THE RIEMANN–STIELTJES INTEGRAL

R. E. Bradley

Adelphi University

Abstract. Riemann–Stieltjes integration is an optional topic for a first course in real analysis. In this paper, we examine some of the pedagogical reasons in favor of its inclusion and some of the technical anachronisms associated with it.

1. Introduction. As the name suggests, Riemann–Stieltjes (RS) integration is a notion of integration properly generalizing Riemann integration — the type of integration covered in freshman calculus. If an undergraduate encounters this somewhat arcane topic at all, it will be in an upper–level course in real analysis. More likely, a student will encounter RS integration in a first graduate course in real analysis. In this setting, students arrive with widely varying backgrounds. Covering RS integration gives the instructor a way to review the usual development of Riemann integration for those with deficient backgrounds while at the same time offering something new for those whose undergraduate programs did include a rigorous treatment of the Riemann integral. Torchinsky has pointed out this pedagogical benefit in [5].

When studying Riemann integration, one may make use of either Riemann sums or upper and lower sums; the class of functions so defined is the same. This rather subtle point can be brought into sharp focus when one discovers that the same does not hold for RS integration, at least when the distribution function is discontinuous. Far from being pathological, the discontinuous distribution function is the vehicle by which RS integration unites the study of Riemann integration with that of probability theory or numerical integration.

The curriculum of a mathematics major usually includes a course in calculus–based probability theory. A student who has mastered this material is in a good position to appreciate the power of the RS integral. RS integration not only unites the apparently disconnected topics of discrete and continuous probability distributions, but facilitates the study of mixed–type distributions.

A student who has studied numerical methods will be familiar with Simpson's rule and other numerical integration schemes. RS integration underlines a connection between numerical quadrature and Riemann integration which is seldom mentioned. For those students of analysis who are unfamiliar with numerical methods, RS integration may even be used as a segue to the study of numerical integration.

2. Two Competing Definitions. The symbol F will always be used to represent a *distribution function* on a non–empty closed interval $I = [a, b]$ of the real line $\mathbb{R}$. This is a real–valued non–decreasing function, necessarily bounded on the interval I.

A *partition* of the interval I is a set $\mathcal{P} = \{x_0, x_1, \ldots, x_n\}$ with $a = x_0 < x_1 < \cdots < x_n = b$. Let $I_k = [x_k, x_{k+1}]$ for $k = 0, 1, 2, \ldots, n-1$. Define the *mesh* of $\mathcal{P}$, $||\mathcal{P}||$, to be the minimum value of $\Delta x_k = x_{k+1} - x_k$ for $k = 0, 1, \ldots, n-1$. Given a distribution function F, let $\Delta F_k = F(x_{k+1}) - F(x_k)$ for $k = 0, 1, \ldots, n-1$.

Suppose that g is a bounded real–valued function defined on I. Let

$$m_k = \inf\{g(x) | x \in I_k\} \quad \text{and} \quad M_k = \sup\{g(x) | x \in I_k\}.$$

We define the *lower* and *upper sums* of g corresponding to $\mathcal{P}$ with respect to F by

$$S_*(g, F, \mathcal{P}) = \sum_{k=0}^{n-1} m_k \Delta F_k \quad \text{and} \quad S^*(g, F, \mathcal{P}) = \sum_{k=0}^{n-1} M_k \Delta F_k.$$

We say the partition $\mathcal{P}_2$ *refines* $\mathcal{P}_1$ (or that $\mathcal{P}_2$ is *finer* than $\mathcal{P}_1$) if $\mathcal{P}_1 \subseteq \mathcal{P}_2$. By induction on the number of points in $\mathcal{P}_2 - \mathcal{P}_1$, it is easy to show that if $\mathcal{P}_2$ refines $\mathcal{P}_1$, then for any g and F

$$S_*(g, F, \mathcal{P}_1) \leq S_*(g, F, \mathcal{P}_2) \leq S^*(g, F, \mathcal{P}_2) \leq S^*(g, F, \mathcal{P}_1).$$

