AN UPCROSSING INEQUALITY WITH APPLICATIONS

Errett Bishop

The theorem of Lebesgue that a function of bounded variation has a derivative almost everywhere, the ergodic theorem of Birkhoff, and the martingale theorem have the common property of being nonconstructive; they are not true in intuitionist mathematics. (See Brouwer [1] for an example of a function of bounded variation that does not have a derivative almost everywhere.) This paper developed from an attempt to save the phenomena.

The martingale theorem has in fact already been saved: the upcrossing inequality of Doob [2, p. 314], which almost trivially implies the martingale theorem, is (with slight modifications) constructively valid. Moreover, Doob's inequality contains important information that can't be gotten from the martingale theorem itself. Thus to a formalist Doob's inequality is interesting because it improves the martingale theorem, and to a constructivist it has the additional merit of making empirical sense.

In this paper we prove an upcrossing inequality (Theorem 1) that stands in the same relation to Lebesgue's theorem as Doob's inequality stands to the martingale theorem: it is a constructively valid inequality which in the formal system constitutes a generalization of Lebesgue's theorem. By a very simple argument, Theorem 1 leads to another constructively valid inequality (Theorem 2), which in turn constitutes in the formal system a generalization of Birkhoff's ergodic theorem. Finally, by an argument unfortunately not so simple, we derive Doob's upcrossing inequality from Theorem 1.

The theorem of Lebesgue, Birkhoff's ergodic theorem, and the martingale theorem are thus consequences of a single inequality (Theorem 1); regarded in this way, these theorems gain both depth and empirical validity.

Although, as we have indicated, it is possible to develop our results within a strictly constructive framework, at present this would lead too far afield; therefore we stay within the formal system.

Definition 1. For each x in R^2 we let x_1 be the abscissa and x_2 the ordinate of x, so that $x = (x_1, x_2)$. For each α in R and x in R^2, we define

$$\sigma_\alpha(x) = x_2 - \alpha x_1 .$$

Definition 2. If $x \in R^2$, $\alpha, \beta \in R$, and $\alpha < \beta$, then points $u^1, v^1, \cdots, u^k, v^k$ in R^2 are said to have property $P(x, \alpha, \beta)$ if

$$x_1 < u_1^1 \leq v_1^1 < u_1^2 \leq v_1^2 < \cdots < u_1^k \leq v_1^k,$$

$$\sigma_\alpha(u^i) < \sigma_\alpha(x), \quad \sigma_\beta(v^i) > \sigma_\beta(x) \qquad (1 \leq i \leq k).$$

Definition 3. Let $x \in R^2$, $\alpha, \beta \in R$, $\alpha < \beta$, $\Gamma \subset R^2$, $t \in R$. Then

Received July 30, 1965.

The author wishes to thank the Miller Foundation of the University of California and the National Science Foundation for the support of this work.

$$\omega(x, \alpha, \beta, \Gamma) = \sup\{k: \exists\, u^1, v^1, \cdots, u^k, v^k \text{ in } \Gamma \text{ with property } P(x, \alpha, \beta)\},$$

and

$$\omega(t, \alpha, \beta, \Gamma) = \sup\{\omega(x, \alpha, \beta, \Gamma): x_1 = t\}.$$

Definition 4. A finite subset $\Gamma = \{x^1, y^1, \cdots, x^n, y^n\}$ of R^2 is *elementary* relative to α if

$$x_1^1 = y_1^1 < x_1^2 = y_1^2 < \cdots < x_1^n = y_1^n,$$

$$x_2^i < y_2^i \quad (1 \le i \le n), \qquad \sigma_\alpha(x^{i+1}) \le \sigma_\alpha(y^i) \quad (1 \le i \le n-1).$$

The integer n is called the *power* of Γ.

The following combinatorial lemma underlies our upcrossing inequality.

LEMMA 1. *If* Γ *is elementary and* $\alpha < \beta$, *then*

$$(*) \qquad \int \omega(t, \alpha, \beta, \Gamma)\, dt \le (\beta - \alpha)^{-1} \sum_{i=1}^{n} (y_2^i - x_2^i).$$

Proof. Notice first that the integral exists, because $\omega(t, \alpha, \beta, \Gamma)$ as a function of t is a finite linear combination of the characteristic functions of finite intervals.

If $u^1, v^1, \cdots, u^k, v^k$ are points in Γ having property $P(x, \alpha, \beta)$, there is no loss of generality in assuming that $u^i \in \{x^1, \cdots, x^n\}$, since, if $u^i = y^j$ for some j, we can replace u^i by x^j. Similarly, we may assume that $v^i \in \{y^1, \cdots, y^n\}$.

