Chapter 1

Prologue

1.1 About This Book

This is a self-contained study of a Riemann sum approach to the theory of random variation, assuming only some familiarity with probability or statistical analysis, basic Riemann integration, and mathematical proof. The primary idea of the book, and the reason why it is different from other treatments of random variation, is its use of non-absolute convergence. The series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$ diverges to infinity. On the other hand, the oscillating series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ converges—but only on condition that the terms are added up in the order in which they are written, without rearranging them. This convergence is called *conditional* or *non-absolute*.

What has this got to do with the theory of random variation? Any conception or understanding of the random variation phenomenon hinges on the notions of probability and its mathematical representation in the form of probability distribution functions. The central, recurring theme of this book is that, provided a non-absolute method of summation is used, **every** finitely additive function of disjoint intervals is integrable. In other words, every distribution function is integrable.

In contrast, more traditional methods in probability theory **exclude** significant classes of such functions whose integrability cannot be established whenever only absolute convergence is considered. Examples of this include:

- The Feynman "probability measure" (which is not a measure and not a probability)—the *probability amplitudes* used in the Feynman path integrals of quantum mechanics. This book presents a framework in which the Feynman path integrals are actual integrals. In effect, the missing pieces of Feynman's original paper [64] are provided here; and then used to express *Feynman diagrams* as convergent series of integrals—as they were originally conceived.

- The increments in the sample paths of Brownian motion—these have infinite variation in every interval, and their integrals (in the usual absolute

A Modern Theory of Random Variation: With Applications in Stochastic Calculus, Financial Mathematics, and Feynman Integration. First Edition. By Pat Muldowney

sense) are therefore divergent. But these increments are integrable in the non-absolute sense, so the *stochastic calculus* of Brownian motion can be put on a simpler footing.

Incorporating these innovations in the theory of random variation entails a radical reformulation of the subject. It turns out that the standard theory of probability or random variation can be simplified and extended provided non-absolute summation procedures are used.

Reformulation and extension of the theory involves some changes and re-interpretations in the standard concepts and notations. Unnecessary changes have been avoided, and as far as possible the text is consistent with more traditional versions. Therefore, with due caution and attention to definitions of terminology and notation, the text can be read in that spirit. An outline and overview are presented in Chapters 1 and 2.

Chapter 7 is the main part of this book, with Chapter 6 providing introductory material, and Chapter 8 some consequences. The book presents a new sphere of application of probability theory by means of the conception of random variation which is elaborated in Chapter 5.

Ralph Henstock's general theory of integration, as extended in [162] (Muldowney, 1987), is the basis for this reformulation of the traditional theory of probability and random variation, and is presented in Chapter 4.

Even though Henstock's theory is different from standard integration theory, many of the results are similar. Therefore Chapter 4 can be regarded as a kind of appendix to subsequent chapters, providing technical background in the manner of many books on probability theory in which measure and integration are appended to the main part of the text. Included in this chapter are results for non-absolutely integrable functions which are not available in traditional integration theory.

A fundamental modification and extension of the Riemann integral was introduced by R. Henstock and, independently, by J. Kurzweil in the 1950s. In Henstock [93] this was designated as the Riemann-complete[1] integral.

The work of Kurzweil has transformed the theory of differential equations—see, for instance, Schwabik [129, 207]. Henstock went on to develop a general theory of integration [85, 93, 94, 103, 105], which includes as special cases the integrals of Riemann, Stieltjes, Lebesgue, Perron, Denjoy, Ward, Burkill, Henstock-Kurzweil, and McShane (see [82]). This is the Henstock integral on which this book is based.

The Henstock integral is not so well known as the Lebesgue integral. Also, the Riemann sum approach to probability theory is new. Therefore the main ideas of this book are introduced in a relatively informal way in Chapters 1 and 2, while Chapter 3 brings forward some notation and definitions from Chapter 4, in advance of the fuller exposition of the main theorems and proofs in the theory of the integral—the Burkill-complete integral—provided in Chapter 4.

[1]This is the Henstock-Kurzweil or HK-integral, also known as the generalized Riemann integral, the Kurzweil integral, the Henstock integral, and the gauge integral.

Chapter 4 can be read as a stand-alone account of the Stieltjes-complete and Burkill-complete versions of the Henstock integral, with emphasis on those parts of integration theory which are important in the study of random variation.

It is possible to get the gist of this book by reading Chapters 1, 2, and 3 in conjunction with Chapter 9's numerical exploration of observable processes, stochastic processes, Brownian motion, and Itô calculus.

The book contains a new approach to several topics. There have to be good reasons for going to the trouble of engaging with a new approach to subjects for some of which there already exist tried and tested methods. As the occasion arises such reasons are pointed out in the text.

Much detail is provided in exposition, explanation, commentary, and proof; with a view to transparency and, not least, facilitation of error detection, error correction, and the like. A degree of repetition is present, for the same purposes.

The text contains examples which illustrate the material of the text with solutions to less difficult issues. They can be regarded as exercises or solved problems and can be used as models for devising further exercises and problems. The numerical calculations in Chapter 9 are intended to illustrate notation and to clarify concepts. Also, as a rich source of insight, motivation, and grounding, there is endless scope for further practical numerical exercises of this kind.

The book builds on the work of numerous authors, many of whom are listed in the text and in the bibliography. The generous help of many colleagues in bringing the material to publication is gratefully acknowledged.

1.2 About the Concepts

An *integrand* generally involves a point function $f(x)$ multiplied by an *integrator*[2] function $F(I)$. Many treatises on integration focus strongly on the properties of $f(x)$, such as continuity and differentiability, or their absence. In mathematical analysis the integrator is often taken to be $F(I) = |I|$, the length of the interval I, with less attention given to alternative integrator functions.

But random variation is not so much concerned with the more difficult manifestations of point function integrands $f(x)$. In this book much more emphasis is placed on properties of probability distribution functions $F(I)$. This is one of the reasons why the book gives much attention to the properties of *variation*[3] of interval functions $F(I)$, a concept which it possible to extend distribution functions F defined on intervals to outer measure defined on arbitrary sets.

In addition, the classical form of an integrand function is a product $f(x)F(I)$ of a point function multiplied by an interval function. But it turns out that Henstock integration is most naturally formulated with integrands of the form

[2]If the integrator $F(I)$ is additive, but is not the length function $|I|$, then the integral is *Stieltjes*-type. If the integrator is a non-additive function $h(I)$, then the integral is *Burkill*-type.

[3]The word *variation* (referring to the ranges of actual values taken by a deterministic function) also occurs in *random variation* (referring to uncertainty in potential data values arising in experiments). To prevent confusion, *random variability* is substituted for the latter.

$h(x, I)$ which are not necessarily the product of a point function times an interval function.

The wording and symbols used in the theory of random variation, as presented in this book, are consistent with or similar to those already in general use and, for the most part, can be understood in the usual way. A note of caution, however. The symbol X is traditionally used to denote a random variable, in the sense of a measurable function. But in this book X denotes a mathematical representation of an "experiment" for which a range of potential data values x is known in advance. And a random variable is a calculation $f(X)$ based on the potential data values x. The symbol X will usually denote a process of joint observation of several unpredictable occurrences. The occurrences or outcomes are actual joint data x, where x is a 'tuple of real numbers such as the observed values of an experiment consisting of repeated throws of a die.

A determination $f(X)$ derived from this experiment or joint observation X could consist of the value of a payout made on the first occasion when ten successive sixes are thrown. X can be thought of as an experiment, an "observable", or a "random variation".

Both X and $f(X)$ involve potential data, x and $f(x)$, respectively, generated by an act of measurement—often joint measurement. Thus X (or $f(X)$) refers to unpredictable potential data, in advance of actual observation. The corresponding x (or $f(x)$) is the actual datum selected by the process of measurement or observation—in other words, the observed value or occurrence.

It is therefore helpful to think of X as the experiment, observation or measuring process which selects a datum x. Similarly, $f(X)$ represents potential data, in advance of actual measurement (or observation or determination), and in advance of calculation of a datum $f(x)$. We can think of $f(X)$ as consisting of *potential data* in association with their *potentialities of occurrence*, the latter consisting of *likelihood* that an actual datum x will belong to any set I of potential values in a *sample space* Ω_X. Such likelihood or *probability* will be denoted by $F_X(I)$; and it can be thought of as the *accuracy potential* of the observation $f(X)$.

There is a "before" and "after" aspect to this. There is unpredictability or uncertainty before observation, but not after. Therefore part of the meaning of x, X (or $f(x)$, $f(X)$) is dependent on the point in time at which they are being considered.

In advance of determination, by measurement or observation, of a datum x (or $f(x)$), we speak of a random variable or observable X (or $f(X)$); by which is meant the potential data values that may be observed, subject to some measure F_X of their potentialities or likelihoods or accuracy.

In light of these various considerations, the expression $f(X)$, as used in this book, is abbreviation for a notation involving several components:

- $f(x)$ represents any deterministic calculation involving a data-value x observed in the experiment;
- Ω_X represents the sample space, or domain of potentially observable data-values x; and

- F_X represents a distribution of probabilities (accuracy potentialities or likelihoods) to which potentially observable data-values x are subject.

Thus, an observable or random variable $f(X)$ is denoted by a triple

$$f(x)\,[\Omega_X, F_X]\,;$$

and $f(X)$ can be thought of as potential data $f(x)$ in association with their potentialities or likelihoods $F_X(I)$, the latter being the likelihood that the datum (or joint datum) x will, in advance of actual observation, belong to any set I of potential data-values. The function F_X enables us to quantify the accuracy potential or degree of unpredictability of prior estimates of datum $f(x)$, in advance of actual measurement.

Since X represents joint observations (possibly infinitely many), stochastic processes are subsumed within a general theory of joint variation.

An important class of stochastic processes, including Brownian motion, is defined by the properties of the increments of the process at successive instants of time. In the case of Brownian motion, almost all infinite series of successive increments diverge absolutely, but all such series are conditionally (or non-absolutely) convergent. Since the method of summation (or integration) used in this book is non-absolute, the stochastic calculus of Brownian motion is significantly simplified.

Other features of this study of random variation may also appear strange, initially. For example, random variables are defined here in such a way that measurability of the variables is a consequence, and not a pre-condition, of the definition. Another unfamiliar aspect of this presentation is that the calculus of probabilities, in the usual sense, is not fundamental to it. Instead the basic properties of probability are deduced (see Section 5.12) as a consequence of the meaning ascribed to random variables. And, in place of probability-measure functions defined on measurable subsets of a sample space, the more fundamental role is taken by distribution functions defined, not on measurable sets, but on intervals.

When these distribution functions are assumed to take only non-negative values the resulting theory is equivalent to the classical or axiomatic theory of probability and random variation. But when they are allowed to take complex values, a significant extension of the classical theory emerges. Of course, the notion that "probability" can manifest itself in anything other than non-negative real values is a conceptual challenge; one that is addressed and rationalized at various points in the book—in Section 2.16, for instance.

