Exceptional Lebesgue densities and random Riemann sums

Jack Grahl

Submitted in partial fulfilment of the requirements of candidature to Doctor of Philosophy of University College London, 2011

I, Jack Grahl, confirm that the work presented in this thesis is my own. Where the work of others has been used, I confirm that this has been indicated in the thesis.

Signed on 30th June 2011

Abstract

We will examine two topics in this thesis. Firstly we give a result which improved a bound for a question asking which values the Lebesgue density of a measurable set in the real line must have (joint work with Toby O'Neil and Marianna Csörnyei). We also show how this result relates to the results obtained by others. Secondly, we give several results which indicate when a Lebesgue measurable function has a random Riemann integral which converges, in either the weak and strong sense.

A Lebesgue measurable set A, subset of $\mathbb{R}$, has density either 0 or 1 at almost every point. Here the density at some point x refers to the proportion of a small ball around x which belongs to A, in the limit as the size of the ball tends to 0. Suppose that A is not either a nullset, which has density 0 at every single point, or the complement of a nullset, which similarly has density 1 everywhere. Then there are certain restrictions on the range of possible values at those exceptional points where the density is neither 0 nor 1. In particular, it is now known that if $\delta < 0.268486\ldots$, where the exact value is the positive root of $8\delta^3 + 8\delta^2 + \delta 1 = 0$, then there must exist a point at which the density of A is between δ and $1 - \delta$, and that this does not remain true for any larger value of δ.

This was proved in a recent paper by Ondřej Kurka. Previous to his work our result given in this thesis was the best known counterexample. We give the background to this, construct the counterexample, and discuss Kurka's proof of the exact bound.

The random Riemann integral is defined as follows. Given a Lebesgue mea-

surable function $f : [0,1] \to \mathbb{R}$ and a partition of $[0,1]$ into disjoint intervals, we can choose a point belonging to each interval, independently and uniformly with respect to Lebesgue measure. We then use these random points to form a Riemann sum, which is itself a random variable. We are interested in knowing whether or not this random Riemann sum converges in probability to some real number. Convergence in probability to r means that the probability that Riemann sum differs from r by more than ε, is less than ε, provided that the maximum length of an interval in the partition is sufficiently small.

We have previously shown that this type of convergence does take place provided that f is Lebesgue integrable. In other words, the random Riemann integral, defined as the limit in probability of the random Riemann sums, has at least the power of the Lebesgue integral. Here we prove that the random Riemann integral of f does not converge unlesss $|f|^{1-e}$ is integrable for $e > 0$ arbitrarily small. We also give another, more technical, necessary condition which applies to functions which are not Lebesgue integrable but are improper Riemann integrable.

We have also done some work on the question of almost sure convergence. This works slightly differently. We must choose, in advance, a sequence of partitions $(\mathcal{P}_n)_{n=1}^{\infty}$, with the size of the intervals of $\mathcal{P}_n$ tending to zero. We form a probability space on which we can take random Riemann sums independently on each partition of the sequence. Almost sure convergence means that the sequence of random Riemann sums converges to some (unique) limit with probabililty 1 in this space. There are two complementary results; firstly that almost sure convergence holds if the function is in L^p and the sequence of partition sizes is in l^{p-1} for some $p \geq 1$. Secondly, we have a partial converse which only applies to nonnegative functions, and if the ratio between the lengths smallest and biggest intervals in each partition is bounded uniformly. This says that if for some $p \geq 1$ f is not in L^p and the partition sizes are not in l^{p-1}, then the sequence of Riemann sums diverges with probability 1.

Contents

Chapter 1

Introduction

This thesis examines two topics in real analysis and measure theory. The connection between the two is the Lebesgue density theorem for measurable sets in $\mathbb{R}$ and its analogue for measurable functions. This theorem establishes that measurable sets and functions can be well approximated, in measure theoretical terms, by regular sets or functions, for example by open sets and continuous or even analytic functions.

The first part of the thesis is about Lebesgue measure on the real line. Chapter 2 provides the background on Lebesgue measure and the Lebesgue integral. The Lebesgue density theorem and Lebesgue differentiation theorem are explained. We remind the reader of these classical results in order to prepare the way for the later chapters. The final section of this chapter places a limit on the strength of the Lebesgue density theorem. We show here that there is no bound on the speed of convergence of Lebesgue densities which holds uniformly for all measurable sets. This is true even if we require bounds that hold not almost everywhere but only on a set arbitrarily close to full measure. There are two reasons why we consider this to be important. Firstly we will use this result, in the second part of the work, to prove that various constructions involving Riemann sums do not converge. Secondly, we are interested in the question of which functions are the densities of measurable sets or of measurable functions,

and we hope that understanding the convergence of densities will help with this.

This last question is a fascinating topic which raises many questions. Of course from a measure-theoretic perspective it is completely solved, since all measurable functions (and no others) are almost everywhere equal to the Lebesgue density of some function, namely themselves. However it appears that little is known concerning what topological characterization such densities might have. In fact, as well as Lebesgue densities, other limiting procedures of critical importance to analysis give rise to the same functions; for example the (upper, Césaro) sums of Fourier series. In each case the weak limit is understood perfectly and the pointwise limit much less well. Although it might seem to be a confusion to look at the everywhere (as opposed to almost everywhere) values of constructions which are firmly part of measure theory, we believe there may be interesting facts here to be discovered. We should also remember that measure theory itself, along with the notions of a.e. and L^p limits, came out of the desire of analysts originally to understand exactly what functions can be expressed as limits in this way.

Chapter 3 deals with just a small part of this question. A Lebesgue measurable set A has density either 0 or 1 at almost every point. Suppose that A is not a nullset, which has density 0 at every single point, or the complement of a nullset, which similarly has density 1 at every point. Then there are certain restrictions on the values at those exceptional points where the density is neither 0 nor 1. In particular, it is now known that if $\delta < 0.268486\ldots$, where the exact value is the positive root of $8x^3 + 8x^2 + x - 1 = 0$, then there must exist a point at which the density of A is between δ and $1 - \delta$, and that this is not true for any larger value of δ.

This was proved in a recent paper by Ondřej Kurka. Previous to this result, several advances were made, building on work by András Szenes. These included an improvement to the upper bound which was proved by the author together with Marianna Csörnyei and Toby O'Neil. To understand our result, it is necessary to be aware of the reformulation of the problem given by Szenes.

The proof of Kurka's theorem, which is very complex, also uses notations, concepts and results from Szenes. Therefore we have considered it advisable to give an exposition of all the work leading up to Kurka's result, and to discuss some of the ideas involved in that result. We hope that in doing so we have made it somewhat easier to understand all the literature on this problem.

The second part of the thesis begins with Chapter 4. This chapter provides the background on Riemann sums, and their use in defining integrals of functions on the real line. As well as defining and motivating the key notions, and discussing the classical Riemann, Kurzweil-Henstock and McShane integrals, we also mention some less well-known integrals which are constructed using Riemann sums. By doing so we hope to highlight which properties of a measurable function we need to be aware of when we try to prove or disprove the convergence of Riemann sums in later chapters.

The following chapter defines the notion that is central to the second part of the thesis; the random Riemann integral. The random Riemann integral is defined as follows. Given a Lebesgue measurable function $f : [0,1] \to \mathbb{R}$ and a partition of $[0,1]$ into disjoint intervals, we can choose a point belonging to each interval, independently and uniformly with respect to Lebesgue measure. We then use these random points to form a Riemann sum with the tag point of each interval being the random point which belongs to it. This Riemann sum is itself a random variable. We can ask two questions about this random Riemann sum. Firstly whether or not it converges in probability to some real number. Convergence in probability to r means that the probability that the Riemann sum differs from r by more than ε, is less than ε, provided that the partition in question is sufficiently small. Secondly, whether given a sequence of partitions there is some number to which the random Riemann sums converge with probability 1. In this case we say that almost sure convergence holds.

In fact, we have previously shown that convergence in probability does hold if the function f is in L^1, and that in this case the limit is the Lebesgue integral of f. This is an easy consequence of the Lebesgue density theorem that was

included in the author's MSci thesis. We repeat the proof here for completeness. Following this, we give several results about the functions and sequences of partitions for which almost sure convergence holds. In particular, if the sequence of sizes of the partitions is in ℓ^{p-1} and the function is in L^p for some value of p, then the sequence of random Riemann sums converges almost surely. The limit is the same as for convergence in probability – the Lebesgue integral of f. We go on to show that this is not a necessary condition. The function can be much larger than this and convergence may still hold, provided that the sets where f takes values roughly equal to the reciprocal of the partition sizes are not too big. We make this statement precise with a theorem giving a necessary condition for convergence not to hold, and a counterexample where it does hold.

The second part of the chapter returns to the question of convergence in probability. We prove that the random Riemann sums do not converge in probability unless $|f|^{1-e}$ is integrable for any $e > 0$. We also give another, more technical, necessary condition which applies to functions which are not Lebesgue integrable but are improper Riemann integrable. This last result stands out – most of the other theorems in the chapter, while apparently concerning functions on the unit interval, can be adapted to any reasonable probability space. This last theorem however, uses topological properties of the interval, which means that in this case we cannot regard the Riemann sums simply as a convenient way of expressing a result about probability distributions in general.

The next two chapters describe several constructions which are similar to the random Riemann integral. In Chapter 6 we outline the so-called Mycielski Riemann sum. Like the random Riemann sum, this is a random variable in the space of Riemann sums on some function f. Whereas the random Riemann sum depends on the deterministic choice of a partition, both the intervals and the sample points of the Mycielski Riemann sum are random. The tag points are chosen at random in some probability space, here the unit interval, and the partition of the interval is given by the Voronoi tesselation using these points. The Voronoi tesselation is a division of a space into regions, each of which

consists of all the points which are closer to a given member of a finite set than to any other member of that set. So the random Mycielski Riemann sum expresses the effect of approximating $f(x)$ by the value of f at the closest point to x among some finite collection of sample points. Again we can ask whether these Riemann sums converge in probability or almost surely. In this chapter we explain the work of Mycielski and Fremlin on this topic. We also make some remarks which we believe are relevant to an unsolved problem – whether or not for all measurable $f : [0,1] \to \mathbb{R}$, the Mycielski step functions converge almost surely in L^1 to f.

Chapter 7 is about the first-return integral. The first-return integral also given by a limit of Riemann sums. We take some sequence $(x_n)_{n\in\mathbb{N}}$, where the point set $\{x_n : n \in \mathbb{N}\}$ is dense in $[0,1]$. Given a partition, the tag point of each interval is the first term of (x_n) which belongs to that interval. This is the first-return Riemann sum. A function is first-return integrable with respect to a sequence of partitions if the Riemann sums converge almost surely. There are connections to both the previous chapters. First of all, a necessary condition for first-return integrability is first-return recoverability for the same sequence of partitions. This turns out to be equivalent to pointwise convergence of the Mycielski step functions. We say something about those sequences and functions for which first-return recovery holds and/or the first-return integral converges. Then we consider the probabilistic versions of both of these notions. This means that we ask when they hold not for a single sequence, but for almost all sequences in a natural probability space.

The probabilistic version yield the connection between this chapter and Chapter 5. In fact for a random sequence in this probability space, the distribution of the first-return Riemann sum on a fixed partition is the same as that of the random Riemann sum. Therefore some of the results of Chapter 5 can be transferred to this context. Since however we do not have independence between Riemann sums on different partitions in the sequence, almost sure first-return integrability is not equivalent to random Riemann integrability. In fact,

we can establish here that random Riemann integrability is stronger than first-return integrability, which is stronger than Mycielski integrability. The chapter ends with a section on the Kieffer-Stanojević integral, a special case of the probabilistic first-return integral, for which it is straightforward to prove almost sure convergence.