We prove Lemma 1 by induction on the length n of Γ. In case $n = 1$, $\omega(x, \alpha, \beta, \Gamma)$ is either 0 or 1. Consider a value of x for which $\omega(x, \alpha, \beta, \Gamma) = 1$. There exist u^1, v^1 in Γ with property $P(x, \alpha, \beta)$. By default, $u^1 = x^1$ and $v^1 = y^1$. Therefore $x_1 < x_1^1$. Also, $\sigma_\alpha(x) > \sigma_\alpha(x^1)$, or

$$x_2 - \alpha x_1 > x_2^1 - \alpha x_1^1.$$

Similarly

$$x_2 - \beta x_1 < y_2^1 - \beta y_1^1.$$

Subtraction gives the inequality

$$(\beta - \alpha) x_1 > x_2^1 - y_2^1 + (\beta - \alpha) x_1^1.$$

Therefore

$$0 < x_1^1 - x_1 < (\beta - \alpha)^{-1} (y_2^1 - x_2^1).$$

It follows that $0 < x_1^1 - t < (\beta - \alpha)^{-1}(y_2^1 - x_2^1)$ whenever $\omega(t, \alpha, \beta, \Gamma) = 1$. Since otherwise $\omega(t, \alpha, \beta, \Gamma) = 0$, the integral of $\omega(t, \alpha, \beta, \Gamma)$ is at most $(\beta - \alpha)^{-1}(y_2^1 - x_2^1)$, and the inequality (*) is valid.

Next assume that $n > 1$ and (*) is valid for all elementary sets of power at most $n - 1$. There are three cases to consider. In case 1,

$$\sigma_\alpha(x^1) \ge \cdots \ge \sigma_\alpha(x^n) \quad \text{and} \quad \sigma_\beta(y^1) \le \cdots \le \sigma_\beta(y^n).$$

For each x, choose points $u^1, v^1, \cdots, u^k, v^k$ in Γ having property $P(x, \alpha, \beta)$ for which $\omega(x, \alpha, \beta, \Gamma) = k$. For $1 \leq j \leq n$, write $\Gamma_j \equiv \{x^j, y^j\}$, so that Γ_j is an elementary set of power 1. For each i $(1 \leq i \leq k)$, there exists a unique integer $j = j(i)$ with $v^i = y^j$, and the integers $j(1), \cdots, j(k)$ are distinct. Because

$$\sigma_\beta(y^j) = \sigma_\beta(v^i) > \sigma_\beta(x) \quad \text{and} \quad \sigma_\alpha(x^j) \leq \sigma_\alpha(u^i) < \sigma_\alpha(x),$$

the points $x^{j(i)}$ and $y^{j(i)}$ have property $P(x, \alpha, \beta)$, so that $\omega(x, \alpha, \beta, \Gamma_{j(i)}) \geq 1$ for $1 \leq i \leq k$. Therefore

$$\omega(x, \alpha, \beta, \Gamma) = k \leq \sum_{j=1}^{n} \omega(x, \alpha, \beta, \Gamma_j).$$

Taking maxima, we obtain the inequality

$$\omega(t, \alpha, \beta, \Gamma) \leq \sum_{j=1}^{n} \omega(t, \alpha, \beta, \Gamma_j).$$

Now (*) is valid for Γ_j, because the power of Γ_j is 1. Therefore integration gives

$$(*) \qquad \int \omega(t, \alpha, \beta, \Gamma)\,dt \leq \sum_{j=1}^{n} \int \omega(t, \alpha, \beta, \Gamma_j)\,dt \leq \sum_{j=1}^{n} (\beta - \alpha)^{-1}(y_2^j - x_2^j).$$

In the remaining cases, there exists m $(1 \leq m \leq n - 1)$ such that

$$\sigma_\alpha(x^{m+1}) \geq \sigma_\alpha(x^{m+2}) \cdots \geq \sigma_\alpha(x^n) \quad \text{and} \quad \sigma_\beta(y^{m+1}) \leq \sigma_\beta(y^{m+2}) \leq \cdots \leq \sigma_\beta(y^n)$$

but either $\sigma_\alpha(x^m) < \sigma_\alpha(x^{m+1})$ or $\sigma_\beta(y^m) > \sigma_\beta(y^{m+1})$. The case $\sigma_\beta(y^m) > \sigma_\beta(y^{m+1})$ will be called case 2, and the case $\sigma_\alpha(x^m) < \sigma_\alpha(x^{m+1})$, $\sigma_\beta(y^m) \leq \sigma_\beta(y^{m+1})$ will be called case 3.

Consider next case 2. We define $\Omega \equiv \Gamma - \{x^{m+1}, y^{m+1}\}$. Since Γ is elementary, to verify that Ω is elementary we need only note that

$$\sigma_\alpha(x^{m+2}) \leq \sigma_\alpha(x^{m+1}) \leq \sigma_\alpha(y^m).$$

The power of Ω is $n - 1$. By the inductive hypothesis,

$$\int \omega(t, \alpha, \beta, \Omega)\,dt \leq (\beta - \alpha)^{-1} \left\{ \sum_{i=1}^{n} (y_2^i - x_2^i) - (y_2^{m+1} - x_2^{m+1}) \right\}.$$

We wish to compare $\omega(x, \alpha, \beta, \Gamma)$ with $\omega(x, \alpha, \beta, \Omega)$. To this end, choose elements $u^1, v^1, \cdots, u^k, v^k$ in Γ having property $P(x, \alpha, \beta)$ with $\omega(x, \alpha, \beta, \Gamma) = k$.

(a) If $u^j \neq x^{m+1}$ and $v^j \neq y^{m+1}$ for all j, then $\omega(x, \alpha, \beta, \Omega) = k$.