These amendments to the classical formulation of probability theory make it possible to bring the Feynman theory of the path integrals of quantum mechanics within the scope of a theory of random variation; and they simplify the theory of stochastic calculus. Also, proofs in the basic theory of probability are simplified. This is because, instead of P-measurable sets A of a probability space $(\Omega, \mathcal{A}, P)$, probabilities are estimated with finitely additive functions $F_X(I)$ of intervals I.

It is not the purpose of the book to give exhaustive or in-depth treatments of the various themes. Instead, it dwells on the relative simplicity, power, and

versatility of the Riemann sum approach to these subjects. Once the method is grasped, it is relatively easy to work out any missing elements.

1.3 About the Notation

Notation for random variables has already been mentioned. Another important issue is notation for integrals. The *integral* concept implies the following elements.

- **Domain of integration**, for example,

$$[0,1], \qquad [a,b], \qquad [a,b]^n, \qquad \mathbf{R}, \qquad \mathbf{R}\times\mathbf{R}\times\cdots.$$

  Traditionally, this is variously written

$$\int_0^1, \quad \int_{[0,1]}, \quad \int_a^b, \quad \int_{[a,b]}, \quad \int_{[a,b]^n}, \quad \int_{\mathbf{R}}, \quad \int_{\mathbf{R}\times\mathbf{R}\times\cdots}$$

  and so on.

- **Expression to be integrated** (or **integrand**), usually involving points x and intervals I in the domain of integration. So an integrand could have the form $f(x)|I|$ where $x \in I$ and the *integrator* $|I|$ is the length of the interval I in one dimension. If I is two-dimensional, then integrator $|I|$ denotes area of I. Integrals involving $|I|$ come under the heading of Riemann integration. The integrand can also have Stieltjes form $f(x)F(I)$ where the integrator function F is some additive function defined on intervals of the domain. Or the integrand could be a function $f(x)h(I)$ where the integrator h is not additive. (With $f(x)$ identically 1, the integral of *non-additive* $h(I)$ is known as the *Burkill integral*—see [25, 26, 103, 202].) Or the integrand could be a joint function $h(x, I)$ of points and intervals, a formulation which includes the Riemann, Stieltjes and Burkill integrands. This suggests a notation of the form $\int_{\mathbf{R}} h(x, I)$. Sometimes an integrand $h(x, I)$ may, in addition to dependence on x and I, also depend on other point and/or interval parameters y and J, say; giving a function $h(x, y, I, J)$. In that case the notation $\int h(x, y, I, J)$ can be an ambiguous notation for the integral. Which of the parameters are "integrated on"? Which remain fixed[4] in the integration? If such ambiguity arises it is removed by notation of the form

$$\int_{I\in\mathbf{I}(\mathbf{R})}^{x\in I^*} h(x, y, I, J).$$

[4] In [162] (Muldowney, 1987), elements additional to x and I are introduced as Riemann sum variables. For the purpose of analyzing random variation, an additional variable of this kind is introduced in Chapter 4.

The meaning of the various parts of this notation is fairly obvious, and precise meanings will be given later. But the integral it denotes has a meaning different from the following integral:

$$\int_{J\in\mathbf{I}(\mathbf{R})}^{y\in I^*} h(x,y,I,J),$$

in which y and J are "integrated on", while the parameters x and I are held constant in the integration. If the integrand $h(x,I)$ is a point function $f(x)$ multiplied by an interval function $F(I)$, then the integral of the product $f(x)F(I)$ may be denoted

$$\int f(x)F(I) \quad \text{or} \quad \int f(x)dF.$$

- **Riemann sums** $\sum f(x)|I|$ which approximate to the integral $\int f(x)|I|$. Thus, if the domain of integration is the real interval J, the integral

$$\int_J f(x)|I|, \quad = \int_{I\in\mathbf{I}(J)}^{x\in I^*} f(x)|I|$$

may be estimated or approximated by Riemann sums

$$\sum f(x)|I| = \sum \{f(x)|I| : I \in \mathcal{P}\},$$

where $\mathcal{P} = \{I\}$ denotes a partition of the interval J, and, for each $I \in \mathcal{P}$, the evaluation point x is contained in I or the closure of I.

Riemann sums are the prevailing theme of this book, and a shorter notation on the lines of the following is used throughout:

$$(\mathcal{P})\sum f(x)|I|, \quad \text{representing} \quad \sum \{f(x)|I| : I \in \mathcal{P}\}.$$

Occasionally the expression *integral of* $f(x)$ is used without reference to any integrator, weighting function, interval function, or measure. In this case the integral should be understood in the traditional way. In other words, *integral of* $f(x)$ should be understood as $\int f(x)dx$, $\int_a^b f(x)dx$, or the like, depending on the context. Formally, the integral of $f(x)$ on $[a,b]$ is

$$\int_a^b f(x)dx, \quad = \int_{I\in\mathbf{I}([a,b])}^{x\in I^*} f(x)|I|.$$

A glossary of symbols is provided in pages xiii–xvi.

1.4 Riemann, Stieltjes, and Burkill Integrals

This section demonstrates simple Riemann sum calculations of Riemann, Stieltjes, and Burkill integrals.

Consider $\int_0^1 f(x)dx$ with $f(x) = 4x^3$. In basic calculus it is observed that $4x^3$ has primitive (or anti-derivative) x^4, and the indefinite integral is $F(x) = x^4 + c$ where c is any constant. Thus basic calculus gives definite integral

$$\int_0^1 f(x)dx = [F(x)]_0^1 = (1+c) - (0+c) = 1. \tag{1.1}$$

If $F(x)$ is written in its incremental or Stieltjes form $F(\,]u,v]) = F(v) - F(u)$, this becomes

$$\int_0^1 f(x)dx = F(\,]0,1]) = F(1) - F(0) = 1. \tag{1.2}$$

This is the **calculus integral**, also called the **Newton integral**.

Example 1 *To evaluate this integral by Riemann sums, then, with benefit of the preceding calculation (1.1), take 1 as the candidate value*[5] *of the Riemann integral, and consider expressions*

$$\left|1 - (\mathcal{P})\sum 4x^3|I|\right|, \;= \left|1 - \sum_{r=1}^{n} 4x_r^3|I_r|\right|,$$

where $\mathcal{P} = \{I_r\}$ *is a partition of* $]0,1]$ *with*

$$I_r = [u_{r-1}, u_r[, \qquad u_{r-1} \le x_r \le u_r, \qquad |I_r| = u_r - u_{r-1},$$

for $r = 1, 2, \ldots, n$, $u_0 = 0$, $u_n = 1$. *Let* $\varepsilon > 0$ *be given. By uniform continuity of the function* $4x^3$ *in* $[0,1]$, *there exists* $\delta > 0$ *so that, for any interval* $I =]u,v] \subset]0,1]$ *satisfying* $|I| = v - u < \delta$, *and for any* x, y *satisfying* $u \le x \le v$, $u \le y \le v$, *then*

$$\left|4x^3 - 4y^2\right| < \varepsilon.$$

Choose a partition $\mathcal{P} = \{I_r\}_{r=1}^n = \{]u_{r-1}, u_r]\}_{r=1}^n$ *satisfying*

$$|I_r| = u_r - u_{r-1} < \delta$$

for $1 \le r \le n$. *Then, by the mean value theorem, for each* r *there exists* y_r *satisfying* $u_{r-1} < y_r < u_r$ *with*

$$u_r^4 - u_{r-1}^4 = 4y_r^3\left(u_r - u_{r-1}\right).$$

Taking the Riemann sum over the partition $\mathcal{P}$, *we have*

$$1 = \sum_{r=1}^{n}\left(u_r^4 - u_{r-1}^4\right)$$

and

[5] A simple application of the triangle inequality (as in Theorem 8) shows that if a pair of such "candidates" satisfy the Riemann sum condition then they must be equal.

$$\begin{aligned}
\left|1-(\mathcal{P})\sum 4x_r^3|I_r|\right| &= \left|\sum_{r=1}^{n}\left(u_r^4-u_{r-1}^4\right)-4x_r^3(u_r-u_{r-1})\right| \\
&= \left|\sum_{r=1}^{n}\left(\left(u_r^4-u_{r-1}^4\right)-4y_r^3\left(u_r-u_{r-1}\right)\right)\right| \\
&\quad + \left|\sum_{r=1}^{n}\left(4y_r^3\left(u_r-u_{r-1}\right)-4x_r^3\left(u_r-u_{r-1}\right)\right)\right| \\
&= \left|\sum_{r=1}^{n}\left(4y_r^3\left(u_r-u_{r-1}\right)-4x_r^3\left(u_r-u_{r-1}\right)\right)\right| \\
&\leq \sum_{r=1}^{n}\left|4y_r^3\left(u_r-u_{r-1}\right)-4x_r^3\left(u_r-u_{r-1}\right)\right| \\
&= \sum_{r=1}^{n}\left|4y_r^3-4x_r^3\right|\left(u_r-u_{r-1}\right) \\
&< \varepsilon\sum_{r=1}^{n}\left(u_r-u_{r-1}\right)=\varepsilon.
\end{aligned}$$

This holds for every such partition $\mathcal{P}$, so

$$\int_0^1 4x^3dx=1, \quad \textit{or} \quad \int_{]0,1]}4x^3|I|=1,$$

as required. ○

Thus, in this case, the calculus integral and the **Riemann integral** give the same result. The function defined by (2.13), page 53 of Chapter 2, shows that existence of the calculus integral does not guarantee existence of the corresponding Riemann integral.

Stieltjes integration is "integration of a point function $f(x)$ with respect to a point function $g(x)$". Suppose $g(x)$ is a point function defined for real numbers x. For intervals $I=]u,v]$ define the interval function $F(I)$ by

$$F(]u,v])\coloneqq g(v)-g(u).$$

The function F is additive on disjoint, adjoining intervals $]u,v]$, $]v,w]$:

$$\begin{aligned}
F(]u,w]) &= g(w)-g(u) \\
&= (g(w)-g(v))+(g(v)-g(u)) \\
&= F(]u,v])+F(]v,w]).
\end{aligned}$$

(Conversely, given an additive interval function $F(I)$, a corresponding point function g can be defined by $g(x) := F(]-\infty, x])$. Additivity[6] of F ensures that g is well defined.) Then the Stieltjes (or Riemann-Stieltjes) integral of f with respect to g on $]0,1]$ is

$$\int_0^1 f(x)dg, \;=\; \int_{]0,1]} f(x)F(I).$$

Example 2 *To illustrate the calculation of a Stieltjes integral using Riemann sums, suppose $f(x) = 2x^2$ and $g(x) = x^2$. Then, for $I =]u, v]$,*

$$F(I) = v^2 - u^2, \qquad \int_0^1 f(x)dg = \int_{]0,1]} f(x)F(I).$$

Take 1 as the candidate for the value of this integral. To test this candidate value, consider Riemann sums

$$\sum_{r=1}^{n} f(x_r)\left(g(u_r) - g(u_{r-1})\right) \;=\; \sum_{r=1}^{n} 2x_r^2\left(u_r^2 - u_{r-1}^2\right) \;=\; \sum_{r=1}^{n} 2x_r^2 F(I_r),$$

with a view to establishing a relation

$$\left|1 - \sum_{r=1}^{n} 2x_r^2\left(u_r^2 - u_{r-1}^2\right)\right| < \varepsilon \tag{1.3}$$

for partitions $\mathcal{P} = \{]u_{r-1}, u_r] \;:\; r = 1, 2, \ldots, n\}$. Choose $\varepsilon > 0$, and note that

$$u_r^4 - u_{r-1}^4 = \left(u_r^2 + u_{r-1}^2\right)\left(u_r^2 - u_{r-1}^2\right) = \left(u_r^2 + u_{r-1}^2\right)F(I_r).$$