These three chapters all involve the Lebesgue density theorem. In several cases convergence in probability can be proved quickly by using this theorem. Measurable functions are approximately continuous close to almost every point. If we have some stochastic method of choosing tag points, this ensures that we do not pick those points for which approximate continuity does not hold. We then obtain convergence of the integral, under weaker conditions on the partitions that those required for the Kurzweil-Henstock or McShane integrals. We can then consider the almost sure convergence of each of these Riemann sum constructions as a statement similar to but stronger than the conclusion of the Lebesgue density theorem. The exact construction will determine what type of function we need for almost sure convergence not to hold. In some cases a function of a particular size is required, elsewhere a bounded function with sufficiently slowly converging densities, as was shown to exist in Chapter 2, is enough.

Finally in Chapter 8, we describe the author's joint work with Togo Nishiura on another problem. This concerns permutations of the unit interval, or functions from $[0,1]$ to $[0,1]$ which act as translations on each interval of a partition of $[0,1]$. We are interested here in those measurable functions which are the almost everywhere pointwise limits of a sequence of such permutations. Our result shows that these are exactly the measure-preserving functions, and therefore that any measurable function can be expressed (to within a Lebesgue null set) as the composition of such a function with its distribution function. This work was motivated by the study of random Riemann sums – it lays the way open for a theory which would give convergence theorems for a general class of constructions, including all those of Chapters 5, 6 and 7. The final section of

the chapter explains this motivation and gives some suggestions as to how this could be carried out.

Chapter 2

Lebesgue measure and integral

In this chapter we remind the reader of some of the basic definition and properties related to Lebesgue measure and the space of Lebesgue measurable functions. The most important results are the Lebesgue density theorem and the Lebesgue differentiation theorem. These will be used throughout the remainder of the thesis. The final section shows that can be no asymptotic bounds on the convergence of the Lebesgue density theorem. All the results in this chapter are classical and well known; except for those of the final section. We are not aware of a reference for the results in the final section but they are simple and undoubtedly already known.

2.1 Preliminaries

2.1.1 Lebesgue measure

We assume that the reader is familiar with the definitions of $\mathbb{R}$, of a *Lebesgue measurable set* $A \subset \mathbb{R}$ and the *Lebesgue measure* of such a set, denoted interchangeably by $\lambda(A)$ or by $|A|$. 'Almost everywhere' always means outside a

Lebesgue null set and 'measurable' means Lebesgue measurable.

We denote the unit interval $[0,1]$ by $\mathbb{I}$.

2.1.2 The Lebesgue integral

We use the standard definitions of a real-valued function on $\mathbb{I}$ being *Lebesgue integrable*, of its *Lebesgue integral* and of L^1, the space of Lebesgue integrable functions under Lebesgue equivalence. When we write $\int f$ it refers to the Lebesgue integral on $\mathbb{I}$.

We will use some more general notions of measure and integral, particularly in the context of probability measures and random variables. See [26] for a guide to probability theory or [21] and [22] for a reference on all required aspects of measure theory.

As well as the traditional definition of L^p and of L^p convergence for $p \geq 1$ we will also use the notion of a weak L^p space.

Definition 2.1.1 A Lebesgue-measurable function is in *weak* L^1 if

$$\lambda(\{x : |f(x)| > y\}) = \mathsf{O}(1/y).$$

The *weak* L^1 *norm* is

$$\| f \|_1^{weak} := \sup_{y>0} y\lambda\{x : |f(x)| > y\}.$$

Weak L^p is defined analogously. Clearly if a function is in weak L^p, it is in $L^{p-\varepsilon}$ for every $\varepsilon > 0$.

Remark The *descriptive definition* of the Lebesgue integral is the following. A function f is Lebesgue integrable iff there exists some absolutely continuous function F such that F', the derivative of F, is almost everywhere equal to f. The function F is called a *primitive* of f. In fact the *indefinite integral of* f, defined by

$$F(x) = \int_0^x f$$

is also a primitive.

The indefinite integral of a function in L^1 is absolutely continuous. The indefinite integral of a function in L^p, for $1 < p < \infty$, is Hölder continuous with Hölder exponent $1 - \frac{1}{p}$, and as can easily be seen, the indefinite integral of an L^∞ function is Lipschitz.

2.2 Lebesgue density and derivative

2.2.1 Lebesgue density

Definition 2.2.1 If A and B are both Lebesgue measurable sets, the *relative measure of A in B*, $\lambda(A|B)$, is defined by

$$\lambda(A|B) := \frac{\lambda(B \cap A)}{\lambda(B)}.$$

Definition 2.2.2 If A is a Lebesgue measurable set and x is a point in $\mathbb{R}$, the *upper density of A at x* is

$$\overline{d}(x, A) := \limsup_{\varepsilon \to 0} \lambda(A|[x - \varepsilon, x + \varepsilon])$$

and similarly the *lower density of A at x* is

$$\underline{d}(x, A) := \liminf_{\varepsilon \to 0} \lambda(A|[x - \varepsilon, x + \varepsilon]).$$

The *density of A at x*, $d(x, A)$, exists if the lower density and upper density of A at x are equal, and is equal to their common value.

The Lebesgue density theorem is a fundamental fact about measurable sets.

Theorem 2.2.3 (Lebesgue density theorem) *Let A be a Lebesgue measurable subset of $\mathbb{I}$. The set of points $\{x \in A : \underline{d}(x, A) < 1\}$ is Lebesgue null.*

Corollary *Let A be a Lebesgue measurable subset of $\mathbb{I}$. The Lebesgue density of A exists almost everywhere, and it is equal to 1 at almost every point of A, and to 0 at almost every point of the complement of A* □

We refer to points at which the density of A is 1 as *density points of A*.

2.2.2 The Lebesgue derivative

Definition 2.2.5 Analogous to Lebesgue density for measurable sets is the *Lebesgue derivative* for measurable functions. This is defined for locally integrable functions as

$$Df(x) := \lim_{\substack{|B|\to 0 \\ B\ni x}} \frac{1}{|B|}\int_B f(x),$$

where B ranges over the intervals which include x. It is the derivative of the indefinite integral of f.

The Lebesgue differentiation theorem says that this limit exists and is equal to $f(x)$ for almost every point x. It is a consequence of Hardy's maximal theorem.

Remark We have defined the Lebesgue density to be the *symmetric* density and the Lebesgue derivative conversely to be the *bilateral* derivative. The first means that we take limits over intervals centered at x whose length tends to zero, the second that the limit is over all intervals which contain x, as their length tends to zero. Clearly we can define both density and derivative in either of these two ways, and the existence of the bilateral density (derivative) will imply the existence of the symmetric density (derivative).

In fact, Theorem 2.2.3 could also be proved for the bilateral Lebesgue density. In what follows, we give properties of the bilateral Lebesgue derivative, since these are strong enough to imply the corresponding properties of the symmetric version. However, we persist in using only the symmetric version of the density, since it is the subject of the next chapter.

Remark If A is a measurable set, an application of Lebesgue density theorem for both A and $\mathbb{I}\backslash A$ shows that this theorem is true for the indicator function

χ_A, which is 1 on A and 0 on the complement of A.

Theorem 2.2.8 (Lebesgue differentiation theorem) *For all $f \in L^1$, $Df(x)$ exists and is equal to $f(x)$ for almost every x.*

Definition 2.2.9 The *Hardy-Littlewood maximal operator* is the operator given by

$$Mf(x) := \sup_{B \ni x} \frac{1}{|B|} \int_B f(x).$$

Here B ranges over the set of intervals containing x. The operator maps integrable functions $\mathbb{R} \to \mathbb{R}$ to measurable functions $\mathbb{R} \to \mathbb{R}$.

Theorem 2.2.10 (Hardy's maximal theorem) *If f is in L^1, then Mf is in weak L^1 and*

$$\lambda(\{z : Mf(z) > x\}) \leq \frac{c \int f}{x}$$

for all x and some absolute constant c.

We can express this theorem by saying that M is of weak type $(1,1)$ – it is continuous from L^1 to weak L^1.

Proof (Lebesgue differentiation theorem). Since $f \in L^1$, there exists a continuous function g such that $\int |f - g| < \varepsilon$. Let $c > 0$. We wish to estimate the size of the set of points x on which:

$$\lim_{r \to 0} \sup_{\substack{B \ni x \\ |B| < r}} \left| \frac{1}{|B|} \int_B f - f(x) \right| > c. \tag{2.1}$$

We can estimate this by the union of the sets

$$\left\{ x : \lim_{r \to 0} \sup_{\substack{B \ni x \\ |B| < r}} \left| \frac{1}{|B|} \int (f - g) \right| > \frac{c}{3} \right\} =: A_1,$$

$$\left\{ x : \lim_{r \to 0} \sup_{\substack{B \ni x \\ |B| < r}} \left| \frac{1}{|B|} \int g - g(x) \right| > \frac{c}{3} \right\} =: A_2$$

and

$$\left\{x : |f(x) - g(x)| > \frac{c}{3}\right\} =: A_3.$$

By Markov's inequality the third set has measure at most $\frac{3}{c}\int |f-g| \leq \frac{3\varepsilon}{c}$. The second set is empty since g is continuous, and the first set is contained within

$$\left\{x : \sup_{B \ni x} \frac{1}{|B|}\int |f-g| > \frac{c}{3}\right\}$$

which by Hardy's maximal theorem has measure at most $\frac{c'\int |f-g|}{\frac{c}{3}} \leq \frac{3c'\varepsilon}{c}$, where c' is the constant from the theorem. Therefore the set of points for which (2.1) holds has measure in $\mathsf{O}(\varepsilon)$ as $\varepsilon \to 0$. Since c is arbitrary this proves that

$$\left|\frac{1}{|B|}\int_B f - f(x)\right| \to 0 \text{ as } |B| \to 0$$

for almost every x. ■

There is an L^p version of the maximal theorem for $1 < p < \infty$ which says that Mf is in L^p if f is in L^p. This leads naturally to a version of the Lebesgue differentiation theorem, which says that if f is in L^p, Df converges in L^p to f. However neither the maximal theorem nor the Lebesgue differentiation theorem hold for the L^1 norm.

In fact, for functions defined on the whole line $\mathbb{R}$, a counterexample is simply any nonnegative function, which is not almost everywhere 0. If $\int_A f = c > 0$ for some bounded set A, say $|x| < D$ for $x \in A$, then $Mf(x) > \frac{c}{2y}$, for x such that $B(x,y) \supset A$, and so $\int_{\mathbb{R}} Mf(x) \geq \int_{\mathbb{R}} \frac{c'}{x+D}$. For functions on $\mathbb{I}$ however, it turns out that Mf is integrable iff $f(x)\log(|f(x)|)$ is integrable.[1]

The necessity of this condition is shown by the example of a nonnegative decreasing function, where

$$\int_\varepsilon^1 Mf(x) \geq \int_\varepsilon^1 \frac{1}{2x}\int_0^x f(t)dtdx = -\frac{1}{2}\log\varepsilon\int_0^\varepsilon f(t)dt - \frac{1}{2}\int_\varepsilon^1 \log(x)f(x)dx$$

[1] This is exactly the condition that the entropy of f is finite. This space of functions is sometimes called $L\log(L)$.

by integration by parts. Since both terms are positive, both must be finite in the limit as $\varepsilon \to 0$, and their sum is clearly estimated from below by $-\frac{1}{2}\int_0^1 \log(x)f(x)$.

We can see that the condition that $\log(x)f(x)$ must be integrable is equivalent to $f(x)\log(f(x))$ being integrable for this class of functions. Consider that $f(x) = \frac{1}{x}\cdot a(x)$, where $a(x) \to 0$ as $x \to 0$, and we can assume that convergence of $a(x)$ to 0 is slower than any positive power of x, since otherwise f would be in L^p. So

$$\log(f(x)) = -\log(x) + \log(a(x))$$

and $\log(a(x)) = \mathsf{o}(\log(x))$, and so $\int \log(x)f(x)$ and $\int \log(f(x))f(x)$ are comparable.

