HENSTOCK ON RANDOM VARIATION

Pat Muldowney

Received September 29, 2007

Abstract. The theory of integration (including measure) is the basis for the study of probability and random variation. Thus Henstock's Riemann-type integration theory has relevance to our understanding of random variation. Henstock addressed this issue in many of his published works, in which he gave interpretations of probability, of the statistical analysis of data, and of random processes. His analysis of Feynman's non-absolute integrals in quantum mechanics brings this subject properly into the domain of random variation.

1 Introduction In a personal report [3] in 1984, Henstock described his research interests as follows:

My research is in the pure mathematical fields of

1. *summability of series and integrals,*

2. *integration theory,*

and especially in problems that link the two. In 1. an essential tool is often the Banach-Steinhaus theorem of functional analysis, with Sargent's modification. For example, if $\int_a^b f dg$ *exists for every Baire- or Borel-measurable function, to prove that* g *is of bounded variation on* $[a, b]$*. Putting a summability factor into the definition of the integral leads to a generalization of Burkill's Cesaro-Perron integrals and the Marcinkiewicz-Zygmund integral. These are of Perron type, defined by inequalities of the type*

$$\int_x^{x+h} \{F(t) - F(x)\}\, d_t N(x, h; t) \geq \int_x^{x+h} f(x)(t - x) d_t N(x, h; t),$$

with similar inequalities for $[x-h, x]$*, and the problem is to find the necessary conditions on* N*. Out of this came the variational integral which then led to the Riemann-complete or generalized Riemann integral, the so-called Kurzweil-Henstock integral. This integral includes the Riemann, Riemann-Stieltjes, Lebesgue, Radon, Denjoy special, and Perron integrals, using Riemann's original sums but a different limit. Set-valued functions, the integrals of which have applications in economics. Wiener-type integration has applications to various stochastic processes such as white noise. Feynman-type integration has applications in quantum theory. An integral that includes the Paley-Wiener-Zygmund integral has applications for stochastic integration, as does an integral equivalent to the Itô integral. The most general form of the generalized Riemann integral can be used to define all these integrals except those defined by the functions* N*.*

The theory of integration with which Henstock is associated arose from his study of the problem of summability described above, but did not completely resolve this problem to his satisfaction. So in a sense, his successful mathematical accomplishments are a by-product

2000 *Mathematics Subject Classification.* 01A60, 01A70, 28-03, 60A99.
Key words and phrases. Ralph Henstock, Riemann integration, Probability, Feynman integration.

of a different project which he felt was incomplete. Such things happen in a life of high achievement.

Another ancillary problem which he addressed at various times in the course of his life was the mathematical analysis of random variation. He first encountered this subject in a very practical way when, in November 1943, he was withdrawn temporarily from his mathematical studies at St. John's College, Cambridge, and assigned to the British Ministry of Supply to do statistical work.

His experience as a civil servant generated in him a visceral dislike of working for the government. But he retained a life-long interest in the analysis of random data. So he took a course of study in stochastic theory from M.S. Bartlett in 1947, and in 1958 he was appointed Fellow of the Royal Statistical Society.

Henstock's approach to the theory of integration builds on the nineteenth century theory of Riemann, and is conceptually different from and independent of the early twentieth century integration theory of Lebesgue. So one can imagine that the mathematical theory of probability founded on the work of Kolmogorov (as described in, for instance, [15], *Foundations of the Theory of Probability*, 1933) could conceivably have been based on a Riemann-type integration rather than the Lebesgue approach used by Kolmogorov and his successors in probability theory.

Henstock's writings give a strong sense of how such an alternative development of probability theory should be accomplished. His 1963 book (*Theory of Integration* [5]) includes a chapter on probability. In it, Henstock reviews three different interpretations, including Kolmogorov's, of the probability concept. Placing emphasis on the role of actual statistical data he discusses the classification or partitioning of numerical data into disjoint real intervals $\{I\}$, which is often the first practical step in the numerical analysis of such data. He provides two ways to define the probability that a numerical measurement x takes a value in a set X. Accordingly, $\text{Prob}(X)$ can be taken to be

$$\text{Prob}(X) = \int_{-\infty}^{\infty} \mathbf{1}_X(x)dP = \int_X dP, \quad \text{or} \tag{1}$$

$$\text{Prob}(X) = \text{V}\,(P\,;\,X)\,. \tag{2}$$

If we take (2) (the P-variation of the set X, see [19] page 26) as the definition of probability, then every set X has a mathematical probability. If we take (1) (the Henstock integral of P in the set X) as the definition, then, as in the Kolmogorov theory, only certain sets X have a probability measure (and, for those sets, (2) gives the same value). In that case, if the function P is a probability measure in the sense of Kolmogorov, then Henstock's approach gives exactly the same measurable sets as Kolmogorov's.

Is it the case, then, that by means of (1), Henstock provides us with a view of probability which is different from that of of Kolmogorov, but not different to any significant extent? No; the purpose of this article is to examine those aspects of Henstock's work which provide substantial new insights into the theory of probability and random variation.

Both (1) and (2) involve Riemann sums which require only that a probability function $P(I)$ exists for the real intervals I into which statistical data is classified or partitioned. Henstock's approach has no *a priori* requirement that probabilities $P(X)$ be defined on all measurable sets X. These are obtained as a consequence of (1). And if we take the approach of (2), probabilities are defined for all sets X.

In his presentation of the subject in [5] it appears that he would have liked to be able to deduce the probability function $P(I)$ from the relative frequency with which the measured data take values in the interval I. But he proves that this is problematic, and proceeds with

probability functions $P(I)$ which are given *a priori* for intervals I, just as Kolmogorov's probability function $P(X)$ is *a priori*, or axiomatic, for measurable sets X.

Again showing a strong bias towards statistical analysis of data, Henstock devotes the rest of this chapter on probability (in [5]) to a study of significance testing.

His 1988 book [6] (*Lectures on the Theory of Integration*) was intended by Henstock to replace and bring up to date his out-of-print 1963 book [5]. The 1988 book also included a chapter on probability, which repeats the earlier study of the probability concept, and has, in addition, a discussion of the correlation of joint random variables.

2 Countable Additivity Between 1963 and 1988 Henstock produced substantial further investigations of subjects which come within the scope of probability. In addition, he spelt out clearly some implications of his Riemann-style approach to integration for alternative formulations of the theory of probability. At the September 1973 Symposium of the Greek Mathematical Society [4], Henstock expanded on his 1963 interpretation of probability:

It has been assumed by the majority of mathematical analysts that the so-called Lebesgue limit theorems follow because Lebesgue measure is countably additive. ... [But] it is no longer necessary, when defining measure ... to give a proof of the countable additivity of the measure nor to have elaborate constructions in the production of such a measure. To see why this is not merely of academic interest to mathematical analysts, but has wider implications, we look at the probability theory of statistics. Note that probabilists claim that probability theory is older than measure theory. [Consider] Kolmogorov's axioms for statistics. The statistician takes a space T of all possible statistical events in a particular system, divides T into a finite number, say, $T_1, \ldots, T_n$, of mutually exclusive subsets, and then takes a sample *(i.e. a finite set) in which each member is a realization of one of the statistical events. If the sample contains q members, of which q_j are realizations of events in T_j, then $\frac{q_j}{q}$ is the* relative frequency *in the sample, of the subset T_j for $j = 1, 2, \ldots, n$. These relative frequencies lie between 0 and 1, ..., and they are* finitely additive. *This last means that the relative frequency in any union of some of the T_j is the sum of the relative frequencies of the separate subsets T_j of the union. Experimentally it is found that if repeated independent random samples are taken, the relative frequencies tend to become stabilized. But in order to say mathematically what many terms in the previous sentence mean, we would be involved in a thorough discussion of the mathematical basis of statistics. Here it is enough to say that the Kolmogorov functions p of sets, the* probabilities, *that obey the same rules, are clearly the natural mathematical models to be used in statistics. Paragraphs like this form part of every course of statistics that mentions probability.*

To ensure that these set functions p are Lebesgue or Radon measures, Kolmogorov also assumed that they are countably additive. *Here, if $U_1, U_2, \ldots$ is a sequence of mutually exclusive subsets of T, with union $U \subseteq T$, then*

$$p(U) = \sum_{j=1}^{\infty} p(U_j). \tag{3}$$

In practical cases p is originally defined only over a family of subsets I of T, from which, by taking finite unions, intersections, and set differences, we obtain a ring *of sets and can then construct the Borel sets. For example, if T is an N-dimensional Euclidean space, p is often originally defined over geometrically simple objects such as N-dimensional rectangles. In order to construct the Lebesgue-Radon integral of even a continuous function by using measure theory, it is vital to define p at least over every open set G of T, using* (3) *with $U = G$. Further if $\mathbf{1}_G(x)$ is the* characteristic or indicator function *of G (the function which*

is 1 when x is in G and 0 otherwise) in Lebesgue-Radon theory we have

$$p(G) = \int_T \mathbf{1}_G(\cdot)dp. \tag{4}$$

Conversely from (4) *and the theorems on the reversal of order of integration and limit, we can prove* (3)*, at least when* $U = G$ *and the* U_j *are members of the pre-ring. Because of this it has been assumed that that Kolmogorov's extra axiom* (3) *is necessary, as well as being sufficient, for* p *to be a probability function.*

But the object of this paper is to show that such a view is erroneous. The generalized Riemann integral ... [is] more powerful than those of Lebesgue and Radon; [it does] not need the countable additivity (3)*, and yet the limit theorems still hold.*

There is a strand among experts in random variation for whom Kolmogorov's Axiom ((3), above) is problematic. For instance, each of us acquires naturally some intuitive sense of the meaning of probability as a real-world phenomenon. But what intuitive meaning can we give to the mathematical abstraction (3) which would help us to accept it as an axiom?