By uniform continuity, $\delta > 0$ can be chosen so that, if $u_r - u_{r-1} < \delta$, then

$$\left|\left(u_r^2 + u_{r-1}^2\right) - 2x_r^2\right| < \varepsilon$$

for any x_r satisfying $u_{r-1} \le x_r \le u_r$. Therefore, for any collection $\mathcal{P} = \{I_r\}$ partitioning $]0,1]$ with $|I_r| < \delta$ for $1 \le r \le n$,

$$\begin{aligned}
\left|1 - (\mathcal{P})\sum 2x_r^2 F(I_r)\right| \;&=\; \left|\sum_{j=1}^{r}\left(\left(u_r^4 - u_{r-1}^4\right) - 2x^2\left(u_r^2 - u_{r-1}^2\right)\right)\right| \\
&=\; \left|\sum_{j=1}^{r}\left(\left(u_r^2 + u_{r-1}^2\right) - 2x^2\right)\left(u_r^2 - u_{r-1}^2\right)\right| \\
&\le\; \sum_{j=1}^{r}\left|\left(\left(u_r^2 + u_{r-1}^2\right) - 2x^2\right)\left(u_r^2 - u_{r-1}^2\right)\right| \\
&<\; \varepsilon\sum_{j=1}^{r}\left(u_r^2 - u_{r-1}^2\right) \;=\; \varepsilon,
\end{aligned}$$

[6] By Theorem 10, for $f(x)$ constant (with value 1, say), every Stieltjes integrator function $F(I)$ is integrable. This fundamental point is a central theme of this book.

so $\int_0^1 f(x)dg(x), = \int_{]0,1]} 2x^2 F(I), = 1$. *In this case too the Riemann sum calculation (1.3) required that the candidate value 1 for the Stieltjes integral be available for testing. The solution to the problem had to be known in advance of solving the problem, so to speak. Where did the candidate value come from? In this case the integrand*

$$f(x)F(I) = 2x^2\left(v^2 - u^2\right) = 2x^2(v+u)(v-u) = 2x^2(v+u)|I|$$

has a form which is fundamentally similar to the integrand in (1.1). Therefore the integral value 1 is worth testing. And, as demonstrated, it satisfies the required Riemann sum inequality. ○

If an interval function $F(I)$ is additive on any finite number of disjoint, adjoining intervals I we designate it as a *Stieltjes cell function* or *Stieltjes integrator.*

The **Burkill integral** (Burkill [25, 26], Henstock [103]) has integrands of the form $h(I)$ which are **not** additive.

Example 3 *For* $I =]u, v]$ *let* $h(I) = 4u^2v(v-u)$. *Then, with* $u < v < w$, $J =]u, w]$, $I_1 =]u, v]$, $I_2 =]v, w]$, *we have* $J = I_1 \cup I_2$,

$$h(J) = 4u^2w(w-u), \quad h(I_1) = 4u^2v(v-u), \quad h(I_2) = 4v^2w(w-v),$$

and $h(J) \neq h(I_1) + h(I_2)$. *A Riemann sum calculation gives* $\int_{]0,1]} h(I) = 1$. *To see this, consider a partition* $\mathcal{P} = \{]u_{r-1}, u_r]$ *of* $]0, 1]$ $(r = 1, \ldots, n)$, *so that, with*

$$\eta = \left|1 - (\mathcal{P})\sum h(I_r)\right|,$$

we have

$$\begin{aligned}\eta &= \sum_{r=1}^{n}\left(\left(u_r^4 - u_{r-1}^4\right) - 4u_r u_{r-1}^2(u_r - u_{r-1})\right)\\ &= \sum_{r=1}^{n}\left(\left(u_r^3 + u_r^2 u_{r-1} + u_r u_{r-1}^2 + u_{r-1}^3\right) - 4u_r u_{r-1}^2\right)(u_r - u_{r-1}).\end{aligned}$$

Let $\varepsilon > 0$ *be given. The expression* $\left(u_r^3 + u_r^2 u_{r-1} + u_r u_{r-1}^2 + u_{r-1}^3\right) - 4u_r u_{r-1}^2$ *is a difference of the functions*

$$s^3 + s^2 t + st^2 + t^3 \quad \text{and} \quad 4st^2.$$

By uniform continuity in both variables of these two functions, a number $\delta > 0$ *can be found so that*

$$\left|\left(u_r^3 + u_r^2 u_{r-1} + u_r u_{r-1}^2 + u_{r-1}^3\right) - 4u_r u_{r-1}^2\right| < \varepsilon$$

whenever $|I_r| = u_r - u_{r-1} < \delta$, *giving*

$$\eta < \varepsilon \sum_{r=1}^{n}(u_r - u_{r-1}) = \varepsilon$$

The result follows from this. ○

Strictly speaking, Burkill integrands $h(I)$ do not contain any element of dependence on points x, and depend—in a non-additive way—only on intervals (or cells) I. For the purposes of this book, however, it is convenient to extend the meaning of Burkill integration to include dependence on points x, so the Burkill integrand is

$$f(x)h(I) \quad \text{or} \quad h(x, I).$$

Thus a Burkill integrand can be a product of a point function f multiplied by an interval function h. Or it can be an integrand h which depends jointly on points x and cells or intervals I. If, further, it is **not** stipulated that $h(I)$ is non-additive, then Burkill integrands $f(x)h(I)$ include, as a special case, additive interval functions of the Stieltjes kind. Viewed this way, Burkill integration is a generalization of Stieltjes integration. The latter, in turn, is a generalization of Riemann integration, with $h(I) = F(I) = |I|$.

Generally speaking, the convention[7] in this book is to use a capital letter such as F to indicate additive interval functions $F(I)$; while lower case letters such as h are used for potentially non-additive interval functions $h(I)$.

The following is an example of a point-interval Burkill integrand $h(x, I)$ which is not a product $f(x)h(I)$.

Example 4 *For $I =]u, v]$ and $u \leq x \leq v$, write*

$$h(x, I) = 2v^2 - x(v - u) - u(u + v).$$

This function is integrable on $]0, 1]$, with integral value 1. To see this, rewrite the integrand as

$$h(x, I) = (v^2 - u^2) + (v - x)(v - u);$$

and, with $\varepsilon > 0$ given, take $\delta = \varepsilon$. Then, for

$$0 = u_0 < u_1 < u_2 < \cdots < u_n = 1, \qquad \eta = \left| 1 - \sum_{r=1}^{n} h(x_r, I_r) \right|,$$

with $|I_r| = u_r - u_{r-1} < \delta$, the Riemann sum satisfies

$$\begin{aligned} \eta &= \left| 1 - \sum_{r=1}^{n} \left((u_r^2 - u_{r-1}^2) + \sum_{r=1}^{n} (u_r - x)(u_r - u_{r-1}) \right) \right| \\ &\leq \varepsilon \sum_{r=1}^{n} (u_r - u_{r-1}) = \varepsilon; \end{aligned}$$

so $\int_{]0,1]} h(x, I) = 1$. ○

[7] An exception to this convention is to be found in stochastic integration (Chapter 8). In that case the incremental, or interval function, form of a point function $x(t)$ is denoted by $\mathbf{x}(]t, s]) = x(s) - x(t)$, in order to define a Stieltjes integrand with respect to the point function x (or with respect to the Stieltjes integrator $\mathbf{x}$).

Integrable functions do not have to be products of point functions and interval functions. Interval functions $h(]u,v])$ need not be additive, and need not depend explicitly on the numbers u, v or $v-u$. In fact, $h(]u,v])$ need not even be monotone: it is not required that $J \supset I$ should imply that $h(J) \geq h(I)$, as the following Burkill integrand shows.

Example 5 *For* $I =]u,v]$ *let*

$$h(I) := \begin{cases} 1 & \text{if } u = \frac{1}{2} \text{ or } v = \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$h\left(\left]0,\tfrac{1}{2}\right]\right) + h\left(\left]\tfrac{1}{2},1\right]\right) = 2, \quad h(]0,1]) = 0, \quad h\left(\left]0,\tfrac{1}{2}\right]\right) = 1.$$

Let δ *be any positive number less than or equal to* $\frac{1}{2}$. *Let*

$$I_1 = \left]\tfrac{1}{2} - \tfrac{\delta}{2}, \tfrac{1}{2} + \tfrac{\delta}{2}\right], \quad I_2 = \left]\tfrac{1}{2} - \delta, \tfrac{1}{2}\right],$$

and let $\mathcal{P}_1$, $\mathcal{P}_2$ *be partitions of* $]0,1]$ *containing* I_1 *and* I_2, *respectively:*

$$\mathcal{P}_1 = \{\cdots, I_1, \cdots\}, \quad \mathcal{P}_2 = \{\cdots, I_2, \cdots\}.$$

Then $\mathcal{P}_2$ *contains an interval* $]\frac{1}{2}, v]$, *and*

$$(\mathcal{P}_1)\sum h(I) = 0, \quad (\mathcal{P}_2)\sum h(I) = 2.$$

Therefore $h(I)$ *is not Burkill integrable on* $]0,1]$ *in the basic sense of Riemann sums. But in Chapter 2 it is shown that it is possible to constrain the formation of partitions* $\mathcal{P}$ *of the domain* $]0,1]$ *in such a way that every partition has the form* $\mathcal{P}_2$. *(See, for instance, Example 15.) In this constrained system of integration,* $h(I)$ *is said to be* Burkill-complete *integrable, with integral 2; that is,* $\int_{]0,1]} h(I) = 2$. ○

Since $|I|$ is an additive function of intervals I, Riemann integrands can be taken to be Stieltjes integrands. Also, Stieltjes integrands can be taken to be Burkill integrands as presented here. Thus the formulation $h(x,I)$ can represent not just a Burkill integrand but also Riemann and Stieltjes integrands. If an integrand is Riemann integrable it is Stieltjes integrable; and, likewise, Stieltjes integrability implies Burkill integrability.

In the workings of the examples above, *indefinite integrals* appear. Letting $h(x,I)$ represent, in turn, Riemann, Stieltjes, and Burkill integrands, an indefinite integral of $h(x,I)$ is an additive interval function $H(J)$ whose value on every interval J equals the integral of $h(x,I)$ on J. The indefinite integral $H(I)$ is thus a Stieltjes cell function; and, as a Stieltjes integrand it is itself integrable—in the Riemann sum sense—on every bounded interval J, with integral $H(J)$.