To give at least a suggestion of why the condition is sufficient, we remind that the original proof of the maximal theorem relies on the fact that expressions such as

$$\int_0^1 \frac{1}{x}\int_0^x f(t)dtdx$$

are maximized for f decreasing, and that the proof of this fact is essentially combinatorial. Hardy and Littlewood's proof is given in [27]. The above remarks are based on the explanation in [38].

2.2.3 Slowly converging densities

We know that the density of a set at almost every point of that set must converge to 1. However, it is possible for this convergence to be very slow, in fact as slow as we like.

Lemma 2.2.11 *Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence of real numbers in $\mathbb{I}$ tending to 0. We can find a Lebesgue measurable set $A \subset \mathbb{I}$ with $\lambda(A) > 0$ such that for every $x \in A$, we have*

$$1 - \lambda(A|B(x, 1/n)) > x_n$$

for all but finitely many n.

Proof. We can assume without loss of generality that (x_n) is a decreasing sequence. Let (x_n) be such a sequence of real numbers. We will define A as a so-called fat Cantor set, one which has positive Lebesgue measure. Choose a sequence of integers (a_n) such that $\sum_{i=1}^{\infty} \frac{1}{a_i}$ is finite, this means that $\prod_{i=1}^{\infty}(1-\frac{1}{a_i})$ is strictly between 0 and 1.

Let $[0,1]$ be the only interval of the zeroth stage of our construction, and set $A_0 := \mathbb{I}$, $B_0 := \{\mathbb{I}\}$. The n^{th} stage will be described by the set $A_n := \bigcup_{I \in B_n} I$, with B_n a finite collection of non-overlapping closed intervals contained in $\mathbb{I}$. For every n, each interval in B_n will be a proper subset of some interval in B_{n-1}, this will mean that $A_n \subset A_{n-1}$. So A_n is a decreasing sequence and A will be its intersection.

At the n^{th} stage of our construction, we divide each interval from B_{n-1} into $M_n a_n$ pieces of equal length. The sequence of natural numbers M_n will be determined later. Counting from the left, all of these pieces except every a_n^{th} one, will belong to the collection of intervals B_n. The set B_n will consist of all the intervals so selected, for every interval in B_{n-1}. The measure of A_n will therefore be $\lambda(A_n) = \frac{a_n - 1}{a_n}\lambda(A_{n-1})$ and the measure of the limiting set of the procedure will be

$$\lambda(A) = \prod_{i=1}^{\infty}(1 - \frac{1}{a_i}).$$

We will call the length of each interval in B_n

$$l_n = \frac{1}{\prod_{i=1}^{n} a_n M_n}.$$

For convenience, we can equivalently express A as the intersection of the nondecreasing sequence of sets $\bigcap_{i=1}^{\infty} A_i^*$, where

$$A_i^* := \bigcup_{j=0}^{\frac{M_n}{l_{n-1}} - 1} \left[(a_n j) l_n, (a_n j + a_n - 1) l_n\right].$$

Note that for all n, $A_n = \bigcap_{i=1}^{n} A_i^*$, which implies of course that $A_n \subset A_n^*$.

Now suppose that we take a point $x \in A$ and examine a ball around x of radius d. We want to find some $n(d)$ so that d is at least as big as l_n but smaller than l_{n-1}. This is always possible since the sequence (l_n) is decreasing.

We will show that there exists a sequence $(C_n)_{n\in\mathbb{N}}$ with $C_n < 1$ for all n, such that for any x and any d, $\lambda(A|B(x,d))$ is less than $C_{n(d)}$.

It is clear that

$$\lambda(A|B(x,d)) \leq \lambda(A_{n+1}|B(x,d)) \leq \lambda(A^*_{n+1}|B(x,d)).$$

The set A^*_{n+1} consists of connected components of length $(a_{n+1}-1)l_{n+1}$ separated by connected components of its complement each of length l_{n+1}. The relative measure $\lambda(A^*_{n+1}|B(x,d))$ can be no more than

$$1 - \frac{K-1}{K}\frac{1}{a_{n+1}}$$

where K is the number of connected components of A^*_{n+1} which $B(x,d)$ intersects. We know that $K \geq 2M_{n+1}$ by virtue of the fact that $2d$ is at least $2l_n = 2M_{n+1}a_{n+1}l_{n+1}$. Provided that $M_n > 4$ for all n, we have $K > 8$ and so we can take

$$C_n = 1 - \frac{7}{8a_{n+1}}$$

and we have seen that $\lambda(A|B(x,d)) \leq C_n$.

Now our proof is complete provided that we can choose the sequence (M_n) to make sure that

$$C_n < 1 - x_k \tag{2.2}$$

whenever

$$\frac{1}{k} < l_{n-1}$$

or equivalently whenever

$$k > \prod_{i=1}^{n-1} a_i M_i. \tag{2.3}$$

We can guarantee this for $n = 2$ by choosing k_1 so big that $1 - x_{k_1}$ is bigger

than C_2, and then choosing M_1 big enough that $a_1 M_1$ is at least k_1. This will mean that (2.3) only holds for values of k so big that (2.2) is also true.

We can continue this procedure for k_2, k_3 etc. At each step we can find a value of k for which C_{n+1} is smaller than $1 - x_k$, since $1 - x_k$ tends to 1. Then we can increase M_n (which is completely unrestricted, except that it must be greater than 4) to make the right hand side of (2.3) large enough that we only must consider at least that value of k and none smaller. ■

We should clarify that this condition on densities converging slowly is enough to imply a much more general condition.

Proposition 2.2.12 *If $(a_k)_{k\in\mathbb{N}}$ and $(d_k)_{k\in\mathbb{N}}$ are two sequences of positive numbers tending to 0, then we can find a Lebesgue measurable set $A \subset \mathbb{I}$ such that for every point $x \in A$, and for all but finitely many values of k, we have*

$$a > a_k \implies 1 - \lambda(A|B(x,a)) > d_k. \tag{2.4}$$

Proof. First we may reorder the a_n without loss of generality to be decreasing. Now we may also assume the d_n are decreasing. This is because if for a pair of natural numbers $i > j$ we have $a_i < a_j$ but $d_i > d_j$ then the condition given by the pair a_j, d_j is redundant.

We wish to find a sequence $(b_n)_{n\in\mathbb{N}}$ so that if $\lambda(A|B(x, \frac{1}{n})) > b_n$ for all n then the condition (2.4) holds for all k. We do so as follows. Take a natural number n, and define c_n to be

$$c_n := \max\left\{ d_k : [a_k, a_{k-1}] \cap \left[\frac{1}{n+1}, \frac{1}{n}\right] \neq \emptyset \right\}.$$

This means that if $\lambda(A|B(x,r)) > c_n$ for all r in the interval $\left[\frac{1}{n+1}, \frac{1}{n}\right]$, then $\lambda(A|B(x,r)) > d_k$ also holds on this interval for all k such that $a_k \leq r$.

We will define $b_{n+1} := \frac{n+1}{n} c_n$ for $n \in \mathbb{N}$ (and set $b_1 := b_2$). First observe that $c_n \to 0$ and therefore also $b_n \to 0$. Secondly that each c_n corresponds to only a finite number of lower bounds d_k, therefore if the density fails to exceed

c_n on $[\frac{1}{n+1}, \frac{1}{n}]$ for only finitely many values of n, then the condition (2.4) also only fails for finitely many values of k. So if we can show that

$$\lambda(A|B(x, \frac{1}{n+1})) > b_{n+1} \implies \lambda(A|B(x,r)) > c_n \text{ for all } r \in \left[\frac{1}{n+1}, \frac{1}{n}\right]$$

then we are done.

This is obvious; the measure of $A \cap B(x,r)$ must exceed the measure of $A \cap B(x, \frac{1}{n+1})$ if $r \in \left[\frac{1}{n+1}, \frac{1}{n}\right]$. This means that

$$\frac{\lambda(A|B(x, \frac{1}{n+1}))}{n+1} \leq r\lambda(A|B(x,r)) \leq \frac{1}{n}\lambda(A|B(x,r)) \tag{2.5}$$

and

$$\lambda(A|B(x, \frac{1}{n+1})) \leq \frac{n+1}{n}\lambda(A|B(x,r))$$

as required. ■

Remark The density of a measurable set A in small intervals, considered as a function of two variables x and r

$$K_A(x,r) := \frac{1}{2r}\lambda\left(A \cap [x-r, x+r]\right) \tag{2.6}$$

satisfies two types of relations which restrict how fast its values can change as x and r change. Both of these are straightforward consequences of the fact that Lebesgue measure is nonnegative on any set.

Firstly if we fix r then $K_A(x,r)$ is Lipschitz in x, with the value of the Lipschitz constant depending on the value of r. Secondly there are the bounds like that of (2.5), where we keep x fixed and consider changes in r. These are slightly more complicated. If $K_A(x,r_0) = c$, then for $r < r_0$ we have

$$K_A(x,r) \in \left[1 + \frac{c - r_0}{r}, \frac{r_0 c}{r}\right] \tag{2.7}$$

and similarly for $r > r_0$. Since the derivatives of $\frac{r_0 c}{r}$ and $\frac{c}{r} - \frac{r_0}{r}$ are bounded,

close to r_0 the graph of the function is contained in a cone through $(r_0, K_A(x, r_0))$. However the function is not necessarily Lipschitz.

It seems plausible that from these two relations we could draw conclusions about the space of possible functions $K(x, r)$. This could lead to information about the set of functions which are the Lebesgue densities of some measurable set, the preoccupation of the next chapter. (The next chapter does not explicitly use any such techniques.)

Chapter 3

Exceptional densities

This chapter discusses the determination of the constant δ_H, defined at the end of the first section. Proposition 3.4.2, the work of the author together with M. Csörnyei and T. O'Neil, gave an upper bound for this constant. The exact value of δ_H is now known, and we felt it would be worthwhile to place our result in this context by giving a brief exposition of the background and of the new result.

3.1 Basic properties

Let A be a Lebesgue measurable subset of $\mathbb{R}$. In this chapter we will look at the lower density $\underline{d}(x, A)$, upper density $\overline{d}(x, A)$ and density $d(x, A)$ of A at x, all given in Definition 2.2.2, considered as functions of x. All three take values in $[0, 1]$; the first two are defined on $\mathbb{R}$ and the third on some subset of $\mathbb{R}$.

If A is a nullset, then all three functions are identically 0 on $\mathbb{R}$. Similarly if A is of full measure, then the densities are 1 everywhere. We shall call a measurable set a *nontrivial* subset of $\mathbb{R}$ if neither of these cases hold.

We can express the density of a measurable set as the *symmetric derivative* of some function, the indefinite integral of χ_A.

Definition 3.1.1 Suppose that f is a function $\mathbb{R} \to \mathbb{R}$. Given $x \in \mathbb{R}$, we say

that f'_s is the *symmetric derivative* of f at x if for h close to 0

$$f(x+h) - f(x-h) = 2f'_s(x)h + o(h).$$

To discuss the symmetric derivative, it will be helpful to define the Dini derivatives.

Definition 3.1.2 The four *Dini derivatives* of f are given by

$$\begin{aligned} D^+f(x) &= \limsup_{h\to 0, h>0} \frac{f(x+h)-f(x)}{h} \\ D^-f(x) &= \limsup_{h\to 0, h<0} \frac{f(x+h)-f(x)}{h} \\ D_+f(x) &= \liminf_{h\to 0, h>0} \frac{f(x+h)-f(x)}{h} \\ D_-f(x) &= \liminf_{h\to 0, h<0} \frac{f(x+h)-f(x)}{h} \end{aligned}$$

If $D^+f = D_+f$ then their common value is the right-sided derivative of f. Similarly if $D^-f = D_-f$ this is the left-sided derivative. If both left- and right-sided derivatives exist then the symmetric derivative is equal to their arithmetic mean.

