less restrictive conditions than previously.

Let $H_k(p_1, p_2, \cdots, p_k)(\sum p_j = 1$; all p's $> 0)$ be a measure of the information provided by the performance of an experiment with k possible outcomes of probabilities $p_1, p_2, \cdots p_k$ (c.f. Shannon [4] or Khinchin [3]).

We assume

(i) that H_k is permutation-symmetric for $k = 2, 3$; i.e. $H_2(t, 1 - t) = H_2(1 - t, t) = h(t)$, say, for $0 < t < 1$, and $H_3(p_1, p_2, p_3) = H_3(p_{\pi_1}, p_{\pi_2}, p_{\pi_3})$ for (π_1, π_2, π_3) any permutation of $(1, 2, 3)$ and any $p_1, p_2, p_3 > 0$ such that $p_1 + p_2 + p_3 = 1$;

(ii) that $h(\cdot)$ is a finite real-valued Lebesgue measurable function defined on $(0, 1)$ and that $h(\frac{1}{2}) = 1$ (previous authors have assumed $h(\cdot)$ continuous on $[0, 1]$, see Fadeev [1], monotone on $(0, \frac{1}{2})$ and on $(\frac{1}{2}, 1)$, see Kendall [2], or Lebesgue integrable on $[0, 1]$, see Tveberg [5]);

(iii) that for $0 < t < 1$, $k > 1$ and $p_1, p_2, \cdots p_k > 0$, $\sum p_j = 1$, we have $H_{k+1}(tp_1, (1 - t)p_1, p_2, \cdots, p_k) = H_k(p_1, \cdots, p_k) + p_1 H_2(t, 1 - t)$, so that $H_3(p_1, p_2 \ \ p_3) = h(p_1 + p_2) + (p_1 + p_2)h(p_1/p_1 + p_2)$.

From (i) and (iii) we see that $h(\cdot)$ must satisfy the functional equations

(iv) $h(t) = h(1 - t)$,

(v) $h(p_1) + (1 - p_1)h(p_2/1 - p_1) = h(p_1 + p_2) + (p_1 + p_2)h(p_1/p_1 + p_2)$

(vi) $h(p_1) + (1 - p_1)h(p_2/1 - p_1) = h(p_2) + (1 - p_2)h(p_1/1 - p_2)$.

We shall show under assumptions (ii), (iv), and (v) that $h(t) = -t \lg t - (1 - t) \lg (1 - t)$ (lg denotes logarithm to base 2, log denoting logarithm to base e). It follows that H_k is uniquely determined for all k; in fact

$$H_k(p_1, p_2, \cdots, p_k) = -\sum p_i \lg p_i .$$

2. Simple lemmas. As in Zaanen [6] (Section 36, Theorem 1 and Lemma γ), we observe that if μ denotes Lebesgue measure in R_1, and if $\phi(\cdot)$ is a continuously differentiable increasing function with a strictly positive derivative which maps an open interval I onto an open interval $\phi(I)$, then ϕ maps Lebesgue subsets Q of I to Lebesgue subsets $\phi(Q)$ of $\phi(I)$, and

$$\mu(\phi(Q)) = \int_Q \phi'(t)\, dt.$$

From this we deduce:

Received 5 August 1963.

LEMMA 1. *Let* $J \subset (0, 1)$ *be a Lebesgue set of measure c. Then*

$$K_y = \{x/[1 - (1 - x)y]; x \varepsilon J\}$$

is a Lebesgue set for each $y \varepsilon [a, b] \subset (0, 1)$ *and* $\mu(K_y) \geqq (1 - b)c$.