In (1.1) the indefinite integral is $F(x) = x^4 + c$; and, for the same Riemann integrand in Example 1, the same indefinite integral is written as the additive interval function (or Stieltjes cell function—a cell is an interval)

$$F(I) = F(\,]u, v]) = v^4 - u^4,$$

this being the Stieltjes increment of the point function $F(x)$ for which $F'(x) = f(x) = 4x^3$.

Each of Examples 2 and 3 also has indefinite integral $v^4 - u^4$; while Example 4 has indefinite integral $v^2 - u^2$. In Example 5, the indefinite integral $H(I)$ of the Burkill integrand $h(x, I)$ does not actually appear in the workings, but a few moments' examination should be sufficient to see that the indefinite integral in this case is the Stieltjes cell function

$$H(I) = \begin{cases} 0 & \text{for} \quad I = \,]u, v]\,, \quad v < \frac{1}{2}, \\ 1 & \text{for} \quad I = \,]u, \frac{1}{2}]\,, \\ 2 & \text{for} \quad I = \,]u, v]\,, \quad u < \frac{1}{2} < v, \\ 1 & \text{for} \quad I = \,]\frac{1}{2}, v]\,, \\ 0 & \text{for} \quad I = \,]u, v]\,, \quad u > \frac{1}{2}. \end{cases}$$

Riemann, Stieltjes, and Burkill integrals feature in this book, but mainly in the form of Riemann-complete, Stieltjes-complete, and Burkill-complete integrals.

This section has focussed on determining the definite and indefinite integrals of given integrands. Though it will not feature in this book, another aspect of integration is the converse problem of determining an integrand $h(x, I)$ from an indefinite integral $H(I)$, or from a differential equation satisfied by an indefinite integral. To illustrate simply, if a function $F(x)$ is differentiable then it is an indefinite integral of its derivative $f(x), = F'(x)$.

1.5 The -Complete Integrals

The Riemann, Stieltjes, and Burkill integrals presented in Section 1.4 are "incomplete" in various ways. For instance, it is not possible to specify broad conditions for which the limit of a sequence of integrands is integrable, with the integral of the limit equal to to the limit of the corresponding sequence of integrals. This makes it difficult to justify, for instance, differentiation under the integral sign, and many other similarly useful calculations on integrals.

From the beginning of the twentieth century the Lebesgue integral has partially remedied this, providing strong conditions under which it is possible to take limits under the integral sign. However it was apparent that the Lebesgue integral is itself "incomplete" in the sense that, just like the basic Riemann integral whose deficiencies needed to be remedied, not every derivative could be integrated by the new method. It is possible for a function with an indefinite integral to **not** have a definite integral. This is the case for the function defined by (2.13) on page 53 in Chapter 2, which is calculus integrable but not Riemann

integrable or Lebesgue integrable. In other words the *fundamental theorem of calculus* is not always valid for Lebesgue integration; even though, by definition, it is valid for the basic calculus or Newton integral.

This issue is explored further in Chapters 2 and 4 where it is shown that, in the -complete system of integration, an integrand has a definite integral if and only if it has an indefinite integral. In advance of that, note the following.

- Any interval function which is additive on every finite collection of disjoint, adjoining intervals is integrable in a Stieltjes sense based on Riemann sum calculation.
- Tautologically, every derivative $f(x) = F'(x)$ has an anti-derivative $F(x)$.
- Provided the partitions used to form Riemann sums $\sum f(x)|I|$ are suitably constrained (as indicated in Example 5 above), the incremental or Stieltjes form of the anti-derivative, $F(I) = F(]u, v]) = F(v) - F(u)$, is an indefinite integral for the integrand $f(x)|I|$.
- Then the finite additivity (or Stieltjes integrability) of $F(I)$ ensures the integrability (i.e., existence of the definite integral) of $f(x)|I|$.

Thus, with "constrained" Riemann sum formation the fundamental theorem of calculus holds for integrands $f(x)|I|$. Therefore, for integrands $f(x)|I|$, it is reasonable to designate this type of integration as *Riemann-complete*.

The fundamental theorem of calculus is especially important in areas such as differential and integral equations. But it is not so important in investigations of random variability, a subject which involves a class of Stieltjes cell functions which is broader than the the class of indefinite integrals formed from anti-derivatives.

Henstock [93] applied the term Riemann-complete to Stieltjes-complete and Burkill-complete integrands. The reason this book makes a distinction between these kinds of integrands is, in part, because of the lesser significance of the fundamental theorem of calculus in this subject area, and greater significance of other kinds of Stieltjes integrands and Stieltjes integrator functions.

The evaluations in Section 1.4 show that a key step in integrating any function is identification of its indefinite integral—an additive interval function or Stieltjes cell function. So Stieltjes-complete integration is the link between the various kinds of integrand. Chapter 4 shows that an integrand $f(x)|I|$ (or $h(x, I)$) is integrable if and only if it is "almost" (in some sense) identical to a Stieltjes cell function $H(I)$.

1.6 Riemann Sums in Statistical Calculation

Elementary statistical calculation is often learned by performing exercises such as the following. *"A sample of 100 individuals is selected, their individual weights are measured, and the results are summarized in Table 1.1. Estimate the mean weight and standard deviation of the weights in the sample."*

Weights (kg)	Proportion of sample
0 – 20	0.2
20 – 40	0.3
40 – 60	0.2
60 – 80	0.2
80 – 100	0.1

Table 1.1: Relative frequency table of distribution of weights.

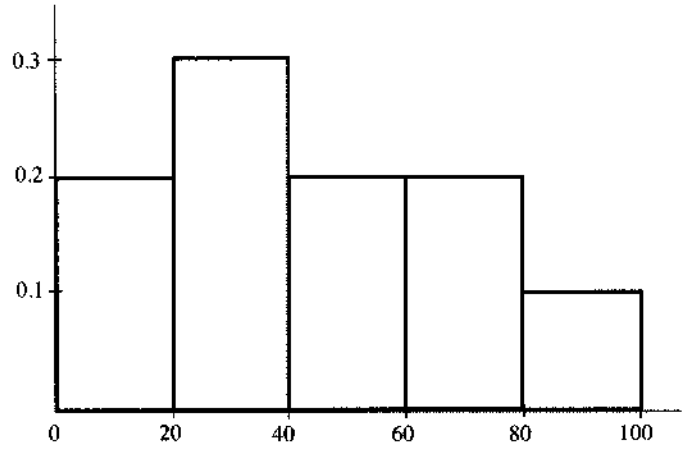

Figure 1.1: Histogram for distribution of weights.

I	$F(I)$	x	$f(x)$	$xF(I)$	$f(x)F(I)$
0 – 20	0.2	10	100	2	20
20 – 40	0.3	30	900	9	270
40 – 60	0.2	50	2500	10	500
60 – 80	0.2	70	4900	14	980
80 – 100	0.1	90	8100	9	810

Table 1.2: Calculation of mean and standard deviation.

Figure 1.1 is the histogram for distribution Table 1.1. Sometimes calculation of the mean and standard deviation is done by setting out the workings as in Table 1.2. The observed weights of the sample members are grouped or classified in intervals I, and the proportion of weights in each interval I is denoted by $F(I)$. A representative weight x is chosen from each interval I. The function $f(x)$ is x^2 since, in this case, these values are needed in order to estimate the variance. Completing the calculation, the estimate of the arithmetic mean weight in the

sample is

$$\sum xF(I) = 44 \text{ kg},$$

while the variance of the weights is approximately

$$\sum x^2F(I) - (44)^2 = 2580 - 1936 = 644.$$

The latter calculation, involving $\sum x^2F(I)$, has the form $\sum f(x)F(I)$ with $f(x) = x^2$. The expressions $\sum xF(I)$ and $\sum f(x)F(I)$ have the form of Riemann sums, in which the interval of real numbers $[0, 100]$ is partitioned by the intervals I, and where each x is a representative data-value in the corresponding interval I. Thus the sums

$$\sum xF(I) \quad \text{and} \quad \sum f(x)F(I)$$

are approximations to the Stieltjes (or Riemann–Stieltjes) integrals

$$\int_J x\,dF \quad \text{and} \quad \int_J f(x)dF, \quad \text{respectively:}$$

the domain of integration [0,100] being denoted by J.

1.7 Random Variability

If X refers to the potential data-values x arising from an experiment corresponding to a weighing of a single individual member of the population under investigation, it can reasonably be declared that the calculation $\sum xF(I)$ above is an estimate of the expected value of X, denoted $\mathrm{E}[X]$. The actual datum x obtained when the single measurement has been completed is the *outcome* of the experiment. The datum x can also be called an *observation* or *occurrence*. In that case, each entry in the column headed "Proportion of sample" in Table 1.1 represents an estimate of the potentiality or probability that the single observation x will lie within a particular range of possible values.

In Table 1.2 a calculation $f(x) = x^2$ is performed on the measured value x. Accordingly, denote by $f(X)$ some function of the *random variability* in the experiment X, such as $f(X) = X^2$ where $x^2 = f(x)$ is the outcome of $f(X)$; and then the calculation $\sum f(x)F(I)$ is an estimate of the expected value of $f(X)$, denoted $\mathrm{E}[f(X)]$. Call $f(X)$ a *contingent random variable*, dependent on the *elementary* random variable X.

The expression *random variable* has been used above without explanation or definition. In Kolmogorov's book [123], the expression is used as a synonym for *experiment*. Intuitively "experiments", "trials", or "random variables" can be recognized and understood as in the following examples.

Example 6 *Measuring the weight of an individual member of a given population.*

Example 7 *Observing the amount of electric current emitted by a photoelectric cell when a beam of light of given intensity is directed on the cell.*

Example 8 *Throwing a die and observing which of the numbers 1 to 6 lands uppermost.*

Example 9 *Throwing a die and noting the square of the number which lands uppermost.*

Example 10 *Throwing a pair of dice and, whenever the sum of the numbers observed exceeds 10, paying out a wager equal to the sum of the two numbers thrown, and otherwise receiving a payment equal to the smaller of the two numbers observed.*

Example 11 *A gambling game in which the gambler pays one cent for each successive throw of a single die, and receives a thousand euro if 100 successive sixes are thrown.*

In each case there is some experiment or trial involving the observation and measurement of some unpredictable value. Underlying factors are the source of the unpredictability of the outcome, and this phenomenon is designated *random variability.*

Example 12 *Calculating the maximum value of the end-of-day prices of a barrel of crude oil observed over thirty consecutive days.*

In Example 12, a value is generated by performing a calculation f (the maximum value calculation) on 30 observable quantities $x_1, \ldots, x_{30}$. So

$$f(x_1, \ldots, x_{30}) = \max\{x_1, \cdots, x_{30}\}$$

is the outcome, depending on the unpredictable basic joint outcome $x_1, \ldots, x_{30}$, each element of the basic joint outcome being itself the elementary outcome of an experiment or trial X_j, for each $j = 1, \ldots, 30$. Thus there are 30 "joint basic random variations": $X_1, \ldots, X_{30}$ corresponding to observable end-of-day prices $x_1, \ldots, x_{30}$, and a "contingent (or **dependent**) random variation" $f(X_1, \ldots, X_{30})$ corresponding to the maximum value calculation, and whose value **depends** on the basic joint outcome composed of 30 elementary outcomes.