However the finite additivity of the relative frequencies of actual data suggests strongly that a probability function should be finitely additive. But this presents no problem for either of the definitions (1) or (2), whose construction depends on only finite Riemann sums involving $P(I)$.

Historically, the probability function P is the concept which enables us to make mathematical sense of the subjects of *measurement, approximation* or *estimation*, all of which can be grouped under the heading of *random variation.* Fundamental to this is the concept of a *random variable* which, in the Kolmogorov theory, is a P-measurable function $\mathcal{X}$ mapping a sample space Ω to a space of values $X = \{\mathcal{X}(\omega) : \omega \in \Omega\}$, corresponding to some measurement or estimate of a real or physical quantity. So writing $x = \mathcal{X}(\omega)$, the *expected value* of $\mathcal{X}$ is

$$\int_\Omega \mathcal{X}(\omega)dP \quad \text{or} \quad \int_{\mathbf{R}} xdF_X, \tag{5}$$

where F_X is the probability distribution function defined by

$$F_X(I) = P(\mathcal{X}^{-1}(I)) \tag{6}$$

for real intervals I, and hence for Borel sets in $\mathbf{R}$.

Sometimes we have to deal with some mathematical function of the actual measured or estimated values $X = \{\mathcal{X}(\omega)\}$. For instance, if we want to estimate the variance or variability, we need the values $X^2 = \{(\mathcal{X}(\omega))^2\}$. So for suitable functions f we can treat the set $f(X) = \{f(x)\}$, or more precisely, the function $f(\mathcal{X})$, as a random variable with expected value

$$\int_\Omega f(\mathcal{X}(\omega))dP \quad \text{or} \quad \int_{\mathbf{R}} f(x)dF_X. \tag{7}$$

If either of the integrals of (7) (or indeed (5)) exists as a Lebesgue integral then the other one also exists as a Lebesgue integral. And if the latter integral, $\int_{\mathbf{R}} f(x)dF_X$, exists as a Lebesgue integral it also exists as a generalized Riemann integral:

$$\int_{\mathbf{R}} f(x)P(\mathcal{X}^{-1}(I)), \text{ or } \int_{\mathbf{R}} f(x)F_X(I). \tag{8}$$

The latter notation denotes generalized Riemann or Henstock integration, and, for a non-negative finitely additive distribution function $F_X(I)$ of standard probability theory, it exists

whenever the corresponding Lebesgue integral $\int_{\mathbf{R}} f(x)dF_X$ does, and the two are equal. Thus it is a relatively simple matter to re-interpret the Kolmogorov theory of probability in generalized Riemann terms by means of (7) or (8). All we have to do is to translate each statement of the standard Kolmogorov theory from Kolmogorov terms into Riemann terms by means of (6) and (8).

But for certain certain kinds of distribution function F which are outside the usual scope of probability theory, the $\int_{\mathbf{R}} f(x)F(I)$ of (7) may exist as a generalized Riemann integral, whilst the corresponding $\int_{\mathbf{R}} f(x)dF$ fails to exist as a Lebesgue integral. So for such F the probability measure of a set or outcome A may exist in the generalized Riemann sense but not in the Lebesgue sense. We will encounter such a function in Section 7 below. Therefore it is possible that probability theory can be extended outside of its usual scope if we approach the subject in generalized Riemann terms. As we shall see later in this article, this insight underlies the main body of Henstock's studies of random variation from the 1960's to the 1990's.

3 Sample Space From the Kolmogorov point of view it is natural to take

$$\int_{\Omega} \mathcal{X}(\omega)dP \tag{9}$$

as the conceptual starting point. While the sample space Ω is assumed to have a given, fixed probability function P, we often have to deal with several, perhaps infinitely many, random variables $\mathcal{X}, \mathcal{Y}, \dots$, each with its own distribution function

$$F_X(I) = P(\mathcal{X}^{-1}(I)), \quad F_Y(I) = P(\mathcal{Y}^{-1}(I)), \quad \dots \quad .$$

The variables $\mathcal{X}(\omega), \mathcal{Y}(\omega), \dots$ are a mathematical representation of real world occurrences or measurements which can generally be represented as real numbers. As discussed in the Introduction above, it is the usual practice in statistical analysis of data to classify (or group, or partition) such values $\mathcal{X}(\omega)$ into the intervals $\{I\}$ of real numbers in which they can potentially occur. (If we are analyzing joint occurrences of $\mathcal{X}, \mathcal{Y}, \dots$, we look at the real intervals $I \times J \times \cdots$ in which random values $(\mathcal{X}(\omega), \mathcal{Y}(\omega), \dots)$ can occur jointly. See Section 7 below.)

So a mathematical representation or abstraction of this practice should be easy to understand, since only real intervals are involved. But it is not so easy to understand a process of grouping these potential values, not just into intervals I of (8), but into the measurable sets of (5).

For instance, it is easy enough to envisage the likelihood that tomorrow's mean temperature will be between, say, ten and twenty degrees celsius. But what if we try to envisage the temperature taking a value in some Cantor subset contained in the interval 10 to 20 degrees?

In the Kolmogorov programme this is mathematically necessary. Not only that, but if we are dealing with many random variables $\mathcal{X}, \mathcal{Y}, \dots$, we must repeat this intuitive effort in each of the ranges of the $\mathcal{X}, \mathcal{Y}, \dots$. And it gets even worse when we are trying to understand joint variation of these random variables.

However, if we encounter this analytical hurdle only in the form of abstract measurable subsets of an abstract sample space Ω—a merely technical device to construct (9)—we put the least possible strain on our intuitive capacity. By doing this, we need only ever consider the abstract measurable subsets of the mathematical abstraction Ω for which there need be no physical interpretation or meaning. Thus we can avoid the corresponding, intuition-challenging, measurable subsets of the real-world ranges in $\mathbf{R}$ of the actual measurements $\mathcal{X}, \mathcal{Y}, \dots$.

In contrast, the only demand made on our intuition in a Riemann approach based on (8) is to envisage, not measurable sets, but only the real intervals $I, J, \dots$ into which the values of each of the random variable or measurement $\mathcal{X}, \mathcal{Y}, \dots$ are classified in the usual manner of elementary statistical calculation. So our imagination is not required to stretch beyond intervals such as 10 - 20 degrees and the like, as described above.

Thus, in the Kolmogorov approach, the concept Ω serves as a mathematical safety valve removing unnecessary burdens from our intuition of the real-world events that we are analyzing. But beyond that, what is the role of Ω? It is called the *sample space*. The actual events in the real world that we are concerned with are generators of actual, real-world, numerical values. Think of an event as being a measurement, in which the result or value of the measurement is a real number. We are led to the belief that such events may be *random*, because we sense from experience that if the measurement were repeated, a different result or value might be obtained, so we cannot predict exactly which of several possible results may be obtained by the measurement. Therefore we are dealing with a *variable*.

Again from experience, we sense that there may be a certain likelihood that a measurement will give one particular value rather than some other value. Likewise, we may sense a likelihood that the measurement will give a value belonging to some interval range of real numbers.

Thus it is reasonable to regard a random variable as simply a list or set of the possible values X of a measurement, in conjunction with the likelihood of occurrence of the possible values. We have already discussed the standard Kolmogorov interpretation of a random variable as a function $\mathcal{X}$. So we have two interpretations, and we will try to keep both of them in mind.

