INROADS

Real Analysis Exchange
Vol. 37(1), 2011/2012, pp. 221–242

Brian S. Thomson, Department of Mathematics, Simon Fraser University, B.C., Canada V5A1S6. email: thomson@sfu.ca

ON RIEMANN SUMS

Abstract

If a sum of the form

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

is used without the familiar requirement that the sequence of points $a = x_0, x_1, \ldots, x_n = b$ is increasing, do we still get a useful approximation to the integral? With a suitable set of hypotheses the answer is yes. We give applications to change of variable formulas and the problem of characterizing derivatives.

1 A theorem of H. E. Robbins

In discussions of an integral

$$\int_a^b f(x)\,dx$$

(in a variety of different senses) one often employs Riemann sums

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}).$$

If one drops the usual requirement that the sequence of points

$$a = x_0, x_1, x_2, \ldots, x_n = b$$

Mathematical Reviews subject classification: Primary: 26A24, 26A42
Key words: Riemann sums, Riemann integral, Lebesgue integral, Stieltjes integral, Henstock-Kurzweil integral, change of variable, derivative
Received by the editors February 8, 2011
Communicated by: Paul D. Humke

must be chosen to be increasing, but places instead an upper bound on the variation of the sequence, then the resulting sums can still be used in a familiar manner.

This is probably well-known and likely can be found even in early literature. The only reference I was able to find is a note of Robbins [7] in a 1943 issue of the MONTHLY. Robbins[1] assumes the function f is continuous and gives an unnecessarily awkward proof. We reproduce his theorem here with a more natural proof that would be accessible to beginning students of integration theory.

Theorem 1 (Robbins)**.** *Let $f : [c, d] \to \mathbb{R}$ be a continuous function and let a, $b \in [c, d]$, $\epsilon > 0$, and $C > 0$ be given. Then there exists a positive number δ with the property that*

$$\left| \int_a^b f(x)\, dx - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \epsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c, d]$ with these four properties:

1. *$a = x_0$ and $b = x_n$.*
2. *$0 < |x_i - x_{i-1}| < \delta$ for all $i = 1, 2, \ldots, n$.*
3. *ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.*
4. *$\sum_{i=1}^{n} |x_i - x_{i-1}| \le C$.*

PROOF. Take δ sufficiently small that

$$|f(x) - f(y)| < \epsilon/C$$

whenever x and y are points of $[c, d]$ for which $|y - x| < \delta$. Write

$$F(x) = \int_c^x f(t)\, dt \qquad (c \le x \le d)$$

and observe that, if $c \le x \le \xi \le y \le d$ and $0 < y - x < \delta$, then

$$\left| \frac{F(y) - F(x)}{y - x} - f(\xi) \right| = \left| \frac{1}{y - x} \int_x^y f(t)\, dt - f(\xi) \right|$$

[1]Herbert E. Robbins (1915-2001) had a long and distinguished career as a mathematical statistician. He started as a pure mathematician, obtaining a Harvard PhD in 1938 under Hassler Whitney, and then turned to statistics during the war while he was in the Navy.

$$= \left| \frac{1}{y-x} \int_x^y [f(t) - f(\xi)]\, dt \right| < \frac{\epsilon}{C}$$

so that

$$|F(y) - F(x) - f(\xi)(y-x)| < \frac{\epsilon}{C}(y-x).$$

Then, for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c, d]$ with the four properties in the statement of the theorem,

$$\left| \int_a^b f(x)\, dx - \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) \right| = \left| F(b) - F(a) - \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) \right|$$

$$= \left| \sum_{i=1}^n [F(x_i) - F(x_{i-1}) - f(\xi_i)(x_i - x_{i-1})] \right|$$

$$\leq \sum_{i=1}^n |F(x_i) - F(x_{i-1}) - f(\xi_i)(x_i - x_{i-1})| < \frac{\epsilon}{C} \sum_{i=1}^n |x_i - x_{i-1}| \leq \epsilon.$$

□

By the same methods one can prove a similar result that uses an infinite sequence of points $\{x_n\}$. The conclusion would be

$$\left| \int_a^b f(x)\, dx - \sum_{i=1}^\infty f(\xi_i)(x_i - x_{i-1}) \right| < \epsilon$$

and the assumptions would include

$$a = x_0, \quad b = \lim_{n \to \infty} x_n, \text{ and } \sum_{i=1}^\infty |x_i - x_{i-1}| \leq C.$$

This is not likely, however, to be of any great interest.

2 A pointwise version of Robbins's theorem

The theorem of Robbins is true for all derivatives, not just continuous ones, if one uses a familiar refined version of closeness for the points. For our second closely-related theorem the integral must exist at least in the Henstock-Kurzweil sense (although it is possible for F' here to be integrable also in a narrower sense). Note that it is equally elementary as Robbins's uniform version with an equally trivial proof.