Example 11 has a basic joint (or *joint-basic*) random variation composed of an infinite series of elementary basic random variations $\{X_j\}$ whose observable values x_j are $1, 2, \ldots, 6$; and a contingent (or dependent) random variation $Y = f(X_1, X_2, X_3, \ldots)$, whose observable value is

$$y = \begin{cases} 1000 \text{ if there exists } j \text{ such that } x_j = x_{j+1} = \cdots = x_{j+99} = 6, \\ 0 \text{ otherwise.} \end{cases}$$

Example 10 has two "basic random variations" X_1 and X_2 corresponding to the numbers x_1 and x_2 thrown for each of the pair of dice, and the wager (or

contingent random variation) $f(X_1, X_2)$ given by the calculation

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & \text{if } x_1 + x_2 > 10, \\ \min\{x_1, x_2\} & \text{if } x_1 + x_2 \leq 10. \end{cases}$$

Example 10 has a contingent random variation $f(X)$ where X is a *joint-basic observable* or joint measurement (X_1, X_2). Example 12 has $f(X)$ where $X = (X_1, \ldots, X_{30})$. If, as in Example 11, X consists of a joint observation of infinitely many values the observable X is traditionally called a *process* or *stochastic process*. Thus a process can consist of a family (X_t), where each t belongs to some infinite domain such as the unit interval $[0, 1]$.

There is a distinction to be made between joint observation of, on the one hand, a finite number of values and, on the other hand, an infinite number of values. But in this book both are encompassed in a single theory.

Thus the intuitive meaning of *random variable* is, firstly, that it involves the generation of a value or *datum* resulting from measurement(s) or observation(s); secondly, in advance of measurement or observation, this value is not certain or definite but can be one of a range of possible occurrences or observations; and, thirdly, that sometimes it is possible to associate some measure of potentiality or likelihood with the possible outcomes or data values that may be observed. In other words, in advance of actual measurement or observation, the datum can be predicted with a degree of accuracy given by some measure of accuracy potentiality or likelihood.

If the possible outcomes or occurrences are discrete, then the potentialities or probabilities are associated with each of the possible values. If the possible outcomes belong to a continuous domain, as in Table 1.2 above, then the potentialities or probabilities are quantities $F(I)$ associated with intervals I of possible outcomes of the measurement. Provided the function F is *atomic*, then $F(I)$ can also be used to represent the probabilities of discrete values.

Thus the intuitive conception of random variation implies a number of elements:

- the generation of a value or datum resulting from observation of one of
- a set of potential data-values or occurrences combined with
- a set of accuracy potentialities or likelihoods.

The first element will be denoted by a symbol such as x; or by $f(x)$ if some deterministic calculation is performed on the measured or observed value x. The second element corresponds to the *sample space* Ω_X for the random variation. The third component corresponds to the *probability measure* (or *potentiality distribution function*) $F_X(I)$ for the random variation. The notations for random variation adopted in this book make reference to these three elements, with the tabular layout and histogram of Table 1.2 and Figure 1.1 as their intuitive basis.

The *sample space* corresponds to the original source of the unpredictability in the value generated by the experiment. In Example 8 the sample space is the

set $\Omega = \{1, 2, \ldots, 6\}$. There is often some flexibility in how the sample space can be designated. Provided the distribution function value $F(I) = \frac{1}{6}$ whenever the real interval I contains one of the values 1 to 6, then we can, for example, take the sample space for this experiment to be the line interval 1 to 6, or the whole real line $\mathbf{R}$. Many other choices of sample space are available. Similarly in Example 10 the sample space can be taken to be any one of various sets such as

$$\{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}, \quad \mathbf{R} \times \mathbf{R}, \quad \{1, 2, 3, \ldots, 35, 36\}, \quad \text{or } \mathbf{R}.$$

Specification of the potentiality distribution function for the experiment will depend on which set is chosen as sample space.

An experiment, measurement, or "random variation" can be represented or specified by an expression involving factors $[\Omega_X, F_X]$, where Ω_X and F_X are suitably chosen mathematical constructions which enable us to represent, describe, and analyze the random variation in the experiment X. Thus a basic random observation or measurement can be denoted by a symbol X and can be expressed, in the chosen representation, by

$$X \simeq x[\Omega_X, F_X].$$

The random variation in Example 8 could be specified in various alternative (but equivalent) ways, such as

Ω_X	F_X
1	$\frac{1}{6}$
$\vdots$	$\vdots$
6	$\frac{1}{6}$

, or $\left[\{i\}, \{F_X(\{i\} = \frac{1}{6}\}\right]_{i=1}^{6}$, or $[\mathbf{R}, F_X]$;

where, in the latter specification, intervals $I \subset \mathbf{R}$ have $F_X(I)$ equal to a sixth if I contains just one of the numbers one to six, with $F_X(I)$ equal to zero otherwise.

In contrast to such a *basic* random variable X, a random variable $f(X)$ can be expressed in *contingent* or dependent form, where some deterministic calculation f is performed on the basic observed value x from the sample space Ω_X. A contingent random variable is denoted by $f(X)$, and written as

$$f(X) \simeq f(x)[\Omega_X, F_X]$$

to specify contingent random variation with outcome $f(x)$. The underlying or basic random variation involved in this is the basic $X \simeq x[\Omega_X, F_X]$.

An alternative approach here would be to denote the set of possible outcomes or occurrences $\{f(x) : x \in \Omega_X\}$ by Ω_Y, and deduce a distribution function F_Y on the intervals of Ω_Y, so

$$Y \simeq y[\Omega_Y, F_Y].$$

Provided Ω_Y is the set of real numbers $\mathbf{R}$ we say that Y is the *elementary* form of the contingent random variable $f(X)$. Accordingly, two possible representations for the random variation in Example 9 could be

$$X^2 \simeq x^2 \left[\{i\}, \left\{F_X(i) = \tfrac{1}{6}\right\}\right]_{i=1}^{6} \tag{1.4}$$

with underlying $X \simeq x\left[\{i\}, \left\{F(i) = \frac{1}{6}\right\}\right]_{i=1}^{6}$; or, alternatively,

$$Y \simeq y\left[\{1^2, 2^2, \ldots, 6^2\},\ \ F_Y(1) = F_Y(4) = \ldots = F_Y(36) = \tfrac{1}{6}\right]. \tag{1.5}$$

The former representation has a "contingent" form (involving the deterministic function of squaring an observed basic value i), while the latter has "elementary" form y.

1.8 Contingent and Elementary Forms

Now consider an experiment X involving observation of the pair of numbers which fall uppermost when a pair of dice is thrown (or when a single die is thrown twice). The result is a single outcome x composed of a pair of joint occurrences (x_1, x_2) where x_1 is the number falling uppermost for the first die and x_2 is the number falling uppermost for the second die. Thus X is (X_1, X_2), where X_r is the observation of die r $(r = 1, 2)$; with

$$X_r \simeq x_r\left[\{k\},\ \ F(k) = \tfrac{1}{6}\right]_{k=1}^{6}.$$

The joint datum is $x = (x_1, x_2)$, and experiment can be represented as

$$X \simeq x\left[\{i\}_{i=1}^{6} \times \{j\}_{j=1}^{6},\ \ F_X(i,j) = \tfrac{1}{36},\ i = 1, 2, \ldots, 6,\ j = 1, 2, \ldots, 6\right],$$

the sample space being

$$\{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}.$$

With $\mathbf{R}$ denoting the set of all real numbers, an alternative way of expressing the joint observation is

$$X \simeq x\left[\mathbf{R} \times \mathbf{R},\ \ F_X(I) = F_X(I_1 \times I_2) = \tfrac{1}{36} \text{ if } I_1 \cap \mathbf{R} = \{i\},\ I_2 \cap \mathbf{R} = \{j\}\right]$$

where F_X is atomic. Now suppose a single datum is generated from the joint observation X by calculating the sum of the two numbers observed to fall uppermost when the pair of dice is thrown. The resulting random variable $f(X) = f(X_1, X_2)$ can be represented as follows:

$$f(X) \simeq f(x)\left[\mathbf{R} \times \mathbf{R},\ \ F_X(I_1 \times I_2)\right], \tag{1.6}$$

where $f(x_1 + x_2) = x_1 + x_2$ (i.e., $f(i,j) = i + j$), and

$$F_X(I_1 \times I_2) = \tfrac{1}{36} \text{ if } I_1 \cap \mathbf{R} = \{i\},\ I_2 \cap \mathbf{R} = \{j\}.$$

This experiment can also be represented as

$$Y \simeq y\left[\{2, 3, \ldots, 12\},\ \ F_Y(2) = \tfrac{1}{36},\ F_Y(3) = \tfrac{1}{18},\ F_Y(4) = \tfrac{1}{12}, \ldots\right].$$

Thus, using an atomic form of distribution function, with $F_Y(J) = \frac{1}{36}$ if $J \cap \mathbf{R} = \frac{1}{j}$ for $j = 2, 3, \ldots, 12$, the experiment $f(X)$ can be expressed as

$$Y \simeq y\left[\mathbf{R}, F_Y\right], \quad \text{with} \quad y = f(x); \tag{1.7}$$

and we can write

$$Y = f(X).$$

In representation (1.6), $f(X)$ has explicitly contingent form and sample space $\mathbf{R} \times \mathbf{R}$; while (1.7) has elementary form Y with sample space $\mathbf{R}$. In (1.7) the contingency or dependence of Y on the joint-basic observation $X = (X_1, X_2)$ is not explicit. Each basic observation X_r is itself an elementary observation since its sample space is $\mathbf{R}$ for each of $r = 1, 2$. The key relationship between the two representations, contingent and elementary, is

$$y = f(x) = x_1 + x_2, \qquad Y = f(X) = X_1 + X_2.$$

The example demonstrates how this relationship enables the distribution values F_Y to be deduced from the values of F_X, and vice versa. Also, there is some loss of information in converting a contingent form to an elementary form, in that the individual components x_1 and x_2 can no longer be seen.

This example illustrates an important point in the representation of a random variable. Knowledge of the likelihood distribution function enables us to glean information about the potential datum values, such as mean and variance. In other words, the distribution function carries information about the accuracy of estimates of the datum.

On the other hand, knowledge of data occurrences, obtained, for instance, by repeated replication of the experiment or measurement enables us to estimate distribution function values, as in Table 1.1 above. And knowledge of the functional relationship $y = f(x)$ between different representations of the same experiment can sometimes enable us to deduce the corresponding likelihood values F_Y and F_X from each other.

The function F_X carries information—in advance of occurrence—about accuracy of estimates of the measurement or datum x. And F_Y does the same for the datum y. The elementary-contingent relationship $y = f(x)$ carries information about the relationship between F_Y and F_X, so the former is deducible from the latter.

The notation is intended to highlight the various perspectives from which particular instances of random variability can be viewed. This can be seen in (1.6) and (1.7) above. The representation in (1.6) shows the underlying random variation as $\{(i,j)\}_{i,j=1}^{6}$, with each instance or occurrence having a likelihood of $1/36$; and the potential data values being then obtained by the further deterministic calculation $i + j$ which is shown to the left of the square brackets. This is the contingent form.