How does Ω help us to understand random variation? One of the ways in which the concept of the sample space Ω is explained to us is by describing it as representing the *possible states of the world*, all possible eventualities in the world—or at least in that part of the world with which we are directly concerned. Whatever happens in the sample space can be thought of as the "cause" of a random variable (or actual measurement) taking one particular value rather than another. The sample space is, in some mathematical sense, the *generator of randomness* in the measurement or random variable, the reason why a random variable can take many different values rather than just one single, deterministic value.

We can think of the sample space as encompassing a mathematical roulette wheel in which a mathematical mechanism sets a mathematical ball in motion and determines one particular outcome rather than another. Furthermore if the possible events in Ω (the subsets of Ω, the sections of the imaginary celestial roulette wheel colored red, black, green and so on—corresponding to the potential "states of the world") each has an associated probability, then these probabilities determine the likelihoods that our random variable or measurement takes particular values corresponding to these colors or "states of the world".

If we are analyzing several, perhaps infinitely many, random variables $\mathcal{X}, \mathcal{Y}, \dots$, we may be concerned with their *joint variation*. For instance, at a particular time of day we may measure jointly—as a single, joint occurrence represented by a pair of values—the temperature and the atmospheric pressure, with a view to investigating whether these two random variables are related in some way; or whether, perhaps, one of them exhibits some dependence on the other. Or what effect the joint values have on some quantity $f(\mathcal{X}, \mathcal{Y}, \dots)$ which depends on each of them, or on all of them jointly.

In general terms, a particular "eventuality" ω of the set of all possible eventualities Ω results in, or "determines", the particular set of joint outcomes $\{\mathcal{X}(\omega), \mathcal{Y}(\omega), \dots\}$. The latter can be thought of as a 'tuple of measured values $(x, y, \dots)$, which can in turn be thought of as an element in $\mathbf{R} \times \mathbf{R} \cdots$. Thus, a single joint occurrence of the random variables $\mathcal{X}, \mathcal{Y}, \dots$ corresponds to a single element $(x, y, \dots)$ of $\mathbf{R} \times \mathbf{R} \times \cdots$, and the set of

all such possible joint occurrences corresponds to $X \times Y \times \cdots \subseteq \mathbf{R} \times \mathbf{R} \times \cdots$ along with some appropriate likelihood function F for the joint possibilities. We will return to this topic in Section 7 below.

4 Random Variable So we have an intuitive notion of a random variable $\mathcal{X}$ as a set of potential values or results X, along with some measure of the degree of potentiality or likelihood, call it F_X, of a particular value (real or complex) actually occurring. What the Kolmogorov theory does is to put this intuitive notion onto a solid mathematical footing.

This is accomplished by defining a random variable to be, not just a set of potential-values-plus-likelihoods, but a function $\mathcal{X}$ from the abstract sample space Ω into the set of potential values. (The values $\mathcal{X}(\omega)$ are generally taken to belong to $\mathbf{R}$. A random variable may also be a composite function $f(\mathcal{X})$ on Ω, and in this case the random values are usually allowed to be real or complex numbers.)

In this standard view, the likelihoods F_X are determined, not just by the function $\mathcal{X}$, but also by the pre-existing and *a priori* probability function P, independent of $\mathcal{X}$ but inherited by $\mathcal{X}$ through the equation

$$F_X(I) = P(\mathcal{X}^{-1}(I)).$$

So a different function $\mathcal{X}'$ could have the same range of values in $\mathbf{R}$ as $\mathcal{X}$ does; while mathematically, as a function different from $\mathcal{X}$ it would be a different random variable. This is somewhat counter-intuitive. Furthermore, this $\mathcal{X}'$ might even generate the same likelihoods in $\mathbf{R}$ as $\mathcal{X}$ does, with $F_X(I) = F_{X'}(I)$ for each I in the range of values, yet still be a "different" random variable from $\mathcal{X}$.

To illustrate, suppose our real-world measurement is to observe the outcome n ($1 \leq n \leq 6$) of a single throw of a die. Suppose the abstract, mathematical sample space Ω corresponds to a "great roulette wheel in the sky", which, for illustrative purposes, has six colors—red, green, black, white, pink and yellow; determining the real-world outcomes of 1 to 6, respectively, whenever the die is thrown. In Kolmogorov terms, the random variable is the mapping $\mathcal{X}$ which makes dice-throw 1 correspond to red, 2 to green, and so on. Consider the probability function P on Ω. First, suppose P is uniform, with $P(\text{red}) = \frac{1}{6}$, and so on. This corresponds to a fair or balanced die. Now suppose that Ω has a different set of probabilities P' defined on it: P' with $P'(\text{red}) = \frac{1}{2}$, and

$$P'(\text{green}) = \cdots = P'(\text{yellow}) = \frac{1}{10}.$$

We have here two different experiments, measurements, or random variables in the intuitive sense; the first being with a balanced die and the second with an unbalanced one. But formally speaking, and in traditional Kolmogorov terms, we have the same sample space Ω in both cases, the same range of values or outcomes $n = 1, \ldots, 6$ generated by the random variation, and hence the same random variable (in the sense of mapping from Ω into $\mathbf{R}$). Two intuitively different random variables are mathematically (in the Kolmogorov sense) the same.

Now suppose Ω and P are as we have just described. But suppose we define a different mapping $\mathcal{X}'$ which sends yellow to 1, pink to 2, and so on. Technically, this is a different mapping from $\mathcal{X}$, but it describes exactly the same experiment—a single throw of a fair or balanced die. So the formally different (in the Kolmogorov sense) $\mathcal{X}$ and $\mathcal{X}'$ are intuitively the same random variable.

Though these difficulties might be avoided if a little care was taken with definitions, many presentations of the standard Kolmogorov theory are open to such anomalies. But

they do not arise at all in the more intuitive "set-of-values-with-likelihoods" conception of random variable. Nonetheless we can, without too much risk, regard these difficulties as a quibble, and not likely to cause us serious trouble; at least not enough to seriously undermine the Kolmogorov approach. So it is reasonably safe to take either approach; or, indeed, to keep both approaches in mind whenever we study random variables.

Suppose we have a random variable $\mathcal{X}$, considered either as a mapping-from-a-sample-space-endowed-with-probability-P, as prescribed by Kolmogorov; or as (X, F_X), a set-of-values-X-with-likelihoods-F_X in the intuitive manner. Now suppose we are interested in the variance of $\mathcal{X}$, so we form the new random variable $\mathcal{Y} = \mathcal{X}^2$. What likelihoods attach themselves to the various possible values of $\mathcal{Y}$? Conveniently, in this case if $\mathcal{X}$ takes a value in an interval I, then $\mathcal{Y}$ takes its value in the interval J, where $J = \{y : y = x^2, x \in I\}$; and then the likelihood of $\mathcal{Y}(\omega)$ being in J is the same as the likelihood of $\mathcal{X}$ being in I, $F_Y(J) = F_X(I)$. (But two distinct intervals in the range of $\mathcal{X}$ can map to a single interval in the range of $\mathcal{Y}$.)

So here we are forming a new random variable $\mathcal{Y} = f(\mathcal{X})$, with its range—its set of possible values—generated, not by measuring something in the real world, but by the mathematical operation of squaring the measured value. The next step is to consider the probability distribution function F_Y of the random variable $\mathcal{Y}$.

5 Distribution Function and Expectation If we follow the Kolmogorov approach for $\mathcal{Y} = f(\mathcal{X})$, we start with a probability measure P defined on subsets of Ω and obtain likelihood distributions F_X, F_Y by

$$F_X(A) := P(\mathcal{X}^{-1}(A)), \quad F_Y(B) := P(\mathcal{X}^{-1}(f^{-1}(B)))$$

for relevant sets A, B of the values taken by $\mathcal{X}$, $\mathcal{Y}$ respectively.

The sets A and B can be real intervals I, J respectively, which are what we require in the Riemann approach. The sets X and Y are all the possible values of the two measurements $\mathcal{X}$ and $\mathcal{Y}$, and $F_X(I)$, $F_Y(J)$ are the *a priori* probabilities that the first measurement takes a value in I and the second measurement takes a value in J.