Definition 3.1.3 The *upper symmetric derivative* of f at x is

$$\overline{f'_s}(x) := \limsup_{h\to 0} \frac{f(x+h)-f(x-h)}{2h}.$$

The *lower symmetric derivative* is

$$\underline{f'_s}(x) := \liminf_{h\to 0} \frac{f(x+h)-f(x-h)}{2h}.$$

Of course $\underline{f'_s}(x) \le f'_s(x) \le \overline{f'_s}(x)$. An easy calculation shows that

$$\overline{f'_s}(x) \le \frac{1}{2}\left(D^-f(x) + D^+f(x)\right)$$

and

$$\underline{f'_s}(x) \geq \frac{1}{2}\left(D_+f(x) + D_-f(x)\right).$$

Symmetric derivatives have been extensively studied and possess many important properties analogous to those of the usual derivative. We are interested in an extremely special case where the function in question $F(x) := \int_0^x \chi_A$ has symmetric derivative equal to either 0 or 1 at almost every point of $\mathbb{R}$.

The functions $F'(x) = d(x, a)$ and $\overline{F}'(x) = \overline{d}(x, A)$ take the values 0 and 1 almost everywhere, therefore neither can be continuous. We will use the notation

$$d(x, r, A) := \lambda(A|B(x - r, x + r)).$$

The fact that the Lebesgue density of A is not necessarily continuous means that $d(x, r, A)$ does not converge uniformly in x as $r \to 0$.

A simple example is when A is the union of disjoint intervals, with only finitely many in any bounded set. Then all three densities are equal everywhere, and equal to 0 or 1 nearly everywhere, with $\frac{1}{2}$ on the remaining points. This density is not Darboux and we do not know if it is possible for the density of a nontrivial set to be Darboux.

In the following we examine a property of the set of (upper, lower) densities which was proved by O. Kurka in [33]. We will recap the work leading up to his proof, which we will then discuss briefly.

By an abuse of notation, we will say that $d(x, A) \in I$ for $I \subset \mathbb{I}$ an interval exactly when $\underline{d}(x, A) \in I$ and $\overline{d}(x, A) \in I$. This includes the case where $d(x, A)$ is undefined.

Theorem 3.1.4 *Let δ_K be the real positive solution to the polynomial equation*

$$8x^3 + 8x^2 + x = 1. \tag{3.1}$$

(The value of δ_k is roughly $0.26849\ldots$)

Then

(I) If $\delta > \delta_K$, there exists some nontrivial set A so that $d(x, A) \notin [\delta, 1-\delta]$ for all $x \in \mathbb{R}$.

(II) If $\delta < \delta_K$ then for any nontrivial set A there is a point $x \in \mathbb{R}$ such that $d(x, A) \in [\delta, 1-\delta]$.

There can be only one number for which both (I) and (II) hold. If $\zeta \in [0, \frac{1}{2}]$, then we will say that $U(\zeta)$ holds if

$$\forall \delta > \zeta, \exists \text{ nontrivial } A \text{ s.t. } d(x, A) \notin [\delta, 1-\delta] \; \forall x \in \mathbb{R}$$

and that $L(\zeta)$ holds if

$$\forall \delta < \zeta, \forall \text{ nontrivial } A, \exists x \in \mathbb{R} \text{ s. t. } d(x, A) \in [\delta, 1-\delta].$$

If we define δ_H to be the supremum of those x for which $L(x)$ holds, equivalently the infimum of those x for which $U(x)$ holds, then we can rephrase Theorem 3.1.4 by saying that

$$\delta_H = \delta_K.$$

3.2 The lower bound $\delta = \frac{1}{4}$

Before Kurka's proof of this theorem, several authors proved results giving upper or lower bounds for δ_H, in other words they proved that $U(x)$ or $L(y)$ hold, for various constants $x > \delta_K$ and $y < \delta_K$. The first of these results was given by Kolyada in [31]. He proved that $L(\frac{1}{4})$ and also that $U(\frac{\sqrt{17}-3}{4})$. We will obtain the first of these, that $\frac{1}{4} \leq \delta_H$, from a generalization by the author which was not previously published.

Lemma 3.2.1 *Let A be a nontrivial set, and $t \in (0,1)$ be any constant. There exists a point of $\mathbb{R}$ such that $d(x, A) \in [\frac{t}{2}, \frac{t+1}{2}]$.*

This will follow from a lemma about symmetric derivatives of Lipschitz functions.

Lemma 3.2.2 *Suppose that f is Lipschitz with coefficient M, and that $\overline{f}'_s$ and $\underline{f}'_s$ are the upper and lower symmetric derivatives of f respectively. Suppose that a, b, m and n are real numbers with $a < b$ and $m < n$. If $\overline{f}'_s(a) > m$ and $\underline{f}'_s(b) < n$, then for every $\varepsilon > 0$ and every $u \in (m, n)$ there exists a point c in $(a - \varepsilon, b + \varepsilon)$ at which $\overline{f}'_s(c) \leq \frac{u+M}{2}$ and $\underline{f}'_s(c) \geq \frac{u-M}{2}$.*

Proof. Assume without loss of generality that $\varepsilon < b-a$. If the upper symmetric derivative of f at a is greater than m, then we can find a positive number $\delta_a < \varepsilon$ such that $f(a + \delta_a) - f(a - \delta_a) > 2m\delta_a$. Similarly, if $\underline{f}'_s(b) < n$, then we can find a δ_b with $0 < \delta_b < \varepsilon$ such that $f(b + \delta_b) - f(b - \delta_b) < 2n\delta_b$. Now define $h(x) := \delta_a + \frac{x-a}{b-a}(\delta_b - \delta_a)$ for $x \in [a, b]$, further define

$$g(x) := \frac{f(x + h(x)) - f(x - h(x))}{2h(x)},$$

for x in the interval $[a, b]$. We know that $g(a) > m$ and that $g(b) < n$, furthermore g is continuous. So $g(\xi) = u$ for some $\xi \in (a, b)$, in other words

$$f(\xi + h(\xi)) - f(\xi - h(\xi)) = 2uh(\xi)$$

and since $h(\xi) < \max(\delta_a, \delta_b) < \varepsilon$ we know that

$$[\xi - h(\xi), \xi + h(\xi)] \subset (a - \varepsilon, b + \varepsilon).$$

Now, the function k defined by $k(x) := f(x) - f(\xi - h(\xi)) - u(x - \xi + h(\xi))$ is 0 at $\xi - h(\xi)$ and at $\xi + h(\xi)$. It has either a local minimum or a local maximum at some point of $(\xi - h(\xi), \xi + h(\xi))$. Suppose that c is such a point and that it is a maximum.

Then for all $y \in (c-d, c)$ for some small $d > 0$ we have that

$$k(y) \leq k(c)$$
$$f(y) - uy \leq f(c) - uc$$
$$f(y) - f(c) \leq u(y - c).$$

Therefore $D_- f(c) \geq u$. Since $D_+ f(c) \geq -M$ by the Lipschitz condition we have that $\underline{f'_s} \geq \frac{u-M}{2}$. If we consider that for $y \in (c, c+d)$ we similarly have $f(y) - f(c) \leq u(y-c)$ it is easy to see that $D^+ f(c) \leq u$ and therefore that $\overline{f'_s}(c) \leq \frac{u+M}{2}$.

If c is a local minimum, we proceed the same way to show that $D^- f(c) \leq u$ and $D_+ f(c) \geq u$ and to draw the same conclusion. ■

Proof (Lemma 3.2.1). Let $u := \frac{t}{2} + \frac{1}{4}$. Note that $t \in (0,1)$ implies $u \in (0,1)$. Since A is nontrivial, there exist points a and b such that $d(a, A) = 1$, $d(b, A) = 0$. Assume that $a < b$. If we define f by

$$f(x) := \begin{cases} \lambda(A \cap (0,x)) - \frac{x}{2} & x \geq 0 \\ -\lambda(A \cap (x,0)) - \frac{x}{2} & x < 0 \end{cases}$$

then f is Lipschitz with constant $\frac{1}{2}$, and $\underline{f'}(a) = \frac{1}{2}$, $\overline{f'}(b) = -\frac{1}{2}$. So by Lemma 3.2.2 since $u - \frac{1}{2}$ is between $-\frac{1}{2}$ and $\frac{1}{2}$ there is some point c such that $\underline{f'_s}(c) \geq u - \frac{1}{2} - \frac{1}{4}$ and $\overline{f'_s}(c) \leq u - \frac{1}{2} + \frac{1}{4}$. This implies immediately that $\underline{d}(c, A) \geq u - \frac{1}{4} = \frac{t}{2}$ and $\overline{d}(c, A) \leq u + \frac{1}{4} = \frac{t+1}{2}$ as required. ■

Corollary *If A is a nontrivial set then there exists a point $x \in \mathbb{R}$ such that $d(x, A) \in [\frac{1}{4}, \frac{3}{4}]$.* □

Remark Given a density point of A and a density point of its complement we can require that the point x lies in between the two. This is clear from the proof since if $d(x, A) = 1$ and $d(y, A) = 0$ then there exists a density point of A and a density point of $\mathbb{R} \backslash A$ contained in (x, y).

Remark We can interpret Lemma 3.2.2 as a very weak Darboux property for the symmetric derivative, which holds even when the symmetric derivative does not exist everywhere on $\mathbb{R}$. Note that although it appears to follow from a Darboux-like property for the left or right derivatives, these are not Darboux even if they exist everywhere.

3.3 Szenes' and Kurka's reductions

In this section we give the demonstrations by Szenes in [39] and Kurka in [32], [33] that the question of determining δ_H can be reduced to either of two other questions, which both concern related properties of finite collections of intervals.

Definition 3.3.1 Given some $\delta \in (0, \frac{1}{2})$, an *S-configuration for* δ (for configuration in the sense of Szenes) is a set C consisting of finitely many disjoint intervals contained within the unit interval, for which the following property holds.

Define $C^* := (-\infty, 0] \cup C$. Given any endpoint[(1)] x of C^*, there exists some positive number $r(x)$ such that

$$d(x, r(x), C^*) \notin (\delta, 1-\delta).$$

It is clear that if C is an S-configuration for some δ, it is also an S-configuration for all larger numbers.

Proposition 3.3.2 *The infimum of those δ for which some S-configuration exists is δ_H.*

Proof. We prove this in two stages.

Suppose that A is a nontrivial set such that $d(x, A) \notin (\delta, 1-\delta)$ for all $x \in \mathbb{R}$. We will show that given any small $\varepsilon > 0$ we can find an S-configuration for $\delta + \varepsilon$.

[(1)] We could equally say, "Given any real number x", since of course any real number which is not an endpoint is the center of some ball on which C^* has either full or empty measure

Choose some point a with $\underline{d}(a, A) > 1-\delta$ and some point b with $\overline{d}(b, A) < \delta$. Since both A and its complement have positive measure this must be possible. By taking an affine image of A if necessary, assume without loss of generality that $a = 0$ and $b = 1$. If we define

$$A^* := (-\infty, 0] \cup (A \cap (0, 1]),$$

it remains true for A^* as for A that $d(x, A^*) \notin (\delta, 1-\delta)$ for all $x \in \mathbb{R}$. We can see this since every negative number is now a density point, every number greater than 1 is now a density point of the complement, and the density at every point of $(0, 1)$ is unchanged. Lastly, the density at 0 can only increase and the density at 1 can only decrease in going from A to A^*.

For each point x in $[0, 1]$, find some radius $r(x)$ such that A^* has relative measure either less than $\delta + \frac{\varepsilon}{3}$ or greater than $1 - \delta - \frac{\varepsilon}{3}$ in $B(x, r(x))$.

