A RIEMANN APPROACH TO RANDOM VARIATION

Patrick Muldowney, Northern Ireland

(Received October 6, 2005)

Dedicated to Prof. J. Kurzweil on the occasion of his 80th birthday

Abstract. This essay outlines a generalized Riemann approach to the analysis of random variation and illustrates it by a construction of Brownian motion in a new and simple manner.

Keywords: Henstock integral, probability, Brownian motion

MSC 2000: 28A20, 60A99, 60G05

1. Introduction

Measurement, estimation and forecasting generally involve elements of approximation. The term "error" is used in this context to denote the difference between the true or actual value and the estimated, measured or forecast value. For many centuries, mathematics has sought to develop concepts and methods of analysis to solve a range of problems which present themselves in this context. This endeavor can be described as the study of the phenomenon of "random variation".

One of the greatest achievements of twentieth century mathematics is the formulation of a rigorous theory of random variation, beginning with the work [4] of A. N. Kolmogorov, in which a calculus of probabilities is developed, leading to a treatment of random variables and their expectations based on Lebesgue's theory of the integral. In this essay the mathematical content and conceptions of the Kolmogorov model are probed, and an alternative mathematical approach is presented.

The Lebesgue integral has good properties, such as the Dominated Convergence Theorem, which make possible the formulation of a rigorous theory of probability. But Lebesgue's just happened to be the first of a number of such investigations into the nature of mathematical integration during the twentieth century.

Subsequent developments in integration, by Denjoy, Perron, Henstock and Kurzweil, have similar properties and were devised to overcome shortcomings in the Lebesgue theory. See [1] for detailed comparison of modern theories of integration. However, theorists of probability and random variation have not yet really "noticed", or taken account of, these developments in the underlying concepts. There are many benefits to be reaped by bringing these fundamental new insights in integration or averaging to the study of random variation.

In fact it is possible to formulate a theory of random variation and probability on the basis of a conceptually simpler Riemann-type approach, and without reference to the more difficult theories of measure and Lebesgue integration. See [6] for an essay in this approach.

2. Averaging: Riemann and Lebesgue

To motivate our discussion of integration, or averaging, we review the elementary calculation of an arithmetic mean as encountered in a first course in simple statistics. Suppose the sample space is the set of real numbers, or a subset of them. Thus, an individual random occurrence, measurement or item of data is a real number x.

While x is the underlying random variable, we are often concerned with some deterministic function f of x; as, for example, in the estimation of the variance of x. Then $f(x)$ is random or unpredictable because x is.

If successive instances of the measurement x are obtained, we might partition the resulting set of data into an appropriate number of classes; then select a representative element of the data from each class; multiply each of the representatives by the relative frequency of the class in which it occurs; and add up the products. This familiar procedure gives an estimate of the mean value of the measurement x.

Likewise, we can estimate the mean or expected value of the random variable $f(x)$. The following scheme (1) illustrates the procedure. The sample space (or domain of measurements) is partitioned into intervals $I^{(j)}$ of the sample variable (or occurrence or measurement) x, the random variable is $f(x)$, and the relative frequency of the class $I^{(j)}$ is $F(I^{(j)})$:

(1)

Classification of the data values	Function $f(x)$ of the data values x	Relative frequency F of the data class
$I^{(1)}$	$f(x^{(1)})$	$F(I^{(1)})$
$I^{(2)}$	$f(x^{(2)})$	$F(I^{(2)})$
$\vdots$	$\vdots$	$\vdots$
$I^{(m)}$	$f(x^{(m)})$	$F(I^{(m)})$

For each j, the measurement value $x^{(j)}$ is a representative element selected from $I^{(j)}$ (or from its closure). The resulting estimate of the mean value of the random variable $f(x)$ is $\sum_{j=1}^{m} f(x^{(j)})F(I^{(j)})$. Note that the sample variable of elementary occurrences x can itself be regarded as a random variable, with mean value estimated as $\sum_{j=1}^{m} x^{(j)}F(I^{(j)})$.

The approach to random variation that we are concerned with in this paper consists of a formalization of this relatively simple Riemann sum technique which puts at our disposal powerful results in analysis such as the Dominated Convergence Theorem.

In contrast the Kolmogorov approach requires, as a preliminary, an excursion into abstract measurable subsets A^j of the sample space:

(2)

Classification of the underlying variable x	Values of the function $f(x)$	Probability measure P
A^1	y^1	$P(A^1)$
A^2	y^2	$P(A^2)$
$\vdots$	$\vdots$	$\vdots$
A^m	y^m	$P(A^m)$

Here, x is again a representative member of a sample space Ω which corresponds to the various potential occurrences or states in the "real world" in which measurements or observations are taking place on a variable $f(x)$ whose values are unpredictable and which can only be estimated beforehand to within a degree of likelihood. (In practice, Ω is often identified with the real numbers or some proper subset of them; or with a Cartesian product, finite or infinite, of such sets.) If we follow the method of (2), numbers y^j are chosen in the range of values of the random variable $f(x)$, and A^j is $f^{-1}([y^{j-1}, y^j[)$. The resulting $\sum_{j=1}^{m} y^j P(A^j)$ is an estimate of the expected value of the random variable $f(x)$. While the sets A^j are usually intervals, or unions of intervals, in principle they are P-measurable sets. Such sets can be mathematically abstruse, and they can place heavy demands on the understanding and intuition of anyone who is not well-versed in mathematical analysis. For instance, it can be difficult for a non-specialist to visualize a Cantor set in terms of laboratory, industrial or financial measurements of some real-world quantity.

In contrast, the data classes $I^{(j)}$ of elementary statistics in (1) are easily understood as real intervals, of one or more dimensions, which arise in actual measurements; and these are the basis of the Riemann approach to random variation.

3. Further points of contrast

We now examine some further aspects of the Lebesgue-Kolmogorov approach in order to highlight some of the points of difference in the two approaches. While the sample space Ω is an abstract conception conveying a minimal core of abstract mathematical structure, it is, as mentioned above, frequently identified with some finite or infinite Cartesian product of the real numbers, or of subsets of these. And a probability measure P on Ω is, in practice, frequently the measure generated by a probability distribution function $F_X(I)$ associated with a particular random variable X. So, in practice, the mathematical structures have quite concrete interpretations.

To illustrate, suppose X is a normally distributed random variable in a sample space Ω. Then we can represent Ω as $\mathbb{R}$, the set of real numbers; with X represented as the identity mapping $X\colon \mathbb{R} \to \mathbb{R}$, $X(x) = x$; and with distribution function F_X defined on the family $\mathcal{I}_{\mathbb{R}}$ of intervals I of $\mathbb{R}$, $F_X\colon \mathcal{I}_{\mathbb{R}} \to [0,1]$:

$$F_X(I) = \frac{1}{\sqrt{2\pi}} \int_I \mathrm{e}^{-\frac{1}{2}s^2}\,\mathrm{d}s. \tag{3}$$

Then, in the Lebesgue-Kolmogorov approach, we generate, from the interval function F_X, a probability measure $P_X\colon \mathcal{A}_{\mathbb{R}} \to [0,1]$ on the family $\mathcal{A}_{\mathbb{R}}$ of Lebesgue measurable subsets of $\Omega = \mathbb{R}$. So the expectation $E^P(f)$ of any P_X-measurable function f of x is the Lebesgue integral $\int_\Omega f(x)\,\mathrm{d}P_X$. With Ω identified as $\mathbb{R}$, this is just the Lebesgue-Stieltjes integral $\int_{\mathbb{R}} f(x)\,\mathrm{d}F_X$, and, since $x \in \mathbb{R}$ is just the standard normal variable of (3), the latter integral reduces to the Riemann-Stieltjes integral—with Cauchy or improper extensions, since the domain of integration is the unbounded $\mathbb{R} =]-\infty, \infty[$.