To illustrate, here are examples of familiar random variables. Take I to be the real interval $[u, v]$ and, without loss of generality, $v - u < 1$:

- Dice throwing, unbiassed:

$$\begin{array}{rcl} X & = & \{1,2,3,4,5,6\}; \\ F_X(I) & = & \begin{cases} 0 & \text{if} \quad I \cap X = \emptyset; \\ \frac{1}{6} & \text{if} \quad x \in X \text{ and } x - 1 < u \leq x \leq v < x + 1. \end{cases} \end{array}$$

- Binomial distribution, probability of x successes in n independent trials where the probability of success in a single trial is p :

$$\begin{array}{rcl} X & = & \{1,2,\ldots,n\}; \\ F_X(I) & = & \begin{cases} 0 & \text{if} \quad I \cap X = \emptyset; \\ \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & \text{if} \quad x \in X \text{ and } I \cap X \neq \emptyset. \end{cases} \end{array}$$

- Standard normal distribution (mean 0, variance 1):

$$\begin{array}{rcl} X & = & \mathbf{R}; \\ F_X(I) & = & \frac{1}{\sqrt{2\pi}} \int_I \exp\left(-\frac{1}{2} y^2\right) dy. \end{array}$$

Taking the alternative or intuitive approach to the notion of random variable, our fundamental random variable $\mathcal{X}$ is a set (not a function!) of potential measurement values, combined with an associated likelihood function F_X defined on real intervals I. But that is sufficient for us to get the analysis of the random variation off the ground, provided we adopt the generalized Riemann approach of Henstock. For instance, the expected value of the random variable (X, F_X) is $\int_{\mathbf{R}} xF_X(I)$ where the integral, if it exists, is generalized Riemann. In this approach, it is the existence of $\int_{\mathbf{R}} xF_X(I)$ that gives $\mathcal{X}$ the mathematical status of random variable, whereas in the Lebesgue approach $\mathcal{X}$ is a random variable if it is a measurable function.

In the examples of familiar random variables given above, it is not necessary to specify the X of $\mathcal{X} = (X, F_X)$, since X is implied in the specification of $F_X(I)$. But intuitively, the set X of possible outcomes or measurement values is what we first envisage, followed by consideration of the related probabilities $F_X(I)$. So it is helpful to retain X explicitly in our notation $\mathcal{X} = (X, F_X)$. Also, if we wished to calculate the variance of, for instance, the dice-throwing random variable, in place of the values $X = \{1, \ldots, 6\}$ we would have to calculate the mean value of values $f(X) = \{(x - 3.5)^2 : x \in X\}$, so we are dealing with a random variable $(f(X), F_X)$.

This is an example of a dependent random variable $f(\mathcal{X})$. What mathematical meaning can we give to this, following the Riemann approach? First we have the set $Y = f(X)$ of function values $f(x)$ where each $x \in X$ is a value of the random variable $\mathcal{X}$. Then, denoting $f(\mathcal{X}) = f((X, F_X))$ by $\mathcal{Y} = (Y, F_Y)$, the Henstock approach of (1) and (7) gives the associated likelihood function $F_Y(J)$, for intervals J, by

$$F_Y(J) = \int_{\mathbf{R}} \mathbf{1}_{f^{-1}(J)}(x) F_X(I) = \int_{f^{-1}(J)} F_X(I), \tag{10}$$

whenever this generalized Riemann integral exists. Thus the expected value of the random variable $(\mathcal{Y}, F_Y) = (f(X), F_X)$ is

$$\int_{\mathbf{R}} yF_Y(J) = \int_{\mathbf{R}} f(x)F_X(I)$$

where each of these integrals exists if and only if each of the others exists.

To illustrate, consider a standard normal random variable $\mathcal{X}$ as defined above. Then the *variance* of $\mathcal{X}$ is the mean of $\mathcal{X}^2$:

$$\int_{\mathbf{R}} x^2 F_X(I), \text{ easily shown to be equal to } \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} y^2 \exp\left(-\frac{1}{2}y^2\right) dy.$$

Should we consider a random variable corresponding to the measurements $f(X)$ to be, on the one hand, a function $f(\mathcal{X})$ defined on some sample space; or, on the other hand, a set-of-values-with-likelihoods $(f(X), F_X)$? We will use the latter approach, as it fits better with our Riemann approach to random variation. But either approach is perfectly feasible, and it is often helpful to keep both in mind.

In the Kolmogorov approach, $\mathcal{X}$ or $f(\mathcal{X})$ are considered to be random variables only if they are measurable functions on Ω. Likewise, in the Henstock approach, (X, F_X) or $(f(X), F_X)$ are considered to be random variables only if the expected values $\int_{\mathbf{R}} xF_X(I)$, $\int_{\mathbf{R}} f(x)F_X(I)$ exist as generalized Riemann integrals.

6 Henstock compared with Kolmogorov What meaning should we assign to Ω if we are adopting a Riemann approach to random variation, but, instead of the intuitive set-of-values-with-likelihoods concept of random variable, we prefer to adopt the Kolmogorov measurable-function-defined-on-sample-space approach?

If the random variable is $\mathcal{X} = (X, F_X)$ or $f(\mathcal{X}) = (f(X), F_X)$, then—as in the familiar examples given in Section 5 above—the sample space Ω can be taken to be $\mathbf{R}$, with each of the elementary events ω identified with one of the potential measurements x of $\mathbf{R}$. Considered as a mapping from $\Omega = \mathbf{R}$, the random variable $\mathcal{X}$ is the identity mapping to $\mathbf{R}$.

If we have several (perhaps infinitely many) random variables $\mathcal{X}, \mathcal{Y}, \cdots$ to be considered jointly, then a random variable has the form of a real- or complex-valued function $f(\mathcal{X}, \mathcal{Y}, \cdots)$, and the appropriate sample space is the Cartesian product $\mathbf{R}^B$, where the cardinality of B corresponds to the number of random variables $\mathcal{X}, \mathcal{Y}, \cdots$. Thus, if we are dealing with a finite number n of random variables (including $n = 1$), the sample space can be taken to be $\mathbf{R}^n$. (We return to this issue in Section 7. Note that we are using $\mathbf{R}^B$ to denote $\prod_B \mathbf{R}$—the set of all functions from B to $\mathbf{R}$; even though $B^{\mathbf{R}}$ is the usual notation for this, not $\mathbf{R}^B$.)

The likelihoods are then determined by a given, or *a priori*, probability distribution function F (just as probabilities are determined by an *a priori* function P in the Kolmogorov theory). In Section 7 we investigate the distribution function of joint likelihoods for a random variable $f(\mathcal{X}, \mathcal{Y}, \dots)$. Since B can be taken to be set containing a single element (giving $\mathbf{R}^B = \mathbf{R}$), the domain $\mathbf{R}^B$ can be taken as a universal sample space for every random variable.

In practice what we have just described is the approach that is used in the Kolmogorov theory whenever Ω and P actually have to be specified in reality (as opposed to abstract exposition of the theory), with $P(\mathcal{X}^{-1}(A))$ taken to be $F_X(A)$ for various sets A. This is clear from the examples of Section 5 above, from which we can make a connection between abstract exposition and practical calculation.

In place of sample space $\mathbf{R}$ (or $\mathbf{R}^B$) we might choose the sample space in an *ad hoc* way. For instance, in throwing dice we might take Ω to be the finite set $\{1, \dots, 6\}$, with $\mathcal{X}$ the identity mapping, and F_X (or P) specified in the obvious way. But, as illustrated in the the first two of the examples of Section 5 above, it is no great burden to embed such a sample space in $\mathbf{R}$ or $\mathbf{R}^B$, with F_X (or P) adapted accordingly. Also, in the study of Brownian motion, it is found useful to employ as sample space, instead of $\mathbf{R}^B$, the set of continuous functions defined on a real interval $]\tau', \tau] = B$. A construction of this sample space by Riemann methods is described in [23]. This construction corresponds to the well-known *continuous modification* of $\mathbf{R}^B$ for Brownian motion.

In fact it is quite feasible to adopt as universal sample space, for all random variables and processes, the real interval $]0, 1[$. This is because $\mathbf{R}$ (and even $\mathbf{R}^B$) can be mapped in a suitable way to $]0, 1[$. But we then lose the natural identification of elements of the sample space $\mathbf{R}$ (or $\mathbf{R}^B$) with the actual values assumed by a random variable $\mathcal{X}$ (or by $\mathcal{X}, \mathcal{Y}, \dots$) jointly.

Our discussion has the ambition of teasing out the implications of Henstock's comments on probability and statistical analysis. Though he did ground-breaking work on a number of important problems, it would be fair to say that Henstock did not himself undertake a comprehensive study of random variation as such. But, as is clear from the quotations given earlier, he did point out some of the major implications of generalized Riemann integration for the theory of probability. Merely by noticing that the right-hand integral of (7) above can be understood as a Henstock integral, we can argue that the existing formulation of the whole Kolmogorov theory of probability—concepts, theorems, proofs, the whole works—can, by a simple act of translation, be re-interpreted in generalized Riemann rather than Lebesgue terms. But this would forego the many benefits and insights which we could reasonably expect to gain by formulating the theory in the new Riemann rather than the existing Lebesgue terms. Just as the pure theory of integration is found to have a quite

different flavor in each of the two versions, Riemann and Lebesgue.