By considering a common refinement $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, we see that for any partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ we have

$$S_*(g, F, \mathcal{P}_1) \leq S^*(g, F, \mathcal{P}_2).$$

Thus, the quantities

$$L(g, F, I) = \sup\{S_*(g, F, \mathcal{P})\} \quad \text{and} \quad U(g, F, I) = \inf\{S^*(g, F, \mathcal{P})\}$$

are well–defined, the supremum and infimum being taken over all partitions $\mathcal{P}$ of I. Furthermore, $L(g, F, I) \leq U(g, F, I)$.

Definition 1. We say that g is *Darboux–Stieltjes* integrable on I with respect to F, denoted $g \in \mathcal{R}_1(F, I)$, if $L(g, F, I) = U(g, F, I)$.

Given a partition $\mathcal{P}$ of I, choose $c_k \in I_k$ for each $k = 0, 1, \ldots, n-1$. Let $\mathcal{C} = \{c_0, c_1, \ldots, c_{n-1}\}$. The *Riemann sum* of g corresponding to $\mathcal{P}$ and $\mathcal{C}$ with respect to F is

$$S(g, F, \mathcal{P}, \mathcal{C}) = \sum_{k=0}^{n-1} g(c_k)\Delta F_k.$$

Definition 2. Suppose there is a real number A with the property that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $||\mathcal{P}|| < \delta$ then $|A - R(g, F, \mathcal{P}, \mathcal{C})| < \varepsilon$ for any choice of $\mathcal{C}$. We say that g is *Riemann–Stieltjes integrable* on I with respect to F and denote this $g \in \mathcal{R}_2(F, I)$.

3. Two Classes of Functions.

Theorem 1. $\mathcal{R}_2(F, I) \subseteq \mathcal{R}_1(F, I)$.

Proof. Suppose that $g \in \mathcal{R}_2(F, I)$ and let $\varepsilon > 0$ be given. Let A be the number given in Definition 2 and choose $\delta > 0$ corresponding to $\varepsilon/2$. Clearly, $L(g, F, I) \leq A \leq U(g, F, I)$. Let $\mathcal{P}$ be a partition satisfying $||\mathcal{P}|| < \delta$. If $F(a) = F(b)$, then F is constant on I and so all sums evaluate to zero. Otherwise, choose a point c_k in each I_k so that

$$g(c_k) - m_k \leq \frac{\varepsilon}{2(F(b) - F(a))}.$$

It then follows that

$$S(g, F, \mathcal{P}, \mathcal{C}) - S_*(g, F, \mathcal{P}) < \frac{\varepsilon}{2}.$$

Then

$$\begin{aligned}|A - S_*(g, F, \mathcal{P})| &\leq |A - S(g, F, \mathcal{P}, \mathcal{C})| + |S(g, F, \mathcal{P}, \mathcal{C}) - S_*(g, F, \mathcal{P})| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.\end{aligned}$$

Similarly, we can find a partition $\mathcal{P}'$ so that $|A - S^*(g, F, \mathcal{P}')| < \varepsilon$. As ε is arbitrary, we have $L(g, F, I) = U(g, F, I) = A$.

If $F(x) = x$, then the two definitions coincide and we have the usual notion of Riemann integrability. We can say even more.

Theorem 2. If F is continuous on I then for any bounded function g on I, $g \in \mathcal{R}_2(F, I)$ if and only if $g \in \mathcal{R}_1(F, I)$.