Thus, although the final result is relatively simple in mathematical terms, to get there from (3) we are obliged to wade through quite deep mathematical waters. In presenting this outline we have omitted many steps, the principal ones being the probability calculus and the construction of the probability measure P. It is precisely these steps which cease to be necessary preliminaries if we take a generalized Riemann approach, instead of the Lebesgue-Kolmogorov one, in the study of random variation.

Because the generalized Riemann approach does not specify an abstract measurable space Ω as the sample space, from here onwards we will take as given the identification of the sample space with $\mathbb{R}$ or some subset of $\mathbb{R}$, or with a Cartesian product of such sets, and take the symbol Ω as denoting such a space. Accordingly we will drop the traditional notations X and $f(X)$ for denoting random variables. Instead an elementary random occurrence will be denoted by the variable (though unpredictable) element x of the (now Cartesian) sample space, and a general random variable will be denoted by a deterministic function f of the underlying variable x. The associated likelihoods or probabilities will be given by a distribution function

$F(I)$ defined on intervals (which may be Cartesian products of one-dimensional intervals) of Ω. Whenever it is necessary to relate the distribution function F to its underlying, elementary random occurrence or outcome x, we may write F as F_x.

4. Foundations of a Riemann approach

The standard approach starts with a probability measure P defined on a sigma-algebra of measurable sets in an abstract sample space Ω; it then deduces probability density functions F. These distribution functions (and not some abstract probability measure) are the practical starting point for the analysis of many actual random variables—normal (as described above in (3)), exponential, Brownian, geometric Brownian, and so on.

In contrast, the generalized Riemann approach posits the probability distribution function F as the starting point of the theory, and proceeds along the lines of the simpler and more familiar (1) instead of the more complicated and less intuitive (2).

To formalize these concepts a little more, we have some domain Ω of potential occurrences which we call the *sample space*. The elements x of Ω are the elementary occurrences or events, each of which can be thought of as a measurement (or combination of joint measurements) which gives unpredictable results. The domain Ω will be identified with $S^B = \prod\{S\colon\ B\}$ where S is $\mathbb{R}$ or some subset of $\mathbb{R}$, and B is an indexing set which may be finite or infinite. In some basic examples such as throwing dice, S may be a set such as $\{1,2,3,4,5,6\}$, or, where there is repeated sampling or repeated observation, a Cartesian product of such sets. A likelihood function F is defined on the data intervals of $\prod\{S\colon\ B\}$. A general random variable (or *observable*) is taken to be a function $f(x)$ defined for $x \in \Omega$.

In Section 13 on Brownian motion we will show how to deal with a sample space which is not itself a Cartesian product, but is a proper subset of a Cartesian product $\mathbb{R}^B$.

The Lebesgue-Kolmogorov approach develops distribution functions F from probability measures $P(A)$ of measurable sets A. Even though probability distribution functions are often the starting point in practice (as in (3) above), Kolmogorov gives primacy to the probability measures P, and they are the basis of the calculus of probabilities, including the crucial relation

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \tag{4}$$

for disjoint P-measurable sets A_j. Viewed as an axiom, the latter is a somewhat mysterious statement about rather mysterious objects. But it is the lynch-pin of the

Lebesgue-Kolmogorov theory, and without it the twentieth century understanding of random variation would have been impossible.

The generalized Riemann approach starts with probability distribution functions F_x defined only on intervals I of the sample space $\Omega = S^B$. We can, as shown below (13), deduce from this approach probability functions P_x defined on a broader class of "integrable" sets A, and a calculus of probabilities which includes the relation (4)—but as a theorem rather than an axiom. So instead of being a starting point, (4) emerges at a later stage in the Riemann approach to random variation.

What, if any, is the relationship between these two approaches to random variation? There is a theorem [8] which states that every Lebesgue integrable function (in $\mathbb{R}^B$) is also generalized Riemann integrable. In effect, this guarantees that every result in the Lebesgue-Kolmogorov theory also holds in the generalized Riemann approach. So, in this sense, the former is a special case of the latter.

The key point in developing a rigorous theory of random variation by means of generalized Riemann integration is, following the scheme of (1) above, to partition the domain or sample space $\Omega = S^B$, in an appropriate way, as we shall proceed to show. (Whereas in the Lebesgue-Kolmogorov approach we step back from (1), and instead use (2) supported by (4). The two approaches part company at the (1) and (2) stages.)

In the generalized Riemann approach we focus on the classification of the sample data into mutually exclusive classes or intervals I. In mathematical language, what is involved in this is the partitioning of the sample space $\Omega = S^B$ by intervals I.

In the first lesson of elementary statistics, the usual practice is to divide up the domain of measurements, or the data, into equal classes, and then perform the averaging operation described in (1) above, to obtain the estimated mean value or expectation of the random variable $f(x)$.

Later on in elementary studies of statistics, a little computational sophistication may be applied to the classification of the data (or partitioning of the sample space). Often this leads to the use of quantile points to classify or code the data. In this case, unequal classes are obtained if the distribution of data is not uniform. So unequal classes are arrived at in order to improve efficiency of computation and accuracy of estimates. If quantile points are used to form the intervals $I^{(j)}$ in (1), then for each representative instance or occurrence $x^{(j)} \in \mathrm{Cl}I^{(j)}$, the random variable value $f(x^{(j)})$ is multiplied by $F(I^{(j)})$, where the $F(I^{(j)})$ are equal for $j = 1, 2, \ldots, n$; giving us data classes of equal likelihood rather than equal size.

In pursuing a rigorous theory of random variation along these lines, this basic idea of partitioning the sample space by intervals is the key. Instead of retreating to the abstract machinery of (2), we find a different way ahead by carefully selecting the intervals $I^{(j)}$ which partition the sample space $\Omega = S^B$.

An idea of what is involved in this can be obtained by recalling the role of Riemann sums in basic integration theory. Suppose for simplicity that the sample space Ω is the interval $[a, b[\subset \mathbb{R}$ and the observable $f(x)$ is given by $f\colon \Omega \to \mathbb{R}$; and suppose $F\colon \mathcal{I} \to [0,1]$ where $\mathcal{I}$ is the family of subintervals $I \subseteq \Omega = [a, b[$.

We can interpret F as the probability distribution function of the underlying occurrence or measurement x, so $F(I)$ is the likelihood that $x \in I$. As a distribution function, F is finitely additive on $\mathcal{I}$.

Probability is a notoriously contentious and difficult concept. But the simplest intuition of likelihood—as something intermediate between certainty of non-occurrence and certainty of occurrence—implies that likelihoods must be representable as numbers between 0 and 1. So we can plausibly infer that the functions F are finitely additive in $\mathcal{I}$ and are thus distribution functions. By making this our starting point we lift the burden of credulity that (4) imposes on our naive or "natural" sense of what probability or likelihood is.