Theorem 2. *Let $F : [c, d] \to \mathbb{R}$ be a differentiable function and let a, $b \in [c, d]$, $\epsilon > 0$, and $C > 0$ be given. Then there is a positive function $\delta : [c, d] \to \mathbb{R}^+$ with the property that*

$$\left| \int_a^b F'(x)\, dx - \sum_{i=1}^n F'(\xi_i)(x_i - x_{i-1}) \right| < \epsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c, d]$ with these four properties:

1. *$a = x_0$ and $b = x_n$.*
2. *$0 < |x_i - x_{i-1}| < \delta(\xi_i)$ for all $i = 1, 2, \ldots, n$.*
3. *ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.*
4. *$\sum_{i=1}^n |x_i - x_{i-1}| \le C$.*

PROOF. Note that the integral

$$\int_a^b F'(x)\, dx = F(b) - F(a)$$

exists in the sense of the Henstock-Kurzweil integral, although it may also be integrable in the Riemann sense, the improper Riemann sense or the Lebesgue sense. For each point ξ in $[c, d]$ take $\delta(\xi)$ sufficiently small that

$$\left| \frac{F(y) - F(x)}{y - x} - F'(\xi) \right| < \frac{\epsilon}{C}$$

whenever x and y are points in $[c, d]$ for which $x \le \xi \le y$ and $0 < y - x < \delta(\xi)$. This gives us

$$|F(y) - F(x) - F'(\xi)(y - x)| < \frac{\epsilon}{C}(y - x).$$

Then, for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c, d]$ with the four properties of the statement of the theorem,

$$\left| \int_a^b F'(x)\, dx - \sum_{i=1}^n F'(\xi_i)(x_i - x_{i-1}) \right|$$
$$= \left| \sum_{i=1}^n [F(x_i) - F(x_{i-1}) - F'(\xi_i)(x_i - x_{i-1})] \right|$$
$$\le \sum_{i=1}^n |F(x_i) - F(x_{i-1}) - F'(\xi_i)(x_i - x_{i-1})| < \frac{\epsilon}{C} \sum_{i=1}^n |x_i - x_{i-1}| \le \epsilon.$$

□

3 Converses

We now show that these two statements have converses. Our first theorem (the converse of Robbin's theorem) shows that the assumption of continuity in the theorem is essential. Thus continuity can be characterized as a kind of strong integrability requirement, a *super-Riemann integrability* as we might perhaps express it.

Theorem 3. *A function $f : [a, b] \to \mathbb{R}$ is continuous if and only if it has the following strong Riemann integrability property: there is a number I so that, for any choice of numbers $\epsilon > 0$ and $C > 0$, there exists a positive number δ such that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \epsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[a, b]$ fulfilling these conditions:

1. *$a = x_0$ and $b = x_n$.*
2. *$0 < |x_i - x_{i-1}| < \delta$ for all $i = 1, 2, \ldots, n$.*
3. *ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.*
4. *$\sum_{i=1}^{n} |x_i - x_{i-1}| \leq C$.*

Proof. Robbin's theorem shows that every continuous function does have this stronger version of the Riemann integrability property. Let us then suppose that f is a function possessing this property. Note first that such a function f with these properties would have to be Riemann integrable and that the number I in the statement would necessarily be

$$I = \int_a^b f(x)\, dx.$$

Suppose that there is a point z of discontinuity of f in the interval. Then there must be a positive number $\eta > 0$ so that any interval $[z, z + t]$ (or any interval $[z - t, z]$) has points z_1 and z_2 for which $f(z_1) - f(z_2) > \eta$. We suppose it is the former case and from this we will obtain a contradiction to the statement in the theorem.

We give the details assuming this and that $a < z < b$. Now we apply the strong Riemann integrability property, using $\epsilon < \eta/4$ and $C = b - a + 4$, to obtain a choice of δ that meets the conditions of the theorem. Choose a number $0 < t < 1$ smaller than δ and so that $z + t < b$. Let r be the least integer so that $rt > 1$. Note that, consequently,

$$1 < rt = (r-1)t + t \leq 1 + t < 2.$$

Using these values of t and r, we construct a sequence that takes advantage of our assumption about the point z, i.e., the assumption that the function f is discontinuous on the right at z. Begin by choosing points c_1, c_2 from the interval $[z, z+t]$ so that

$$f(c_1) - f(c_2) > \eta.$$

Then write

$$u_0 = z,\ u_1 = z + t,\ u_2 = z,\ u_3 = z + t, \dots,\ u_{2r} = z$$

and $\upsilon_{2i-1} = c_1$ and $\upsilon_{2i} = c_2$ for each i. Note that

$$\sum_{i=1}^{2r} |u_i - u_{i-1}| = 2rt < 4$$

and

$$\sum_{i=1}^{2r} f(\upsilon_i)(u_i - u_{i-1}) = \sum_{i=1}^{r} [f(c_1) - f(c_2)]t > \eta rt > \eta.$$

Now construct a sequence

$$a = z_0 < z_1 < \dots z_p = z$$

along with associated points ζ_i so that $0 < z_i - z_{i-1} < \delta$ and so that

$$\left| \int_a^z f(x)\,dx - \sum_{i=1}^{p} f(\zeta_i)(z_i - z_{i-1}) \right| < \eta/4.$$

We also need a sequence

$$z = w_0 < w_1 < \dots w_q = b$$

along with associated points ω_i so that $0 < w_i - w_{i-1} < \delta$ and so that

$$\left| \int_z^b f(x)\,dx - \sum_{i=1}^{q} f(\omega_i)(w_i - w_{i-1}) \right| < \eta/4.$$

Both of these just use the Riemann integrability of f.