If the relative measure of some set E in a ball $B(s_0, t)$ is $d(s_0, t, E)$, then for all s we have

$$|d(s_0, t, E) - d(s, t, E)| \leq \frac{|s_0 - s|}{2t}.$$

This is because

$$d(s_0, t, E) - d(s, t, E) = \frac{1}{2t}\left(\lambda(E \cap B(s_0, t)) - \lambda(E \cap B(s, t))\right) \leq \frac{1}{2t}\lambda(E \cap [s_0 + t, s + t])$$

and likewise

$$d(s, t, E) - d(s_0, t, E) = \frac{1}{2t}\left(\lambda(E \cap B(s, t)) - \lambda(E \cap B(s_0, t))\right) = \frac{1}{2t}\lambda(E \cap [s - t, s_0 - t]).$$

So observe that for y in the interval $(x - \frac{2r(x)\varepsilon}{3}, x + \frac{2r(x)\varepsilon}{3})$ we have

$$|d(y, r(x), A^*) - d(x, r(x), A^*)| \leq \frac{2r(x)\varepsilon}{6r(x)} = \frac{\varepsilon}{3}$$

and so $d(y, r(x), A^*) \notin (\delta + \frac{2\varepsilon}{3}, 1 - \delta - \frac{2\varepsilon}{3})$. We can find a finite collection of points $\{x_i : 1 \leq i \leq k\}$ so that the balls $B(x, \frac{2r(x)\varepsilon}{3})$ cover the compact set $[0, 1]$.

Then set r to be the minimum of $r(x_i)$ and set $r^*(x)$ to be $r(x_i)$ for one of the i for which $x \in B(x_i, \frac{2r(x_i)\varepsilon}{3})$. Now we know that for all $x \in [0,1]$

$$d(x, r^*(x), A) \notin (\delta + \frac{2\varepsilon}{3}, 1 - \delta - \frac{2\varepsilon}{3}).$$

Choose C to be some finite collection of intervals contained in $[0,1]$ such that

$$\lambda((C \triangle A) \cap [0,1]) < \frac{2}{3} r\varepsilon.$$

This means that the relative measures of $C \cup (-\infty, 0]$ and A^* on any interval of length more than $2r$ will differ by no more than $\frac{\varepsilon}{3}$. This proves the claim.

Now we turn to the other part of the proof; the claim that if we have an S-configuration for δ and $\varepsilon > 0$ is given then we can find a nontrivial set with $d(x, A) \notin (\delta + \varepsilon, 1 - \delta - \varepsilon)$ for all x in $\mathbb{R}$.

We call C the S-configuration in question. We write $A_0 := (-\infty, 0] \cup C$. We will define an increasing series of measurable sets $(A_n)_{n \in \mathbb{N}}$ and then show that $U(\delta + \varepsilon)$ holds for their union.

For each endpoint of A_0 we have some radius $r(x)$ given by Definition 3.3.1, so that the relative measure of A_0 on the ball $B(x, r(x))$ is not contained in $(\delta, 1-\delta)$. Furthermore, if $x \in \partial A_0$ is such a point, and y is a point in $B(x, \frac{\varepsilon r(x)}{2})$ then the relative measure of A_0 on $B(y, r(x))$ will not be contained in $(\delta + \frac{\varepsilon}{2}, 1 - \delta - \frac{\varepsilon}{2})$. Setting $m := \min\{r(x) : x \in \partial A_0\}$ we call $B(x, \frac{\varepsilon m}{2})$ the 'good region' around the point x.

Let us write ∂_L for the set of left endpoints of A_0 and ∂_R for the set of right endpoints. Then we can define

$$A_{n+1} := A_0 \cup \bigcup_{x \in \partial_L} (-\alpha(A_n \cap [0,1]) + x) \cup \bigcup_{x \in \partial_R} (\alpha(A_n \cap [0,1]) + x).$$

Here α is a constant $0 < \alpha \ll \varepsilon$ which we will determine later. The sets A_n are increasing in n and we take their union to be the set A.

Let k be the number of endpoints of A_0.

We choose α so that the following four conditions will be satisfied

- The set $A \backslash A_0$ is made up of several pieces, each an affine copy of $A \cap [0,1]$, one for each endpoint of A_0. Each piece lies within $\frac{\alpha}{1-\alpha}$ of the corresponding endpoint. We need to ensure that the whole piece is contained within the 'good region' around that endpoint. This gives us

$$\frac{\alpha}{1-\alpha} < \frac{\varepsilon m}{2}.$$

- We want to treat each piece of $A \backslash A_0$ as an affine copy of A itself. A piece belongs to an interval $[y,z]$ which is a connected component of $[0,1] \backslash A$. Suppose that the piece corresponds to the endpoint y. We want to make sure that for any x which belongs to this piece, $B(x, r(v))$ does not overlap with $(\frac{y+z}{2}, z)$. Here v can be any endpoint of A_0. So we require

$$\frac{\alpha}{1-\alpha} + \alpha M < \frac{m_1}{2}$$

where $M := \max\{r(x) : x \in \partial A_0\}$ and m_1 is the minimum length of an interval of the complement of C.

- Now suppose that (u,y) is an interval of A_0 and x is in the piece of $A \backslash A_0$ associated with y. We need that $B(x, \alpha r(v)) \not\supseteq (u,y)$ for any $v \in \partial A_0$. So

$$\alpha M < m_2$$

where m_2 is the minimum length of an interval of C.

- Finally we want to make sure that the total measure of $A \backslash A_0$ is less than $\frac{\varepsilon m}{2}$. So

$$\frac{k\alpha}{1-k\alpha} < \frac{\varepsilon m}{2}.$$

Clearly we can find some positive α which meets all of these.

Now we have seen that for each point x in the 'good region' around an

endpoint y of A_0, we have $d(x, r(y), A_0) \notin (\delta + \frac{\varepsilon}{2}, 1 - \delta - \frac{\varepsilon}{2})$. We know by the first condition on α that each point of ∂A lies inside such a good region. By the fourth we see that for all such points we also have $d(x, r(y), A) \notin (\delta+\varepsilon, 1-\delta-\varepsilon)$.

We now look at the points of $A \backslash A_1$ and apply an identical argument. Next to each endpoint of A_1 there is a piece of $A \backslash A_1$, which is the affine image under a contraction by α of a piece of $A \backslash A_0$. If the corresponding endpoint of A_0 was y, then we use the radius $\alpha r(y)$ for all points of the good region around the endpoint of A_1. By the second and third conditions, the relative measure of A on balls of this size is the same as for balls of radius $r(y)$ in the good region around y. As before we know that the union of such good regions contains the whole of ∂A.

It is clear that we can repeat this argument for $A \backslash A_n$ for any $n \in \mathbb{N}$. So for every point x of ∂A we can find a sequence of radii, $\alpha^n r(y_n)$ such that $B(x, \alpha^n r(y_n))$ is not contained in $(\delta + \varepsilon, 1 - \delta - \varepsilon)$. This proves the theorem.■

Two upper bounds for δ_H were given as examples of S-configurations, in [39] and [10]. We will however give Kurka's further reduction of the problem, then explain the upper bounds in his language.

Definition 3.3.3 A *K-configuration for* δ, or configuration in the sense of Kurka, is a finite collection of intervals G contained in $[0, 1]$, with the following properties.

1. For every endpoint x of $G + \mathbb{Z}$, there exists some radius $R(x)$ such that $d(x, R(x), G + \mathbb{Z})$ is not contained in $(\delta, 1 - \delta)$.

2. There exists a collection of intervals $(I_n)_{n \in \mathbb{N}}$, each centered at an endpoint of $G + \mathbb{Z}$, whose union covers $\mathbb{R}$, such that $\lambda(G + \mathbb{Z} | I_n) > 1 - \delta$ for all n.

3. For all $b \in \mathbb{R}$, there exist u and v, $u < b < v$ such that $d(u, r, G + \mathbb{Z} \cap (-\infty, b]) \in (\delta, 1 - \delta)$ and $d(v, r, G + \mathbb{Z} \cap [b, \infty)) \in (\delta, 1 - \delta)$ for all $r > 0$.

4. There do not exist any two points a and b such that $(a, b) \cap (G + \mathbb{Z})$ is the affine image of an S-configuration.

For every $\delta \in (0, \frac{1}{2})$ we can find a K-configuration, since we can take G to have measure greater than $1 - \delta$ so that conditions 1 and 2 are satisfied by taking large enough intervals. Then we can ensure that the endpoints of $G + \mathbb{Z}$ are sufficiently close together that any b in condition 3 lies close to one of them. Then it will be seen that no portion of $\mathbb{R}$ can be mapped affinely onto $[0, 1]$ to form an S-configuration, since if x was the supremum of this S-configuration than any ball around x would have relative measure close to $\frac{1}{2}$.

In fact we are interested in the existence of *good K-configurations* for δ – those which have Lebesgue measure less than $\frac{4\delta^2}{1-2\delta}$. This distinction is only important if $\frac{4\delta^2}{1-2\delta} < 1 - \delta$, since otherwise the considerations of the previous paragraph apply. We will assume this inequality in the remainder of this section.

Kurka showed in his first paper on the question, [32], that the infimum of δ for which S-configurations exist is equal to the infimum of δ for which good K-configurations exist. We are going to sketch the proof of this reduction of the question of densities to the question of the existence of good K-configurations. The details can be found in [32] which uses notation and results from [39], or in [33], an almost identical approach which is self-contained.

Proposition 3.3.4 *Suppose that δ is given and that there exists some good K-configuration G. Then for any $\varepsilon > 0$ there is some S-configuration for $\delta + \varepsilon$.*

This direction is straightforward; in fact it would be uncomplicated to combine the following with the corresponding direction of Proposition 3.3.2 to prove directly that

$$\exists \text{ a good K-configuration for } \delta \implies U(\delta).$$

Proof. Let δ, G and ε be given. Call $(c_n)_{n\in\mathbb{N}}$ the endpoints of $G + \mathbb{Z}$; to each corresponds a positive number $R(c_n)$ such that $d(c_n, R(c_n), G + \mathbb{Z})$ is either less than δ or greater than $1 - \delta$. We know that $\mathbb{R}$ is covered by intervals on which $G + \mathbb{Z}$ has relative measure at least $1 - \delta$, therefore by Lemma 6 of [39] (see Lemma 3.5.1 below), G must have measure at least $\frac{1-\delta}{1+\delta}$. Using the same lemma, it is not possible for $\mathbb{R}$, nor any half-line in $\mathbb{R}$, to be covered by intervals on all

of which $G + \mathbb{Z}$ has relative measure less than δ.

We should observe that $R(c_n)$ has some maximum over all $n \in \mathbb{N}$, call it R_M. This is because G has measure between $\frac{1-\delta}{1+\delta}$ and $\frac{4\delta^2}{1-2\delta}$, therefore between δ and $1-\delta$. So $G+\mathbb{Z}$ has relative measure between δ and $1-\delta$ in sufficiently large intervals.

So choose some large integer M, and choose two points $a, b \in \mathbb{R}$, such that $b - a > M$, and such that neither a nor b is contained in any ball $B(c_n, R(c_n))$ on which the relative measure of $G+\mathbb{Z}$ is less than δ. There might however be some collection of $c_n \in (a,b)$ for which $c_n - R(c_n) < a$ and with the relative measure of $G+\mathbb{Z}$ on $B(c_n, R(c_n))$ greater than $1-\delta$. If M is at least $2R_M$ then this collection is nonempty, since by Property 4 of Definition 3.3.3 there must exists some $c_n > a$ such that $B(c_n, R(c_n))$ contains a. Let a' be defined as the minimum of $c_n - R(c_n)$ for all of these. Likewise set b' to be the maximum of $c_n + R(c_n)$ for those $c_n \in (a,b)$ with balls $B(c_n, R(c_n))$ that overlap (b, ∞), on each of which necessarily the relative measure of $G+\mathbb{Z}$ is greater than $1-\delta$.

Now the set $G_{a,b} := [a', a] \cup (G \cap (a,b)) \cup [b, b']$ is such that if x is an endpoint of $G+\mathbb{Z}$ contained with (a,b), then $B(x, R(x)) \subset [a', b']$ and $d(x, R(x), G_{a,b})$ is not contained in $(\delta, 1-\delta)$.