(b) If $u^j \neq x^{m+1}$ for all j but $v^j = y^{m+1}$ for some value of j, replace v^j by $\tilde{v}^j = y^m$. Then

$$u_1^j \leq \tilde{v}_1^j < u_1^{j+1} \quad \text{and} \quad \sigma_\beta(\tilde{v}^j) = \sigma_\beta(y^m) > \sigma_\beta(y^{m+1}) = \sigma_\beta(v^j) > \sigma_\beta(x).$$

Thus the points $u^1, v^1, \cdots, \tilde{v}^j, \cdots, v^k$ in Ω have property $P(x, \alpha, \beta)$. Therefore $\omega(x, \alpha, \beta, \Omega) = k$.

(c) If $v^j \neq y^{m+1}$ for all j, but $u^j = x^{m+1}$ for some value of j, replace u^j by $\tilde{u}^j = x^{m+2}$. Then

$$v^{j-1} < \tilde{u}^j \leq v^j \quad \text{and} \quad \sigma_\alpha(\tilde{u}^j) = \sigma_\alpha(x^{m+2}) \leq \sigma_\alpha(x^{m+1}) = \sigma_\alpha(u^j) < \sigma_\alpha(x).$$

Thus the points $u^1, v^1, \cdots, \tilde{u}^j, \cdots, v^k$ in Ω have property $P(x, \alpha, \beta)$. Therefore $\omega(x, \alpha, \beta, \Omega) = k$.

(d) If $u^j = x^{m+1}$, $v^j = y^{m+1}$ for some j, then $\omega(x, \alpha, \beta, \Omega) \geq k - 1$, $\omega(x, \alpha, \beta, \Gamma_{m+1}) \geq 1$. Therefore

$$\omega(x, \alpha, \beta, \Gamma) = k \leq \omega(x, \alpha, \beta, \Omega) + \omega(x, \alpha, \beta, \Gamma_{m+1}).$$

This last inequality therefore holds in all cases. Taking maxima, we see that

$$\omega(t, \alpha, \beta, \Gamma) \leq \omega(t, \alpha, \beta, \Omega) + \omega(t, \alpha, \beta, \Gamma_{m+1}).$$

Integration now gives the inequality (*):

$$\int \omega(t, \alpha, \beta, \Gamma)\,dt \leq (\beta - \alpha)^{-1} \left\{ \sum_{i=1}^{n} (y_2^i - x_2^i) - (y_2^{m+1} - x_2^{m+1}) \right\}$$

$$+ (\beta - \alpha)^{-1} (y_2^{m+1} - x_2^{m+1}) = (\beta - \alpha)^{-1} \sum_{i=1}^{n} (y_2^i - x_2^i).$$

In the final case, case 3, we obtain Ω from Γ by leaving out the points x^m and y^m and replacing x^{m+1} by the point $\tilde{x}^{m+1} \equiv (x_1^{m+1}, x_2^m + \alpha(x_1^{m+1} - x_1^m))$. Since Γ is elementary, to verify that Ω is elementary we need only note that

$$\begin{aligned} y_2^{m+1} - \tilde{x}_2^{m+1} &= y_2^{m+1} - x_2^m - \alpha(x_1^{m+1} - x_1^m) \\ &= y_2^{m+1} - x_2^{m+1} + \sigma_\alpha(x^{m+1}) - \sigma_\alpha(x^m) > y_2^{m+1} - x_2^{m+1} > 0 \end{aligned}$$

and

$$\begin{aligned} \sigma_\alpha(y^{m-1}) - \sigma_\alpha(\tilde{x}^{m+1}) &\geq \sigma_\alpha(x^m) - \sigma_\alpha(\tilde{x}^{m+1}) \\ &= x_2^m - \alpha x_1^m - \{x_2^m + \alpha(x_1^{m+1} - x_1^m) - \alpha x_1^{m+1}\} = 0. \end{aligned}$$

By the inductive hypothesis,

$$\int \omega(t, \alpha, \beta, \Omega)\,dt \leq (\beta - \alpha)^{-1} \left\{ \sum_{i=1}^{n} (y_2^i - x_2^i) - (y_2^m - x_2^m) + x_2^{m+1} - \tilde{x}_2^{m+1} \right\}$$

$$= (\beta - \alpha)^{-1} \left\{ \sum_{i=1}^{n} (y_2^i - x_2^i) - (y_2^m - x_2^m - x_2^{m+1} + x_2^m + \alpha(x_1^{m+1} - x_1^m)) \right\}$$

$$= (\beta - \alpha)^{-1} \sum_{i=1}^{n} (y_2^i - x_2^i) - (\beta - \alpha)^{-1} (\sigma_\alpha(y^m) - \sigma_\alpha(x^{m+1})).$$

Consider any x in R^2. To compare $\omega(x, \alpha, \beta, \Gamma)$ with $\omega(x, \alpha, \beta, \Omega)$, choose points $u^1, v^1, \cdots, u^k, v^k$ in Γ with property $P(x, \alpha, \beta)$ such that $\omega(x, \alpha, \beta, \Gamma) = k$. Define

$$\tilde{u}^i = \begin{cases} u^i & \text{whenever } u^i \neq x^m \text{ and } u^i \neq x^{m+1}, \\ \tilde{x}^{m+1} & \text{whenever } u^i = x^m \text{ or } u^i = x^{m+1}. \end{cases}$$