We put these three sequences together in this way

$$a = z_0 < z_1 < \cdots < z_p = z = u_0, u_1, \dots, u_{2r} = z = w_0 < w_1 < \dots w_q = b$$

to form a new sequence $a = x_0, x_1, \dots, x_N = b$ for which $|x_i - x_{i-1}| < \delta$ and for which

$$\sum_{i=1}^{N} |x_i - x_{i-1}| = (z - a) + 2rt + (b - z) = b - a + 2rt < b - a + 4 = C.$$

We use ξ_i in each case as the appropriate intermediate point used earlier: thus associated with an interval $[z_{i-1}, z_i]$ we had used ζ_i; associated with an interval $[w_{i-1}, w_i]$ we had used ω_i; while associated with a pair (u_{i-1}, u_i) we had used c_1 or c_2 depending on whether the interval goes forward or backwards.

Consider the sum

$$\sum_{i=1}^{N} f(x_i)(x_i - x_{i-1})$$

taken over the entire sequence thus constructed. Because the points satisfy the conditions of the theorem for the δ selected we must have

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}) \right| < \epsilon < \eta/4.$$

On the other hand

$$\left[\int_a^z f(x)\,dx + \int_z^b f(x)\,dx - \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}) \right]$$

$$= \left[\int_a^z f(x)\,dx - \sum_{i=1}^{p} f(\zeta_i)(z_i - z_{i-1}) \right]$$

$$+ \left[\int_z^b f(x)\,dx - \sum_{i=1}^{q} f(\omega_i)(w_i - w_{i-1}) \right]$$

$$+ \left[\sum_{i=1}^{r} f(\upsilon_i)(u_i - u_{i-1}) \right].$$

From this we deduce that

$$\sum_{i=1}^{r} f(\upsilon_i)(u_i - u_{i-1}) < 3\eta/4.$$

and yet we recall that

$$\sum_{i=1}^{r} f(v_i)(u_i - u_{i-1}) > \eta r t > \eta.$$

This contradiction completes the proof. □

Our second theorem (the converse of Theorem 2) shows that the assumption that f is an exact derivative in the theorem is essential. A single point where this fails would not be allowed. In fact then the property of being an exact derivative can be characterized as a kind of strong integrability requirement, a *super-Henstock-Kurzweil integrability* as we would be inclined to express it.

Theorem 4. *A function $f : [a,b] \to \mathbb{R}$ is an exact derivative if and only if it has the following strong integrability property: there is a number I so that, for any choice of $\epsilon > 0$ and $C > 0$, there must exist a positive function $\delta : [a,b] \to \mathbb{R}^+$ with the property that*

$$\left| I - \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \right| < \epsilon$$

for any choice of points $x_0, x_1, \ldots, x_n$ and $\xi_1, \xi_2, \ldots, \xi_n$ from $[c,d]$ with these four properties:

1. *$a = x_0$ and $b = x_n$.*
2. *$0 < |x_i - x_{i-1}| < \delta(\xi_i)$ for all $i = 1, 2, \ldots, n$.*
3. *ξ_i belongs to the interval with endpoints x_i and x_{i-1} for $i = 1, 2, \ldots, n$.*
4. *$\sum_{i=1}^{n} |x_i - x_{i-1}| \le C$.*

Proof. The proof is structured so as to be similar in many details to the proof of Theorem 3. First we observe that Theorem 2 shows that every derivative does have this stronger version of the Henstock-Kurzweil integrability property. Let us then suppose that f is a function possessing this property. Under the hypotheses here, f is evidently Henstock-Kurzweil integrable on this interval and so there is an indefinite integral F with the property that

$$F(b) - F(a) = \int_a^b f(t)\,dt = I$$

where I is the number stated in the theorem.

Suppose that there is a point z in the interval at which it is not true that $F'(z) = f(z)$. One possibility is that this is because the upper right-hand (Dini) derivative at z exceeds $f(z)$ by some positive value $\eta > 0$. Another is that the value $f(z)$ exceeds the upper right-hand (Dini) derivative at z by some positive value $\eta > 0$. There are six other possibilities, corresponding to the other three Dini derivatives under which $F'(z) = f(z)$ might fail. It is sufficient for a proof that we show that this first possibility cannot occur. From this we will obtain a contradiction to the statement in the theorem.

Thus we will assume that there must be a positive number $\eta > 0$ so that we can choose an arbitrarily small positive number t so that the interval $[z, z+t]$ has this property:

$$\frac{F(z+t) - F(z)}{t} > f(z) + \eta$$

and hence so that

$$F(z+t) - F(z) > f(z)t + \eta t.$$

We give the details assuming this and that $a < z < b$. Now we apply the theorem using $\epsilon < \eta/4$, and $C = b - a + 6$ to obtain a choice of positive function δ that meets the conditions of the theorem. Choose a number $0 < t < 1$ for which $t < \delta(z)$ and $z + t < b$ and with the property that

$$F(z+t) - F(z) > f(z)t + \eta t.$$

Let s be the least integer so that $st > 2$. Note that, consequently,

$$2 < st = (s-1)t + t \leq 2 + t < 3.$$

We first select a sequence of points

$$z = u_0 < u_1 < u_2 < \cdots < u_{k-1} = z + t$$

and points υ_i from $[x_{i-1}, x_i]$ so that $0 < u_i - u_{i-1} < \delta(\upsilon_i)$ and

$$\left| F(z+t) - F(z) - \sum_{i=1}^{k-1} f(\upsilon_i)(u_i - u_{i-1}) \right| < \eta t/2$$

This is possible simply because f is Henstock-Kurzweil integrable on the interval $[z, z+t]$. Now we add in the point $u_k = z$ and $\upsilon_k = z$.