With f a deterministic function of the underlying random variable x, the random variation of $f(x)$ is the object of our investigation. In the first instance we wish to establish $E(f)$, the expected value of $f(x)$, as, in some sense, the integral of f with respect to F, which is often estimated as in (1).

Following broadly the scheme of (1), we first select an arbitrary number $\delta > 0$. Then we choose a finite number of disjoint intervals $I^1, \dots, I^m$; $I^j = [u^{j-1}, u^j[$, $a = u^0 < u^1 < \ldots < u^m = b$, with each interval I^j satisfying

$$|I^j| := u^j - u^{j-1} < \delta. \tag{5}$$

We then select a representative $x^j \in \mathrm{Cl} I^j$; that is, $u^{j-1} \leqslant x^j \leqslant u^j$, $1 \leqslant j \leqslant m$.

(For simplicity we are using superscript ^{j} instead of $^{(j)}$—for labelling, not exponentiation. The reason for not using subscript $_j$ is to keep such subscripts available to denote dimensions in multi-dimensional variables.)

Then the Riemann (or Riemann-Stieltjes) integral of f with respect to F exists, with $\int_a^b f(x)\,\mathrm{d}F = \alpha$, if, given any $\varepsilon > 0$, there exists a number $\delta > 0$ so that

$$\left| \sum_{j=1}^{m} f(x^j) F(I^j) - \alpha \right| < \varepsilon \tag{6}$$

for every such choice of x^j, I^j satisfying (5), $1 \leqslant j \leqslant m$.

If we could succeed in creating a theory of random variation along these lines, then we could reasonably declare that the expectation $E^F(f)$ of the observable $f(x)$

relative to the distribution function $F(I)$, is $\int_a^b f(x)\,\mathrm{d}F$ whenever the latter exists in the sense of (6). (In fact this statement is true, but a justification of it takes us deep into the Kolmogorov theory of probability and random variation. A different justification is given in this paper.)

But (5) and (6) on their own do not yield an adequate theory of random variation. For one thing, it is well known that not every Lebesgue integrable function is Riemann integrable. So in this sense at least, (2) goes further than (1) and (6).

More importantly, any theory of random variation must contain results such as Central Limit Theorems and Laws of Large Numbers, which are the core of our understanding of random variation, and the proofs of such results require theorems like the Dominated Convergence Theorem, which are available for (2) and Lebesgue integrals, but which are not available for the ordinary Riemann integrals of (1) and (6).

However, before we take further steps towards the generalized Riemann version of (6) which gives us what we need, let us pause to give further consideration to data classification.

6. Aspects of data classification

Though the classes I^j used in (6) above are not required to be of equal length, it is certainly consistent with (6) to partition the sample data into equal classes. To see this, choose m so that $(b-a)/m < \delta$, and then choose each u^j so that $u^j - u^{j-1} = (b-a)/m$. Then $I^j = [u^{j-1}, u^j[$ $(1 \leqslant j \leqslant m)$ gives us a partition of $\Omega = [a, b[$ in which each I^j has the same length $(b-a)/m$.

We could also, in principle, obtain quantile classification of the data by this method of δ-partitioning. Suppose we want decile classification; that is, $[a, b[= I^1 \cup \ldots \cup I^m$ with $F(I^j) = 0.1$, $1 \leqslant j \leqslant m$, so $m = 10$. This is possible, since the function $F(u) := F([a, u[)$ is monotone increasing and continuous for almost all $u \in]a, b[$, and hence there exist u^j such that $F(u^j) = j/10$ for $1 \leqslant j \leqslant 10$. So if δ happens to be greater than $\max\{u^j - u^{j-1} \colon 1 \leqslant j \leqslant 10\}$, then the decile classification satisfies $|I^j| = u^j - u^{j-1} < \delta$ for $1 \leqslant j \leqslant 10$. (This argument merely establishes the existence of such a classification. Actually determining quantile points for a particular distribution function requires *ad hoc* consideration of the distribution function in question.)

In fact, this focus on the system of data classification is the avenue to a rigorous theory of random variation within a Riemann framework, as we shall now see.

In the previous sections we took the sample space to be $[a, b[$. Henceforth we will take the sample space Ω to be $\mathbb{R}$, or a multiple Cartesian product $\mathbb{R}^B$ of $\mathbb{R}$ by itself. There is no loss of generality in doing this, as we can, in effect, obtain "smaller" sample spaces, whenever they are required, by defining the distribution function so that it has zero support outside the "smaller" set. In Section 13 below on Brownian motion we show how to deal with a sample space which is not a Cartesian product though it is a subset of $\mathbb{R}^B$.

For the moment we take B to be a finite set with n elements. An interval I of $\mathbb{R}^B = \mathbb{R}^n$ is an n-times Cartesian product of real intervals of dimension one. For each elementary occurrence $x \in \Omega = \mathbb{R}^B = \mathbb{R}^n$, let $\delta(x)$ be a positive number. Then an admissible classification of the sample space, called a *δ-fine division* of Ω, is a finite collection

$$\mathcal{E}_\delta := \{(x^j, I^j)\}_{j=1}^m \tag{7}$$

so that $x^j \in \mathrm{Cl}I^j$, the I^j are disjoint with union Ω, and the lengths of the edges (or sides) of each I^j are bounded by $\delta(x^j)$, in the sense of (12) below.

So, referring back to Section 1 and the table (1) of elementary statistics, what we are doing here is selecting the data classification intervals I^j along with a representative value x^j from I^j. The pair (x^j, I^j) then describes a set x^j of measurements, $x^j = (x_1^j, x_2^j, \ldots, x_n^j)$, with the individual real measurements x_r^j jointly occurring in the real intervals I_r^j, $1 \leqslant r \leqslant n$.

It is convenient (though not a requirement of the theory) that the representative value x^j should be a vertex of I^j, and that is how we shall proceed.

The Riemann sum corresponding to (7) is

$$(\mathcal{E}_\delta) \sum f(x)F(I) := \sum_{j=1}^m f(x^j)F(I^j). \tag{8}$$

We say that f is *generalized Riemann integrable* with respect to F, with $\int_\Omega f(x)F(I) = \alpha$, if, for each $\varepsilon > 0$, there exists a function $\delta\colon\ \Omega \to]0, \infty[$ so that, for every $\mathcal{E}_\delta$,

$$\left|(\mathcal{E}_\delta) \sum f(x)F(I) - \alpha\right| < \varepsilon. \tag{9}$$

With this step we overcome the two previously mentioned objections to the use of Riemann-type integration in a theory of random variation. Firstly, every function f which is Lebesgue-Stieltjes integrable in Ω with respect to F is also generalized Riemann integrable, in the sense of (9). See [1] for a proof of this. Secondly, we have theorems such as the Dominated Convergence Theorem (see, for example, [1]) which enable us to prove Laws of Large Numbers, Central Limit Theorems and other results which are needed for a theory of random variation.