Proof. The implication follows from Theorem 1. For the converse, suppose that $g \in \mathcal{R}_1(F, I)$ and that $\varepsilon > 0$ is given. Suppose that $\mathcal{P}$ is an $n+1$ point partition of I such that $S^*(g, F, \mathcal{P}) - S_*(g, F, \mathcal{P}) < \varepsilon/2$. Let M be an upper bound for $|g|$ on I. Because F is continuous on a closed interval, it is uniformly continuous. Therefore, there is a $\delta > 0$ so that

$$|F(x) - F(y)| < \frac{\varepsilon}{2Mn} \quad \text{whenever} \quad |x - y| < \delta.$$

Let Q be any partition of I such that $\|Q\| < \delta$. Let $J_0, J_1, \ldots, J_{m-1}$ be the subintervals of I determined by Q and let $d_l \in J_l$ for $l = 0, 1, \ldots, m-1$. Let $\mathcal{D} = \{d_0, d_1, \ldots, d_{m-1}\}$. There are, at most, $n-1$ intervals J_l which contain points from $\mathcal{P}$ in their interiors. Call these the bad intervals of Q and the remainder the good intervals. The total contribution to $S(g, F, Q, \mathcal{D})$ from the bad intervals is bounded in absolute value by $\varepsilon/2$. Let $\mathcal{U} = Q \cup \mathcal{P}$ and choose a sequence of points $\mathcal{E}$ from the intervals so determined. If we do this in such a way that the points in $\mathcal{E}$ coincide with those in $\mathcal{D}$ in all of the good intervals of Q, then

$$|S(g, F, Q, \mathcal{D}) - S(g, F, \mathcal{U}, \mathcal{E})| < \frac{\varepsilon}{2}.$$

Now every interval of $\mathcal{U}$ is a subset of some I_k, so for the corresponding element $e \in \mathcal{E}$, we have $m_k \leq g(e) \leq M_k$. From this it easily follows that

$$S_*(g, F, \mathcal{P}) \leq S(g, F, \mathcal{U}, \mathcal{E}) \leq S^*(g, F, \mathcal{P}).$$

Therefore, if we let A be the common value of $L(g, F, I)$ and $U(g, F, I)$, we have shown that

$$|A - S(g, F, Q, \mathcal{D})| < \varepsilon \quad \text{whenever} \quad \|Q\| < \delta.$$

The inclusion in Theorem 1 is proper when F is discontinuous; consider the following example.

Example 1. Suppose that

$$F = \begin{cases} 0 & \text{if } x \in [0, 1) \\ 1 & \text{if } x \in [1, 2] \end{cases} \quad \text{and} \quad g = \begin{cases} 0 & \text{if } x \in [0, 1] \\ 1 & \text{if } x \in (1, 2]. \end{cases}$$

It is easy to check that $S_*(g, F, \mathcal{P}) = 0$ for any partition $\mathcal{P}$ of $[0, 2]$ and that $S^*(g, F, \mathcal{P})$ is equal to 0 when $1 \in \mathcal{P}$ but is equal to 1 when $1 \notin \mathcal{P}$. Therefore $g \in \mathcal{R}_1(F, I)$. However, no matter how small $\|\mathcal{P}\|$ is, there will always be choices $\mathcal{C}$ containing the point 1 and others which do not. Hence, $g \notin \mathcal{R}_2(F, I)$. We note that if we let $g = F$ in this example, then g is in neither $\mathcal{R}_1(F, I)$ nor $\mathcal{R}_2(F, I)$.

The class of Riemann integrable functions on I, denoted $\mathcal{R}(I)$, is the class $\mathcal{R}_1(x, I)$ ($= \mathcal{R}_2(x, I)$, by Theorem 2). Most of the familiar properties of $\mathcal{R}(I)$ are true of both $\mathcal{R}_1(F, I)$ and $\mathcal{R}_2(F, I)$. In particular, all continuous functions on I are in both $\mathcal{R}_1(F, I)$ and $\mathcal{R}_2(F, I)$, and both classes are linear spaces. As well, if J is a closed interval with $J \subseteq I$, then $\mathcal{R}_1(F, I) \subseteq \mathcal{R}_1(F, J)$ and similarly for $\mathcal{R}_2$. The converse to this last statement is not true.