3. The fundamental lemma.

LEMMA 2. *Let E be a Lebesgue set of positive measure in* [0, 1] *which is symmetrical about the point* $\frac{1}{2}$. *Then there exist* $k \geqq 2, 0 < a < b < 1$, *and* $c > 0$ *such that any* $y \varepsilon [a, b]$ *can be expressed as* $y = \gamma/1 - \eta$ *with* $\gamma, \eta \varepsilon E \cap (1/k, 1 - 1/k)$ *for a Lebesgue set of distinct values of* η *of measure at least c.*

PROOF. Choose a k such that $G = E \cap (1/k, 1 - 1/k)$ has positive measure. For any continuous function f on [0, 1] one has $\lim_{y \uparrow 1} f(yx) = f(x)$ uniformly in x, (as f must be uniformly continuous on [0, 1]), thus

$$\lim_{y \uparrow 1} \int |f(yx) - f(x)|\mu(dx) = 0.$$

Because the continuous functions are dense in L^1_μ, (Zaanen [6], Section 30, Ex. 10) this implies

$$\lim_{y \uparrow 1} \int |\chi_G(yx) - \chi_G(x)|\mu(dx) = 0$$

(χ_G being the indicator of G). Hence

$$\lim_{y \uparrow 1} \mu\{\eta:\eta \varepsilon G, (1 - \eta)y \varepsilon G\} = \lim_{y \uparrow 1} \mu\{x: x \varepsilon G, xy \varepsilon G\}$$

$$= \lim_{y \uparrow 1} \int_G \chi_G(xy)\mu(dx) = \int_G \chi_G \, d\mu.$$

On writing $(1 - \eta)y = \gamma$, $\frac{1}{2}\mu(G) = c$, the lemma is proved, for suitable a and b. (N.B. We could here have taken $b = 1$, but in the application to follow we shall want $b < 1$.)

4. Boundedness of solutions. We now show that: *Every measurable solution of* (iv) *and* (vi) *is bounded on some interval.*

PROOF. $\{\xi:0 < \xi < 1, |h(\xi)| > n\}$ is measurable and of finite measure ($\leqq 1$), and decreases to the null set as $n \uparrow \infty$, so there exists N such that, when $E = \{\xi:0 < \xi < 1, |h(\xi)| \leqq N\}$, $\mu(E) \geqq \frac{1}{2}$. Choose k, a, b, and c as in Lemma 2, noting that $E = 1 - E$. If $y \varepsilon [a, b]$, then, as in Lemma 2 we may write $y = \gamma/1 - \eta$ in a great many ways with $\gamma, \eta \varepsilon E \cap (1/k, 1 - 1/k)$. The values of $\eta/1 - \gamma$ for such representations are all positive and less than unity (as $\gamma < 1 - \eta \Rightarrow \eta < 1 - \gamma$), and the set of such values is the set of values of $\eta/(1 - (1 - \eta)y)$, hence (Lemma 1) covers a Lebesgue set whose measure is at least $(1 - b)c$.

We now observe that there exists $M \geqq N$ such that $\mu(F) < \frac{1}{2}(1 - b)c$, where $F = \{\xi:0 < \xi < 1, |h(\xi)| > M\}$. It follows that for any $y \varepsilon [a, b]$ there is at least one representation $y = \gamma/1 - \eta$, such that $\eta/1 - \gamma \not\varepsilon F$, but $\gamma, \eta \,\varepsilon\, E \cap (1/k, 1 - 1/k)$. Hence, by (vi),

$$|h(y)| = |h(\gamma/1 - \eta)| = (1 - \eta)^{-1}|h(\gamma) + (1 - \gamma)h(\eta/1 - \gamma) - h(\eta)|$$
$$\leqq (1 - \eta)^{-1}\{|h(\gamma)| + |h(\eta/1 - \gamma)| + |h(\eta)|\} < 3kM \quad \text{as} \quad \eta \,\varepsilon\, (1/k, 1 - 1/k).$$

This completes the proof.

The above is now extended to become: *Each measurable solution of* (iv) *and* (vi) *is bounded on every compact subset of* (0, 1).