We compute that

$$\sum_{i=1}^{k} f(\upsilon_i)(u_i - u_{i-1}) = -f(z)t + \sum_{i=1}^{k-1} f(\upsilon_i)(u_i - u_{i-1})$$

$$> -[F(z+t) - F(z) - \eta t] + \sum_{i=1}^{k-1} f(\upsilon_i)(u_i - u_{i-1}) > \eta t/2.$$

while at the same time

$$\sum_{i=1}^{k} |x_i - x_{i-1}| = 2t.$$

Repeat this sequence

$$z = u_0 < u_1 < \cdots < u_{k-1} > u_k = z$$

exactly s times so as to produce a sequence

$$z = u_0, u_1, \ldots u_{r-1}, u_r = z$$

with the property that

$$\sum_{i=1}^{r} f(\upsilon_i)(u_i - u_{i-1}) > \eta st/2 > \eta$$

while at the same time

$$\sum_{i=1}^{r} |u_i - u_{i-1}| = 2st < 6.$$

Now construct a sequence

$$a = z_0 < z_1 < \cdots < z_p = z$$

along with associated points ζ_i so that $0 < z_i - z_{i-1} < \delta(\zeta_i)$ and so that

$$\left| \int_a^z f(x)\,dx - \sum_{i=1}^{p} f(\zeta_i)(z_i - z_{i-1}) \right| < \eta/4.$$

We also need a sequence

$$z = w_0 < w_1 < \ldots w_q = b$$

along with associated points ω_i so that $0 < w_i - w_{i-1} < \delta(\omega_i)$ and so that

$$\left| \int_z^b f(x)\,dx - \sum_{i=1}^{q} f(\omega_i)(w_i - w_{i-1}) \right| < \eta/4.$$

Both of these just use the Henstock-Kurzweil integrability of f.

Now we put these three sequences together in this way

$$a = z_0 < z_1 < \cdots < z_p = z = u_0, u_1, \ldots, u_r = z = w_0 < w_1 < \ldots w_q = b$$

to form a new sequence $a = x_0, x_1, \ldots, x_N = b$ for which $|x_i - x_{i-1}| < \delta(\xi_i)$ and for which

$$\sum_{i=1}^{N} |x_i - x_{i-1}| = (z - a) + 2st + (b - z) = b - a + 2st < b - a + 6 = C.$$

We use ξ_i in each case as the appropriate intermediate point used earlier: thus associated with an interval $[z_{i-1}, z_i]$ we had used ζ_i; associated with an interval $[w_{i-1}, w_i]$ we had used ω_i; while associated with a pair (u_{i-1}, u_i) we use υ_i.

Consider the sum

$$\sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1})$$

taken over the entire sequence thus constructed. Because the points satisfy the conditions of the theorem for the δ function selected we must have

$$\left| \int_a^b f(x)\,dx - \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}) \right| < \epsilon < \eta/4.$$

On the other hand

$$\left[\int_a^z f(x)\,dx + \int_z^b f(x)\,dx - \sum_{i=1}^{N} f(\xi_i)(x_i - x_{i-1}) \right]$$
$$= \left[\int_a^z f(x)\,dx - \sum_{i=1}^{p} f(\zeta_i)(z_i - z_{i-1}) \right]$$
$$+ \left[\int_z^b f(x)\,dx - \sum_{i=1}^{q} f(\omega_i)(w_i - w_{i-1}) \right]$$
$$+ \left[\sum_{i=1}^{r} f(\upsilon_i)(u_i - u_{i-1}) \right].$$

From this we deduce that

$$\sum_{i=1}^{r} f(\xi_i)(u_i - u_{i-1}) < 3\eta/4$$

and yet we recall that

$$\sum_{i=1}^{r} f(\xi_i)(u_i - u_{i-1}) > \eta st/2 > \eta.$$

This contradiction completes the proof. □

Characterizing derivatives? Theorem 4 could be considered to be a characterization of the property that a function is the exact derivative of some other function. Since we are imposing a condition that transparently is stronger than integrability then this answers (formally at least and expressed in terms of Riemann sums) the question of what integrable functions are derivatives. We can interpret Theorem 4 as providing a necessary and sufficient that a given function f on an interval would be the derivative of some other function there.

But does this indeed constitute a characterization of derivatives? It was an old problem of W. H. Young to determine, if possible, necessary and sufficient conditions on a function f in order that it should be the derivative of some other function. Elementary students know only one sufficient condition (that f might be continuous) and perhaps one necessary condition (that f should have the intermediate value property). Advanced students know a number of others but there is to date no completely satisfactory statement of a condition that is both necessary and sufficient. (See the discussion in Bruckner and Thomson [1] for further background and history and the full quote from Young.)

The characterization of Theorem 4 is closely related to a similar one given by Chris Freiling [3]. The discussion the author gives in that paper as to what constitutes a logical and meaningful characterization should be consulted for its clarity and insight. As Chris has put it:

> "We also argue that the only way to characterize derivatives is by using some object or procedure which is at least as complicated as an integral."

