A GENERALISATION OF THE RADON-NIKODYM THEOREM

P. D. FINCH

(received 10 January 1964, revised 6 August 1964)

1. Introduction

Let $\mathscr{X}$ be a space of points x, $\mathscr{M}$ a σ-field of subsets of $\mathscr{X}$ and μ a σ-finite measure on $\mathscr{M}$. The elements of $\mathscr{M}$ will be called measurable sets and all the sets considered in this paper are measurable sets. A real-valued point function $\mathfrak{t}(x)$ on $\mathscr{X}$ will be said to be measurable if, for each real α, the set $\{x : \mathfrak{t}(x) \leqq \alpha\}$ is measurable. Let $\mathscr{M}(S)$, $S \subset \mathscr{X}$ denote the σ-field of all measurable subsets of S. A real-valued function $f(\cdot)$ on $\mathscr{M}$ will be called a set function.

In Finch [1] a theory of integration of set functions $f(M)$, $M \in \mathscr{M}$ with respect to the measure μ is developed. In that theory the integral

$$I_f(S) = (\Pi) \int_{\mathscr{M}(S)} f(M)\mu(M) \tag{1.1}$$

is, when it exists, the limit of the approximating sums

$$F_{\Pi}(S) = \sum_{\Pi(S)} f(M)\mu(M) \tag{1.2}$$

where the summation is over all elements with positive μ-measure, of the partition $\Pi(S)$ of S by elements of $\mathscr{M}(S)$ and the limit is taken in the sense of Moore-Smith convergence as the partitions spread. For details of the theory we refer to Finch [1] where it is shown that the Π-integral (1.1) is, when it exists, a σ-additive set function on $\mathscr{M}$, that is,

$$I_f\{\sum_{j=1}^{\infty} M_j\} = \sum_{j=1}^{\infty} I_f(M_j) \tag{1.3}$$

whenever the sets M_j are mutually disjoint elements of $\mathscr{M}$. Thus I_f is a signed measure on $\mathscr{M}$ and it follows from (1.2) that it is absolutely continuous with respect to μ. It follows from the Radon-Nikodym theorem that there is a measurable point function $i_f(x)$ on which is finite except possibly on a set of μ-measure zero, such that

$$I_f(S) = (L) \int_S i_f(x) d\mu(x) \tag{1.4}$$

https://doi.org/10.1017/S1446788700025829 Published online by Cambridge University Press

where the L-integral is the Lebesgue integral of $i_f(x)$ with respect to the measure μ. Further if $j_f(x)$ is any other measurable point function satisfying (1.4) then

$$\mu\{x : j_f(x) \neq i_f(x)\} = 0.$$

It is of interest therefore to examine the relationship between the Π-integrable set function f and the associated point function i_f. A partial solution to this problem is provided by the following theorem to whose proof this paper is devoted.

THEOREM 1. *Let $\mathscr{X}$ be a space of points x, $\mathscr{M}$ a σ-field of subsets of $\mathscr{X}$ and μ a σ-finite measure on $\mathscr{M}$. Let ν be a σ-finite signed measure on $\mathscr{M}$ and let $g(\xi)$ be a real-valued function of bounded variation of the variable ξ. Write*

$$f(M) = g\{\nu(M)/\mu(M)\}, \quad M \in \mathscr{M}, \quad \mu(M) > 0, \tag{1.5}$$

then there exists a real-valued measurable point function $\theta(x)$ on $\mathscr{X}$ which is finite, except possibly on a set of μ-measure zero, such that for each S

$$(\Pi)\int_{\mathscr{M}(S)} f(M)\mu(M) = (L)\int_S g\{\theta(x)\}d\mu(x) \tag{1.6}$$

whenever either integral exists.

REMARKS. Since $\mathscr{X}$ is the countable union of disjoint elements of $\mathscr{M}$ on which μ and ν are each finite it is sufficient to prove the theorem when μ and ν are each finite. Secondly it is clearly sufficient to prove the theorem when the function g is monotonic and non-negative. From here on, therefore, we shall assume that μ is a finite measure, ν is a finite signed measure and that g is monotonic non-decreasing and non-negative.