Theorem 3. Suppose $I = [a, b]$ and $c \in (a, b)$. If $g \in \mathcal{R}_1(F, [a, c])$ and $g \in \mathcal{R}_1(F, [c, b])$ then $g \in \mathcal{R}_1(F, I)$. The corresponding statement concerning $\mathcal{R}_2$ is false.

Proof. Let $L' = \sup\{S_*(g, F, Q)\}$ and $U' = \inf\{S^*(g, F, Q)\}$, with the supremum and infimum taken over all partitions Q of I containing c. Since g is contained in both $\mathcal{R}_1(F, [a, c])$ and $\mathcal{R}_1(F, [c, b])$, it follows that $L' = U'$. From this it follows that $L(g, F, I) = U(g, F, I)$, as the supremum and infimum here are taken over a larger class of partitions, and so $g \in \mathcal{R}_1(F, I)$.

For the second statement, consider Example 1, where $g \in \mathcal{R}_2(F, [0, 1])$ and $g \in \mathcal{R}_2(F, [0, 2])$.

4. Which is the Right Definition? Our terminology is based on [2]. It has the unfortunate side effect of suggesting that Definition 2 is the correct one for the RS integral. The use of 'Riemann' in the name of the integral is to distinguish it from the Lebesgue–Stieltjes integral. The use of 'Riemann' in the notion of integrability reflects the use of Riemann sums in the definition, as opposed to upper and lower sums. In fact, there appears to be no consensus in the literature as to which of the two classes $\mathcal{R}_1$ or $\mathcal{R}_2$ is the natural one.

Most contemporary authors present only one definition and use the term 'Riemann–Stieltjes integrable' whether they have chosen Definition 1 (e.g. [4], [5]), or Definition 2 (e.g. [3]). It is clear from the proof of Theorem 1 that when $g \in \mathcal{R}_2(F, I)$ then the value A in

Definition 2 is the common value of $L(g, F, I)$ and $U(g, F, I)$ from Definition 1. Thus the value of the integral, denoted

$$\int_a^b g dF \quad \text{or} \quad \int_a^b g(x) dF(x)$$

is unambiguously defined whichever definition of integrability is adopted.

History is on the side of Definition 2, as this was the one considered by Stieltjes (1856-1894) himself in 1894. He used the sort of sum which Riemann (1826-1866) considered in his rigorous treatment of the integral in 1854, based on values $g(c_k)$ of the integrand. It was Darboux (1842-1917) who first showed in 1875 that upper and lower sums could be used in place of Riemann sums in the ordinary integral.

On the other hand, Definition 1 has the advantage of being simpler to use than Definition 2. As well, those who find the property of $\mathcal{R}_2$ in Theorem 3 to be pathological may prefer the class $\mathcal{R}_1$.

Definition 1 is not free from pathology either. The case of equally–equally spaced partitions illuminates an odd property of Definition 1.

When introducing the Riemann integral, some texts consider equally spaced partitions only; that is, partitions where $\Delta x_k = (b-a)/n$ for each $k = 0, 1, \ldots, n-1$. This simplification gives rise to the same class of functions $\mathcal{R}(I)$, and it can be shown that the same is true for $\mathcal{R}_2(F, I)$ (and hence, also for $\mathcal{R}_1(F, I)$ when F is continuous). The same simplification does not carry over to $\mathcal{R}_1(F, I)$ in general. Consider Example 1, with the role of the number 1 played instead by an irrational number $x_0 \in [0, 2]$. Then we still have $f \in \mathcal{R}_1(F, [0, 2])$, whereas for any equally spaced partition $\mathcal{P}$, $S_*(g, F, \mathcal{P}) = 0$, while $S^*(g, F, \mathcal{P}) = 1$ since $x_0 \notin \mathcal{P}$.