That certainly applies to Theorem 4 which is essentially an integration method.

4 Change of variables

We include now a discussion of change of variables formulas for integrals on the real line. Theorems 1 and 2 give particularly easy and transparent versions.

The traditional change of variables formula is

$$F(G(b)) - F(G(a)) = \int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

where

$$F(x) = \int_{G(a)}^{x} f(s)\,ds \quad\text{and}\quad G(t) = \int_a^t g(u)\,du.$$

All or part of this formula may be given and only some of the integrals are assumed or proved to exist in some sense (either Riemann, improper Riemann, Lebesgue, or Henstock-Kurzweil).

Our first lemma shows that the part of this identity that asserts that

$$\int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

is available trivially in most cases. This is most likely well-known, but we supply the simple proof for convenience.

Lemma 5. *Let g be a Henstock-Kurzweil [Riemann] integrable function with an indefinite integral G on an interval $[a,b]$ and suppose that f is a real-valued function on $G([a,b])$. Then*

$$\int_a^b f(G(x))dG(x) = \int_a^b f(G(t))g(t)\,dt$$

where if one of the integrals exists in the Henstock-Kurzweil [Riemann] sense so too does the other and the stated identity is valid.

Proof. If f is assumed to be bounded then a rather simple proof can be given (both for the Riemann and Henstock-Kurzweil integrals). For unbounded functions we first partition the interval $[a,b]$ as follows. For each $m = 1, 2, 3, \dots$ let

$$X_m = \{t \in [a,b] : m-1 \le |f(G(t))| < m\}.$$

Let $\epsilon > 0$ and, for each $m = 1, 2, 3, \dots$, choose a positive function δ_m on $[a,b]$ so that for any points $a = t_0 < t_1 < \cdots < t_n = b$ and $t_{i-1} \le \tau_i \le t_i$ for which $0 < t_i - t_{i-1} < \delta_m(\tau_i)$ we must have

$$\sum_{i=1}^{n} |G(t_i) - G(t_{i-1}) - g(\tau_i)(x_i - x_{i-1})| < \epsilon m^{-1} 2^{-m}.$$

This just uses the fact that G is an indefinite Henstock-Kurzweil integral of g.

Define, for each $t \in [a,b]$, $\delta^*(t) = \delta_m(t)$ provided $G(t) \in X_m$. Suppose now that we have any points $a = t_0 < t_1 < \cdots < t_n = b$ and $t_{i-1} \le \tau_i \le t_i$ for which $0 < t_i - t_{i-1} < \delta^*(\tau_i)$.

Note that

$$\sum_{i=1}^{n} |f(G(\tau_i))[\,G(t_i) - G(t_{i-1})\,] - f(G(\tau_i))g(\tau_i)[\,x_i - x_{i-1}\,]| < \sum_{m=1}^{\infty} \epsilon 2^{-m} = \epsilon$$

by summing separately the terms for which τ_i belongs to X_m. We also have then

$$\left| \sum_{i=1}^{n} f(G(\tau_i))\,[G(t_i) - G(t_{i-1})] - \sum_{i=1}^{n} f(G(\tau_i))g(\tau_i)\,[x_i - x_{i-1}] \right| < \epsilon.$$

This means the Riemann sums for the two integrals are arbitrarily close together and this can be used to prove that the two integrals

$$\int_a^b f(G(x))dG(x) \text{ and } \int_a^b f(G(t))g(t)\,dt$$

exist as Henstock-Kurzweil integrals if one of them exists and that they are equal.

This completes the proof for the Henstock-Kurzweil integral. For the Riemann case one assumes that f is bounded and uses a constant δ in place of a function. Otherwise the proof is unchanged. □

We now apply Theorem 1 to give what may be the simplest nontrivial version of a change of variables formula. Robbins gave no applications of his adjusted Riemann sums result in [7], contenting himself with a brief statement and proof amounting to little more than a single page. His paper alludes, however, to this idea occurring during an investigation of change of variables formulas. Thus, no doubt, the theorem which follows uses the method he had in mind.

Lemma 6. *Let G be a continuous function of bounded variation on an interval $[a, b]$ and suppose that f is continuous on $G([a, b])$. Then*

$$\int_{G(a)}^{G(b)} f(x)\,dx = \int_a^b f(G(t))dG(t)$$

where the integrals exist in the Riemann and Riemann-Stieltjes senses respectively.