Define $\tilde{v}^i = v^i$ whenever $v^i \neq y^m$, and $\tilde{v}^i = y^{m+1}$ whenever $v^i = y^m$. The points $\tilde{u}^i$ and $\tilde{v}^i$ are all in Ω and satisfy the inequalities

$$\sigma_\alpha(\tilde{u}^i) \leq \sigma_\alpha(u^i) < \sigma_\alpha(x), \qquad \sigma_\beta(\tilde{v}^i) \geq \sigma_\beta(v^i) > \sigma_\beta(x).$$

Now $\tilde{u}_1^i \leq \tilde{v}_1^i$ for all i, and $\tilde{v}_1^i < \tilde{u}_1^{i+1}$ for $1 \leq i \leq n - 1$, except when $v^i = y^m$ and $u^{i+1} = x^{m+1}$. In this exceptional case, $x_1 < x_1^m$ and

$$\sigma_\alpha(x) = x_2 - \alpha x_1 > \sigma_\alpha(x^{m+1}) = x_2^{m+1} - \alpha x_1^{m+1},$$

$$\sigma_\beta(x) = x_2 - \beta x_1 < \sigma_\beta(y^m) = y_2^m - \beta y_1^m.$$

Subtraction gives the inequality

$$\begin{aligned}(\beta - \alpha) x_1 &> \beta y_1^m - \alpha x_1^{m+1} + x_2^{m+1} - y_2^m \\ &= (\beta - \alpha) x_1^m + \alpha(x_1^m - x_1^{m+1}) + x_2^{m+1} - y_2^m \\ &= (\beta - \alpha) x_1^m + \sigma_\alpha(x^{m+1}) - \sigma_\alpha(y^m).\end{aligned}$$

Therefore

$$0 < x_1^m - x < (\beta - \alpha)^{-1} (\sigma_\alpha(y^m) - \sigma_\alpha(x^{m+1})).$$

Thus, unless x_1 satisfies this inequality, the points $\tilde{u}^1, \tilde{v}^1, \cdots, \tilde{u}^k, \tilde{v}^k$ have property $P(x, \alpha, \beta)$, so that $\omega(x, \alpha, \beta, \Omega) \geq \omega(x, \alpha, \beta, \Gamma)$. Therefore

$$\omega(t, \alpha, \beta, \Omega) \geq \omega(t, \alpha, \beta, \Gamma),$$

except possibly when $0 < x_1^m - t < (\beta - \alpha)^{-1} (\sigma_\alpha(y^m) - \sigma_\alpha(x^{m+1}))$, and in any case $\omega(t, \alpha, \beta, \Omega) \geq \omega(t, \alpha, \beta, \Gamma) - 1$. Integration gives

$$\begin{aligned}\int \omega(t, \alpha, \beta, \Gamma)\, dt &\leq (\beta - \alpha)^{-1} (\sigma_\alpha(y^m) - \sigma_\alpha(x^{m+1})) + \int \omega(t, \alpha, \beta, \Omega)\, dt \\ &\leq (\beta - \alpha)^{-1} \sum_{i=1}^{n} (y_2^i - x_2^i).\end{aligned}$$

The inequality (*) has been shown to hold in all cases, and the lemma is proved.

Definition 5. If f is of bounded variation on [a, b], we define

$$V^+(f) = \sup\left\{\sum_{i=1}^{n} (f(y_i) - f(x_i)) : a \leq x_1 < y_1 < x_2 < y_2 < \cdots < x_n < y_n \leq b\right\}.$$

The following upcrossing inequality is our main result.

THEOREM 1. *If* f *is of bounded variation and* $\alpha < \beta$, *then the function* $\omega(t, \alpha, \beta, f)$ *of* t *is integrable, and*

$$\int \omega(t, \alpha, \beta, f)\,dt \leq (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}),$$

where id *is the identity function* $x \to x$ *from* R *to* R.

Proof. First consider the case where f is a continuous and piecewise linear function of bounded variation on a finite interval $[a, b]$. Here there exist real numbers $a \leq r_1 < s_1 \leq r_2 < s_2 \leq \cdots \leq r_n < s_n \leq b$ such that $f - \alpha\,\mathrm{id}$ is linear and increasing on each of the intervals $[r_i, s_i]$ and is nonincreasing on each of the intervals $[s_i, r_{i+1}]$, $[a, r_1]$, $[s_n, b]$. For $1 \leq i \leq n$ we define

$$x^i \equiv (s_i, f(r_i) + \alpha(s_i - r_i)), \qquad y^i \equiv (s_i, f(s_i)).$$

Write

$$\Gamma \equiv \{x^1, y^1, \cdots, x^n, y^n\}.$$

To see that Γ is elementary, note that

(a) $y_2^i - x_2^i = f(s_i) - f(r_i) - \alpha(s_i - r_i) = f(s_i) - \alpha\,\mathrm{id}(s_i) - (f(r_i) - \alpha\,\mathrm{id}(r_i)) > 0$

and

(b) $$\sigma_\alpha(y^i) - \sigma_\alpha(x^{i+1}) = f(s_i) - \alpha s_i - (f(r_{i+1}) + \alpha(s_{i+1} - r_{i+1}) - \alpha s_{i+1})$$
$$= f(s_i) - \alpha\,\mathrm{id}(s_i) - (f(r_{i+1}) - \alpha\,\mathrm{id}(r_{i+1})) \geq 0.$$