So we can legitimately use the usual language and notation of probability theory. Thus, the expectation of the observable $f(x)$ with respect to the probability distribution function $F(I)$ is

$$E^F(f) = \int_\Omega f(x)F(I).$$

To preserve consistency with the standard terminology of probability theory, it would be appropriate to designate $f(x)$ as a *random variable* only whenever $E^F(f)$ exists. We have assumed for the moment that B is a finite set. But whenever B consists of a single element, so $\Omega = \mathbb{R}$, the underlying variable $x \in \mathbb{R}$ may itself be a random variable in this sense, provided $E^F(x) := \int_{\mathbb{R}} xF(I)$ exists.

To recapitulate, elementary statistics involves calculations of the form (1) of Section 1, often with classes I of equal size, or classes of different sizes but equal likelihood. We refine this method by carefully selecting the data classification intervals I. In fact our Riemann sum estimates involve choosing a finite number of occurrences $\{x^{(1)}, \ldots, x^{(m)}\}$ from Ω (actually, from the closure of Ω), and then selecting associated classes $\{I^{(1)}, \ldots, I^{(m)}\}$, disjoint with union Ω, with $x^{(j)} \in \mathrm{Cl}I^{(j)}$ (or with each $x^{(j)}$ a vertex of $I^{(j)}$, in the version of the theory that we are presenting here), such that for each $1 \leqslant j \leqslant m$, $I^{(j)}$ is δ*-fine*. The meaning of this is as follows.

Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ be $\mathbb{R}$ with the points $-\infty$ and ∞ adjoined. This is what was meant by the closure in the preceding paragraph. In the following paragraph, $x = -\infty$ and $x = \infty$ are given special treatment. Many functions are undefined for $x = \pm\infty$; and if the integrand has points of singularity other than $\pm\infty$, we can make arrangements similar to the following ones.

Let I be an interval in $\mathbb{R}$, of the form

$$]-\infty, v[, \ \ [u, v[, \ \text{ or } [u, \infty[, \tag{10}$$

and let $\delta\colon \overline{\mathbb{R}} \to]0, \infty[$ be a positive function defined for $x \in \overline{\mathbb{R}}$. The function δ is called a *gauge* in $\mathbb{R}$. We say that I is *attached to* x (or *associated with* x) if

$$x = -\infty, \ x = u \text{ or } v, \ x = \infty \tag{11}$$

respectively. If I is attached to x we say that (x, I) is δ*-fine* (or simply that I is δ-fine) if

$$v < \delta(x), \ v - u < \delta(x), \ u > \frac{1}{\delta(x)} \tag{12}$$

respectively.

That is what we mean by δ-fineness in one dimension. What about higher dimensions? We consider next the case where B is finite with more than one element, so $\mathbb{R}^B = \mathbb{R}^n$ with $n \geqslant 2$. (The case of infinite B is considered in Section 11 below.)

Suppose $I = I_1 \times I_2 \times \ldots \times I_n$ is an interval of $\mathbb{R}^n$, each I_j being a one-dimensional interval of form (10). A point $x = (x_1, x_2, \ldots, x_n)$ of $\overline{\mathbb{R}}^n$ is attached to I in $\mathbb{R}^n$ if each x_j is attached to I_j in $\mathbb{R}$, $1 \leqslant j \leqslant n$. Given a function $\delta\colon \overline{\mathbb{R}}^n \mapsto]0,\infty[$, an associated pair (x, I) is δ-fine in $\mathbb{R}^n$ if each I_j satisfies the relevant condition in (12) with the new $\delta(x)$. A finite collection of associated (x, I) is a δ-fine division of $\mathbb{R}^n$ if the intervals I are disjoint with union $\mathbb{R}^n$, and if each of the (x, I) is δ-fine. A proof of the existence of such a δ-fine division is given in [1].

If X is a subset of the domain of integration $\mathbb{R}^B$, we sometimes need to give meaning to integrals such as $\int_X f(x)\,\mathrm{d}F$ or $\int_X f(x)F(I)$. In the Riemann theory, this is done by taking $\int_X f(x)F(I)$ to be $\int_{\mathbb{R}^B} f(x)\mathbf{1}_X(x)F(I)$, where $\mathbf{1}_X(x)$ is the characteristic function or indicator function of the set X in $\mathbb{R}^B$.

8. Where is the calculus of probabilities?

There are certain familiar landmarks in the study of probability theory and its offshoots. Such as the calculus of probabilities, which has not entered into the discussion thus far. The key point in this calculus is the relationship (4) above:

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P\left(A_j\right).$$

In fact the set-functions P and their calculus are not used as the basis of the generalized Riemann approach to the study of random variation. Instead, the basis is the simpler set-functions F, defined only on intervals, and finitely additive on them.

But, as mentioned earlier, a consequence of the generalized Riemann approach is that we can recover set-functions defined on sets (including the measurable sets of the Kolmogorov theory) which are more general than intervals, and we can recover the probability calculus which is associated with them.

To see this, suppose $A \subseteq \Omega$ is such that $E^F(\mathbf{1}_A)$ exists in the sense of (9), so the characteristic function or indicator function $\mathbf{1}_A(x)$, of the set A, is a random variable. Then define

$$P_F(A) := E^F(\mathbf{1}_A) = \int_\Omega \mathbf{1}_A(x)F(I), \tag{13}$$

and we can easily deduce from the Monotone Convergence Theorem for generalized Riemann integrals, that for disjoint A_j for which $P_F(A_j)$ exists,

$$P_F\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P_F(A_j).$$

Other familiar properties of the calculus of probabilities are easily deduced from (13).

Since every Lebesgue integrable function is also generalized Riemann integrable [1], every result obtained by Lebesgue integration is also valid for generalized Riemann integration. So in this sense, the generalized Riemann theory of random variation is an extension or generalization of the theory developed by Kolmogorov, Levy, Itô and others. (On the other hand, unlike the Lebesgue integral on which the classical Kolmogorov approach is based, the generalized Riemann integral is relatively undeveloped for spaces other than Cartesian products of the real numbers.)

However the kind of argument which is natural for Lebesgue integration is different from that which would naturally be used in generalized Riemann integration, so it is more productive in the latter case to develop the theory of random variation from first principles on Riemann lines. Some pointers to such a development are given in [6].

Many of the standard distributions (normal, exponential and others) are mathematically elementary, and the expected or average values of random variables, with respect to these distributions—whether computed by means of the generalized Riemann or Lebesgue methods—often reduce to Riemann or Riemann-Stieltjes integrals. Many aspects of these distributions can be discovered with ordinary Riemann integration. But it is their existence as generalized Riemann integrals, possessing properties such as the Dominated Convergence Theorem and Fubini's Theorem, that gives us access to a full-blown theory of random variation.

Here are some useful and convenient terms, which correspond with standard usages. If $E^F(\mathbf{1}_A) = 0$ we say that A is *null* (or F-null) in Ω, and its complement $\Omega \setminus A$ is *of full F-likelihood* in Ω.

9. Joint variation and marginal distributions

Random variation often involves joint variation of several (possibly infinitely many) variables or measurements. Thus our theory of random variation must enable us to analyze the properties of joint variation involving infinitely many random occurrences.