Note that the theorem does not assert that the function $g\{\theta(x)\}$ is L-integrable with respect to μ, in fact a necessary and sufficient condition for this is the existence of the Π-itegral in (1.6). Note also that the statement of the theorem does not assert that the signed measure ν is absolutely continuous with respect to μ. However the L-integrability of $\theta(x)$ or equivalently the existence of the Π-integral (1.6) when $g(\xi) \equiv \xi$ is a necessary and sufficient condition for the absolute continuity of ν with respect to μ.

To see this observe that when $g(\xi) \equiv \xi$ the approximating sum (1.2) to the Π-integral (1.6) is

$$F_\Pi(S) = \sum_{\Pi(S)} \nu(M) \tag{1.7}$$

where the summation is over those elements M of the partition $\Pi(S)$ with $\mu(M) > 0$. If ν is absolutely continuous with respect to μ then $F_\Pi(S) = \nu(S)$ since $\mu(M) = 0$ implies $\nu(M) = 0$ and the Π integral exists and has the value $\nu(S)$.

Conversely if the Π-integral exists, that is, if the Π-limit of (1.7) exists this limit is unique. Choosing a sequence of partitions $\{\Pi_n(S)\}$ of S with $\Pi_{n+1}(S)$ finer than $\Pi_n(S)$ and such that each element of the partition $\Pi_n(S)$ has positive μ-measure we see that this limit is $\nu(S)$. Let S_0 be any element of $\mathscr{M}$ with $\mu(S_0) = 0$ and write $S_1 = S \cup S_0$. Choosing a sequence of partitions $\{\Pi_n(S_1)\}$, of S_1 such that each element of $\Pi_n(S_1)$ has positive μ-measure and $\Pi_{n+1}(S_1)$ is finer that $\Pi_n(S_1)$ we obtain the limit $\nu(S_1)$. Since the Π-integral has a unique value $\nu(S_1) = \nu(S)$, that is, $\nu(S_0) = 0$ and this shows that ν is absolutely continous with respect to μ.

It follows from the above that theorem 1, contains the Radon-Nikodym theorem as a particular case and for this reason our proof of it does not depend on the Radon-Nikodym theorem. An example showing that the theorem can be true when ν is not absolutely continuous with respect to μ, in fact when μ is absolutely continuous with respect to ν is given in section 3. One use of theorem 1 is that it reduces the calculation of the Π-integral to that of an L-integral, such a use is illustrated in section 4 by application to a problem in information theory.

2. Some preliminary results

In this section we state some preliminary results which are required for the proof of theorem 1.

LEMMA (2.1). *Let R denote the set of real numbers and let $\{\alpha_j\}$ be a sequence of real numbers which is dense in R. Suppose that $\{M(\alpha_j)\}$ is a family of elements of $\mathscr{M}$, indexed by the dense sequence α_j and such that*

(i) $M(\alpha_i) \subset M(\alpha_j)$ *if* $\alpha_i < \alpha_j$

(ii) $M(\alpha_i) = \bigcap_{\alpha_j > \alpha_i} M(\alpha_j)$

For any real α define

$$M(\alpha) = \bigcap_{\alpha_j > \alpha} M(\alpha_j)$$

then there exists a real-valued measurable point function $\theta(x)$ on $\mathscr{X}$ such that

$$M(\alpha) = \{x : \theta(x) \leqq \alpha\}.$$

If further

(iii) $\lim_{\alpha\to\infty} \mu\{\mathscr{X} - M(\alpha)\} = 0$, $\lim_{\alpha\to\infty} \mu\{M(-\alpha)\} = 0$,

then $\theta(x)$ is finite except possibly on a set of μ-measure zero.

This lemma is proved easily by writing

$$\theta(x) = \inf\{\alpha : x \in M(\alpha)\}. \tag{2.1}$$

Using lemma (2.1) one may prove

LEMMA (2.2). *If μ is a finite measure on $\mathscr{M}$, ν is a finite signed measure on $\mathscr{M}$ then there exists a measurable point function $\theta(x)$ on $\mathscr{M}$ which is finite* *except possibly on a set of μ-measure zero, such that if for each real α,*

(2.2) $$M(\alpha) = \{x : \theta(x) \leqq \alpha\}$$

then

(2.3) $$\begin{aligned} \nu(M) &\leqq \alpha\mu(M), \qquad M \subset M(\alpha) \\ \nu(M) &\geqq \alpha\mu(M), \qquad M \subset \mathscr{X} - M(\alpha). \end{aligned}$$