Lemma 1 implies that

$$\int \omega(t, \alpha, \beta, \Gamma)\,dt \leq (\beta - \alpha)^{-1} \sum_{i=1}^{n} (y_2^i - x_2^i)$$
$$= (\beta - \alpha)^{-1} \sum_{i=1}^{n} (f(s_i) - f(r_i) - \alpha(s_i - r_i)) = (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}).$$

To compare $\omega(x, \alpha, \beta, \Gamma)$ with $\omega(x, \alpha, \beta, f)$, choose points $u^1, v^1, \cdots, u^k, v^k$ in the graph of f having property $P(x, \alpha, \beta)$ and such that $\omega(x, \alpha, \beta, f) = k$. The set of points z^i in the graph of f for which

$$u_1^i \leq z_1^i \leq v_1^i \quad \text{and} \quad \sigma_\alpha(z^i) \leq \sigma_\alpha(u^i)$$

is closed. Therefore there exists such a point z^i for which z_1^i is a maximum, and $\sigma_\alpha(z) > \sigma_\alpha(u^i)$ whenever z is on the graph of f and $z_1^i < z_1 \leq v_1^i$. If we replace each u^i by the point z^i, we reduce the argument to the case where $\sigma_\alpha(z) > \sigma_\alpha(u^i)$

whenever z is on the graph of f and $u_1^i < z_1 < v_1^i$. Similarly, we reduce the argument further to the case where

$$\sigma_\alpha(z) > \sigma_\alpha(u^i) \qquad \text{and} \qquad \sigma_\beta(z) < \sigma_\beta(v^i)$$

whenever z is on the graph of f and $u_1^i < z_1 < v_1^i$. Thus the derivative from the right of f at u_1^i is greater than α, and the derivative from the left of f at v_1^i is greater than β. It follows that there exist integers $j = j(i)$ and $m = m(i)$ with

$$u_1^i \in [r_j, s_j), \qquad v_1^i \in (r_m, s_m].$$

Clearly, $s_j \le s_m$. Write $\tilde{u}^i \equiv x^j$, $\tilde{v}^i \equiv y^m$. Then $\tilde{u}_1^i \le \tilde{v}_1^i$. Since $f'(t) > \beta > \alpha$ for all $t \in (v_1^i, s_m)$, we have the relations

$$\sigma_\alpha(z) - \sigma_\alpha(x) > \sigma_\alpha(v^i) - \sigma_\alpha(x) = \sigma_\beta(v^i) - \sigma_\beta(x) + (\beta - \alpha)(v_1^i - x_1) > 0$$

whenever z is in the graph of f and $z_1 \in (v_1^i, s_{m(i)})$. Therefore $u_1^{i+1} > s_{m(i)}$, so that

$$\tilde{u}_1^{i+1} \ge u_1^{i+1} > s_{m(i)} = \tilde{v}_1^i.$$

Also, $x_1 < u_1^1 \le \tilde{u}_1^1$, so that

$$x_1 < \tilde{u}_1^1 \le \tilde{v}_1^1 < \tilde{u}_1^2 \le \tilde{v}_1^2 < \cdots < \tilde{u}_1^k \le \tilde{v}_1^k.$$

Moreover,

$$\sigma_\alpha(\tilde{u}^i) = \sigma_\alpha(x^j) = f(r_j) + \alpha(s_j - r_j) - \alpha s_j = f(r_j) - \alpha r_j < \sigma_\alpha(u^i) < \sigma_\alpha(x)$$

and

$$\sigma_\beta(\tilde{v}^i) = \sigma_\beta(y^k) > \sigma_\beta(v^i) > \sigma_\beta(x).$$

Therefore the points $\tilde{u}^1, \tilde{v}^1, \cdots, \tilde{u}^k, \tilde{v}^k$ of Γ have property $P(x, \alpha, \beta)$. Thus $\omega(x, \alpha, \beta, \Gamma) \ge k = \omega(x, \alpha, \beta, f)$. Since this is true for all x,

$$\int \omega(t, \alpha, \beta, f)\,dt \le \int \omega(t, \alpha, \beta, \Gamma)\,dt \le (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}).$$

Next we consider the general case. Let $\{t_n\}$ be a dense sequence of points in the domain of f that includes all points of discontinuity of f. For each positive integer n, let f_n be the function whose domain is the smallest closed interval that contains the points $t_1, \cdots, t_n$, agrees with f at these points, and is linear on each complimentary interval. By the case already considered,

$$\int \omega(t, \alpha, \beta, f_n)\,dt \le (\beta - \alpha)^{-1} V^+(f_n - \alpha\,\mathrm{id}) \le (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}).$$

For each x in R^2,

$$\omega(x, \alpha, \beta, f_1) \le \cdots \le \omega(x, \alpha, \beta, f_n) \le \cdots$$

and

$$\lim_{n\to\infty} \omega(x, \alpha, \beta, f_n) = \omega(x, \alpha, \beta, f).$$

Taking suprema, we see that the same inequalities hold for each real number t. Lebesgue's monotone convergence theorem now gives the desired inequality:

$$\int \omega(t, \alpha, \beta, f)\,dt = \lim_{n\to\infty} \int \omega(t, \alpha, \beta, f_n)\,dt \leq (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}).$$