For simplicity we set $\gamma := 2\delta$ and $\lambda := \lambda(G)$. Observe that since G is a good K-configuration, we have $\lambda \leq \frac{\gamma^2}{1-\gamma}$.

Define $\alpha := \frac{1}{\gamma+\lambda}$. Since

$$\lambda > \frac{1-\delta}{1+\delta} > 1 - 2\delta$$

this is less than 1. Since both γ and λ are less than 1, it is more than $\frac{1}{2}$. Let ψ be the affine mapping that sends a' to $m := 1-\alpha$ and b' to 1. Note that $m < \frac{1}{2}$. The S-configuration is given by the image $\psi(G_{a,b}) =: C$. We will check that it is an S-configuration for $\delta + \varepsilon$. All the endpoints contained strictly between $\psi(a)$ and $\psi(b)$ have radii with the correct relative measures, since they were preserved by the affine mapping. There remain at most five endpoints, 0, m, $\psi(a)$, $\psi(b)$

and 1. (At most five and not five since $\psi(a)$ and $\psi(b)$ might not be endpoints of $\psi(G_{a,b})$.) We will set radii $r(0) = r(1) = r(\psi(b)) = 1$ and $r(m) = r(\psi(a)) = \alpha$.

First of all we see that $B(0,1) \supset B(m,\alpha)$ and $B(0,1)\backslash B(m,\alpha) \subset (-\infty, 0]$. So if we show that the relative measure of $C \cup (-\infty, 0]$ in the latter ball is greater that $1-\delta$, the same will hold for the former ball. We shall show then, that the relative measures in the ball $B(m,\alpha)$ and $B(\psi(a),\alpha)$ are large, and the relative measures in $B(1,1)$ and $B(\psi(b),1)$ are small.

Also note that the differences $|d(\psi(a),\alpha,C\cup(-\infty,0]) - d(m,\alpha,C\cup(-\infty,0])|$ and $|d(\psi(b),1,C\cup(-\infty,0]) - d(1,1,C\cup(-\infty,0])|$ are both $\mathsf{O}(\frac{1}{M})$. This follows from $|\psi(a)-m|$ and $|\psi(b)-1|$ also being $\mathsf{O}(\frac{1}{M})$. Furthermore $\lambda(C\cap(m,1)) = \alpha\lambda + \mathsf{O}(\frac{1}{M})$. We will consider it enough to show that $d(m,\alpha,C^*\cup(-\infty,0]) > 1-\delta$ and that $d(1,1,C^*\cup(-\infty,0]) < \delta$, for C^* some set contained in $(m,1)$ of measure exactly $\alpha\lambda$. We leave to the reader the limiting argument which proves that for large M we can obtain the correct relative measures in all four balls for the actual set C, provided that we accept $\delta+\varepsilon$ in place of δ.

The calculations are now straightforward. The measure of $B(1,1)\cap C^*$ is

$$\alpha\lambda = \frac{\lambda}{\gamma+\lambda}.$$

We require this to be less than $2\delta = \gamma$ and this is true because

$$\lambda < \frac{\gamma^2}{1-\gamma}.$$

Furthermore the measure of the complement of $B(m,\alpha)\backslash(C^*\cup(-\infty,0])$ is

$$1-\alpha\lambda = \frac{\gamma}{\gamma+\lambda} = \alpha\gamma$$

and this last term is $2\delta\alpha$ as required. ■

Proposition 3.3.5 *Suppose that δ is given and that there is some S-configuration C for δ. Then there is some good K-configuration for δ.*

Proof (outline). Let δ and C be given. Moreover suppose that C consists of the lowest number of intervals of any S-configuration for δ. Let C^* be $C \cup (-\infty, 0]$ as usual.

Szenes showed in [39] several properties of the S-configuration with a minimal number of intervals which we will use here.

Lemma 3.3.6 (Lemma 7 of [39]) *If x is an endpoint of C in $[0, \frac{1}{2}]$ and if $d(x, r, C^*) = 1 - \delta$, where r is the maximal radius for which this holds, then either $B(x, r) \subseteq [0, 1]$ or $[0, 1] \subseteq B(x, r)$. Likewise x is an endpoint of C in $[\frac{1}{2}, 1]$ and $d(x, r, C^*) = \delta$, where r is the maximal radius for which this holds, then either $B(x, r) \subseteq [0, 1]$ or $[0, 1] \subseteq B(x, r)$.*

To show this (we consider only the first statement), assume that for some x it is not true. This means that $x < r < 1 - x$, where r is the maximal radius such that C^* has relative measure at least $1 - \delta$ in $B(x, r)$. We could then take $C \cap [0, x + r]$ to be an S-configuration for δ, one containing fewer intervals than C. The reason this is possible is that if any ball $B(y, r_y)$ contains relative measure of C^* of at least $1 - \delta$, and $y + r_y > x + r$, then we could increase the radius around x by $y + r_y - (x + r)$, keeping the relative measure above $1 - \delta$. This contradicts r being maximal. So the assumption $x < r < 1 - x$ contradicts the number of intervals in C being minimal.

Now, let $v_B := \max\{x \in \partial C \cap [0, \frac{1}{2}] : d(x, r, C^*) \geq 1 - \delta \text{ for some } r \geq 1 - x\}$ and $v_W := \min\{x \in \partial C \cap [\frac{1}{2}, 1] : d(x, r, C^*) \leq \delta \text{ for some } r \geq x\}$. Now if $x \in \partial C \cap (v_B, v_W)$ and $d(x, r, C^*) > 1 - \delta$ for some r, let $r(x)$ be the radius for which $d(x, r(x), C^*)$ is maximal. Define SB_1 to be the union of $B(x, r(x))$ for such points, and SB the set formed by adjoining to SB_1 all the intervals of C which it overlaps. If x is any other point in $\partial C \cap (v_B, v_W)$, then let $r(x)$ be such that $d(x, r(x), C^*)$ is minimal (and necessarily below δ). Define SW_1 to be the union of $B(x, r(x))$ for such points, and SW the set formed by adjoining to SB_1 all the intervals of $[0, 1] \backslash C$ which it overlaps.

Kurka proved the following using Szenes' Lemma 12 and Corollary 13:

Lemma 3.3.7 (Lemma 1.2 of [32]) *Both of SB and SW are intervals and one contains the other.*

The proof relies on showing that if a component of SB and a component of SW overlap, but neither contains the other, then (an affine image of) the place where they overlap forms an S-configuration, which must contain fewer intervals than C.

Suppose without loss of generality then that $SB \supset SW$. We take a final result from Szenes' analysis of minimal S-configurations.

Lemma 3.3.8 (Lemma 8 of [39]) *If $\rho := \lambda(C)$ then*

$$\frac{1-\rho}{2(1-v_B)} \leq \delta \text{ and } \frac{\rho}{2v_W} \leq \delta.$$

The first inequality is true since there is a ball of radius $(1-v_B)$, which contains all of $[0,1]$, and in which $[0,1]\backslash C$ has relative measure less than δ. The second is analogous.

We obtain from this that $\rho < 2\delta$, and that $v_W - v_B \geq \frac{1}{2\delta} - 1$. From this we see that the relative measure of C^* in (v_B, v_W) is no more than $\frac{4\delta^2}{1-2\delta}$.

To construct the K-configuration, we will choose points $q \in (\inf SB, v_B)$, p and p' in (v_B, v_W) with $p < p'$ and $q' \in (v_W, \sup SB)$, and let the set E be given by

$$E := (q,p] \cup (C \cap (p,p')) \cup [p',q).$$

The points p and p' are chosen so that if $x \in (v_B, v_W)$ has a radius $r(x)$ with $d(x, r(x), C^*) \leq \delta$, then $B(x, r(x)) \subseteq [p, p']$. Hence the relative measure of E in the same ball will also be less than δ. The points q and q' are chosen so that if on the other hand $x \in (v_W, v_B)$ has a radius with $d(x, r(x), C^*) \geq 1-\delta$ then $d(x, r'(x), E) \geq 1-\delta$ will also hold, where $r'(x) := \min\{r(x), x-q, q'-x\}$. We do this modification of $C^* \cap SB$ to obtain E in order that v_B and v_W should not be endpoints any longer. This will mean that we can transplant E into another construction, with all the endpoints keeping their correct radii

and relative measures, since all relevant balls are contained in $[q, q']$. We skip the exact construction of q, p, p' and q' as well as the proof that the relative measure of E in $(\frac{q+p}{2}, \frac{p'+q'}{2})$ is less than $\frac{4\delta^2}{1-2\delta}$. These are given in Section 1 of [32] and in slightly changed form in Section 3 of [33]. Of course the proof of the latter relies on the relative measure of C^* in (v_B, v_W) also being less than $\frac{4\delta^2}{1-2\delta}$.

We should also observe that these constructions are valid only for δ in a limited range. However this range includes the best known upper and lower bounds for δ_H prior to Kurka's work, and so the construction works for all values of δ which we could practically require.

It is now possible to define G. Let ϕ_1 be the affine map that sends $\frac{q+p}{2}$ to 0 and $\frac{p'+q'}{2}$ to $\frac{1}{2}$, and ϕ_2 the affine map that sends $\frac{p'+q'}{2}$ to $\frac{1}{2}$ and $\frac{q+p}{2}$ to 1. Then

$$g := \phi_1\left(E \cap \left(\frac{q+p}{2}, \frac{p'+q'}{2}\right)\right) \cup \phi_2\left(E \cap \left(\frac{q+p}{2}, \frac{p'+q'}{2}\right)\right).$$

If β is the common ratio of these contractions, then the intervals of $G+\mathbb{Z}$ which contain points of $\mathbb{Z}$ or $\mathbb{Z}+\frac{1}{2}$ are of length $\beta(p-q)$ and $\beta(q'-p')$ respectively. For each endpoint of $G+\mathbb{Z}$, we can take the radius of the corresponding endpoint of E, scaled by β, to satisfy Property 1 of Definition 3.3.3. The fact that Property 2 is satisfied for G follows from the corresponding property for (v_B, v_W).

To prove the remaining two properties of a K-configuration, we consider the following. Suppose that Property 4 is false, and that $(G+Z) \cap (a, b) =: G_{a,b}$ is an S-configuration (after an affine transformation). If (a, b) is contained in some interval of the form $(\frac{n}{2}, \frac{n+1}{2})$ for integer n, then $G_{a,b}$ contains fewer intervals than any S-configuration, which is a contradiction.

So (a, b) contains $\frac{n}{2}$ for some $n \in \mathbb{Z}$. It is possible to show that either $(a, \frac{n}{2})$ or $(\frac{n}{2}, b)$ is also the affine image of a S-configuration. This is because, if a radius around an endpoint of $(-\infty, b] \cup G_{a,b}$ was such that the relative measure of this set was more than $1-\delta$, then the same will be true for $(-\infty, \frac{n}{2}] \cup G_{a,b}$. If on the other hand for some endpoint there was a ball on which the relative measure was less than δ, it can be shown that this ball does not contain $\frac{n}{2}$. This last

fact follows from the choice of p, q, p' and q' and the related property of E.

So we have demonstrated that there exists some other (affine copy of a) S-configuration, which contains one fewer half-integer point than (a, b). Since we showed that every S-configuration must contain such a point, this is a contradiction.

Property 3 is proved in an identical manner. This completes the proof of the proposition. ■

3.4 Three upper bounds

Having proved the equivalence of the existence of good K-configurations for δ with $U(\delta)$ in the previous section, it is now simple to give the three successive upper bounds on δ_H from [39], [10] and [32].

Proposition 3.4.1 (Szenes' upper bound) *If $\delta > 0.2719\ldots$ then $U(\delta)$ holds.*

The exact value is the positive solution to $8\delta^3 + 4\delta^2 + 2\delta - 1 = 0$.