The same experiment is represented differently in the elementary-form version (1.7). In this representation the manifestations of underlying random variation (shown inside the square brackets) are the possible totals generated by a single throw of a pair of dice. The potential data-values, shown to the left of the square brackets, are the same numbers without any further deterministic calculation.

This book makes use of these alternative perspectives in the various areas of probability theory. In (1.6) and (1.7) it is easy to deduce the likelihood values

F_Y from those of F_X. But, for more complicated forms of contingency f, this step can be difficult; and it is sometimes necessary to resort to sophisticated theory in order to make such deduction. It is helpful to view some of the big themes of probability—such as central limit theorems and Itô's formula—from this point of view, and this is demonstrated in the text.

Expressions (1.6) and (1.7) also illustrate those situations where we seek to examine features of several (perhaps infinitely many) basic random occurrences considered jointly, the occurrences or observations being linked together in some way, as when a pair or more of dice are thrown at the same time; or a single die is thrown repeatedly, at successive instants of time. Both the elementary and explicitly contingent forms of representation are widely used in the study of joint random variation, and justification for this is provided in Theorem 82 of Chapter 5.

Generally, in this book the contingent representation of joint variation is preferred. One reason for this is that, in the contingent representation, it is easier to analyze aspects and consequences of the joint variation, such as the *independence* or otherwise of the basic random variables X_r. Conversion of joint-basic random variability to elementary form involves some degree of concealment or masking of information about the joint variability involved.

Analysis of random variability is concerned with establishing or predicting both the datum and its "degree of accuracy"; and relating each one to the other. "Degree of accuracy" is given by the distribution function. Broadly speaking, establishing the correct distribution function is the key point in specifying an observable, and that is why much of this book is concerned with deducing the elementary-form distribution functions of observables from their contingent form.

1.9 Comparison With Axiomatic Theory

The analysis of random variation in this book is built up from these intuitive conceptions. In contrast, the traditional definition of random variable X (or $f(X)$) given in many textbooks is that it is a measurable function defined[8] on an abstract sample space Ω.

To illustrate the traditional approach, consider Example 4 above—a single throw of a single die, the random variable being, intuitively, the observation of which of the six sides falls uppermost. Since the outcome is random—that is, uncertain or unpredictable—it is possible to envisage or postulate some abstract domain Ω which, somehow, generates the various possible outcomes of throwing the 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, which determine the real-world outcomes

[8]In practice, however, the actual sample space is chosen as in the preceding examples. And Theorem 76 of Chapter 5 reverses the traditional or axiomatic definition of a random variable, by **deducing** the measurability of a random variable.

of 1 to 6, respectively, whenever the die is thrown. In terms of the axiomatic theory of probability, the random variable is the mapping 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, with

$$P'(\text{red}) = \tfrac{1}{2}, \quad \text{and} \quad P'(\text{green}) = \cdots = P'(\text{yellow}) = \tfrac{1}{10}.$$

This counterposes two different experiments, measurements, or random variables in the intuitive sense; the first involving a balanced die and the second an unbalanced one; corresponding to P and P'. But formally speaking, and in traditional axiomatic 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, in the axiomatic sense, the same.

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

But, setting these reservations aside, as long as the probability measure P of a probability space $(\Omega, \mathcal{A}, P)$ is linked to the sample space Ω in the "measurable function $f(X)$" conception of random variable, there is no essential difference between this conception and the intuitive "set-of-potential-data-values-linked-with-accuracy-potentiality-distribution" representation $f(X)$ with $X \simeq x[\Omega_X, F_X]$. As the preceding sections show, the latter approach also allows us to easily choose alternative specification of both the sample space Ω_X and the potentiality function F_X.

The relationship between the two conceptions of random variation can be demonstrated as follows. If the distribution function $F_X(I)$ is deduced from the probability measure P by

$$F_X(I) := P\left(X^{-1}(I)\right),$$

then

$$\mathrm{E}\left[f(X)\right] = \int_{\Omega} f(X(\omega))dP = \int_{\mathbf{R}} f(x)dF_X,$$

where $X(\omega) = x$ and the latter integral is the Lebesgue–Stieltjes integral. And, as will be demonstrated in Section 2.8, the Lebesgue–Stieltjes integral is equal to the Stieltjes-complete integral

$$\int_{\mathbf{R}} f(x)F_X(I),$$

which is the basis of the analysis of random variation presented in this book.

1.10 What Is Probability?

Probability values are sometimes established by reasoning from the specific details of actual measurements. For instance, in tossing coins, throwing dice, or assessing the motion of a pollen particle released into a gas-filled glass vessel, potentialities or likelihoods of particular eventualities may be deduced from observation of the behavior and characteristics of the actual phenomena themselves.

This involves an assumption that probability/likelihood is, in some sense, a "real" phenomenon; and that it is actually present in the events or measurements under consideration—in other words, that it exists, and is "knowable".

In the traditional axiomatic theory of probability, there is an a priori function P defined on an abstract probability space $(\Omega, \mathcal{A}, P)$ from which we purport to deduce the probabilities of actual phenomena by means of mathematical theory.

Similarly, the Riemann sum approach posits the "reality" or objectivity, in some sense, of a notion of "accuracy potentiality"[9] or likelihood $\mathcal{L}$, from which is deduced a mathematical device—the distribution function F_X.

This approach may be close to the way in which supposedly objective potentialities or probabilities/likelihoods $\mathcal{L}$ are linked to their mathematical manifestation in the form of potentiality distribution functions F_X for actual experiments or observations X. Essentially,

$$\mathcal{L}[I] = F_X(I),$$

where F_X is somehow estimated or deduced from "real-world" data or real physical events in which likelihood $\mathcal{L}$ is assumed to exist. This equation is the bridge between a supposedly natural manifestation of likelihood or potentiality, and its mathematical representation in a context of estimating, measuring, or observing a potential datum x.

In contrast the axiomatic approach has

$$F_X(I) := \int_{(X \in I)} dP$$

with probability P postulated mathematically.

Suppose several experiments X_t are considered jointly. Then there are the individual and separate experiments X_t, throws of dice, say; and the joint experiment $X = (X_t)$ with joint outcome $x = (x_t)$—such as throwing 10 sixes in 10 throws of the dice. Assuming each separate throw has a likelihood $\mathcal{L}$, and that the joint measurement or joint observation has its own likelihood, then the assumed objectivity or "reality" of $\mathcal{L}$ means that the separate and joint manifestations of $\mathcal{L}$ cannot contradict each other—they are *consistent*.

Therefore, in constructing the corresponding mathematical entities F_{X_t} and F_X, care must be taken that these too do not contradict each other. In the axiomatic theory a similar point applies regarding the postulated probability

[9] As far as possible this book avoids the term "probability", along with its associated symbol P, because the burden of meaning they carry in traditional usage may cause confusion here.

space $(\Omega, \mathcal{A}, P)$ whenever several random variables X_t have to be considered jointly.

Another complication—one that is not present in the standard, axiomatic theory of probability—is that potentiality $\mathcal{L}$ is permitted to have negative and imaginary values. This is related to those random variation scenarios in nature which involve non-absolute convergence, and is discussed further in Section 2.16 and elsewhere.

What is probability? No answer to this question is offered in this book. But a somewhat broader mathematical conception of accuracy potentiality or likelihood is presented here in its place.

1.11 Joint Variability

Understanding of *joint variability* is one of the primary purposes of the study of random variation. Suppose the individuals in the sample of Table 1.1 above had their heights measured, in addition to their weights, so there are two random variables instead of one. This is described in Table 1.3.

For $i = 1, 2$, denote the two basic random variations (weight and height) by X_i, and their intervals and potentiality distributions by I_i and $F_{X_i}(I_i)$, respectively. Note that the measurements or random occurrences are *jointly observed.* In this case, an observed value in one of the random variations is linked to an observed value of the other random variation, since both observations pertain to a single individual. Call the pair (X_1, X_2) a *joint-basic* experiment.

Then $X = (X_1, X_2)$ represents the joint variability or joint estimation of the weight and height of a single individual. It is reasonable to ask, for each pair I_1, I_2, what proportion of the sample members jointly have weight in the interval I_1 and height in the interval I_2. In other words, what is the likelihood that, after actual measurement of an individual's weight x_1 and height x_2, the joint datum $x = (x_1, x_2)$ will be contained in the set $I = I_1 \times I_2$?

Thus the set I denotes a possible joint outcome of the joint observation X. Let the joint potentialities

$$F(I), = F(I_1 \times I_2) = F_X(I_1 \times I_2) = F_X(I),$$

correspond to these sample proportions, with twenty possible values corresponding to the twenty possible joint intervals $I_1 \times I_2$. Table 1.4 displays the joint-basic

Weights (kg)	Proportion of sample	Heights (cm)	Proportion
0 – 20	0.2	0 – 50	0.3
20 – 40	0.3	50 – 100	0.2
40 – 60	0.2	100 – 150	0.4
60 – 80	0.2	150 – 200	0.1
80 – 100	0.1		

Table 1.3: Joint observables.

					I_1		
		0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	
	0 – 50	⋯					0.3
	50 – 100						0.2
I_2	100 – 150				$F_X(I_1 \times I_2)$		0.4
	150 – 200					⋯	0.1
		0.2	0.3	0.2	0.2	0.1	

Table 1.4: A joint-basic observable.

observable $X = (X_1, X_2)$, showing the sample space of potential joint data values $x = (x_1, x_2)$; some of the possible joint events $I = (I_1, I_2)$ in the form of joint intervals; and indicating the display, or format, of the observable's joint likelihood values (or accuracy potentialities) $F_X(I)$. The entries for the twenty values of $F_X(I_1 \times I_2)$ have been left blank; but the nine *marginal distribution* values $F_{X_1}(I_1)$ and $F_{X_2}(I_2)$ have been included, along the bottom and right-hand margins of the table.

If the height of an individual has no bearing on that individual's weight, we would expect that

$$F_X(I_1 \times I_2), = F_{(X_1,X_2)}(I_1 \times I_2), = F_{X_1}(I_1)F_{X_2}(I_2).$$

If two joint random variations are *independent*, this property holds for all possible choices of I_1 and I_2.

Two potential events are labeled in Table 1.4, where I_1 is the possibility that an individual selected weighs between 60 and 80 kg, while I_2 is the possibility that an individual measures between 100 and 150 cm tall. The joint event $I_1 \times I_2$ is the possibility that an individual selected is between 60 and 80 kg in weight and between 100 and 150 cm in height. If

$$F_X(I_1 \times I_2) = 0.2 \times 0.25$$

then these two particular outcomes are independent possibilities or independent events.

The *marginal potentialities* in Table 1.4 are the numbers appearing on the right-hand edge (or *margin*) and along the bottom line (or *margin*) of the table. They satisfy

$$F_{X_1}(I_1) = F_X(I_1 \times \mathbf{R}), \quad F_{X_2}(I_2) = F_X(\mathbf{R} \times \mathbf{I_2})$$

for each I_1, I_2. If the joint potentialities were given (i.e., if we filled in the blank boxes in the table), then those figures, when added horizontally, should give the totals in the right-hand margin; and if added vertically, they should give the totals along the bottom margin. This property of joint potentialities is called *consistency*.