COROLLARY 1. *Let* f *be a function of bounded variation on* R, *and let* α *and* β *be real numbers* ($\alpha < \beta$). *For each* t *in* R, *let* $\nu(t, \alpha, \beta, f)$ *be the supremum of all positive integers* k *such that there exist points* $v^1, u^2, v^2, \cdots, u^k, v^k$ *in the graph of* f *with*

$$t < v_1^1 < u_1^2 \leq v_1^2 < \cdots < u_1^k \leq v_1^k$$

and

$$\sigma_\alpha(u^i) < \sigma_\alpha(t, f(t)), \qquad \sigma_\beta(v^i) > \sigma_\beta(t, f(t)).$$

Then

$$\int \nu(x, \alpha, \beta, f)dx \leq (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}).$$

Proof. It is clear that if δ is a sufficiently small positive constant, then

$$\nu(t, \alpha, \beta, f) \leq \omega((t - \delta^2, f(t) + \delta), \alpha, \beta, f) \leq \omega(t - \delta^2, \alpha, \beta, f).$$

The inequality now follows from Theorem 1.

COROLLARY 2. *Let* f *be a function of bounded variation, and let* α *and* β ($\alpha < \beta$) *be real numbers. Then*

$$\int \omega((t, f(t)), \alpha, \beta, f)\,dt \leq (\beta - \alpha)^{-1} V(f),$$

where V(f) *is the total variation of* f.

Proof. If $\alpha \geq 0$, Theorem 1 implies that

$$\int \omega(t, \alpha, \beta, f)\,dt \leq (\beta - \alpha)^{-1} V^+(f - \alpha\,\mathrm{id}) \leq (\beta - \alpha)^{-1} V^+(f).$$

If $\beta \leq 0$, then $\alpha \leq 0$ also, and $\omega((t, f(t)), \alpha, \beta, f) = \nu(t, -\beta, -\alpha, -f)$, and therefore, by Theorem 1,

$$\int \omega((t, f(t)), \alpha, \beta, f)\,dt \leq (\beta - \alpha)^{-1} V^+(-f + \alpha\,\mathrm{id}) \leq (\beta - \alpha)^{-1} V^+(-f) \leq (\beta - \alpha)^{-1} V(f).$$

There remains the case $\alpha < 0 < \beta$. Clearly,

$$\omega(t, \alpha, \beta, f) \leq \omega(t, 0, \beta, f) \quad \text{and} \quad \omega(t, \alpha, \beta, f) \leq \omega(t, \alpha, 0, f).$$

From the above cases we also see that

$$\beta \int \omega(t, 0, \beta, f)\,dt \leq V^+(f), \qquad -\alpha \int \omega((t, f(t)), \alpha, 0, f)\,dt \leq V^+(-f).$$

Therefore

$$(\beta - \alpha) \int \omega((t, f(t)), \alpha, \beta, f)\,dt \leq \beta \int \omega(t, 0, \beta, f)\,dt - \alpha \int \omega((t, f(t)), \alpha, 0, f)\,dt$$
$$\leq V^+(f) + V^+(-f) = V(f).$$

As our first application of Theorem 1, we show that a function of bounded variation has a derivative almost everywhere. Indeed, Theorem 1 and its corollaries can be regarded as a strong generalization of this fact. They assert that the difference quotient of a function of bounded variation cannot oscillate too much.

THEOREM (Lebesgue). *A function* f *of bounded variation on a closed interval* [a, b] *has a finite derivative at almost all points* x *of* [a, b].

Proof. Let S be the subset of [a, b] on which the derivative of f from the right does not exist. Then

$$S = \bigcup_{\alpha,\beta} \{x \in [a, b]: \omega(x, \alpha, \beta, f) = \infty\},$$

where the union is taken over all pairs α, β of rational numbers with $\alpha < \beta$. By Theorem 1, $\omega(x, \alpha, \beta, f) < \infty$ almost everywhere, so that S has measure 0. Thus the derivative g of f from the right exists almost everywhere. Let T be the set of all x for which $g(x) = +\infty$. Then $\nu(x, 0, n, f) \geq 1$ for all x in T and all positive integers n. Thus the measure of T is at most

$$\int \nu(x, 0, n, f)\,dx \leq n^{-1} V^+(f),$$

and this tends to 0 as $n \to \infty$. Replacing f by $-f$, we see that the set of all x for which $g(x) = -\infty$ also has measure 0. It follows that g is finite almost everywhere. Similarly, f has a finite derivative h from the left almost everywhere. Thus to each $\varepsilon > 0$ there corresponds a $\delta > 0$ and a measurable set $U \subset [a, b]$ of measure at most ε with the property that when $x \in [a, b] - U$, then

$$\left|\frac{f(y) - f(x)}{y - x} - g(x)\right| < \varepsilon \quad \text{if } x < y \leq x + 2\delta,$$

$$\left|\frac{f(y) - f(x)}{y - x} - h(x)\right| < \varepsilon \quad \text{if } x - 2\delta \leq y < x.$$

Thus, if $x \in [a, b] - U$ and $x + \delta \in [a, b] - U$, then

$$|g(x) - h(x + \delta)| \leq \left|g(x) - \frac{f(x + \delta) - f(x)}{\delta}\right| + \left|h(x + \delta) - \frac{f(x) - f(x + \delta)}{-\delta}\right| < 2\varepsilon.$$

On the other hand,

$$|h(x) - h(x + \delta)| < \left|\frac{f(x) - f(x - \delta)}{\delta} - \left(2\,\frac{f(x + \delta) - f(x - \delta)}{2\delta} - \frac{f(x + \delta) - f(x)}{\delta}\right)\right| + 4\varepsilon = 4\varepsilon.$$

Thus $|h(x) - g(x)| < 6\varepsilon$. Since ε is arbitrary, it follows that $h(x) = g(x)$ almost everywhere, as desired.