When a family of events $\{x_t\}_{t \in B}$ are being considered jointly, their marginal behavior is a primary consideration. This means examining the joint behavior of any finite subset of the variables, the remaining ones (whether finitely or infinitely many) being arbitrary or left out of consideration. Thus we are led to families

$$\{x_t\colon\ t \in N\}_{N \subseteq B}$$

where the sets N belong to the family $\mathcal{F}(B)$ of finite subsets of B, the set B being itself finite or infinite. When B is infinite the object $(x_t)_{t \in B}$ is often called a *process* or *stochastic process*, especially when the variable t represents time. For each t we will write the elementary occurrence x_t as $x(t)$ depending on the context; likewise $x_{t_j} = x(t_j) = x_j$.

Accordingly, for any finite subset $N = \{t_1, t_2, \ldots, t_n\} \subseteq B$, the marginal distribution function of the process $x = x_B = (x_t)_{t \in B}$ is the function

$$F_{(x_1, x_2, \ldots, x_n)}(I_1 \times I_2 \times \ldots \times I_n) \tag{14}$$

defined on the intervals $I_1 \times \ldots \times I_n$ of $\mathbb{R}^N$, which we interpret as the likelihood that the occurrence or measurement x_j takes a value in the one-dimensional interval I_j for each j, $1 \leqslant j \leqslant n$; with the remaining measurements x_t arbitrary for $t \in B \setminus N$.

One of the uses to which the marginal behavior is put is to determine the presence or absence of *independence*. The family of occurrences or measurements $\{x_t\}_{t \in B}$ is *independent* if the marginal distribution functions satisfy

$$F_{(x_1, x_2, \ldots, x_n)}(I_1 \times I_2 \times \ldots \times I_n) = F_{x_1}(I_1) \times F_{x_2}(I_2) \times \ldots F_{x_n}(I_n)$$

for every finite subset $N = \{t_1, \ldots, t_n\} \subseteq B$. That is, the likelihood that the measurements x_{t_1}, x_{t_2}, …, x_{t_n} jointly take values in I_1, I_2 …, I_n (with x_t arbitrary for $t \in B \setminus N$), is the product over $j = 1, 2, \ldots, n$ of the likelihoods of x_{t_j} belonging to I_j (with x_t arbitrary for $t \neq t_j$, $j = 1, 2, \ldots, n$) for every choice of such intervals, and for every choice of $N \in \mathcal{F}(B)$.

Of course, if B is itself finite, it is sufficient to consider only $N = B$ in order to establish whether or not the occurrences $\{x_t\}$ are independent.

10. Cylindrical intervals

When B is infinite (so $x = (x(t))_{t \in B}$ is a stochastic process), it is usual to define the distribution of x as the family of distribution functions

$$\left\{F_{(x(t_1), x(t_2), \ldots, x(t_n))}(I_1 \times I_2 \times \ldots \times I_n)\colon\ \{t_1, t_2, \ldots, t_n\} \subset B\right\} \tag{15}$$

This is somewhat awkward, since up to this point the likelihood function has been given as a single function defined on intervals of the sample space, and not as a family of functions. However we can tidy up this awkwardness as follows.

Firstly, the sample space Ω is now the Cartesian product $\prod_B \mathbb{R} = \mathbb{R}^B$. In the following discussion B can be finite or infinite, but if it is finite, the situation reduces to the earlier one of Section 7. With $\mathcal{F}(B)$ the family of finite subsets $N = \{t_1, t_2, \ldots, t_n\}$ of B, for any N the set

$$I = I[N] := I_{t_1} \times I_{t_2} \times \ldots \times I_{t_n} \times \prod\{\mathbb{R}:\ B \setminus N\} \tag{16}$$

is called a *cylindrical interval* if B is infinite. Taking all choices of $N \in \mathcal{F}(B)$ and all choices of one-dimensional intervals I_j $(t_j \in N)$, denote the resulting class of cylindrical intervals by $\mathcal{I}$. These cylindrical intervals are the subsets of the sample space that we need to define the distribution function F_x of x in $\mathbb{R}^B$:

$$F_x(I[N]) := F_{(x(t_1), x(t_2), \ldots, x(t_n))}(I_{t_1} \times I_{t_2} \times \ldots \times I_{t_n}) \tag{17}$$

for every $N \in \mathcal{F}(B)$ and every $I[N] \in \mathcal{I}$.

By thus defining the distribution function F_x (of the underlying process $x \in \mathbb{R}^B$) on the family of subsets $\mathcal{I}$ (the cylindrical intervals) of $\mathbb{R}^B$, we are in conformity with the system used for describing distribution functions in finite-dimensional sample spaces.

As in the elementary situation (1), it naturally follows, if we want to estimate the expected value of some deterministic function of the process $(x(t))_{t \in B}$, that the joint sample space $\Omega = \mathbb{R}^B$ of the family $\{x(t)\}$ of individual occurrences $x(t)$ should be partitioned by means of cylindrical intervals $I[N]$. In this case, an elementary occurrence $x \in \Omega$ consists of the joint occurrence $(x(t))_{t \in B} \in \mathbb{R}^B$, and we classify or codify all possible such occurrences x into a finite number of mutually exclusive classes, each of which has form $I = I[N]$. The N may be different for different classes or intervals I.

And, as in the finite-dimensional situation, the Riemann sum estimate of the expected value of some observable function f can be improved by "shrinking" (in some sense) the classes or intervals $I[N]$ which form the partition of the joint sample space. (Or, in another terminology, by refining the classification of the joint data.)

There are essentially two different ways in which this shrinking can be produced. Referring to (16), a subset of $I = I[N]$ can be obtained by choosing restricted intervals whose edges are smaller than the edges of I_j $(t_j \in N)$, or we can restrict the cylinder in extra dimensions—that is, choose $I = I[M]$ with $M \supset N$. Also we can combine these two modes of restricting or reducing the class of elementary occurrences.

Of course, if we are dealing with the joint variation of only a finite number of variables then B is a finite set, and in the "shrinking" described above we eventually get $N = B$ for all intervals $I[N]$, so we are back to the situation described in the previous sections.

11. A theory of joint variation of infinitely many variables

To formulate a theory of joint variation of infinitely many variables, we must establish what kind of partitions are to be permitted in forming the Riemann sum approximation to the expected value of an observable function which depends on these variables.

So we now address the formation of Riemann sums over a partition of an infinite-dimensional sample space. The observable to be averaged or integrated will be some deterministic function $f(x)$ of $x = (x(t))_{t \in B}$, corresponding to $f(x)$ in (1), Section 1. The averaging or integration of $f(x)$ will be with respect to some likelihood function, or probability distribution function $F(I[N])$ defined on the cylindrical intervals $I[N]$ from which a partition of $\Omega = \mathbb{R}^B$ is formed, just as the classes $I^{(j)}$ partition the one-dimensional domain of (1), Section 1.

The distribution function $F(I[N])$ is the likelihood or probability that $x \in I[N]$ (that is, $x(t_j) \in I_j$ for $t_j \in N = \{t_1, \ldots, t_n\}$ with $x(t)$ unrestricted for $t \in B \setminus N$). For an arbitrary partition $\mathcal{E}$ the Riemann sum estimate of the expected value of the observable f is

$$(\mathcal{E}) \sum f(x) F(I[N]).$$

Clearly, as we take different terms in this Riemann sum, we have different representative occurrences or processes x and different intervals or data classes $I = I[N]$, and the different $I[N]$ may have different sets N of restricted directions or dimensions.