PROOF. For each real α and each $M \in \mathscr{M}$ write

$$\lambda(M; \alpha) = \nu(M) - \alpha\mu(M)$$

Let $\{\alpha_j\}$ be a dense sequence of real numbers, for each α_j, $\lambda(M; \alpha_j)$ is a finite signed measure on $\mathscr{M}$ and so, by the Hahn decomposition of $\mathscr{X}$ with respect to this signed measure, there exists an element $M(\alpha_j)$ of $\mathscr{M}$ such that

(2.4) $$\begin{aligned} \lambda(M; \alpha_j) &\leqq 0, \qquad M \subset M(\alpha_j) \\ \lambda(M; \alpha_j) &\geqq 0, \qquad M \subset \mathscr{X} - M(\alpha_j). \end{aligned}$$

The proof of lemma (2.2) consists in verifying that we can choose the sets $M(\alpha_j)$ to satisfy the conditions of lemma (2.1). Since this verification uses standard procedures, for example, Royden [3], it will be omitted.

3. Proof of theorem 1

We proceed now to the proof of theorem 1. Since g is non-negative and monotonic non-decreasing the inequality

$$f(M) \leqq \alpha, \quad M \in \mathscr{M}, \quad \mu(M) > 0$$

is equivalent to the inequality

$$\nu(M) - (g^{-1}\alpha)\mu(M) \leqq 0, \quad M \in \mathscr{M}, \quad \mu(M) > 0.$$

Here and in what follows

$$g^{-1}\alpha = \sup\{\xi : g(\xi) \leqq \alpha\}.$$

Thus if $\theta(x)$ is the measurable point function of lemma (2.2) we have,

(3.1) $$\begin{aligned} f(M) &\leqq \alpha \quad \text{if} \quad M \subset M(g^{-1}\alpha) \\ f(M) &\geqq \alpha \quad \text{if} \quad M \subset \mathscr{X} - M(g^{-1}\alpha) \end{aligned}$$

where

$$M(g^{-1}\alpha) = \{x: \theta(x) \leqq g^{-1}\alpha\}$$

and

$$g\{\theta(x)\} = \inf\{\alpha : x \in M(g^{-1}\alpha)\}.$$

Let δ be an arbitrary positive real number and let $\{\delta_j\}$, $j = 0, 1, \cdots$ be a sequence of real numbers with $\delta_0 = 0$, and such that

$$0 < \delta_j - \delta_{j-1} \leqq \delta, \quad j \geqq 1, \quad \text{and } \sup \delta_j = +\infty.$$

Write

$$(3.2) \qquad M_j = \{x : \delta_{j-1} < g\{\theta(x)\} \leqq \delta_j\}, \qquad j \geqq 1.$$

Then for any j such that $\mu(M_j) > 0$ we have

$$(3.3) \qquad \delta_{j-1} \leqq f(M) \leqq \delta_j, \quad M \in \mathscr{M}, \quad M \subset M_j, \quad \mu(M) > 0.$$

It follows that the total variation

$$|f|(M) = \sup\{f(A_1) - f(A_2) : \mu(A_i) > 0, A_i \subset M\}$$

on f on $\mathscr{M}$ with respect to μ does not exceed δ on the measurable subsets of each M_j. Thus

$$(3.4) \qquad |f|(M) \leqq \delta, \quad M \in \mathscr{M}, M \subset M_j; \quad \mu(M) > 0.$$

Write

$$f^{(n)}(M) = \begin{cases} f(M) & \text{if } f(M) \leqq n \\ n & \text{if } f(M) \geqq n \end{cases}$$

for each $M \in \mathscr{M}$ with $\mu(M) > 0$.

Let S be any element of $\mathscr{M}$, then

$$\Pi(S) = \{SM_j\}, \quad j = 0, 1, 2, \cdots$$

is a partition of S. It follows from (3.4) and theorem (3.3) of Finch [1], that $f^{(n)}(\cdot)$ is Π-integrable on $\mathscr{M}(S)$ with respect to μ, that is,

$$(3.5) \qquad \begin{aligned} &(\Pi) \int_{\mathscr{M}(S)} f^{(n)}(M)\mu(M) \\ &\quad = (\Pi) \lim \sum_{j=-\infty}^{+\infty} f^{(n)}(SM_j)\mu(SM_j) \end{aligned}$$

exists.