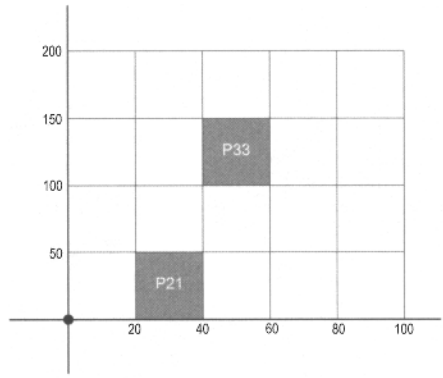

Figure 1.2: Cartesian representation of joint-basic observable.

Independence is a mathematical device or abstraction, rarely or never occurring in practice. In this case, a taller individual is also likely to be heavier, so we would expect, for instance, that more than 0.2×0.25 of the sample in the height range 100 to 150 centimeters will have weight in the range 80 to 100 kg.

It may happen that these two particular joint events are independent of each other; with likelihood of the joint event equal to the product of the two corresponding marginal likelihoods. But independence of the pair of joint observables X_1 and X_2 requires that this relationship should hold for every possible pair of joint events I_1 and I_2, not just for the twenty possibilities displayed in Table 1.4. While independence of joint observables is entertained for mathematical reasons, it seems to be a practical impossibility.

A geometric sense of the relationships between joint potentialities is conveyed by diagrams in which joint observations $x = (x_1, x_2)$ and joint events (intervals) $I = I_1 \times I_2$ are represented in a system of orthogonal axes. The orientation of the vertical axis in Figure 1.2, measuring height, is in the opposite direction to the weight column in Table 1.5 which, instead of running upward in the geometric manner, runs downward to facilitate numerical calculation. Figure 1.2 has the usual Cartesian geometric orientation for the y-axis. P21 corresponds to $F_X\left(I_1^{(2)} \times I_2^{(1)}\right)$, and P33 corresponds to $F_X\left(I_1^{(3)} \times I_2^{(3)}\right)$.

This diagram shows how the domain of the joint random variables is partitioned for Riemann sum calculation of expected value, for instance. Unlike the histogram of Figure 1.1, it does not display the distribution function values, and an act of imagination is required to substitute some visual image of $F_X\left(I_1^{(m)} \times I_2^{(n)}\right)$ for Pmn.

But Figure 1.3 gives a partial histogram for a pair of joint random variables. Even this becomes impractical when there are three or more joint random variables; and then the helpful histogram description of random variation has to be abandoned in favor of purely analytical expressions of the form $X \simeq x[\Omega_X, F_X]$.

For an elementary basic random variation such as the one in Table 1.1, the

	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	
0 – 50	0.06	0.09	0.06	0.06	0.03	0.3
50 – 100	0.04	0.06	0.04	0.04	0.02	0.2
100 – 150	0.08	0.12	0.08	0.08	0.04	0.4
150 – 200	0.02	0.03	0.02	0.02	0.01	0.1
	0.2	0.3	0.2	0.2	0.1	

Table 1.5: Independent joint-basic observable.

table and histogram (Figure 1.1) convey a sense of the random variability: the potential datum x, the sample space Ω_X, and the values of the potentiality distribution function F_X. In other words, all the elements of $X \simeq x[\Omega_X, F_X]$ are displayed, or at least indicated, in the table and histogram.

It is difficult to provide a similarly intuitive display for joint random variability. Suppose there is independence in the joint height–weight data of Table 1.4. In that case the twenty joint likelihoods would be as in Table 1.5. The sample space, sample joint data values, and potentiality distribution function values on joint interval events can be partially illustrated in a two-dimensional histogram, as indicated in Figure 1.3.

Given joint random variation $X = (X_1, X_2)$, a random variable contingent on (X_1, X_2) is

$$f(X) = (X_1 - \mathrm{E}[X_1])\,(X_2 - \mathrm{E}[X_2])\,, \quad f(x) = (x_1 - \mathrm{E}[X_1])\,(x_2 - \mathrm{E}[X_2])\,,$$

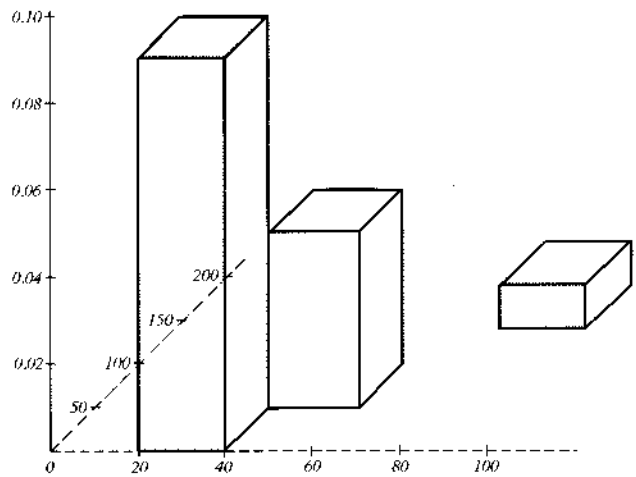

Figure 1.3: Part of histogram for two independent joint random variables.

from which is obtained the *covariance* of $X = (X_1, X_2)$; the covariance being

$$\mathrm{Cov}[X] = \mathrm{Cov}[(X_1, X_2)] = \mathrm{E}[f(X)] = \mathbf{E}\left[f_1(X_1) f_2(X_2)\right],$$

where

$$f_j(X_j) = X_j - \mathbf{E}\left[X_j\right], \quad j = 1, 2.$$

For the joint random variable of Table 1.3, the covariance can be estimated by choosing a sample value (x_1, x_2) in $I_1 \times I_2$ for each of the joint intervals I_1, I_2, and then calculating the sum of twenty terms:

$$\begin{aligned} \sum f(x_1, x_2) F_X(I_1 \times I_2) &= \sum (x_1 - \mathrm{E}[X_1])\,(x_2 - \mathbf{E}[X_2])\, F_{(X_1, X_2)}(I_1 \times I_2), \\ &= \sum (x_1 - \mathrm{E}[X_1]) \left(\sum (x_2 - \mathrm{E}[X_2])\, F_X(I_1 \times I_2)\right); \end{aligned}$$

where

$$\mathrm{E}[X_j] = \sum x_j F_{X_j}(I_j), \quad j = 1, 2.$$

Again, this estimate of the covariance has the form of a Riemann sum approximation to a Riemann–Stieltjes double integral

$$\mathbf{E}[f(X)] = \int_{\mathbf{R}\times\mathbf{R}} f(x) dF_X;$$

that is,

$$\mathrm{E}[f(X_1, X_2)] = \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f(x_2, x_2) dF_{(X_1, X_2)} \right). \tag{1.8}$$

If the two joint random variables are independent as indicated in Table 1.5, then

$$\mathbf{E}[f(X_1, X_2)] = \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f(x_2, x_2) dF_{X_1} \right) dF_{X_2};$$

and the Riemann sum estimate of the covariance can be calculated as

$$\sum \left((x_1 - \mathrm{E}[X_1]) \left(\sum (x_2 - \mathrm{E}[X_2])\, F_{X_2}(I_2) \right) F_{X_1}(I_1) \right),$$

which is

$$\left(\sum (x_1 - \mathrm{E}[X_1])\, F_{X_1}(I_1) \right) \left(\sum (x_2 - \mathrm{E}[X_2])\, F_{X_2}(I_2) \right).$$

By definition of the expectation $\mathbf{E}$, each of the factors is zero, so independence implies covariance zero.

Example 13 *To illustrate numerically, suppose the following sample values are used to calculate Riemann sum estimates for the covariance of the joint random variables in Table 1.5:*

$$\begin{aligned} x_1 &: \quad 10 \quad 30 \quad 50 \quad 70 \quad 90, \\ x_2 &: \quad 25 \quad 75 \quad 125 \quad 175. \end{aligned}$$

Then the corresponding sample joint data consist of twenty pairs:

$$\begin{array}{cccc} (10,25) & (30,25) & \cdots & (90,25) \\ (10,75) & (30,75) & \cdots & (90,75) \\ \vdots & & & \vdots \\ (10,175) & (30,175) & \cdots & (90,175) \end{array}$$

The expected values of X_1 and X_2 are obtained from the marginal distributions in Table 1.5:

$$\begin{aligned} \mathrm{E}[X_1] &= \textstyle\sum_{j=1}^{5} x_1^{(j)} F_{X_1}\left(I_1^{(j)}\right) = 10 \times 0.2 + \cdots + 90 \times 0.1 = 42, \\ \mathrm{E}[X_2] &= \textstyle\sum_{k=1}^{4} x_2^{(k)} F_{X_2}\left(I_2^{(k)}\right) = 25 \times 0.3 + \cdots + 175 \times 0.1 = 90. \end{aligned}$$

The sample estimate of covariance is then given by the Riemann sum calculation $\sum_{j=1}^{5}\sum_{k=1}^{4}\left(x_1^{(j)}-42\right)\left(x_2^{(k)}-90\right)F_{(X_1,X_2)}\left(I_1^{(j)}\times I_2^{(k)}\right)$, so independence implies that $\mathrm{Cov}[(X_1,X_2)]$ is

$$\left(\sum_{j=1}^{5}\left(x_1^{(j)}-42\right)F_{X_1}\left(I_1^{(j)}\right)\right)\left(\sum_{k=1}^{4}\left(x_2^{(k)}-90\right)F_{X_2}\left(I_2^{(k)}\right)\right).$$

Numerical calculation of the Riemann sum in the format of (1.8) gives

$$(10-42)\times(25-90)\times 0.06+\cdots+(90-42)\times(175-90)\times 0.01=0;$$

so $\mathrm{Cov}[(X_1,X_2)]=0$, confirming that independence implies covariance zero. ◯

1.12 Independence

The notion of joint variability extends to arbitrary collections $(X_t)_{t\in T}$ of random variables, and the notion of independence is extended accordingly. In Table 1.11 a third measurement—the age of the individual—could be included along with joint measurement of an individual's weight and height; giving $T=\{1,2,3\}$, and random variables X_j with intervals I_j and potentiality distributions $F_{X_j}(I_j)$, $j=1,2,3$. The joint observation $X=(X_1,X_2,X_3)$ could perhaps be illustrated by means of a table like Table 1.4 above, but it is difficult to display three variables in tabular format, and when there are more than three variables, tables become unmanageable.

Tables can display a single variable in a vertical direction, as in Table 1.1, and two variables can be displayed in vertical and horizontal directions, as in Table 1.4. But that is practically the full extent of the tabular format. For three or more variables, it is possible instead to use analytic Cartesian formulation, and, to a limited extent, the corresponding geometric Cartesian representation.

Observations (occurrences or joint data) $x = (x_1, x_2, x_3)$ and joint events (intervals) $I = I_1 \times I_2 \times I_3$ can be represented by means of orthogonal axes in three dimensions, as in Figure 1.4. Like Figure 1.2, this diagram shows how the domain of the joint random variation is partitioned by sets I for Riemann sum calculation; and it does not attempt histogram-type display of distribution function values $F_X(I)$.