PROOF. Let $\epsilon > 0$ and define $C = \mathrm{Var}(G, [a, b])$. Choose $\delta > 0$ so that the conditions in Theorem 1 are met on the interval $G([a, b])$. Choose $\delta_1 > 0$ so that

$$|G(s) - G(t)| < \delta$$

if s and t are points in $[a,b]$ with $|s-t|<\delta_1$. Choose any points $a=t_0<t_1<\cdots<t_n=b$ and $t_{i-1}\le\tau_i\le t_i$ for which $0<t_i-t_{i-1}<\delta_1$ and consider the Riemann-Stieltjes sum

$$\sum_{i=1}^{n} f(G(\tau_i))[G(t_i)-G(t_{i-1})].$$

Consider the points $x_i=G(t_i)$, $\xi_i=G(\tau_i)$. Note that $x_0=G(a)$, $x_n=G(b)$, $|x_i-x_{i-1}|<\delta$ and

$$\sum_{i=1}^{n}|x_i-x_{i-1}|=\sum_{i=1}^{n}|G(t_i)-G(t_{i-1})|\le \mathrm{Var}(G,[a,b])=C.$$

In order to apply Theorem 1 we would need to know that ξ_i is between the points x_{i-1} and x_i, i.e., that $G(\tau_i)$ is between $G(t_{i-1})$ and $G(t_i)$. This may not be the case. Should one of these fail we return to our original Riemann-Stieltjes sum and replace the offending term by using

$$f(G(\tau_i))[G(t_i)-G(t_{i-1})]=f(G(\tau_i))[G(\tau_i)-G(t_{i-1})]+f(G(\tau_i))[G(t_i)-G(\tau_i)].$$

Having prepared our sum in this way we can then proceed as described and claim, that in each case, ξ_i is between the points x_{i-1} and x_i.

Consequently, applying Theorem 1, we have

$$\left|\int_{G(a)}^{G(b)} f(t)\,dt-\sum_{i=1}^{n} f(G(\tau_i)[G(t_i)-G(t_{i-1})]\right|$$

$$=\left|\int_{G(a)}^{G(b)} f(t)\,dt-\sum_{i=1}^{n} f(\xi_i)(x_i-x_{i-1})\right|<\epsilon.$$

This proves the existence of the Riemann-Stieltjes integral and establishes the formula. □

Lemma 6 holds for all continuous f, but does not hold for Riemann integrable functions. Take G as the Cantor singular function on $[0,1]$ which is continuous, increasing, and with a zero derivative almost everywhere. The corresponding Riemann-Stieltjes integral does not exist for all Riemann integrable f.

Exactly the same method used in Lemma 6 gives another elementary version of a change of variables formula for the Henstock-Kurzweil integral.

Theorem 7. *Let G be a continuous function of bounded variation on an interval $[a,b]$ and suppose that F is differentiable on $G([a,b])$. Then*

$$F(G(b)) - F(G(a)) = \int_{G(a)}^{G(b)} F'(x)\,dx = \int_a^b F'(G(t))dG(t)$$

where the integrals exists in the Henstock-Kurzweil sense.

Finally we give a more formal version that continues the theme and uses a recognizably similar argument.

Theorem 8. *Let g and G be functions defined on an interval $[a,b]$ for which $G'(t) = g(t)$ for a.e. point t in $[a,b]$ and suppose that f and F are functions defined on an interval $[c,d]$ that includes $G([a,b])$ for which $F'(x) = f(x)$ for a.e. point x in $[c,d]$. Then*

$$F(G(b)) - F(G(a)) = \int_a^b f(G(t))g(t)\,dt \tag{1}$$

where

1. *the identity (1) holds in the sense of the Henstock-Kurzweil integral if and only if the composition $F \circ G$ is ACG_* on $[a,b]$.*
2. *the identity (1) holds in the sense of the Lebesgue integral if and only if the composition $F \circ G$ is absolutely continuous on $[a,b]$.*

PROOF. There are a number of different characterizations of the concept ACG_* that plays such an important role in the study of the Henstock-Kurzweil integral (see [8, Chapter VII]). The simplest (the one that we use) merely requires that the function have zero variation on all sets of measure zero. Specifically H is ACG_* on $[a,b]$ if and only if for all $\epsilon > 0$ and all sets $N \subset [a,b]$ of measure zero there is a positive function δ on N so that

$$\sum_{i=1}^{k} |H(q_i) - H(p_i)| < \epsilon$$

if $[p_1, q_1]$, $\dots [p_k, q_k]$ are nonoverlapping subintervals of $[a,b]$ satisfying, for some choice of $\tau_i \in N \cap [p_i, q_i]$, the inequalities

$$0 < q_i - p_i < \delta(\tau_i) \qquad (i = 1, 2, 3, \dots, k).$$

For the second part of the theorem we need, also, to remember that a function is absolutely continuous if and only if it is ACG_* and has bounded variation.

The proof uses only one idea that is not a near trivial manipulation of Riemann sums. We need to know that a function H that has a derivative $H'(t)$ at each point of a set E for which $H(E)$ is of measure zero must have $H'(t) = 0$ at a.e. point of E. See, for example, [9] who also use this fact to prove their version of this theorem.

The condition that the composition $F \circ G$ should be ACG$_*$ is clearly necessary since all indefinite Henstock-Kurzweil integrals have this property. We shall show that it is also sufficient. Thus let us assume that $F \circ G$ is ACG$_*$ on $[a, b]$.

Let $N_1 \subset [a, b]$ be the measure zero set of points $t \in [a, b]$ at which $G'(t)$ does not exist. Let M be the measure zero set of points $x \in [c, d]$ at which $F'(x) = f(x)$ fails. Let N_2 be the set of points t in $[a, b]$ at which $G'(t)$ exists, is not equal to zero and for which $G(t)$ is in M. Since M has measure zero it follows (from our remark above) that N_2 also has measure zero.

Let $g_1(t) = 0$ if t is in either of the sets N_1 or N_2 and let $g_1(t) = g(t)$ at all other values of t. Since g and g_1 agree almost everywhere it is enough to prove the theorem using g_1 instead of g.