Next we apply Theorem 1 to establish an ergodic theorem that gives a precise bound for the integral of the number of upcrossings of the relevant average.

THEOREM 2. *Let* $\{T^t\}_{t \geq 0}$ *be a one-parameter semigroup of measure-preserving transformations on a measure space* X *of total measure* 1, *and let* f *be an integrable function on* X. *Let* α *and* β *be real numbers with* $\alpha < \beta$, *and for each* x *in* X *let* h_x *be the function* $h_x(t) \equiv \int_0^t f(T^t x)dt$ *defined for* $t > 0$. *Write*

$$\omega(x) \equiv \omega((0, 0), \alpha, \beta, h_x)$$

for each x *in* X. *Then*

$$\int \omega(x)\, d\mu(x) \leq (\beta - \alpha)^{-1} \int (f - \alpha)^+ d\mu .$$

Proof. Let r be any positive constant, and let h_x^r be the restriction of h_x to the interval (0, r). By Theorem 1,

$$\int_0^r \omega((t, h_x(t)), \alpha, \beta, h_x^r)\, dt \leq (\beta - \alpha)^{-1} V^+(h_x^r - \alpha\, \mathrm{id}) = (\beta - \alpha)^{-1} \int_0^r (f(T^t x) - \alpha)^+ dt .$$

Therefore

$$\int_X \int_0^r \omega((t, h_x(t)), \alpha, \beta, h_x^r)\, dt\, d\mu(x) \leq (\beta - \alpha)^{-1} r \int (f - \alpha)^+ d\mu .$$

It is clear that

$$\omega((t, h_x(t)), \alpha, \beta, h_x^r) = \omega((0, 0), \alpha, \beta, h_u^{r-t}),$$

where $u = T^t x$. Therefore

$$\int_X r^{-1} \int_0^r \omega((0, 0), \alpha, \beta, h_x^t)\, dt\, d\mu(x) = \int_X r^{-1} \int_0^r \omega((0, 0), \alpha, \beta, h_x^{r-t})\, dt\, d\mu(x)$$

$$= r^{-1} \int_0^r \int_X \omega((0, 0), \alpha, \beta, h_u^{r-t})\, d\mu(x)\, dt$$

$$= r^{-1} \int_0^r \int_X \omega((t, h_x(t)), \alpha, \beta, h_x^r)\, d\mu(x)\, dt$$

$$\leq (\beta - \alpha)^{-1} \int (f - \alpha)^+ d\mu .$$

Now, since $\omega((0, 0), \alpha, \beta, h_x^t) \to \omega(x)$ as $t \to \infty$,

$$r^{-1}\int_0^r \omega((0,\ 0),\ \alpha,\ \beta,\ h_x^t)\,dt \to \omega(x) \qquad \text{as } r \to \infty\ .$$

Therefore, passing to the limit in the last equation, we obtain the desired inequality

$$\int_X \omega(x)\,d\mu(x) \le (\beta - \alpha)^{-1} \int (f - \alpha)^+ d\mu\ .$$

It is possible to establish variants of Theorem 2, based on Corollaries 1 and 2 of Theorem 1. Thus from Corollary 2 it follows easily that

$$\int \omega(x)\,d\mu(x) \le (\beta - \alpha)^{-1} \int |f|\, d\mu\ .$$

As it stands, Theorem 2 is not constructive, because $\omega(x)$, and therefore the quantity $\int \omega(x)\,d\mu(x)$, is not computable. The correct constructive statement is almost, but not quite, that $\int \omega(x)\,d\mu(x)$ is the least upper bound of a bounded set S of real numbers, that $(\beta - \alpha)^{-1} \int (f - \alpha)^+ d\mu$ is the greatest lower bound of a bounded set T of real numbers, and that $s \le t$ for all s in S and t in T.

We conclude with the sketch of a derivation of the upcrossing inequality of Doob [2] that is fundamental to martingale theory. Our motivation is not to give a better proof of Doob's result (we don't), but to establish the relationship of his result to Theorem 1.