The joint potentialities are

$$F_X(I) = F_{(X_1,X_2,X_3)}(I_1 \times I_2 \times I_3),$$

and the triple-joint random variations $X = (X_1, X_2, X_3)$ are said to be independent if

$$F_{(X_1,X_2,X_3)}(I_1 \times I_2 \times I_3) = F_{X_1}(I_1)F_{X_1}(I_2)F_{X_1}(I_3)$$

for all possible choices of the intervals I_1, I_2, I_3 in the domains Ω_{X_j} of the joint random variations.

In this case there are more marginals:

$$F_{X_1}(I_1) = F_X(I_1 \times \mathbf{R} \times \mathbf{R}), \quad F_{(X_1,X_2)}(I_2 \times I_2) = F_X(I_1 \times I_2 \times \mathbf{R}),$$

and so on, in various combinations of marginals. As before, consistency conditions apply to these joint potentialities.

Given a family of joint random variations, sub-families can be independent while the family members as a whole are not independent. Here is an illustration. Suppose two independent random variables each has potential data values -1, 1, with potentialities 0.5, 0.5. The random variation formed by multiplying the two potential data values is independent of each of the first two, but the

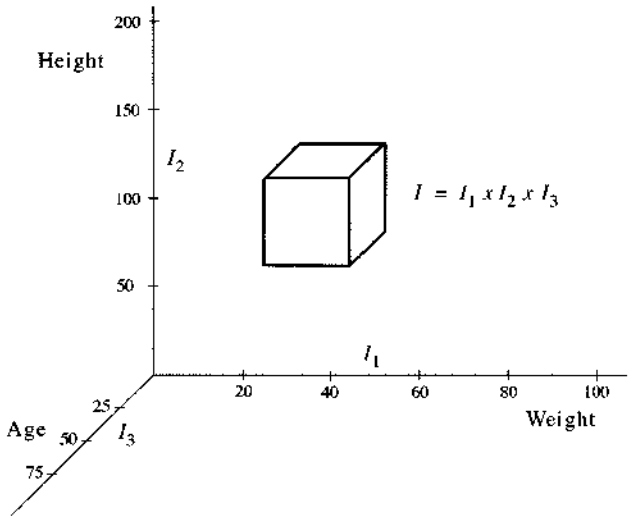

Figure 1.4: Framework for joint random variables.

three random variations, considered jointly, are not independent. Expressing this more formally, for $j = 1, 2$ we have $X_j \simeq x_j[\Omega_{X_j}, F_{X_j}]$ with $\Omega_{X_j} = \{1, -1\}$ and $F_{X_j}(1) = F_{X_j}(-1) = 0.5$.

The random variable $f(X) = f(X_1, X_2) := X_1X_2$ can be constructed, firstly as an observable contingent on the joint observable $X = (X_1, X_2)$, and secondly as an elementary observable Y. Accordingly, X_1X_2 can be regarded as a contingent observable with the following sample space:

$$\begin{aligned}\Omega &= \Omega_{X_1} \times \Omega_{X_2} \\ &= \{1, -1\} \times \{1, -1\} \\ &= \{(1, 1),\ (1, -1),\ (-1, 1),\ (-1, -1)\}\,.\end{aligned}$$

The potentiality of each of these four joint observations is, by the independence of X_1 and X_2,

$$F_{X_1}(i)F_{X_2}(j) = 0.5 \times 0.5 = 0.25$$

for $i, j = \pm 1$. The potential outcomes of the contingent observable X_1X_2 corresponding to each of these four joint observations are, respectively,

$$1, \quad -1, \quad -1, \quad 1,$$

with potentialities, respectively,

$$0.25, \quad 0.25, \quad 0.25, \quad 0.25,$$

and expected value:

$$\mathrm{E}\,[X_1X_2] = (1\times1)\times0.25+(1\times-1)\times0.25+(-1\times1)\times0.25+(-1\times-1)\times0.25 = 0.$$

The *distinct* outcomes of $f(X) = X_1X_2$ are

$$1, \quad -1$$

with potentialities

$$\sum_{i,j}\{F_{X_1}(i)F_{X_2}(j) : ij = 1\} = 0.5, \qquad \sum_{i,j}\{F_{X_1}(i)F_{X_2}(j) : ij = -1\} = 0.5,$$

respectively; thus establishing that the contingent random variable $f(X_1, X_2) = X_1X_2$ can be expressed[10] as an elementary random variable Y whose sample space is

$$\Omega_Y = \Omega = \{1,\ -1\}$$

with potentialities

$$F_Y(1) = 0.5 \quad \text{and} \quad F_Y(-1) = 0.5$$

[10]Theorem 82 of Chapter 5 deals with pointwise equivalence of different representations of random variables. A weaker, non-pointwise form of equivalence is introduced in Theorems 44 and 236.

and expected value

$$\mathrm{E}[Y] = 1 \times 0.5 + (-1) \times 0.5 = 0,$$

giving

$$\mathrm{E}\,[X_1 X_2] = \mathrm{E}[Y],$$

where each of the expected values is calculated using a different sample space and different set of potentialities.

The joint outcomes of X_1 and Y are

$$(1,1),\ \ (1,-1),\ \ (-1,1),\ \ (-1,1)$$

with potentialities 0.25 in each case, so X_1 and Y are independent. Likewise X_2 and Y. The joint outcomes, with potentialities, of X_1, X_2, Y are:

Joint outcome	**Potentiality**
$(1,1,1)$	0.25
$(1,1,-1)$	0
$(1,-1,1)$	0
$(1,-1,-1)$	0.25
$(-1,1,1)$	0
$(-1,1,-1)$	0.25
$(-1,-1,1)$	0.25
$(-1,-1,-1)$	0

(1.9)

Independence of X_1, X_2 and Y would require that the potentiality of each of the joint "possibilities" listed in this table should equal

$$F_{X_1}(i)F_{X_2}(j)F_Y(k) = 0.5 \times 0.5 \times 0.5 = 0.125,$$

for $i, j, k = \pm 1$. This is not the case. In fact, four of the eight "possibilities" or joint "outcomes" listed are not actually possible. Thus X_1, X_2, Y are not independent.

Note that if Y is denoted by X_3, then we obtain a joint observable $X = (X_1, X_2, X_3)$ with representation

$$X \simeq x[\Omega_X, F_X],$$

where $x = (x_1, x_2, x_3) \in \Omega_X$. The sample space Ω_X can consist of the outcomes in (1.9). Or we can regard the sample space as $\mathbf{R} \times \mathbf{R} \times \mathbf{R}$ with the outcomes of (1.9) embedded in it. The potentiality distribution function F_X consists of the potentialities of (1.9) or some equivalent in $\mathbf{R} \times \mathbf{R} \times \mathbf{R}$. Then each basic X_j can be represented in contingent form by

$$X_j = f_j(X), \quad x_j = f_j(x), \quad j = 1,2,3,$$

where each $x = (x_1, x_2, x_3)$ is a joint outcome in (1.9), and the potentiality of $x_j = a$ occurring is

$$\sum\{F_X(x) : f_j(x) = f_j(x_1, x_2, x_3) = x_j = a\}.$$

Thus the contingent variables $f_j(X)$ have representations

$$f_j(X) \simeq f_j(x)[\Omega_X, F_X], \quad j = 1, 2, 3,$$

which are equivalent to the representations

$$X_j \simeq x_j[\Omega_{X_j}, F_{X_j}], \quad j = 1, 2, 3.$$

1.13 Stochastic Processes

For larger (and infinite) collections of joint observables, the Cartesian representation is more difficult to illustrate in a diagram, but the concepts follow the pattern already evident. Suppose we have a joint-basic variation

$$X = (X_t)_{t \in T}$$

of basic observables X_t where each t belongs to a (finite or infinite) set T. Then the classes of *joint events* that should be considered are the possibilities that, for any finite n and any selection of

$$t_1, \ldots, t_n,$$

the observed values (or joint data-values) $x_{t_1}, \ldots, x_{t_n}$ of $X_{t_1}, \ldots, X_{t_n}$ satisfy

$$x_{t_j} \in I_{t_j}, \quad 1 \leq j \leq n,$$

which is an instance of a class $\mathbf{I}$ of events I that can be denoted by

$$I = I_{t_1} \times \cdots \times I_{t_n} \times \mathbf{R}^{T \setminus \{t_1, \ldots, t_n\}}, \tag{1.10}$$

with potentialities $\{F_X(I)\}$. The observables X_t, $t \in T$ are independent if, for each choice of n, $t_1, \ldots, t_n$, and, correspondingly, all choices of I_{t_j}, we have

$$F_X(I) = \prod_{j=1}^{n} F_{X_j}(I_{t_j}).$$

This definition is consistent with the previous definition of independence of joint observables when T consists of just two elements or three elements.

In a sense, when T is an infinite set, **all** potentialities are marginal, since an infinite product

$$\prod_{j=1}^{\infty} F_{X_j}(I_{t_j})$$

will usually be zero, so it is more useful to consider potentialities in the joint *cylindrical intervals* I of (1.10) above. Since the potentiality of any cylindrical

interval is marginal, consistency conditions apply to all of these joint potentiality values.

A random variable in this context is a real- or complex-valued function $f(X)$ of the joint random variation $X = (X_t)_{t\in T}$,

$$f(X) \simeq f(x)\,[\Omega_X, F_X(I)]\,,$$

where $I \in \mathbf{I}$, and, with $x = (x_t)_{t\in T} = x_T \in \mathbf{R}^T = \Omega_X$, the set of possible values of $f(X)$ is

$$\{f\,((x_t)_{t\in T}) : x_t \in \Omega_{X_t},\ \ t \in T\}\,. \tag{1.11}$$

Table 1.2 describes a procedure for calculating the expected value of a random variable involving a finite number of joint observations (T finite). It might be anticipated that, with T infinite, the expected value $\mathbf{E}[f(X)]$ could be estimated by a similar Riemann sum calculation:

$$\sum f\,((x_t)_{t\in T})\,F_X(I),\ \text{ or }\ \sum f(x)F_X(I),$$

where the joint intervals I (cylindrical intervals (1.10)) partition the domain (1.11) of the random variable.

In the following chapters the ideas outlined above are developed in more detail. One objective, as indicated earlier, is to present the main results of probability theory in a Riemann sum format. In other words, the aim is to use Riemann sums (corresponding to the intuitive approach of Tables 1.2 and 1.4 above) and Riemann sum-based integration, instead of the traditional measure theory and Lebesgue integration.

Another objective of this book is to formulate the Feynman path integral theory of quantum mechanics as a branch of probability theory. The formal similarities between the Feynman theory and the theory of Brownian motion are well known. It was demonstrated by Muldowney [162] that these two theories could be expressed in a common framework of Stieltjes-type integrals using Riemann sum constructions. In this book these subjects are formulated in a probability framework based on the Stieltjes-complete and Burkill-complete versions of the Henstock integral [94].