Let $\epsilon > 0$ and for each point t in $[a, b]$ but not in $N_1 \cup N_2$ choose $\delta(t) > 0$ so that

$$|F(G(p)) - F(G(q)) - f(G(t))g(t)(q - p)| < \frac{\epsilon}{2(b - a)}(q - p)$$

if $t \in [p, q]$ and $0 < q - p < \delta(t)$. This just uses the fact that we can compute the derivative of $F \circ G$ at each such point.

For all points t in $N_1 \cup N_2$ choose $\delta(t) > 0$ so that

$$\sum_{i=1}^{k} |F(G(q_i)) - F(G(p_i))| < \frac{\epsilon}{2}$$

if the intervals $[p_1, q_1], \ldots [p_k, q_k]$ are nonoverlapping and satisfy

$$0 < q_i - p_i < \delta(\tau_i)$$

for some choice of $\tau_i \in (N_1 \cup N_2) \cap [p_i, q_i]$. This just uses the fact that $F \circ G$ is ACG$_*$ on $[a, b]$ which we have assumed.

Choose any partition of $[a, b]$ that is finer than δ, i.e., take any points $a = t_0 < t_1 < \cdots < t_n = b$ and $t_{i-1} \le \tau_i \le t_i$ for which $0 < t_i - t_{i-1} < \delta(\tau_i)$. We must have

$$\left| F(G(b)) - F(G(a)) - \sum_{i=1}^{n} f(G(\tau_i))g_1(\tau_i)(t_i - t_{i-1}) \right|$$

$$\leq \sum_{i=1}^{n} |F(G(t_i)) - F(G(t_{i-1})) - f(G(\tau_i))g_1(\tau_i)(t_i - t_{i-1})|$$

$$\leq \sum_{\tau_i \in N_1 \cup N_2} |F(G(t_i)) - F(G(t_{i-1}))|$$

$$+ \sum_{\tau_i \notin N_1 \cup N_2} |F(G(t_i)) - F(G(t_{i-1})) - f(G(\tau_i))g_1(\tau_i)(t_i - t_{i-1})|$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2(b-a)} \sum_{i=1}^{n} (t_i - t_{i-1}) \leq \epsilon.$$

By definition then the identity

$$F(G(b)) - F(G(a)) = \int_a^b f(G(t))g_1(t)\,dt$$

and hence also the identity (1) holds in the sense of the Henstock-Kurzweil integral.

For the second part of the theorem it is enough to recall that a function is absolutely continuous if and only if it is ACG_* and has bounded variation. Since indefinite Lebesgue integrals are absolutely continuous part two follows from part one. □

Theorem 8 is a more general version of a theorem on change of variables given by Serrin and Varberg [9]. In a sense it appears definitive but, on inspecting the proof, it is clear that it is not deep and merely gives a formal condition for the formula. The formula itself is then essentially always true provided one can establish integrability. But, in any application, it might not be so easy or straightforward to determine properties of the composition $F \circ G$.

In general, it is possible for two function F and G to be absolutely continuous and yet the composition $F \circ G$ is not (see [8, p. 286]). When F is Lipschitz (as Corollaries 9 and 10 now illustrate) this is not a difficulty. It is easy to establish that the composition $F \circ G$ is absolutely continuous when F is Lipschitz and G is absolutely continuous.

Corollary 9. *Let g be Lebesgue integrable on $[a,b]$, let G be its indefinite integral, and suppose that F is a Lipschitz function defined on the interval $G([a,b])$. Then*

$$F(G(b)) - F(G(a)) = \int_a^b F'(G(t))g(t)\,dt$$

where the integral exists as a Lebesgue integral.

Corollary 10. *Let g be Henstock-Kurzweil integrable on $[a,b]$, let G be its indefinite integral, and suppose that F is a Lipschitz function defined on the interval $G([a,b])$. Then*

$$F(G(b)) - F(G(a)) = \int_a^b F'(G(t))g(t)\,dt$$

where the integral exists as a Henstock-Kurzweil integral.

5 Theorem of Kestelman, Preiss, and Uher

Theorem 8 can also be used to clarify the situation for the Riemann integral. The complete picture is available in Kestelman [4] and Preiss and Uher [6]. We reproduce this result here with a somewhat new proof.

Theorem 11 (Kestelman-Preiss-Uher)**.** *Suppose that g is Riemann integrable on an interval $[a,b]$ with an indefinite integral*

$$G(t) = \int_a^t g(u)\,du \quad (a \le t \le b)$$

and that f is a bounded function on $G([a,b])$. Then the identity

$$\int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

holds with all integrals interpreted in the Riemann sense provided either f is Riemann integrable on $G([a,b])$, or the second integral exists as a Riemann-Stieltjes integral, or finally the function $(f \circ G)g$ is Riemann integrable on $[a,b]$.

Proof. Suppose first that f is Riemann integrable on $G([a,b])$ and that F is its indefinite integral. By Corollary 9 we have immediately that

$$F(G(b)) - F(G(a)) = \int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))g(t)\,dt$$

where the function $(f \circ G)g$ must be Lebesgue integrable on $[a,b]$.