The project of reformulating the theory of probability in Riemann terms is still a work in progress, and is the subject of a forthcoming book by the present writer.

From the discussion above, one particular advantage of the Riemann approach can be immediately detected. In the classical theory we identify, within the abstract sample space Ω, a class of P-measurable subsets. Subsets outside of this class are assumed not to have a probability value. But this is counter-intuitive. If there is such a thing as likelihood, and if various eventualities (such as non-measurable subsets of Ω) are conceivable, then on what grounds should we suppose that one eventuality possesses a likelihood value, while another does not? Can we make such a counter-intuitive distinction merely on the basis of some abstract mathematical condition, such as whether or not the subset or eventuality belongs to some special sigma-algebra of subsets?

In the Riemann framework, with $\Omega = \mathbf{R}^B$, we can easily avoid this challenge to our intuitive understanding. Because the *variation* $V(F_X\ ;\ A)$ of F_X in A, is defined for *every* subset A of $\mathbf{R}$ or $\mathbf{R}^B$, with $V(F_X\ ;\ A) = \int_A F_X(I) = P(A)$ whenever A is F_X-integrable, we can define the likelihood, or probability, of *any* set or eventuality A to be $V(F_X\ ;\ A)$. (Variation corresponds to outer measure, so we could use the same ploy in the standard theory of probability, provided we replace the abstract sample space Ω by a concrete sample space such as $\mathbf{R}$ or $\mathbf{R}^B$.)

7 Stochastic Process The framework for probability described above has been received from Henstock in a somewhat fragmentary way. But within this implied framework, he made a deep and sustained investigation of some difficult problems of random variation. The most significant of these is [8], in which problems of Brownian motion (involving Wiener integration) and quantum mechanics (Feynman integration) are examined as *stochastic processes* by means of generalized Riemann concepts. Curiously, in a dense and closely argued paper of twenty eight pages, the word "probability" appears just three times in a single introductory paragraph, and otherwise not at all. Henstock treats these subjects primarily as problems of infinite-dimensional integration. But in a way this is not something to be wondered at, since the actual source of the mathematical conquest of random variation is the theories of measure and integration.

Henstock used ideas from Jessen [14] to extend the generalized Riemann integral from finite-dimensional to infinite-dimensional spaces.

The probabilistic context for this development is as follows. Suppose we are analyzing the joint variation of random variables $\mathcal{X}_t$, $t \in B$. In Henstock's original conception, the set B is a labelling or indexing set without structure. But it is useful to assume further that B has an order relation and a metric; and in practice B will be one of the following: a real number singleton $\{t\}$; or a finite set N of real numbers; or a real interval $]\tau', \tau]$.

Each $\mathcal{X}_t$ can be thought of as (X_t, F_t) with $X_t \subseteq \mathbf{R} = \mathbf{R}^{\{t\}}$, and $F_t := F_{X_t}$ defined on the intervals I_t of $\mathbf{R}^{\{t\}}$. For $N = \{t_1, \dots, t_n\} \subset B$, with $t_1 < \cdots < t_n$, the joint variables $\mathcal{X}_N := (\mathcal{X}_{t_1}, \dots, \mathcal{X}_{t_n})$ have a joint distribution function $F_N = F_{X_N} = F_{X_{t_1},\dots,X_{t_n}}$ defined on the intervals

$$I(N) := I_{t_1} \times \cdots \times I_{t_n} \ \subset\ \mathbf{R}^N = R^{\{t_1\}} \times \cdots \times R^{\{t_n\}} = \mathbf{R} \times \cdots \times \mathbf{R} = \mathbf{R}^n,$$

where $F_N(I(N))$ is the probability that each $\mathcal{X}_{t_j}$ takes a value $x_{t_j} \in I_{t_j}$ for $t_j \in N$. Recall that the set of random variables $\mathcal{X}_N := (\mathcal{X}_{t_1}, \dots, \mathcal{X}_{t_n})$ are independent if, for each interval $I(N) \subset \mathbf{R}^N$,

$$F_N(I(N)) = \prod_{t_j \in N} F_{t_j}(I_{t_j}).$$

Our problem is to analyze the joint variation of the random variables

$$\mathcal{X}_B = (\mathcal{X}_t)_{t\in B}$$

where B is an infinite set such as $]\tau', \tau]$. This problem arises when we are dealing with real- or complex-valued random variables $f(\mathcal{X}_B)$ whose expected or mean value is defined by an integral of the form

$$\int_{\mathbf{R}^B} f(\mathcal{X}_B) F_B(I).$$

To give some meaning to this for infinite B, we must establish the following: how the distribution function F_B is to be understood; what the intervals I of $\mathbf{R}^B$ are; and what is meant by the generalized Riemann integral in $\mathbf{R}^B$.

The set of intervals of $\mathbf{R}^B$ is denoted by $\mathcal{I}(\mathbf{R}^B)$, and individual intervals I of $\mathbf{R}^B$ are

$$I = I[N] := \mathrm{Proj}_N^{-1}(I(N)) = I(N) \times \prod\{R^{\{t\}} : t \in B \setminus N\}, \quad N \in \mathcal{F}(B)$$

where $\mathcal{F}(B)$ is the family of finite subsets N of B, and Proj_N is the mapping which, for each $N \in \mathcal{F}(B)$, projects $\mathbf{R}^B$ into $\mathbf{R}^N$: $\mathrm{Proj}_N(x_B) = x_N$, with

$$x_B = (x_t)_{t\in B} \in \mathbf{R}^B,\ x_N = (x_{t_1}, \dots, x_{t_n}) \in \mathbf{R}^n,\ N = \{t_1, \dots, t_n\} \in \mathcal{F}(B).$$

So while rectangular or brick-like intervals $I(N) = I_1 \times \cdots \times I_n$ are used to partition finite-dimensional spaces $\mathbf{R}^N = \mathbf{R} \times \cdots \times \mathbf{R} = \mathbf{R}^n$, cylindrical intervals $I[N]$ are used to partition the space $\mathbf{R}^B$ for Riemann-type integration,

$$\mathbf{R}^B = I^1[N^1] \cup \cdots \cup I^p[N^p].$$

The distribution function $F_B = F_{X_B}$ of the joint variables $\mathcal{X}_B = (\mathcal{X}_t)_{t\in B}$ is defined as

$$F_B(I[N]) := F_{X_N}(I(N)) = F_N(I(N)) \text{ for } I[N] \subset \mathbf{R}^B,\ N \in \mathcal{F}(B). \tag{11}$$

If the random variables $\{\mathcal{X}_t : t \in N\}$ are independent for each $N \in \mathcal{F}(B)$, then the random variables $\mathcal{X}_B = \{\mathcal{X}_t : t \in B\}$ are independent.

If f is a real- or complex-valued function defined on $\{x_t : t \in B\} = X_B$ in $\mathbf{R}^B$, with distribution function $F_{X_B}(I[N])$, then $f(\mathcal{X}_B)$ is a random variable if the expected value exists, defined as:

$$E(f(\mathcal{X}_B)) = \int_{\mathbf{R}^B} f(x_B) F_{X_B}(I[N]), \tag{12}$$

a generalized Riemann integral whose Riemann sum estimates have the form

$$\sum f(x_B) F_{X_B}(I[N]) = \sum_{q=1}^{p} f(x_B^q) F_{X_B}(I^q[N^q]). \tag{13}$$

where $\mathbf{R}^B$ is partitioned by the intervals

$$\{I^1[N^1], \dots, I^p[N^p]\}.$$

Each x_B^q belongs to the closure of $I^q[N^q]$; and the intervals $I^q[N^q]$ are made to depend on the representative values x_B^q in a way which is familiar in generalized Riemann integration. The

dimensions N of the cylindrical intervals $I = I[N]$ should be thought of as expanding sets, sets increasing without limit (other than remaining finite subsets of B), in these Riemann sums. The evaluation point x_B of f in each term of the Riemann sum can conveniently be taken as a vertex of $I[N]$. In other words, x_N can be a vertex of $I(N)$ in each term. Different finite sets N can appear in the various intervals $I[N]$ of any such partition. This part of the theory can be found in [7, 8, 13, 19, 22].) The best definition of the integral (12) is given in [13], pages 797-8.