THEOREM (Doob). *Let* $f_1, \cdots, f_n$ *be a semimartingale, that is, a finite sequence of integrable functions on a measure space* X *of total mass* 1, *such that*

$$\int_S f_i\,d\mu \le \int_S f_j\,d\mu \qquad (1 \le i < j \le n),$$

where S *is any measurable subset of* X *of the form*

$$S = \{x\colon a_1 \le f_1(x) \le b_1,\ \cdots,\ a_i \le f_i(x) \le b_i\}\ .$$

Let $\alpha < \beta$ *be real numbers, and for each* x *in* X *let* $\omega(x)$ *be the number of upcrossings of the interval* $[\alpha, \beta]$ *by the sequence* $f_1(x), \cdots, f_n(x)$, *that is, the maximum value of* k *for which there exist integers*

$$1 \le u^1 < v^1 < \cdots < u^k < v^k \le n$$

with $f_{u^i}(x) < \alpha$, $f_{v^i}(x) > \beta$ $(1 \le i \le k)$. *Then*

$$\int \omega\,d\mu \le (\beta - \alpha)^{-1} \int (f_n - \alpha)^+ d\mu\ .$$

Proof. We first consider the case where each f_i assumes only finitely many values. We show by induction on n that there exists a function g on [0, 1] with the following four properties:

(a) g is piecewise linear and continuous on the right, but not necessarily continuous on the left;

(b) g decreases at each of its finitely many points of discontinuity;

(c) if $X = X_1 \cup \cdots \cup X_N$ is the decomposition of X into the sets on each of which each of the functions $f_1, \cdots, f_n$ is constant, then $[0, 1]$ can be partitioned into disjoint subsets $I_1, \cdots, I_N$, each being a finite union of intervals, such that $\mu(X_k) = |I_k|$ and such that g' is constant on I_k with $g'(t) = f_n(x)$ for all t in I_k and all x in X_k;

(d) for all points t of I_k, except for a subset of arbitrarily small measure, there exist $t < t_n < \cdots < t_1 \leq 1$ such that the numbers

$$(t_i - t)^{-1}(g(t_i) - g(t)) \qquad (1 \leq i \leq n)$$

are arbitrarily near to the numbers $f_1(x), \cdots, f_n(x)$, respectively, where x is any point in X_k.

We first construct g for $n = 1$. For this it suffices to let g be any continuous piecewise linear function on $[0, 1]$ such that g' assumes the same values as f_1, with the same probabilities.

Assume next that we have constructed a function g satisfying (a), (b), (c), and (d) relative to the semimartingale $f_1, \cdots, f_n$. We wish to construct a function G that satisfies (a), (b), (c), and (d) relative to the semimartingale $f_1, \cdots, f_{n+1}$. To this end, we decompose each of the sets I_k into very many small subintervals. Let $[a, b]$ be such a subinterval of I_k. We obtain G by modifying g on each of the intervals $[a, b]$ as follows. First, G is piecewise linear on $[a, b]$ and continuous on $[a, b)$, with $G(a) = g(a)$ and $G(b) = g(b)$. Second, the values that G' assumes on $[a, b]$ are the same as those that f_{n+1} assumes on X_k, and if f_{n+1} assumes the value v on X_k with probability p, then G' assumes the value v on $[a, b]$ with probability p. It is clear that G satisfies requirements (a) and (c). To verify (b), consider the behavior of G at the possible discontinuity b.

We have the relation

$$\lim_{t \to b} G(t) = G(a) + \int_a^b G'(x)\,dx = g(a) + (b - a)(|I_k|)^{-1} \int_{X_k} f_{n+1}\,d\mu$$
$$\geq g(a) + (b - a)(|I_k|)^{-1} \int_{X_k} f_n\,d\mu = g(a) + \int_a^b g'(x)\,dx \geq g(b) = G(b).$$

Thus (b) is satisfied by G.

To verify that G satisfies (d), note first that by making the intervals $[a, b]$ small enough, we bring G uniformly close to g. Thus to verify (d) we need only note that for all except finitely many points t of $[a, b]$, the derivative $G'(t)$ exists and equals $f_{n+1}(x)$, where x is any point in X_k at which $f_{n+1}(x) = G'(t)$.

This completes the construction of g. By (b) and (c),

$$V^+(g - \alpha\,\mathrm{id}) = \int_0^1 (g'(x) - \alpha)^+ dx = \int (f_n - \alpha)^+ \,d\mu .$$

Also, by (d), for all t in I_k, except on a set of arbitrarily small measure,

$$\nu(t, \alpha, \beta, g) \geq \omega(x),$$

where x is any point in X_k. Therefore, by Corollary 1,

$$\int \omega \, d\mu \leq (\beta - \alpha)^{-1} V^+(g - \alpha \, id) = \int (f_n - \alpha)^+ d\mu .$$

Now, in the general case, it is easy to approximate the given semimartingale $f_1, \cdots, f_n$ by a semimartingale $h_1, \cdots, h_n$, on the same space X, in which each of the functions h_i assumes only finitely many values, in the sense that for every x in X either $h_i(x)$ and $f_i(x)$ are very close together or are both very large with $|h_i(x)| \leq |f_i(x)|$. Thus the general result follows by approximation from the special case already obtained.

REFERENCES

1. L. E. J. Brouwer, *Nadere addenda en corrigenda over de rol van het principium tertii exclusi in de wiskunde,* (Further addenda and corrigenda on the role of the *principium tertii exclusi* in mathematics), Nederl. Akad. Wetensch. Proc. Ser. A. 57 = Indagationes Math. 16 (1954), 109-111 (Dutch).

2. J. L. Doob, *Stochastic processes,* Wiley, New York, 1953.

The University of California, San Diego