Thus it is sufficient that we prove that this function is also Riemann integrable as well. Clearly the function is bounded so it is enough to prove that it is continuous a.e. on $[a,b]$ (i.e., to use Lebesgue's criterion for Riemann integrability).

Our analysis[2] is similar to the methods in Kestelman [4]. Roy Davies [2] gave an alternative proof that directly uses Riemann's criterion for integrability.

Let S_1 be the set of points in $[a,b]$ at which g is not continuous. Let N be the set of points in $G([a,b])$ at which f is not continuous. Let S_2 be the set of points t in $[a,b] \setminus S_1$ at which $G(t) \in N$ and $g(t) \neq 0$.

The function $(f \circ G)g$ is continuous at any point t that is not in $S_1 \cup S_2$. The set S_1 is a set of measure zero because g is Riemann integrable. The set S_2 maps by G into the zero measure set N and G is differentiable with a nonzero derivative at each point of S_2.

Recall that we previously used (in the proof of Theorem 8) the fact that a function H that has a derivative $H'(t)$ at each point of a set E for which $H(E)$ is of measure zero must have $H'(t) = 0$ at a.e. point of E. This implies here that S_2 must be a measure zero set. Consequently $(f \circ G)g$ is continuous a.e. in $[a,b]$ as we require.

Let us now suppose that the function $(f \circ G)g$ is Riemann integrable on $[a,b]$ and prove that f must also be Riemann integrable. Then it must follow that

$$\int_{G(a)}^{G(b)} f(s)\,ds = \int_a^b f(G(t))g(t)\,dt$$

by what we just proved. It is sufficient, then, simply to show that f is continuous a.e. in $G([a,b])$.

Our proof is similar to the analysis given in the first part of the theorem. Preiss and Uher [6] directly use the Riemann criterion for integrability. There is also a proof of this fact in Navrátil [5] where he uses the Darboux integral instead. (Both of these papers are in Czech which presents difficulties to some of us.)

Let A be the set of points in $[a,b]$ at which either g is not continuous or $(f \circ G)g$ is not continuous. This is a set of measure zero since both of these functions are Riemann integrable. It is also true that $G(A)$ is a set of measure zero in $G([a,b])$ since G is Lipschitz.

Let B be the set of points t in $[a,b]$ at which g is continuous and $g(t) = 0$. This need not be a set of measure zero but $G(B)$ is a set of measure zero in $G([a,b])$ since G' vanishes on B.

Finally let C be the set of points t in $[a,b]$ at which g is continuous and $(f \circ G)g$ is continuous and $g(t) \neq 0$. We show that f is continuous at every

[2]This is harder than one might think. If $f(G(t))$ is itself Riemann integrable on $[a,b]$ then certainly so too is $f(G(t))g(t)$. Kestelman [4] includes an example to show that even if $f(G(t))g(t)$ is integrable on $[a,b]$ and f is integrable on $G([a,b])$ it may well happen that $f(G(t))$ is not Riemann integrable on $[a,b]$. This, he remarks, is the source of the difficulty for the problem.

point of $G(C)$. Since $[a,b] = A \cup B \cup C$ and since both $G(A)$ and $G(B)$ have measure zero we will have proved that f is a.e. continuous in $G([a,b])$.

Suppose $x = G(t)$ for some $t \in C$ and that $G'(t) = g(t) > 0$. Then G is strictly increasing on an interval containing the point t. If f is discontinuous at x then $(f \circ G)g$ would have to be discontinuous at t which is not the case. The same argument works if $G'(t) = g(t) < 0$.

This completes the proof except for mention of the Riemann-Stieltjes integral in the statement of the theorem. But we have already established in Lemma 5 that

$$\int_a^b f(G(t))dG(t) = \int_a^b f(G(t))g(t)\,dt$$

where the existence of one integral in the Riemann sense implies the existence of the other. □

References

[1] Andrew M. Bruckner and Brian S. Thomson, Real variable contributions of G. C. Young and W. H. Young, Expo. Math. 19 (2001), no. 4, 337–358.

[2] Roy O. Davies, An elementary proof of the theorem on change of variable in Riemann integration, The Mathematical Gazette, Vol. 45, No. 351 (1961), 23–25.

[3] Chris Freiling, On the problem of characterizing derivatives, Real Anal. Exchange 23 (1997/98), no. 2, 805–812.

[4] Hyman Kestelman, Change of variable in Riemann integration, The Mathematical Gazette, Vol. 45, No. 351 (1961), 17–23.

[5] Jiří Navrátil, A note on the theorem on change of variable in a Riemann integral, (Czech. English summary) Časopis Pěst. Mat., Vol. 106 (1981), No. 1, 79–83.

[6] David Preiss and Jaromír Uher, A remark on the substitution for the Riemann integral. (Czech. English summary) Časopis Pěst. Mat., Vol. 95 (1970), No. 4, 345–347.

[7] Herbert E. Robbins, Note on the Riemann integral, American Math. Monthly, Vol. 50, No. 10 (Dec., 1943), 617–618.

[8] Stanisław Saks, *Theory of the Integral*, 2nd revised Ed., Dover, New York, 1964.

[9] James Serrin and Dale E. Varberg, A general chain rule for derivatives and the change of variables formula for the Lebesgue integral, Amer. Math. Monthly, 76 (1969) 514–520.