5. Application – Probability Theory. A *cumulative distribution function* (CDF) is a non–negative, non–decreasing function F defined on the entire real line $\mathbb{R}$ with the properties

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to +\lim} F(x) = 1.$$

F is the CDF for a random variable X if $Pr(X \leq x) = F(x)$.

Undergraduate probability texts usually recognize two types of random variable: discrete and continuous. A discrete random variable takes at most countably many values $x_1, x_2, \ldots$. If we let $p_i = Pr(X = x_i)$ for each i, then

$$F(x) = \sum_{x_i \le x} p_i. \tag{1}$$

A random variable has continuous distribution if it takes all the values in some (bounded or unbounded) interval in the real line, and $Pr(X = x) = 0$ for every $x \in \mathbb{R}$. In practice, the textbooks consider only random variables with CDFs that are differentiable (except possibly at finitely many points). The *probability density function* is then defined by $f(x) = F'(x)$.

The *expectation* of a random variable $g(X)$ is defined to be

$$E(g(X)) = \sum_i g(x_i)p_i \quad \text{or} \quad E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx,$$

respectively, for discrete or continuous random variables. Using the RS integral, we may say that

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)dF(x) \tag{2}$$

in either case. One must first make the obvious extension to an improper RS integral (Stieltjes himself considered the case $b = \infty$ in his 1894 paper). Then for discrete random variables, equation 2 is clearly valid. For continuous random variables, one must use the result that if F is differentiable and $f = F'$, then

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

A random variable is said to have a mixed–type distribution if its range is uncountable and yet there are points with $Pr(X = x) > 0$. Such random variables arise naturally, but are beyond the scope of most undergraduate texts. The RS integral allows such random variables to be dealt with in the same fashion as discrete and continuous ones. In particular, the expectation $E(g(X))$ is defined by equation 2 in this case as well.

Example 2. (Discrete) Suppose X = the number of heads in three tosses of a fair coin. F is given by equation 1.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/8 & \text{if } x \in [0,1) \\ 1/2 & \text{if } x \in [1,2) \\ 7/8 & \text{if } x \in [2,3) \\ 1 & \text{if } x \geq 3. \end{cases}$$

Example 3. (Mixed–type) Suppose a traffic light works on a one minute cycle with 24 seconds of green, 6 seconds of yellow, and 30 seconds of red. Let X = the waiting time at the traffic light, in seconds. For cautious and reckless drivers, respectively, the CDFs are F_1 and F_2, where

$$F_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ x/60 + .4 & \text{if } x \in [0,36] \\ 1 & \text{if } x > 36 \end{cases}$$

and

$$F_2(x) = \begin{cases} 0 & \text{if } x < 0 \\ x/60 + .5 & \text{if } x \in [0,30] \\ 1 & \text{if } x > 30. \end{cases}$$

The expected waiting times are 18 and 15 seconds, respectively.

6. Application – Numerical Integration. Let $h = (b-a)/2$. Simpson's rule states that

$$\int_a^b g(x)\,dx \approx \frac{h}{3}[g(a) + 4g(a+h) + g(b)]$$

$$= \int_a^b g(x) dF(x),$$

where

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ h/3 & \text{if } x \in [a, a+h) \\ 5h/3 & \text{if } x \in [a+h, b) \\ 2h & \text{if } x \geq b. \end{cases}$$

Similar schemes can be given for the trapezoid rule, the midpoint rule, and any of the other familiar numerical integration formulas.

References

1. M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, NY, 1972.
2. E. J. McShane, T. A. Botts, *Real Analysis*, Van Nostrand, Princeton, NJ, 1959.
3. E. R. Philips, *An Introduction to Analysis and Integration Theory,* Intext Educational Publishers, Scranton, NJ, 1971.
4. W. Rudin, *Principles of Mathematical Analysis*, McGraw–Hill, New York, NY, 1953, 1964, 1976.
5. A. Torchinsky, *Real Variables* Addison–Wesley, Reading, MA, 1988.