If B is a finite set, or a singleton, this theory of functions of stochastic processes reduces to the basic theory of random variables described earlier. For instance, if B consists of just two elements we can without loss of generality take N to be B. Then we are dealing with the joint variation of two random variables $\mathcal{X}_1 = (X_1, F_{X_1}(I_1))$ and $\mathcal{X}_2 = (X_2, F_{X_2}(I_2))$; with the joint variable $\mathcal{X}_B = (\mathcal{X}_1, \mathcal{X}_2)$ taking values in $X_1 \times X_2 \subseteq \mathbf{R} \times \mathbf{R}$; and with joint distribution function $F_{X_B}(I_1 \times I_2)$, so that the probability of $\mathcal{X}_B = (\mathcal{X}_1, \mathcal{X}_2)$ having a joint occurrence (x_1, x_2) in $I_1 \times I_2$ is $F_{X_B}(I_1 \times I_2)$. The random variables $\mathcal{X}_1$ and $\mathcal{X}_2$ are independent if $F_{X_B}(I_1 \times I_2) = F_{X_1}(I_1)F_{X_2}(I_2)$ for all I_1 and I_2.

When analyzing joint variation of, for instance, a pair of variables $\mathcal{X}_1$, $\mathcal{X}_2$ with potential values $\{x_1\}$, $\{x_2\}$ respectively, the kind of thing we have to deal with are random variables $f(\mathcal{X}_1, \mathcal{X}_2)$ such as $\mathcal{X}_1 + \mathcal{X}_2$, $\mathcal{X}_1\mathcal{X}_2$ or $\exp(\sqrt{-1}\mathcal{X}_1\mathcal{X}_2)$ taking values $\{x_1 + x_2\}$, $\{x_1x_2\}$, or $\{\exp(\sqrt{-1}x_1x_2)\}$ respectively. So in this case we have random variables $f(\mathcal{X}_B) = f(\mathcal{X}_1, \mathcal{X}_2)$, whose expected value is given by (12) with Riemann sum estimates (13), taking $N = B$.

It is easy to deduce from (12) that the expected value of the random variable $f(\mathcal{X}_B) = \mathcal{X}_1 + \mathcal{X}_2$ whose values are $x_1 + x_2 \in X_1 + X_2$ is the sum of the expected values of $\mathcal{X}_1$ and $\mathcal{X}_2$. Similarly, if $\mathcal{X}_1$ and $\mathcal{X}_2$ are independent then, using Fubini's Theorem, (12) implies that the expected value of the product random variable $f(\mathcal{X}_B) = \mathcal{X}_1\mathcal{X}_2$—whose values are $\{x_1x_2\}$—is the product of the expected values of $\mathcal{X}_1$ and $\mathcal{X}_2$.

In general, whether or not the function $f(\mathcal{X}_B)$ (—or, if we prefer, the set of real or complex values $f(X_B) = \{f(x_B)\}$ with distribution function F_{X_B}) is to be regarded as a random variable depends on whether the integral $\int_{\mathbf{R}^B} f(x)F_B(I)$ exists, where $x = x_B \in \mathbf{R}^B$, $I = I[N] \subset \mathbf{R}^B$, and $F_B(I) = F_{X_B}(I[N])$ is the joint probability distribution function for the process $\mathcal{X}_B$.

This in turn depends on F_{X_B}, which we must examine more closely. To determine whether the expectation or integral (12) exists we must examine the convergence properties of the actual distribution function in question. Anyone who studies basic probability theory will be familiar with various distribution functions in one dimension or in a finite number of dimensions. Some of these have been mentioned in Section 5 above. For instance, the one-dimensional distribution function $F_X(I)$ for a random variable $\mathcal{X}$ following a normal distribution with expected value μ and variance σ^2 is

$$F_X(I) = \frac{1}{\sigma\sqrt{2\pi}} \int_I \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy. \tag{14}$$

The discussion we are engaged in is about developing a Riemann-type theory of such distribution functions, but in unlimited number of dimensions.

The first comprehensive explanation of this problem by Henstock is in [8]. The particular distribution function he deals with is as follows. For $\mathbf{R}^B$ with B infinite, a distribution function $F_B(I[N]) = F_{X_B}(I[N])$ is defined on the cylindrical intervals $\mathcal{I}(\mathbf{R}^B)$ of $\mathbf{R}^B$. For each $N = \{t_1 < \cdots < t_n\} \in \mathcal{F}(B)$, the distribution function $F_N(I(N)) = F_{X_N}(I(N))$ is

defined on the finite-dimensional intervals $I(N) = \prod_{j=1}^{n} I_j$ of $\mathbf{R}^N$; with

$$F_{X_N}(I(N)) = \int_{I_{t_1} \times I_{t_2} \times \cdots \times I_{t_n}} \prod_{j=1}^{n} \left(\frac{\exp \frac{c}{2} \frac{(y_j - y_{j-1})^2}{t_j - t_{j-1}}}{\frac{2\pi}{-c}(t_j - t_{j-1})} \right) dy_1 dy_2 \cdots dy_n. \tag{15}$$

In accordance with (11), the distribution function defined on the cylindrical intervals $I[N]$ of $\mathbf{R}^B$ is

$$F_{X_B}(I[N]) = F_{X_N}(I(N)) \text{ for } I[N] \in \mathcal{I}(\mathbf{R}^B). \tag{16}$$

The function F_{X_N} of (15) is, of course, an example of a finite joint distribution function of the kind we discussed earlier. When $c = -1$, as in the one-dimensional (14), (15) is the probability that the displacement, in one particular direction, of a particle undergoing Brownian motion will, at times t_j, be contained (jointly) within the real intervals I_j for $j = 1, \ldots, n$. Thus, if the particle has displacement y_{j-1} at time t_{j-1}, the displacement y_j at time t_j is the value of a normally distributed random variable with mean y_{j-1} and variance $t_j - t_{j-1}$, corresponding to (14).

To get a basic idea of the way the infinite-dimensional distribution function (16) behaves, take $c = -1$, take B to be the finite set $\{\tau_1, \ldots, \tau_n\}$ and N identically equal to B (without loss of generality), so replacing the variable times t_j by fixed τ_j $(1 \leq j \leq n)$. This is a reversion to (15). Then, with $f(\mathcal{X}_B) = f(x_B)$ identically 1, we can interpret the integral (12) of the expression (16) in $\mathbf{R}^B = \mathbf{R}^N = \mathbf{R}^n$ as simply giving us the probability that, for $j = 1$ to n, the particle displacement y_j takes an arbitrary value in $\mathbf{R}$. So the integral value is 1. This can easily be verified by actual integration, using Fubini's Theorem in $\mathbf{R}^n$ to perform successive integrations from $y_j = -\infty$ to $+\infty$, $j = 1$ to n. (The integration involved uses only elementary methods of calculus. See [19] pages 53 and 85.)

Reverting to infinite B, when $c = \sqrt{-1}$ the expression (16) corresponds in some way to the likelihood that a quantum mechanical particle which is unconstrained by any external forces will satisfy the displacement condition described above for Brownian motion. This is the phenomenon analyzed by Richard Feynman in his PhD thesis and in [2]. It is the basis of quantum physics, and notoriously gives rise to problems of convergence—the so-called renormalization problem.

Our discussion of (16) from this point onwards will be mainly concerned with the case $c = \sqrt{-1}$.

8 Feynman Integration Henstock undertook this particular investigation, culminating in [8] in 1973, following a discussion with a physics researcher, R. Johnson, in Queen's University Belfast in 1962. This discussion appears to have had a quite profound effect on Henstock's subsequent research—but indirectly, since he did not place a special or exclusive focus on Feynman integration, or on any other specific problems of random variation. Instead, he focussed on the broader integration issues that these problems give rise to. For instance, even a superficial examination of (16), with $c = \sqrt{-1}$, shows challenging problems. Because then, if we take the absolute value, we get a factor

$$\prod_{j=1}^{n} (t_j - t_{j-1})^{\frac{1}{2}}$$

which diverges very "strongly" as the set N "expands" and the number of terms in this finite product increases without limit, with each of the $t_j - t_{j-1}$ tending to zero.

One of the ways in which Henstock responded to this challenge to the theory of integration was to rework his Riemann-type integration theory into its fundamental or most abstract form in order to extract the best possible results.

Remember, his mathematical journey had begun with difficult problems of summability or convergence of series and integrals; in particular, the use of convergence factors which could change divergent sums to convergent ones. It was conceivable that the summability problems in (16) would also be amenable to the Riemannian theory of integration which he had developed in the course of this quest. This, in effect, is the question that the physicist R. Johnson put to him in Belfast in 1962.