Proof. The corresponding K-configuration is given by $G := [0, \frac{s}{2}) \cup (1 - \frac{s}{2}, 1]$ where $s := \frac{1}{1+2\delta}$. This is a good K-configuration for all δ such that

$$\frac{1}{1+2\delta} < \frac{4\delta^2}{1-2\delta}$$

which gives us that

$$8\delta^3 + 4\delta^2 + 2\delta > 1.$$

For all endpoints of $G + \mathbb{Z}$ we take the radius s. Since $\frac{1-s}{s} = 2\delta$, Condition 1 is true. Since the balls in question cover $\mathbb{R}$, so is Condition 2. The other two conditions can be checked easily by seeing that each ball is the only one centered on that point which has the correct relative measure, and each ball contains another endpoint to the left and one to the right. So if we were to take a set of the form $(G + \mathbb{Z})\backslash(a, -\infty)$, there would be at least one endpoint with no appropriate radius. ■

Proposition 3.4.2 (Csörnyei, Grahl and O'Neil's upper bound) *If* $\delta > 0.27107\ldots$ *then* $U(\delta)$ *holds.*

The exact value is the positive solution to $2\delta^3 + 2\delta^2 + 3\delta - 1 = 0$.

Proof. Let $\alpha := \frac{1-\delta}{2+\delta}$ and $\beta := \frac{4\delta-1}{4+2\delta}$. We have

$$3\alpha + 2\beta = \frac{3-3\delta}{2+\delta} + \frac{4\delta-1}{2+\delta} = 1.$$

Since $\delta > 0.27107 > \frac{1}{4}$, we know that β is positive.

The K-configuration is $G := [0,\alpha] \cup [\alpha+\beta, 2\alpha+\beta]$. The complement of G in $[0,1]$ consists of two intervals, one of length β and one of length $1-(2\alpha+\beta) = \alpha+\beta$. If an endpoint of $G+\mathbb{Z}$ is contiguous to an interval of $\mathbb{R}\backslash(G+\mathbb{Z})$ of length $\alpha+\beta$, its radius is given by $2\alpha+\beta$. The ball of this radius will contain three intervals of $G+\mathbb{Z}$, each of length α, and we see that

$$\frac{3\alpha}{2(2\alpha+\beta)} = \frac{3\alpha}{1+\alpha} = \frac{3-3\delta}{3} = 1-\delta$$

as required. In case x is an endpoint of the other intervals of $\mathbb{R}\backslash(G+\mathbb{Z})$, those of length β, then we take the ball around x of radius α. This ball contains only that one interval of length β from the complement of $G+\mathbb{Z}$. So we need only that

$$\frac{\beta}{2\alpha} = \frac{4\delta-1}{4-4\delta} \leq \delta$$

to establish Condition 1. This is true if $\delta \leq \frac{1}{2}$.

Condition 2 follows immediately from the lengths of the balls given. The other two conditions follow in the same way as the previous proof.

The measure of G is $2\alpha = \frac{2-2\delta}{2+\delta}$. This is less than $\frac{4\delta^2}{1-2\delta}$ if

$$2\delta^3 + 2\delta^2 + 3\delta > 1.$$

■

Proposition 3.4.3 (Kurka's upper bound) *If* $\delta > \delta_K$ *then* $U(\delta)$.

Proof. To build the configuration, we define four lengths:

$$\alpha := \frac{1 + 2\delta - 4\delta^2}{4 + 12\delta}$$
$$\beta := \frac{8\delta^2}{4 + 12\delta}$$
$$\phi := \frac{4\delta}{4 + 12\delta}$$
$$\psi := \frac{2}{4 + 12\delta}$$

Note that $2\alpha + \beta + 2\phi + \psi = 1$. The set G consists of the intervals

$$[0, \alpha] \cup [\alpha + \beta, 2\alpha + \beta] \cup [1 - \phi - \psi, 1 - \phi],$$

of lengths α, α and ψ respectively, and $[0, 1] \backslash G$ is made up of three intervals, of lengths β, ϕ and ϕ from left to right.

The radii at endpoints of G are as follows:

$$r(0) = r(2\alpha + \beta) := 2\alpha + \beta = \phi + \psi$$
$$r(1 - \phi - \psi) = r(1 - \phi) := \psi$$
$$r(\alpha) = r(\beta) := \alpha$$

with radii at all other endpoints given by that of the equivalent point modulo 1. To check Condition 1 it suffices to check that

$$\beta \leq 2\delta\alpha$$

which is implied by $1 - 4\delta^2 > 2$, or $2\delta < 0.618\ldots$, that

$$\phi \leq 2\delta\psi$$

which is in fact true with equality, and that

$$\phi + \beta \leq 2\delta(\phi + \psi)$$

which is also true with equality, following from the previous inequality and the fact that $\beta = 2\delta\psi$.

As above, the other three conditions can be proved by simple arguments. When the measure of G, $\frac{2-8\delta^2}{4+12\delta}$, is less than $\frac{4\delta^2}{1-2\delta}$ we have a good K-configuration. This is equivalent to saying that δ is greater than the positive solution of (3.1), which yields the correct upper bound. ■

Remark We believe that certain aspects of this example suggest why it is in fact optimal. The optimality was established using a complicated argument about K-configurations which is discussed in the next section. However, the following argument may provide some intuition as well as leading, if correct, to a simpler proof.

Choose an interval $I \in \mathbb{R}\backslash(G+\mathbb{Z})$ which is as large as any other connected component of $\mathbb{R}\backslash(G+\mathbb{Z})$. Define $\rho := |I|$. In the previous example we would have $\rho = \phi$. This interval I must be contained in some ball in which $G+\mathbb{Z}$ has relative measure $1-\delta$. Call this ball B_{max}, and assume that $|B_{max}| \leq 1$, otherwise we could redefine G to be $\frac{1}{2}(G \cup G+1)$ and thus reduce B_{max} by half. It might be possible to show that B_{max} covers all of (some rotated copy of) G, otherwise we could remove the parts of G which lie outside B_{max}, rescale, and somehow obtain a K-configuration of smaller measure[(2)]. So suppose that G is contained within a ball of length $1-\sigma = |B_{max}|$ on which its relative measure is at least $1-\delta$. The minimum possible measure of G is then $(1-\delta)(1-\sigma)$.

Now, $\sigma \leq \rho$ by definition of ρ. In the previous example both ρ and σ are given by ϕ. If we could prove that it is not possible for ρ to take up a greater proportion of B_{max} than $\frac{\phi}{1-\phi}$, we would have shown that the previous example is optimal.

Suppose then that I, the largest interval of $[0,1]\backslash G$, is bigger than in the previous example. At the endpoint of I which is the center of B_{max}, we still have a possibility of finding a radius to satisfy Condition 1, since the relevant

[(2)] For example, a procedure like this would yield Kurka's G from that of Csörnyei, Grahl and O'Neil.

inequality for Kurka's G is not sharp. That is to say

$$\lambda(I|B_{max}) = \frac{\phi}{1-\phi} = \frac{4\delta}{4+8\delta} < \delta.$$

However at the other endpoint of I, a ball of radius $\frac{B_{max}}{2}$ would overlap with the interval $[0,1]\backslash B_{max}$ (or an equivalent set modulo 1), which is contained in $\mathbb{R}\backslash(G+\mathbb{Z})$. This would mean that this ball had relative measure of $G+\mathbb{Z}$ strictly less than in B_{max}. This would suggest that the example is not optimal as it implies $\lambda(G+\mathbb{Z}|B_{max}) > 1-\delta$. If we exclude this possibility, the largest radius possible would seem to be $\frac{B_{max}}{2} - \rho$. But for there to exist a ball of this radius, containing an interval of length ρ from $[0,1]\backslash G$, and in which G has relative measure at least $1-\delta$, we need

$$\rho \leq 2\delta(\frac{|B_{max}|}{2} - \rho) = \delta(1-\rho-2\rho).$$

This yields $\rho = \frac{1}{1+3\delta}$, or exactly ϕ in Kurka's G.

This argument obviously relies on several assumptions which may be false or very difficult to prove. However we feel that it may be possible to complete these steps either as consequences of properties of K-configurations given in [33] or independently, and that this may lead to a proof of Kurka's result $\delta_H = \delta_K$ which is much simpler than that of [33].

3.5 Kurka's lower bound

Before [32], the best lower bound for δ_H was given in [39]. If we rephrase his result in the language of Kurka's papers, Szenes proved that the measure of a K-configuration for δ must be at least $\frac{1-\delta}{1+\delta}$. Therefore it is only possible for a good K-configuration to exist for δ if

$$\frac{1-\delta}{1+\delta} \leq \frac{4\delta^2}{1-2\delta}$$

or

$$4\delta^3 + 2\delta^2 + 3\delta > 1 \tag{3.2}$$

(assuming that $0 < \delta < \frac{1}{2}$). This gives $L(0.2629\ldots)$ with the exact value being the positive root of $4x^3 + 2x^2 + 3x - 1$. The proof of this lower bound on the measure of K-configurations is easy:

Lemma 3.5.1 (Lemma 6 of [39]) *Suppose that $\{I_k : 1 \leq k \leq n\}$ is a collection of intervals whose union is another interval I, and that $B \subset I$ is a measurable set. Let δ be some constant $0 < \delta < \frac{1}{2}$. If $\lambda(B|I_i) > 1 - \delta$ for all i then $\lambda(B|I) > \frac{1-\delta}{1+\delta}$.*

This result immediately yields the fact that K-configurations measure at least $\frac{1-\delta}{1+\delta}$ and thence Szenes' lower bound.

Proof. By removing intervals from the set if necessary, we may assume without loss of generality that no interval is contained in the union of all the others. This means that we can number the intervals in increasing order of their left endpoints, and their right endpoints will be in the same order. So let $I_k := [a_k, b_k]$ where $a_k < a_{k+1}$ and $b_k < b_{k+1}$. Here we assume for simplicity that the intervals are closed.

Each interval is made up of three regions: one which overlaps the left neighbour (if there is one), one consisting of points which belong only to that interval, and one overlapping the right interval, if any. Either of the overlapping parts may be empty or consist of a single point. Let W be the union of the overlapping parts and U be the union of the unique parts. By summing

$$\lambda(B \cap I_i) > (1 - \delta)\lambda(I_i)$$

over all i we obtain

$$\lambda(B \cap U) + 2\lambda(B \cap W) > (1 - \delta)\left(\lambda(U) + 2\lambda(W)\right)$$

since each overlapping part appears twice in the sum. From this we obtain

$$\begin{aligned}
\lambda(B\cap U)+2\lambda(B\cap W)-(1-\delta)\lambda(W) &> (1-\delta)\left(\lambda(U)+\lambda(W)\right)\\
\lambda(B\cap U)+(1+\delta)\lambda(B\cap W) &> (1-\delta)\left(\lambda(U)+\lambda(W)\right)\\
(1+\delta)\left(\lambda(B\cap U)+\lambda(B\cap W)\right) &> (1-\delta)\left(\lambda(U)+\lambda(W)\right)\\
(1+\delta)\lambda(B\cap I) &> (1-\delta)\lambda(I)
\end{aligned}$$

as required. ■

In this last calculation we used only that $\lambda(B\cap W)\leq\lambda(W)$ and that $\lambda(B\cap U)\geq 0$. In fact the optimal case is given when B is exactly W (to within a nullset). So I_i contains a portion of length $\delta|I_i|$ which does not belong to any other interval, or to B, and the remainder of I_i, which overlaps I_{i-1} or I_{i+1}, is contained in B.

Szenes suggested that his bound for δ_H could be improved on since neither his configuration nor Kurka's is covered with an arbitrary set of intervals. Rather, we need there to be one of these intervals centered about each endpoint. It is not possible in the optimal case for the previous lemma to make sure that each point of the boundary of B is the center of some interval I_i.

Were it not for Condition 4 in the definition of a K-configuration, we could place a small affine copy of some S-configuration close to each point of ∂B as in the proof of Proposition 3.3.2. This would allow us to choose a small interval centered about the point of ∂B on which B has sufficiently high relative measure. Thus each endpoint would be the center of some ball B for which $\lambda(G|B)\notin(\delta,1-\delta)$.