In ordinary Riemann integration we form Riemann sums by choosing partitions whose finite-dimensional intervals have edges (sides or faces) which are bounded by a positive constant δ. Then we make δ successively smaller. Likewise for generalized Riemann integration, where the constant δ is replaced by a positive function $\delta(x)$. In any case, we are choosing successive partitions in which the component intervals successively shrink in some sense.

For the infinite-dimensional situation, we seek likewise to shrink the cylindrical intervals $I[N]$ of which successive partitions are composed.

Our earlier discussion provides us with the intuition we need to construct appropriate rules for forming partitions for Riemann sums in infinite-dimensional spaces. That is, the faces (or edges) of the restricted sides I_j of the cylindrical interval $I[N]$ (see (16) are reduced by requiring them to be bounded by some positive function δ, and the set N in which $I[N]$ has restricted faces is increased by requiring that N include some minimal set.

As before, let $\mathcal{F}(B)$ denote the family of finite subsets N of the (possibly infinite) set B. Let a typical $N \in \mathcal{F}(B)$ be denoted $\{t_1, t_2, \ldots, t_n\}$. The sample space is $\Omega = \mathbb{R}^B$. For $N \in \mathcal{F}(B)$, let $\mathbb{R}^N$ be the range of the projection

$$\mathcal{P}_N\colon\ (x(t))_{t\in B} \mapsto (x(t_1), \ldots, x(t_n))\,,\ \Omega \to \mathbb{R}^N\,.$$

Suppose $I_j \subset \mathbb{R}^{\{t_j\}}$ is an interval of type (10) (Section 7). Then $I_1 \times I_2 \times \ldots \times I_n \times \mathbb{R}^{B\setminus N}$ is a cylindrical interval, denoted $I[N]$; and $I[N] = \mathcal{P}_N^{-1}(I_1 \times \ldots \times I_n)$. As before, let $\mathcal{I}$ denote the class of cylindrical intervals obtained through all choices of $N \in \mathcal{F}(B)$, and all choices of intervals I_j of type (10), for each $t_j \in N$. A point $x \in \overline{\mathbb{R}}^N \times \mathbb{R}^{B\setminus N}$ is associated with a cylindrical interval $I[N]$ if, for each $t_j \in N$, the component $x_j = x(t_j)$ is associated with I_j in the sense of (11). A finite collection $\mathcal{E}$ of associated pairs $(x, I[N])$ is a *division* of $\mathbb{R}^B$ if the finite number of the cylindrical intervals $I[N]$ form a partition of $\mathbb{R}^B$; that is, if they are disjoint with union $\mathbb{R}^B$.

Now define functions δ and L as follows. Let $L\colon \overline{\mathbb{R}}^B \to \mathcal{F}(B)$, and for each $N \in \mathcal{F}(B)$ let $\delta\colon \overline{\mathbb{R}}^B \times \mathcal{F}(B) \to\ ]0, \infty[$. The mapping L is defined on the set of associated points of the cylindrical intervals $I[N] \in \mathcal{I}$; and the mapping δ is a function defined jointly on the set of pairs $(x, N) \in \overline{\mathbb{R}}^B \times \mathcal{F}(B)$.

The sets $L(x)$ and the numbers $\delta(x, N)$ determine the kinds of cylindrical intervals, partitioning the sample space, which we permit in forming Riemann sums.

A set $L(x) \in \mathcal{F}(B)$ determines a minimal set of restricted dimensions which must be possessed by any cylindrical interval $I[N]$ associated with x. In other words, we require that $N \supseteq L(x)$. The numbers $\delta(x, N)$ form the bounds on the lengths of the restricted faces of the cylindrical intervals $I[N]$ associated with x. Formally, the role of L and δ is as follows.

For any choice of L and any choice of δ, let γ denote (L, δ). We call γ a *gauge* in $\mathbb{R}^B$. The class of all gauges is obtained by varying the choices of the mappings L and δ.

Given a gauge γ, an associated pair $(x, I[N])$ is γ-*fine* provided $N \supseteq L(x)$, and provided, for each $t_j \in N$, (x_j, I_j) is δ-fine, satisfying the relevant condition in (12) (Section 7) with $\delta(x, N)$ in place of $\delta(x)$. A discussion of the partitioning of $\mathbb{R}^B$ for Riemann sums can be found in [2].

Given an observable function f of x, with a probability distribution function F defined on the cylindrical intervals $I[N]$ of $\mathcal{I}$, the integrand $f(x)F(I[N])$ is integrable in $\mathbb{R}^B$, with $\int_{\mathbb{R}^B} f(x)F(I[N]) = \alpha$, if, given $\varepsilon > 0$, there exists a gauge γ so that, for every γ-fine division $\mathcal{E}_\gamma$ of $\mathbb{R}^B$, the corresponding Riemann sum satisfies

$$\left|(\mathcal{E}_\gamma)\sum f(x)F(I[N]) - \alpha\right| < \varepsilon. \tag{18}$$

If B is finite, this definition reduces to definition (9), because, as each $L(x)$ increases, in this case it is not "without limit"; as eventually $L(x) = B$ for all x, and then (18) is equivalent to (9). Also (18) yields results such as Fubini's Theorem and the Dominated Convergence Theorem (see [5]) which are needed for the theory of joint variation of infinitely many random variables.

12. Random variables

We now extend the notion of a random variable, or observable, as follows. Let f be a deterministic function defined on $\mathbb{R}^B \times \mathcal{F}(B)$, so in definition (18) above, $f(x)$ is replaced by $f(x, N)$, with both x and N variable. The variables of integration in this case are x, N and $I[N]$, and the elements of N may appear explicitly in the integrand of (18).

The expectation of $f(x, N)$ is then $E^F(f) := \int_{\mathbb{R}^B} f(x, N)F(I[N])$ whenever the integral exists in the sense of (18), and f is a *random variable*, or F-*random variable*, or F-*observable*, whenever it has an expected value $E^F(f)$. If B happens to be finite, then this conception of random variable essentially reduces to the one expounded earlier.

But why extend the concept, and why extend it in this particular way? To see the motivation for this, we examine various different ways in which approximation or estimation of some quantity leads us to seek an expected value for such a quantity.

If a sample space Ω is $\mathbb{R}$, with elementary occurrences represented by $x \in \mathbb{R}$, then we conceive of an F-random variable as a deterministic real- or complex-valued function f of x for which $E^F(f)$ exists. In this case, the variable x may itself be a random variable, even though its primary role is to label, in a way which is amenable to mathematical analysis, the various non-mathematical outcomes or "states of the world"; as is done by the ω of an abstract measurable sample space Ω in classical probability theory.

For instance, in a single throw of a fair die, the possible physical outcomes, as the die comes to rest, may be represented by the integers $1, 2, 3, 4, 5, 6$. But the sample space may still be taken to be $\Omega = \mathbb{R}$ provided we take the distribution of likelihoods to be determined by, for instance, $F(]-\infty, r]) = r/6$, $0 \leqslant r \leqslant 6$. (Of course, there are many other valid ways in which this particular instance of random variation can be mathematically modelled, even if we restrict ourselves to the Riemann approach considered here.) By examining some Riemann sums, we see that x is itself a random variable, with $E^F(x) = \int_R xF(I) = 3.5$.