PROOF. Let Λ be the set of positive λ such that $h(\alpha) - \lambda h(\alpha/\lambda)$ is ultimately bounded as $\alpha \downarrow 0$. Plainly (a) $1 \,\varepsilon\, \Lambda$, (b) $\Lambda^2 \subset \Lambda$, (c) $\Lambda^{-1} \subset \Lambda$, so Λ is a multiplicative group. Now let J be an open interval contained in (0, 1) on which $h(\cdot)$ is bounded; we already know that such an interval exists. If $1 - \lambda \,\varepsilon\, J$, then since (vi) gives

$$(1) \qquad h(\alpha) - \lambda h(\alpha/\lambda) = h(1 - \lambda) - (1 - \alpha)h(1 - \lambda/1 - \alpha) \qquad (0 < \alpha < \lambda < 1),$$

we see that $\lambda \,\varepsilon\, \Lambda$. Thus Λ covers a measurable set of positive measure, and so is the whole of $(0, \infty)$ (the latter remark follows on noticing that on taking logarithms we have an additive group, which thus contains its own difference set, hence a non-degenerate interval around zero (Zaanen [6], Section 10, Lemma β)).

From Equation (1) above and from (iv) it now follows that each $t = 1 - \lambda \,\varepsilon\, (0, 1)$ lies in an open interval of boundedness of $h(\cdot)$, so that $h(\cdot)$ is bounded on each compact subset of (0, 1).

5. The uniqueness of $h(\cdot)$. We cannot from the argument above deduce the boundedness of $h(\cdot)$ on the whole of (0, 1); if we could, the desired result would follow from the work of Tveberg [5]. We can, however, adapt one of his arguments to obtain further useful information about $h(\cdot)$. Integration of (vi) (Zaanen [6], Section 36, Theorem 1) gives

$$(\mu - \lambda)h(\alpha) = \int_{\lambda}^{\mu} h(\gamma)\, d\gamma + \alpha^2 \int_{\alpha/1-\lambda}^{\alpha/1-\mu} \gamma^{-3}h(\gamma)\, d\gamma - (1 - \alpha)^2 \int_{\lambda/1-\alpha}^{\mu/1-\alpha} h(\gamma)\, d\gamma$$

for $0 < \alpha < \alpha + \lambda \leqq \alpha + \mu < 1$, and so we see that $h(\cdot)$ is continuous and indeed (by iteration) of class C^∞ at every interior point of the unit interval.

We cannot, however, appeal to Fadeev's theorem [1] since that would require continuity on the closed interval [0, 1], but we can adapt the argument at the end of Kendall's paper [2] (see his Equation (12) et seq.). The argument, which uses (v) and (vi), simplifies drastically because of continuous differentiability; it depends on putting $h'(t) = E(t/1 - t)$ to obtain $E(u) + E(v) = E(uv)$, the unique continuous solution of which is well known. Note that under (iv), (v) $\Leftrightarrow$ (vi). The following has thus been proved:

THEOREM 1. *The only measurable solutions to* (iv) *and* (vi) *are the multiples of*

Shannon's function $h(t) = -t \lg t - (1 - t) \lg (1 - t)$. The uniqueness of H_k for $k > 2$ now follows.

This result is best possible in the sense that there are nonmeasurable solutions of (iv) and (vi) other than Shannon's function, for example $h(t) = -tf(\lg t) - (1 - t)f(\lg (1 - t))$ where f is a non-measurable linear function.

I am greatly indebted to Professor D. G. Kendall and the referee for several helpful comments on this work leading to simpler proofs of the results of Sections 2 and 5, and to the Department of Scientific and Industrial Research for a Research Studentship.

REFERENCES

[1] FADEEV, D. K. (1956). On the concept of entropy of a finite probabilistic scheme (in Russian). *Uspehi Mat. Nauk.* (*N.S.*) **11** 227–31.

[2] KENDALL, D. G. Functional equations in information thèory. *Z. für Wahrscheinlichkeitstheorie* (to appear).

[3] KHINCHIN, A. I. (1957). *Mathematical Foundations of Information Theory*. Dover, New York.

[4] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423 and 623–656.

[5] TVEBERG, H. (1958). A new derivation of the information function. *Math. Scand.* **6** 297–8.

[6] ZAANEN, A. C. (1958). *An Introduction to the Theory of Integration*. North-Holland, Amsterdam.