Because of (3.2) and (3.3) it follows also that the sum on the right-hand side of (3.5) is an approximating sum for the Lebesgue integral

$$(L) \int_S g^{(n)}\{\theta(x)\} d\mu(x)$$

where

$$g^{(n)}(\xi) = \begin{cases} g(\xi) & \text{if } g(\xi) \leqq n, \\ n & \text{if } g(\xi) > n, \end{cases}$$

and hence that

$$(3.6) \qquad (\Pi) \int_{\mathscr{M}(S)} f^{(n)}(M)\mu(M) = (L) \int_S g^{(n)}\{\theta(x)\} d\mu(x)$$

Letting $n \to \infty$ in (3.6) we obtain (1.6) whenever either integral exists. The uniqueness of $\theta(x)$ follows immediately since if $\varphi(x)$ is another such measurable point function

$$(L)\int_S [g^{(n)}\{\theta(x)\} - g^{(n)}\{\varphi(x)\}]d\mu(x) = 0$$

for all $S \in \mathscr{M}$ and each $n > 0$, hence

$$\mu\{x; \phi(x) \neq \theta(x)\} = 0.$$

This completes the proof of the theorem.

As remarked in section one the formulation of theorem 1 does not introduce explicitly the condition that ν should be absolutely continuous with respect to μ, although, as we have shown, if the Π-integral (1.6) exists when $g(\xi) \equiv \xi$ this implies that the signed measure ν is absolutely continuous with respect to μ. To illustrate that meaningful results may be obtained when ν is not absolutely continuous with respect to suppose in fact that ν and μ are both finite measures and that μ is absolutely continuous with respect to ν with density $\phi(x)$, so that

$$\mu(M) = (L)\int_M \phi(x)d\nu(x).$$

Suppose also that $\phi(x)$ belongs to the class $L_p(\nu)$ for some $p > 1$, so that

$$(L)\int_M \{\phi(x)\}^p d\nu(x)$$

exists for each $M \in \mathscr{M}$.

Consider the identity

$$\{\mu(M)/\nu(M)\}^p \nu(M) = \{\nu(M)/\mu(M)\}^{-(p-1)}\mu(M)$$

where $\nu(M) > 0$. By applying theorem 1 to the left-hand side we obtain

$$(3.7) \qquad (\Pi)\int_{\mathscr{M}(S)} \{\mu(M)/\nu(M)\}^p \nu(M) = (L)\int_S \{\phi(x)\}^p d\nu(x)$$

Since μ is absolutely continuous with respect to ν

$$\sum_{\nu(M)>0} \left\{\frac{\mu(M)}{\nu(M)}\right\}^p \nu(M) = \sum_{\mu(M)>0} \left\{\frac{\nu(M)}{\mu(M)}\right\}^{-p-1} \mu(M)$$

where the summations are over the elements of the partition $\Pi(S)$ of $S \in \mathscr{M}$ with positive ν and μ measure respectively. Thus the Π-integral of which the right-hand side is the approximating sum exists and equals the Π-integral of (3.7), that is,

$$(3.8) \qquad (\Pi)\int_{\mathscr{M})S)} \{\nu(M)/\mu(M)\}^{-(p-1)}\mu(M) = (\Pi)\int_{\mathscr{M}(S)} \{\mu(M)/\nu(M)\}^p \nu(M).$$

Since the Π-integral on the left-hand side of (3.8) exists theorem 1 ensures the existence of $\theta(x)$ such that

$$(3.9)\quad (\Pi)\int_{\mathscr{M}(S)} \{\nu(M)/\mu(M)\}^{-(p-1)}\mu(M) = (L)\int_S \{\theta(x)\}^{-(p-1)} d\mu(x).$$

In fact it is clear that $\theta(x) = \{\phi(x)\}^{-1}$ except on a set of ν measure zero. Equation (3.9) is the desired example of theorem 1 when ν is not absolutely continuous with respect to μ.

4. An application to information theory

Let $\mathscr{X}$ be a space of points x, $\mathscr{M}$ a σ-field of subsets of $\mathscr{X}$ and let $\{P(\cdot|\theta_j)\}$, $j = 1, 2, \cdots, k$, be a finite family of probability measures on $\mathscr{M}$. We write $\theta^{(k)} = (\theta_1, \theta_2, \cdots, \theta_k)$, call the θ_j indices or index values and refer to $\theta^{(k)}$ as the indexing set. The elements of $\mathscr{M}$ we refer to as events. For each $\theta \in \theta^{(k)}$ we call the ordered pair $\{\mathscr{X}, P(\cdot|\theta)\}$ a probability space.