If we are investigating the random variation of n elementary occurrences, to be considered jointly, then the joint occurrence $x = (x_1, \ldots, x_n)$ can be taken to be an element of a sample space $\Omega = \mathbb{R}^n$. Again, random variables appear in the form of real- or complex-valued functions $f(x)$; for instance, $f(x) = x_1 + \ldots + x_n$.

If we have infinitely many elementary occurrences $\{x(t)\colon t \in B\}$ to be considered jointly, then a random variable is, again, $f(x)$ where $x = (x(t))_{t \in B} \in \mathbb{R}^B$. A joint occurrence x may sometimes belong to the subset C of those x in $\mathbb{R}^B$ which are continuous functions of t. And, with $B =]0,1]$ and V a continuous function of real numbers, the following function could conceivably be a random variable:

$$f(x) = \begin{cases} \exp\left(-\int_0^1 V(x(s))\,\mathrm{d}s\right) & \text{if } x \in C, \\ 0 & \text{if } x \in \mathbb{R}^B \setminus C. \end{cases} \tag{19}$$

Recall that a distribution function $F(I[N])$ for $\Omega = \mathbb{R}^B$, with B infinite, depends on "viewing" the process x at the "instants" $t_1, \ldots, t_n$ of $N \in \mathcal{F}(B)$. The function (21) below illustrates the explicit appearance of variables $t_1, \ldots, t_n$ in a probability distribution function.

So it is not unnatural that a random variable f might also depend on corresponding "views" of the process x. Thus, following the above example, we might have a random variable

$$f(x, N) = \exp\left(-\sum_{j=1}^{n} V(x(t_j))(t_j - t_{j-1})\right). \tag{20}$$

In fact, random variables often appear in the form $f(x(t_1), \ldots, x(t_n))$ where $N = \{t_1, \ldots, t_n\}$ are the variable "instants" at which both the probability distribution function and the random variable "view" the process.

Bearing in mind that $f(x)$ is fundamentally an estimated or approximated measurement, it might be reasonable in certain circumstances to regard (20) as an equally valid way of estimating the underlying quantity which is also estimated by (19). In fact the discrete version in (20) may, in practice, be the only way in which the underlying quantity can be estimated. A discussion of how (19) and (20) might relate to each other is given in [5].

By designating a random variable as $f(x, N)$ we include random variables of the forms $f(x)$ and $f(x(t_1), \ldots, x(t_n))$, as well as other possible representations and formulations of the measured or approximated quantity.

We now illustrate the Riemann approach to the analysis of random variation by giving a new construction of Brownian motion. (This means constructing a mathematical model or theory which closely represents the more important properties observed in the physical phenomenon itself.)

Let $B =]0, \infty[$. A *Brownian motion* is a random variable $x = (x(t))_{t \in B}$ such that

1. $x(0) = 0$;
2. Each of the random variables $x(t) - x(s)$ $(t \in B,\ s \in B,\ t > s)$ is normally distributed with mean 0 and variance $t - s$;
3. The family of random variables $x(t) - x(s)$ $(t \in B,\ s \in B,\ t > s)$ is independent; and
4. The sample space for joint occurrences x is $\Omega = C$, the subset of continuous functions in $\mathbb{R}^B$.

To construct a process x satisfying the first three of these conditions, we define the following function on the family $\mathcal{I}$ of cylindrical intervals $I[N]$ of $\mathbb{R}^B$. With $N = \{t_1, \ldots, t_n\}$, $0 < t_1 < \ldots < t_n$, and taking t_0 to be 0, refer to (3) in Section 3 above and, with $y_0 = 0$, define $g(I_1 \times \ldots \times I_n)$ to be

$$\prod_{j=1}^{n} (2\pi(t_j - t_{j-1}))^{-\frac{1}{2}} \int_{I_1} \cdots \int_{I_n} \exp\Big(-\frac{1}{2} \frac{(y_j - y_{j-1})^2}{t_j - t_{j-1}} \Big) \, \mathrm{d}y_1 \ldots \mathrm{d}y_n, \tag{21}$$

and then define $G(I[N]) := g(I_1 \times \ldots \times I_n)$ for each $N \in \mathcal{F}(B)$ and each cylindrical interval.

A cylindrical interval can be represented in various ways. For instance, with $N = \{t_1, \ldots, t_n\}$, $N' = N \cup \{t_{n+1}\}$, $I[N] = I_1 \times \ldots \times I_n \times \mathbb{R}^{B \setminus N}$ and $I[N'] = I_1 \times \ldots \times I_n \times \mathbb{R} \times \mathbb{R}^{B \setminus N'}$, then $I[N] = I[N']$. The argument of Proposition 36 of [5] shows that $G(I[N]) = G(I[N'])$; and it can be easily adapted to show that, in general, $G(I[N])$ is well-defined. The fact that $G(I[N])$ is a distribution function, with $\int_{\mathbb{R}^B} f(x, N)G(I[N]) = 1$ whenever f is identically 1, so $E^G(1) = 1$, also follows from the evaluation in Proposition 36 of [5]. These results correspond to the Daniell-Kolmogorov Theorem of the classical theory. They are practically self-evident, because cylindrical intervals and their distribution functions are less complicated than measurable sets and their probability measures.

In order to satisfy condition (4) of Brownian motion, we might aspire to a definition of the expectation of a random variable $f(x)$ $(x \in C)$ as, in some sense, $\int_C f(x) \, \mathrm{d}G$; which, in turn, might be approximated by Riemann sums $\sum f(x)G(I_C[N])$, where each $I_C[N] = I[N] \cap C$ is one of a finite number of sets partitioning C. In fact if X is a subset of $\mathbb{R}^B$ (B finite or infinite), the standard way of defining an expression such as

$\int_X f(x)\,\mathrm{d}G$ by means of generalized Riemann integration is $\int_{\mathbb{R}^B} \mathbf{1}_X(x)f(x)G(I[N])$. But C is a non-measurable (in the classical sense) subset of $\mathbb{R}^B$ (see [3]), and $\int_{\mathbb{R}^B} \mathbf{1}_C(x)f(x)G(I[N])$ does not generally exist.

In effect, even though C is not a "flat", G-null subset of $\mathbb{R}^B$, in the way that $\mathbb{R}^2$ is a "flat" projection of the points of $\mathbb{R}^3$, C is nonetheless too small a subset of $\mathbb{R}^B$. For a discussion of this point, see [3].

The problem is, that in the Riemann sums $\sum \mathbf{1}_C(x)f(x)G(I[N])$, which might be expected to yield close approximations to $\int_{\mathbb{R}^B} \mathbf{1}_C(x)f(x)G(I[N])$ if the integral actually existed, too many terms of the Riemann sum are removed by the factor $\mathbf{1}_C(x)$.

To satisfy (4) we must find some way round this obstacle. The Riemann solution to the problem uses the same feature of Brownian motion that the classical solutions use. But while the latter focusses essentially on a suitable modification of the function G, the approach presented here looks to a modification of the random variable f.