By 1968, along with a parallel investigation of summability [7], Henstock had developed an abstract version of the Riemann-complete (gauge or generalized Riemann) integral in terms of what he called *division spaces*. This theory is also included in [7], along with a very brief introduction to his theory of integration in infinitely many dimensions.

As well as providing a general theory within which practically every system of mathematical integration (Riemann, Lebesgue, Radon, Burkill, Perron, Denjoy) could be located, this general theory of integration exposed the essential, irreducible elements of integration in order to get at the best possible theorems. To put it plainly, if convergence of sums of the very "badly behaved" (16), with $c = \sqrt{-1}$, is to be capable of being proved, only the most delicate of arguments is likely to prevail. In [8], Henstock teased out a convergence result which he hoped might succeed with (16), where the Lebesgue Dominated Convergence Theorem fails. But he also demonstrated in [8] the limitations of this first attempt.

9 Feynman Convergence Theorems A suitable convergence theorem was evolved in stages over the following two decades, reaching its best form in [11]. But by this time Henstock's attention was on other problems, and he was was less concerned with the physics problems which had inspired him to go down this particular route in the first place.

In the meantime, he occasionally paid some attention to the Feynman problem using whatever means that he could devise. One approach was to express the integral in $\mathbf{R}^B$ of (16)—essentially an infinite-dimensional integral—as the limit if finite-dimensional integrals when the number of dimensions increases without limit.

To get an idea of this approach, in (16) think of $t_1, \ldots, t_n$ as fixed numbers, so (16) is an expression in a fixed domain $\mathbf{R}^n$. Then, even when $c = \sqrt{-1}$, (16) is integrable in $\mathbf{R}^n$ without much difficulty—see Muldowney [19], Proposition 68.

The approach which predominated in these later studies by Henstock is one which, when $c = -1$, is called Wiener's Formula, in which the expected value of some random variable dependent on a Brownian motion is the limit as $n \to \infty$ of integrals in $\mathbf{R}^n$. When $c = -1$, (16) is the well-known probability distribution function for Brownian motion, and the result holds trivially with a function $f(\mathcal{X}_B)$ which is not random but constant. And even when $c = \sqrt{-1}$ it is also true in the same sense, provided the integrals are generalized Riemann. (The latter result is the result mentioned above—Proposition 68, page 84 of [19].) The general problem is to prove it when we are calculating the expected value, with respect to the distribution function (16), of some random variable.

In a way, there is something strange about this kind of argument. Consider a function of three variables, integrated in three dimensions, and compare this with the same function integrated with respect to only two of the variables in only two dimensions. In the first case we are dealing with something three-dimensional or solid; in the second, two-dimensional or flat. Yet we talk about "approaching" an infinite-dimensional integral via finite-dimensional integrals of increasing dimension. In what sense can a series of lower-dimensional objects "approach" some higher-dimensional one? How can a square "approach" a cube?

But when we integrate a constant function with respect to (16) with any fixed n, we get the same result as if we integrate the constant function, or other uncomplicated functions, with respect to (16) in an infinite number of dimensions. (See [19] for details of such evaluations.) Therefore the argument from a finite number of dimensions to infinitely many dimensions works out all right in these simple cases.

Having got this far with this line of argument, we could consider step functions, then continuous functions, and so on, in a manner which is familiar in analysis. But in order to see this argument through to the end we would need an appropriate integral convergence theorem, which until [11] was not available.

Before his discovery of this result, Henstock made use of the concept of *backwards martingales* [9]: "*A sequence* (p_n) *of random variables is a* martingale *if* p_{n-1}) *is almost surely the conditional expectation of* p_n *given* $p_1, \dots, p_{n-1}$ *(*$n = 2, 3, \dots$*), and often* p_n *tends almost surely to a pointwise limit as* $n \to \infty$*. The sequence is a* backwards martingale *if [...]* p_n *is the conditional expectation of* p_j *(*$j < n$*), for* $n = 2, 3, \dots$*, and here* p_n *often tends almost surely to a constant as* $n \to \infty$."

On the strength of this idea (previously studied by Jessen [14]), and replacing random sequences by random processes, we are not too far away from the situation described in the previous paragraph and Henstock obtained Jessen- and Wiener-type results even for the extreme case $c = \sqrt{-1}$ which arises in Feynman integration.

One drawback under which this work was done by him was the misapprehension that when $c = \sqrt{-1}$, (16) is not of *generalized bounded variation* (or VBG*). Henstock explained his change of mind on this problem at a lecture he gave at the XV Summer Symposium in Real Analysis in Smolenice, Czechoslovakia, 1991. There is an outline of his thinking in [12]. Essentially, the issue here is whether we can find some special subsets A^k of $\mathbf{R}^B$ so that when we take appropriate Riemann sums of the absolute value of (16), and when we remove every term which is not "tagged" in a particular A^k (that is, has $x_B^q \in A^k$), the resulting sums are always bounded by some constant α_k. Even a cursory examination of (16) with $c = \sqrt{-1}$ will persuade the reader that such sets A^k, if they exist, would have to be very special, very "small". But we require that they be large enough that their countable union is $\mathbf{R}^B$. The problem is discussed at some length in [19]. The fact that the VBG* property holds for (16) means that the kind of divergences arising from the products of $(t_j - t_{j-1})^{-\frac{1}{2}}$, with each $t_j - t_{j-1}$ tending to zero, may after all be manageable. And this property, along with Henstock's advanced limit theorems for integrals from [11], is what makes it possible to bring the Feynman theory of quantum mechanics into the theories of integration and random variation, in a sense which is compatible with our ordinary understanding of these subjects. A fuller account of this can be found in [20].

M.M. Rao [24] commented on these developments as follows: *"It is of interest to present an outline of the non-absolute Feynman integral since Henstock [8] has already extended the (Perron) P-partition method to this case, and a further elaboration is given by Muldowney [19]. This gives an ultimate reason for studying Henstock's method of the generalized Riemann integral which is simpler (and also more general) than many of the existing procedures based on extensions of Lebesgue's integral."*

Before moving on from [8], a final thought. An elusive issue that appears in many accounts of the subject, is to find some equivalent of Lebesgue measure for $\mathbf{R}^B$: some interpretation of $\prod\{dx_t : t \in B\}$ in $\mathbf{R}^B$ corresponding to $dx_1 dx_2 \cdots dx_n$ in $\mathbf{R}^n$. We can attribute the conquest of this outpost to Henstock, in some generalized Riemann sense. In this sense, if I_j is the interval $[u_j, v_j[$, denote the length $v_j - u_j$ by $|I_j|$, so the volume element $\prod_{j=1}^n |I_j| = |I(N)|$ corresponds to $dx_N = dx_1 \cdots dx_n$, or Lebesgue measure.

A result outlined by Henstock in [8], and explained in greater detail in Muldowney [22], showed that we can have integrals in $\mathbf{R}^B$ with $|I(N)|$ or $dx_1 \cdots dx_n$ as integrator instead of

the distribution function (16). This is because (16) can be replaced by

$$\prod_{j=1}^{n}\left(\frac{2\pi}{-c}(t_j-t_{j-1})\right)^{-\frac{1}{2}}\exp\frac{c}{2}\frac{(x_j-x_{j-1})^2}{t_j-t_{j-1}}|I_j|,$$

in which the factor $\prod_{j=1}^{n}|I_j|$ corresponds to the integrator $\prod_j dx_j$. So, given a random variable $f(\mathcal{X}_B)$ in the sample space $\mathbf{R}^B$, the expected value

$$\int_{\mathbf{R}^B} f(x_B)F(I[N])$$

of $f(\mathcal{X}_B)$ can also be expressed as

$$\int_{\mathbf{R}^B} f(x_B)F(x_B,N)|I(N)|,\ \text{or}\ \int_{\mathbf{R}^B} f(x_B)F(x_B,N)dx_N \tag{17}$$

(the latter notation is closer to the more common one), with

$$F(x_B,N)=\prod_{j=1}^{n}\left(\frac{2\pi}{-c}\,(t_j-t_{j-1})\right)^{-\frac{1}{2}}\exp\frac{c}{2}\frac{(x_j-x_{j-1})^2}{t_j-t_{j-1}}.$$

The expression (17) is not using "Lebesgue measure in $\mathbf{R}^B$", but each Riemann sum approximation to (17) has the form

$$\sum f(x_B)F(x_B,N)|I(N)|$$

where $|I(N)|$ is the pre-measure for Lebesgue measure in $\mathbf{R}^N$.