In Finch [2] it is shown that an appropriate measure of the amount of conditional information about the particular probability space $\{\mathscr{X}, P(\cdot|\theta_j)\}$ provided by the occurrence of the event M when it is known that $\theta \in \theta^{(k)}$ is given by

$$(4.1)\quad I[\{\mathscr{X}, P(\cdot|\theta)\} : M|\theta \in \theta^{(k)}] = -\log\left[P(M|\theta)/\textstyle\sum_{j=1}^k P(M|\theta_j)\right], \qquad \theta \in \theta^{(k)}.$$

The quantity

$$(4.2)\quad G(M|\theta^{(k)}) = k^{-1} \textstyle\sum_{j=1}^k P(M|\theta_j),$$

is a probability measure over $\mathscr{M}$ and, according to Finch [2], can be interpreted as the generalised probability that the event M occurs under the logical disjunction of hypotheses $\theta_1 \vee \theta_2 \vee \cdots \vee \theta_k$.

The quantity (4.1) defines an amount of information provided by the occurrence of a particular event $M \in \mathscr{M}$. In order to define an average amount of information it is natural to introduce the quantity

$$(4.3)\quad E \cdot I[\{\mathscr{X}, P(\cdot|\theta)\}|\theta^{(k)}] = (\Pi)\int_{\mathscr{M}} I[\{\mathscr{X}, P(\cdot|\theta)\} : M|\theta \in \theta^{(k)}]\, G(M|\theta^{(k)}),$$

for each $\theta \in \theta^{(k)}$ whenever the Π-integral exists. The quantity (4.3) is the expected amount of conditional information about the probability space $\{\mathscr{X}, P(\cdot|\theta)\}$ provided by an experiment, whose possible outcomes are the events of $\mathscr{M}$, when it is known that $\theta \in \theta^{(k)}$.

The quantity

$$(4.4)\quad a(\theta|M; \theta^{(k)}) = P(M|\theta)/\textstyle\sum_{j=1}^k P(M|\theta_j), \qquad \theta \in \theta^{(k)}$$

is called the acceptability of the index value θ in the light of the occurrence of the event M when it is known that $\theta \in \theta^{(k)}$. In terms of the acceptabilities we may rewrite equation (4.3) in the form

$$
\begin{aligned}
(4.5) \qquad & E \cdot I[\{\mathscr{X}, P(\cdot|\theta)\}|\theta^{(k)}] \\
& \qquad = -(\Pi) \int_{\mathscr{M}} \log \{a(\theta|M; \theta^{(k)}\} G(M|\theta^{(k)}).
\end{aligned}
$$

It follows from theorem 1 that when this Π-integral exists there is a real-valued measurable point function on $\mathscr{X}$, a $(\theta|x; \theta^{(k)})$ which is finite, except possibly on a set of $G(\cdot|\theta^{(k)})$ measure zero, such that

$$
\begin{aligned}
(4.6) \qquad & E \cdot I[\{\mathscr{X}, P(\cdot|\theta)\}|\theta \in \theta^{(k)}] \\
& \qquad = -(L) \int_{\mathscr{X}} \log \{a(\theta|x; \theta^{(k)})\} \cdot G(dx|\theta^{(k)})
\end{aligned}
$$

and where the Lebesgue integral exists if and only if the Π-integral (4.5) exists.

Since the probability measure $P(\cdot|\theta)$ is absolutely continuous with respect to the probability measure $G(\cdot|\theta^{(k)})$ for each $\theta \in \theta^{(k)}$ it follows from the proof of theorem 1 also, that the point function $a(\theta|x; \theta^{(k)})$ is in fact the density of $P(\cdot|\theta)$ with respect to the measure $kG(\cdot|\theta^{(k)})$. Thus theorem 1 reduces the calculation of the Π-integral (4.5) to that of the L-integral (4.6).

References

[1] Finch, P. D., Integration of real-valued set functions in abstract spaces, This Journal 4 (1964), 202—213.

[2] Finch P. D., The theory of information and statistical inference I, Journ. App. Prob. 1 (1964), 121—140.

[3] Royden H. L., Real Analysis. Macmillan, New York (1963).

Monash University