Let us recall the standpoint from which we have chosen to view the problem of Brownian motion. We have to perform some calculation or measurement f which depends on the unpredictable course of a quantity $x(t)$ whose values are continuous with respect to t; and, assuming that the increments $x(t)-x(s)$ are independent and normally distributed with variance $t-s$, we seek to determine the expected value of f. Our difficulty, as in the classical treatment of the subject, is that the theory thus far has led us to a sample space $\Omega = \mathbb{R}^B$, in which our calculation or measurement f is undefined or meaningless outside of the subset C.

However, if M is any fixed, finite subset of B, and if $C(M)$ denotes the set of $x \in \mathbb{R}^B$ which are continuous at each $t \in M$, then, by Proposition 46 of [5], the integral $\int_{\mathbb{R}^B} \mathbf{1}_{C(M)}(x)G(I[N])$ exists and equal 1, for each M. So the expected value of the random variable $\mathbf{1}_{C(M)}$ is 1; $E^G(\mathbf{1}_{C(M)}) = 1$. In other words, $\mathbb{R}^B \setminus C(M)$ is G-null for every fixed, finite $M \subset B$. (The gist of the argument is as follows. In the Riemann sums, the variances $t_{j+1} - t_j$ of the normally distributed $x(t_{j+1}) - x(t_j)$ become arbitrarily small. A discontinuity in x at any one of the $\tau = t_j \in M$ then makes the normal increment $x(t_{j+1}) - x(t_j)$ arbitrarily large compared to its variance $t_{j+1} - t_j$. And the normal distribution then places an arbitrarily small common multiplicative factor in each of the likelihood or G-terms of the Riemann sum approximation to $\int_{\mathbb{R}^B} \mathbf{1}_{C(M)}(x)G(I[N])$.)

From this, it is easy to deduce that the random variable $\mathbf{1}_{C(N)}$, with N variable, has expectation 1; so $\int_{\mathbb{R}^B} \mathbf{1}_{C(N)}(x)G(I[N]) = 1$. Here, the Brownian distribution function $G(I[N])$ "views" the process x at the variable instants $t \in N$; and at each such view, the likelihood G "sees" those x which are continuous at $t \in N$, but which may be discontinuous at any $t \in B \setminus N$. The latter terms, of which there are a vast multitude, are the ones which would have disappeared from the Riemann sum if the factor $\mathbf{1}_C(x)$ had been used instead of $\mathbf{1}_{C(N)}(x)$.

Armed with this insight, we demonstrate a generalized Riemann version of a continuous modification. That is, we show how to meaningfully establish the G-expectation of a measurement or calculation f which is determined only by unpredictable occurrences $x(t)$ which are continuous at each t. In fact we demonstrate the modification when f is determined only by the jagged paths which are commonly used in diagrams to illustrate Brownian motion. (It is easy to adapt the argument for other classes of continuous paths x.)

Suppose Y is the set of polygonal paths in $C \subset R^B$, so for each $N = \{t_1, t_2, \ldots t_n\} \in \mathcal{F}(B)$, $y = y_N \in Y$ satisfies

$$(22) \qquad \begin{aligned} & y(t_j) \in \mathbb{R},\ 1 \leqslant j \leqslant n; \quad \text{and} \\ & y(t) = y(t_{j-1}) + \frac{t - t_{j-1}}{t_j - t_{j-1}} \left(y(t_j) - y(t_{j-1})\right) \quad \text{for} \quad t \in [t_{j-1}, t_j[. \end{aligned}$$

Suppose a function $f(y)$ has a value, real or complex, for each elementary occurrence $y \in Y$ but is otherwise undefined; so the sample space Y is a proper subset of $\mathbb{R}^B$ and is not itself a Cartesian product space. For any $x \in \mathbb{R}^B$ and any $N \in \mathcal{F}(B)$, choose $y_N \in Y$ so that $y_N(N) = x(N)$, with $y_N(t)$ given by (22) for $t \in B \setminus N$. Now define f_Y on $\mathbb{R}^B \times \mathcal{F}(B)$ by

$$(23) \qquad f_Y(x, N) := \begin{cases} f(y_N) & \text{if } x \in C(N), \\ 0 & \text{otherwise.} \end{cases}$$

We define the expectation $E_Y^G(f)$, or "$\int_Y f(y_N)\,\mathrm{d}G$", by

$$(24) \qquad E_Y^G(f) := \int_{\mathbb{R}^B} f_Y(x, N)\mathbf{1}_{C(N)}(x) G(I[N]).$$

Whenever the latter exists we say that f is a random variable with sample space Y. (The factor $\mathbf{1}_{C(N)}(x)$ in the integrand of (24) is redundant, but is inserted as an aid to intuition.) Thus $E_Y^G(f)$ is defined to be $E^G(f_Y)$ whenever the latter exists.

A "random variable" which takes a constant value c with certainty ought to have an expected value of c. Accordingly, if $f(y) = c$ for all $y \in Y$, the earlier discussion shows that

$$E_Y^G(f) = \int_{\mathbb{R}^B} c\mathbf{1}_{C(N)}(x) G(I[N]) = c;$$

and we can use theorems such as the Dominated Convergence Theorem in $\mathbb{R}^B$ to deduce that "well-behaved" functions f are likewise random variables in Y, in this modified sense. Thus, with the modifications (23) and (24), the function G can be interpreted as a probability distribution function on the sample space Y.

Given any $N = \{t_1, \ldots, t_n\}$ of $\mathcal{F}(B)$ and any $x \in C(N)$, the set Y is big enough to enable us to find $y \in Y$ such that $y_N(t_1, \ldots, t_n) = x(t_1, \ldots, t_n)$, but it is not big enough to contain a set of full G-likelihood in $\mathbb{R}^B$. That is why a random variable f which is defined only on Y must be adjusted by means of (23) in order to admit extra terms into the Riemann sum approximation of the expected value of f_Y in (24), thereby enabling us to satisfy condition (4) of Brownian motion.

References

[1] *Gordon, R.*: The Integrals of Lebesgue, Denjoy, Perron, and Henstock. American Mathematical Society, 1994.

[2] *Henstock, R., Muldowney, P., Skvortsov, V. A.*: Partitioning infinite-dimensional spaces for generalized Riemann integration. To appear in Bull. London Math. Soc.

[3] *Karatzas, I., Shreve, S. E.*: Brownian Motion and Stochastic Calculus. Springer, Berlin, 1988.

[4] *Kolmogorov, A. N.*: Foundations of the Theory of Probability, 1933; English translation: Chelsea, New York, 1950.

[5] *Muldowney, P.*: A General Theory of Integration in Function Spaces, Including Wiener and Feynman Integration. Pitman Research Notes in Mathematics no. 153, Harlow, 1987.

[6] *Muldowney, P.*: Topics in probability using generalised Riemann integration. Math. Proc. R. Ir. Acad. *99(A)1* (1999), 39–50.

[7] *Muldowney, P.*: The infinite dimensional Henstock integral and problems of Black-Scholes expectation. J. Appl. Anal. *8* (2002), 1–21.

[8] *Muldowney, P., Skvortsov, V. A.*: Lebesgue integrability implies generalized Riemann integrability in $\mathbf{R}^{[0,1]}$. Real Anal. Exch. *27* (2001/2002), 223–234.

Author's address: *P. Muldowney*, Magee College, University of Ulster, Derry, N. Ireland, e-mail: `p.muldowney@ulster.ac.uk`.