10 Stochastic Calculus In 1974 E.J. McShane [18] published a theory of stochastic calculus which puts the stochastic integrals of Itô on a Riemann-type footing. What is the idea behind this? We start with a process $\{\mathcal{X}_t\}$ (usually Brownian motion) whose increments in the time interval s to $s+\Delta t$ can be denoted by $\Delta\mathcal{X}_s=\mathcal{X}_{s+\Delta s}-\mathcal{X}_s$. These increments are associated with a probability measure in the underlying sample space Ω whose pre-measure is given by (16) with $c=-1$. We are often concerned with a related process whose increments have the form

$$\Delta\mathcal{Y}_s=h_1(\mathcal{X}_s)\Delta\mathcal{X}_s+h_2(\mathcal{X}_s)\Delta s. \tag{18}$$

The problem then is, how to define the process $\{\mathcal{Y}_t\}$ itself. If we know, in some sense, what the "increments" $\Delta\mathcal{Y}_s$ are, for each s, how do we determine what each of the random variables $\mathcal{Y}_t$ is? Following the Riemann-type argument of this article, as pioneered by Henstock, we would, at this point, seek out the probability distribution function $F_{Y_t}(I)$ where the intervals I are the cylindrical intervals $I[N]$ of $\mathbf{R}^B$. (This is how the problem is approached in [21].) But what is the standard approach, as developed by Itô and others?

Formally,

$$\mathcal{Y}_t-\mathcal{Y}_0=\sum_{\Delta_s}(h_1(\mathcal{X}_s)\Delta\mathcal{X}_s+h_2(\mathcal{X}_s)\Delta s)\ \text{or}\ \int_0^t h_1(\mathcal{X}_s)d\mathcal{X}_s+\int_0^t h_2(\mathcal{X}_s)ds.$$

($\mathcal{Y}_0$ can often be taken to be 0, deterministically, so the left hand term is $\mathcal{Y}_t$.) Mathematical meaning was given by Itô to certain expressions of the form $\int_0^t h_1(\mathcal{X}_s)d\mathcal{X}_s$, by means of

the theory of *stochastic integrals*. This theory makes it possible to investigate forms of probability measures for $\{\mathcal{Y}_t\}$ in Ω that are derived from the probabilities (16) but which relate to the values assumed by $\{\mathcal{Y}_t\}$. E.J. McShane [18] provided a Riemann version of the theory of stochastic integration, which has been substantially developed by P.Y. Lee, T.S. Chew and others. See [1, 17].

Another contributor was T.W. Lee who, under Henstock's supervision, wrote a Ph.D. thesis entitled *Integration Theorems for Product Division Spaces*, 1970. This was a study of the concepts, described above, that Henstock was developing in the areas of Wiener and Feynman integration, and which he outlined in his 1967 book *Linear Analysis* [7] prior to the more detailed study of [8]. T.W. Lee brought these ideas together with ideas from R.H. Cameron, in what he called the G-integral and G*-integral. In 1976 T.W. Lee published a paper [16] in which he relates stochastic integrals of the McShane type with the G-integrals, producing a double integral involving both. In 1990, Henstock published a further study [10] of these problems, in which he elaborates a version of Fubini's Theorem for the T.W. Lee scenario.

What is at issue in stochastic calculus is how to determine the "law" of a process or motion $\mathcal{X}_B$ described by equations such as (18). This means finding out the probabilities which govern the values taken by the process $\mathcal{X}_B = \{\mathcal{X}_t : t \in B\}$; thereby enabling us to determine, for instance, the expected value of a random variable $f(\mathcal{X}_B)$. The Itô calculus uses stochastic integrals to provide us with a way of deducing such a "law", or set of probabilities, from the law of the underlying process $\mathcal{X}_B = \{\mathcal{X}_t\}$. When $\mathcal{X}_B$ is Brownian motion, this underlying law $F_{X_B}(I)$ is given by (16) with $c = -1$. The Itô calculus then enables us to deduce from this the probabilities of a related process such as $\mathcal{Y}_B = \exp \mathcal{X}_B$ defined by $Y_t = \exp X_t$ or $y_t = \exp x_t$ for $x_t \in X_t$, $t \in B$.

But when we step back a little, and view this subject from the Riemann standpoint whose development by Henstock has been the subject of this article, we find a route to the law of $\mathcal{Y}_B$ which is simpler and more direct than the route provided by the Itô calculus. The Henstock method hinges on direct examination of the underlying law or pre-measure $F_{\mathcal{X}_B}(I)$ of $\mathcal{X}_B$. The method is illustrated in [21] where the important case of derivative price processes (or Black-Scholes theory) is elaborated using the Riemannian or Henstock method. The key to this is that only pre-measures like (16), and not measures, are needed to carry out the analysis by Riemann methods.

11 Conclusion This article has described how Henstock laid the foundations for a development of a theory of probability and random variation on Riemann rather than Lebesgue lines, providing an alternative to the Kolmogorov approach. We have indicated how the Henstock approach enables us to bring Feynman's theory of quantum mechanics within the scope of random variation, where the Lebesgue approach failed. The theory of stochastic processes can be treated as part of the theory of joint variation of two or more random variables. And we have pointed out how stochastic calculus can be simplified (in comparison with the Itô calculus) by adopting Henstock's approach, and without using Itô's Lemma or any other result of the standard stochastic calculus.

Though he is widely acclaimed for his achievements in other branches of mathematics, the innovations of Henstock in the theory of random variation are equally profound.

References

[1] Chew Tuan Seng, Huang Zhiyuan, Wang Cai Shi, The non-uniform Riemann approach to anticipating stochastic integrals, Stochastic Analysis and Applications 22(2), 429-442, 2005.

[2] R. Feynman, Space-time approach to non-relativistic quantum mechanics, Reviews of Modern Physics 20, 367-387, 1948.

[3] Henstock Archive, University of Ulster.

[4] R. Henstock, Additivity and the Lebesgue limit theorems, 223-241, Proceedings of the C. Caratheodory International Symposium, Greek Mathematical Society, Athens, 1974.

[5] R. Henstock, Theory of Integration, Butterworths, London, 1963.

[6] R. Henstock, Lectures on the Theory of Integration, World Scientific, Singapore 1988.

[7] R. Henstock, Linear Analysis, Butterworths, London, 1967.

[8] R. Henstock, Integration in product spaces, including Wiener and Feynman integration, Proceedings of the London Mathematical Society, 1973.

[9] R. Henstock, Division spaces, vector-valued functions and backwards martingales, Mathematical Proceedings of the Royal Irish Academy 80A(2), 217-233, 1980.

[10] R. Henstock, Stochastic and other functional integrals, Real Analysis Exchange 16, 1990-91, 460-470.

[11] R. Henstock, The General Theory of Integration, Clarendon Press, Oxford, 1991.

[12] R. Henstock, The integral over product spaces and Wiener formula, Real Analysis Exchange 17(1), 1991-92.

[13] R. Henstock, P. Muldowney, V.A. Skvortsov, Partitioning infinite-dimensional spaces for generalized Riemann integration, Bulletin of the London Mathematical Society 38, 795-803, 2006.

[14] B. Jessen, The theory of integration in a space of an infinite number of dimensions, Acta Mathematica 63, 249-323, 1934.

[15] A.N. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitreichnung, Ergebnisse der Mathematik, 1933 (Foundations of the Theory of Probability, Chelsea Publishing Company, New York, 1950).

[16] T.W. Lee, On the generalized Riemann integral and stochastic integral, Journal of the Australian Mathematial Society 21(A), 1976, 64-71.

[17] Lee Peng Yee, Xu Jiagu, Stochastic integrals of Itô and Henstock, Real Analysis Exchange 18(2), 352-366, 1992-93.

[18] E.F. McShane, Stochastic Calculus and Stochastic Models, Academic Press, New York, 1974.

[19] P. Muldowney, A General Theory of Integration, Including Wiener and Feynman Integration, Pitman, Harlow, 1987.

[20] P. Muldowney, Feynman's path integrals and Henstock's non-absolute integration, Journal of Applied Analysis, 6(1), 1-24, 2000.

[21] P. Muldowney, The Henstock integral and the Black-Scholes theory of derivative asset pricing, Real Analysis Exchange, 25(1), 117-132, 2000-01.

[22] P. Muldowney, The infinite dimensional Henstock integral and problems of Black-Scholes expectation, Journal of Applied Analysis 8, 1-21, 2002.

[23] P. Muldowney, A Riemann approach to random variation, Mathematica Bohemica 131(2), 167-188, 2006.

[24] M.M. Rao, Measure Theory and Integration, Marcel Dekker, New York, 2004.

P. Muldowney, Magee College, University of Ulster, N. Ireland
p.muldowney@ulster.ac.uk