So Kurka uses the special quality of a K-configuration and of the set of intervals which satisfy Condition 2 in Definition 3.3.3, to raise Szenes' lower bound on the measure of a K-configuration.

Proposition 3.5.2 (Proposition 6.8 of [33]) *Any K-configuration for δ has*

measure no less than

$$\frac{(1-\delta)(1+2\delta)}{1+3\delta}.$$

This yields both the lower bound $L(\delta_K)$ and and thence the correct value $\delta_H = \delta_K$.

The proof of this proposition given in [33] is very complicated. We only give an indication of the method by quoting an intermediary result.

Lemma 3.5.3 (Lemma 5.1 of [33]) *If G is a K-configuration, write $H := G + \mathbb{Z}$. We call r_0 a* minimal radius for x *if $d(x, r_0, H) = 1 - \delta$ and $d(x, r, H) < 1 - \delta$ for all $r < r_0$. If (u, w) is some interval with $\lambda(H|(u, w)) < \frac{1-\delta}{2}$, then there is some point a with $r(a)$ a minimal radius, such that $B(a, r(a)) \supset (u, w)$, and for each point $v \in (u, w)$ there is a point b_v, with a minimal radius $r(b_v)$, such that $B(b_v, r(b_v)) \supset (v, w)$ and*

$$b - a \geq u - v - \frac{2}{1-\delta}\lambda(H \cap (v, u)). \tag{3.3}$$

To establish this result, we consider that each point x in the complement of H is covered by both of two balls, $B(y, r(y))$ and $B(z, r(z))$, where $y < x < z$ and $r(y)$, $r(z)$ are both minimal. We can choose $s \in (u, w)$ to be a point such that $\lambda(H|(s, s'))$ is less than $1 - 2\delta$ for all $s' \in (s, w)$. Then if $a < s$, and $r(a)$ is a minimal radius for a, $B(a, r(a))$ must contain w if it contains s. This is because otherwise $\lambda(H|(s, a + r(a))) < 1 - 2\delta$, which implies $\lambda(H|B(a, r(a))\backslash B(a, s - a)) < 1 - \delta$ and so $r(a)$ cannot be minimal. Arguments like these allow us to show that $B(a, r(a)) \supset (u, w)$. We also find a point $t \in (v, w)$ such that $\lambda(H|(t', t)) < 1 - 2\delta$ if $t' \in (v, t)$, a point $b > t$ and a minimal radius $r(b)$ for b so that $B(b, r(b)) \supset (v, w)$. The equation (3.3) will follow from the choice of s and t.

Having proved Lemma 3.5.3, Kurka goes on to show how this implies Proposition 3.5.2. This is a much more complicated version of the proof of Lemma 3.5.1, which establishes that for the measure of a K-configuration to be close to the minimum possible, intervals on which the relative measure of H is less

than $1 - 2\delta$ must not be too rare. Each such interval implies the existence of two intervals I_a and I_b, on both of which H has relative measure at least $1 - \delta$. However, unlike the optimal case in Lemma 3.5.1, H has low relative measure where these two overlap, and high relative measure in the parts unique to each. This will allow us to eventually reach the improved lower bound of Proposition 3.5.2.

Chapter 4

Riemann sums

This chapter defines Riemann sums and briefly describes some notions based on them which are relevant to Chapters 5, 6 and 7.

4.1 Notation

Definition 4.1.1 A *partition of an interval* I is a finite collection of nondegenerate intervals $\mathcal{P} = \{I_k : 1 \leq k \leq n\}$ such that any two members of $\mathcal{P}$ are disjoint, except possibly for a shared endpoint (non-overlapping), and such that

$$I = \bigcup_{J \in \mathcal{P}} J. \tag{4.1}$$

We refer to a partition of $\mathbb{I}$ simply as a *partition*. The *size* of a partition, written $|\mathcal{P}|$, is the length of its longest element.

Definition 4.1.2 A *tagged partition of an interval* I is a finite collection of ordered pairs $\mathcal{T} = \{(t_k, I_k) : 1 \leq k \leq n\}$ such that each $t_k \in \mathbb{I}$ and $\{I_k : 1 \leq k \leq n\}$ is a partition of I.

As before we refer to a tagged partition of $\mathbb{I}$ simply as a *tagged partition*. Given a tagged partition $\mathcal{T} = \{(t_k, I_k) : 1 \leq k \leq n\}$, the partition $\{I_k : 1 \leq k \leq n\}$ is

referred to as its *underlying partition*. The *size* of a tagged partition is the size of its underlying partition.

For each interval I in the underlying partition of a tagged partition, its *tag* or *tag point* is the unique point t such that (t, I) is in the tagged partition.

Definition 4.1.3 Suppose that f is a function $\mathbb{I} \to \mathbb{R}$ and that $\mathcal{T}$ is a tagged partition. The *Riemann sum of f on $\mathcal{T}$* is given by

$$\sum_{\mathcal{T}} f := \sum_{(t,I)\in\mathcal{T}} f(t)|I| \tag{4.2}$$

Definition 4.1.4 A *gauge* is a function $\mathbb{I} \to \mathbb{R}^+$. If g is a gauge, we say that a tagged partition $\mathcal{T}$ is *smaller than g* if, for every (t, I) belonging to $\mathcal{T}$, we have

$$|I| \leq g(t). \tag{4.3}$$

Thus it is equivalent to say that the size of a tagged partition is smaller than δ and to say that the tagged partition is smaller than the gauge c_δ, where c_δ is the constant function $c_\delta(x) := \delta$.

The notion of Riemann sum gives rise to the Riemann integral. The definition and the important properties of this integral are widely known.

4.2 The Kurzweil-Henstock integral

The Kurzweil-Henstock integral was discovered by Kurzweil and its most important properties were elaborated by Henstock in [28].

Definition 4.2.1 A function f is *Kurzweil-Henstock integrable* or *KH-integrable* and M is its *Kurzweil-Henstock integral* iff for every $\varepsilon > 0$, there exists a gauge function $\delta(x)$ such that

$$\left|\sum_{\mathcal{T}} f - M\right| < \varepsilon \tag{4.4}$$

holds for any tagged partition $\mathcal{T}$ smaller than $\delta(x)$ and with the inclusion condition.

For this definition to be meaningful, we must be able to show that for any gauge function δ, we can find a tagged partition smaller than δ and with the inclusion condition. The proof of this is essentially a reworking of the proof of the Heine-Borel theorem for the unit interval.

The KH-integral of f on $\mathbb{I}$ is written as $\text{KH}\int f$, and as with the Lebesgue integral, the KH-integral of f on a measurable set $E \subset \mathbb{I}$ is defined as

$$\text{KH}\int_E f := \text{KH}\int \chi_E f.$$

There are several important theorems on the Kurzweil-Henstock integral which we state without proof. The proofs can be found for example in [23].

Theorem 4.2.2 (Extends Lebesgue integral) *Let f be a Lebesgue integrable function. Then f is KH-integrable and the KH-integral of f equals the Lebesgue integral of f.*

Thus the KH-integral is at least as powerful as the Lebesgue integral. Its main advantage over the Lebesgue integral is that every derivative can be integrated.

Theorem 4.2.3 (Fundamental Theorem of Calculus) *Suppose that f is differentiable at every point of $\mathbb{I}$. Then f' is KH-integrable and the indefinite integral of f', $\text{KH}\int_0^x f'$, equals $f(x) + c$ for all $x \in \mathbb{I}$, for some constant c .*

Remark The opposite of the theorem can be proved with only the Lebesgue integral. Suppose that $F(x)$ defined by $\int_0^x f$ is differentiable on $\mathbb{I}$. Then $F'(x) = f(x)$ almost everywhere on $\mathbb{I}$.

Remark It is not possible to prove the Fundamental Theorem for functions f which are only almost everywhere differentiable, as the celebrated counterexample known as the Cantor staircase shows.

We can prove both the two previous theorems directly, using in either case a regularity property of the functions in question. For Lebesgue integrable

functions it is the Lebesgue differentiation theorem, and for derivatives the fact that for every x,

$$F(x) - F(x+h) - hf(x) = \mathsf{o}(h) \text{ as } h \to 0.$$

The limits are not uniform in x in either case, which motivates the use of a gauge function, chosen to reflect the asymptotic behavior of the $\mathsf{o}(\frac{1}{h})$ term for a given x.

Whereas some functions cannot be integrated using proper Riemann or Lebesgue integration, and require the definition of an improper integral, this is never necessary for the KH-integral.

In fact, the following theorem tells us that the improper KH-integral is equivalent to the KH-integral.

Theorem 4.2.6 *Suppose that f is KH-integrable on $(\varepsilon, 1]$ for all $\varepsilon > 0$, and that*

$$\lim_{\varepsilon \to 0} \mathsf{KH}\int_{(\varepsilon,1]} f \tag{4.5}$$

exists. Then f is KH-integrable on $[0,1]$ and

$$\mathsf{KH}\int_{\mathbb{I}} f = \lim_{\varepsilon \to 0} \mathsf{KH}\int_{(\varepsilon,1]} f. \tag{4.6}$$

Like the Lebesgue integral, the indefinite KH-integral is continuous. Furthermore, the Henstock lemma, a technical lemma which is very important in proving some of the above properties of the integral, shows that if a function is KH-integrable on $\mathbb{I}$, then it is KH-integrable on every subinterval of $\mathbb{I}$. (Unlike the Lebesgue integral, it isn't true that the integrability is inherited by *every* measurable subset of $\mathbb{I}$.) These two facts provide a converse to the previous theorem: if we know the KH-integral of f on $\mathbb{I}$, then we know that $\mathsf{KH}\int_{\varepsilon}^{1} f$ exists for all $\varepsilon \in (0,1]$, and that its limit as $\varepsilon \to 0$ is $\mathsf{KH}\int f$.

The indefinite Lebesgue integral is absolutely continuous, but the indefinite KH integral is not. It is however, ACG^*, which gives the descriptive definition of

the Denjoy integral, equivalent to the Kurzweil-Henstock integral. A function is AC^* on some subset S of the domain, if for every set of intervals with endpoints in S, of total length δ, the variation of f on those intervals is at most $\varepsilon(\delta)$, for some function $\varepsilon(\delta)$ which tends to 0 as δ tends to 0. For f to be ACG^* on some set E means that there is a countable collection of sets whose union is E, on each of which f is AC^*.

The fact that every everywhere differentiable function is ACG^* yields Theorem 4.2.3. This can be proved by taking for each n the set E_n of points where the derivative is between $-n$ and n. The function is AC^* on each of these.

4.3 Extending the Riemann integral

Due to the ubiquity and the simplicity of the Riemann integral, there are several constructions which extend its power without requiring the more complicated notions used in the Lebesgue or Kurzweil-Henstock integral.

Definition 4.3.1 (Improper Riemann integral) Suppose that for every natural number n there exists some finite union of intervals $F_n \subsetneq \mathbb{I}$, such that $F_n \subseteq F_m$ if $n < m$, and so that the limit $\bigcup_n^\infty F_n = \mathbb{I}\backslash\{x_1, \ldots, x_k\}$ for some finite collection of points $x_1, \ldots, x_k$. Let f be a function which is Riemann integrable on each of the sets F_n. If the limit $\lim_{n\to\infty} \int_{F_n} f$ exists, and does not depend on the choice of the sets F_n, then this limit is the *improper Riemann integral of f*.

An example of an improper Riemann integrable function is an unbounded monotone function which is in L^1.

It is easy to see that every improperly Riemann integrable function is also Kurzweil-Henstock integrable. The paper [7] showed that an improper Riemann integral is equivalent to a Kurzweil-Henstock integral, if the set of possible gauge functions is restricted to those functions which are continuous outside some finite set of points.

However there do exist functions which are not Lebesgue integrable, but which are improperly Riemann integrable.

Definition 4.3.2 (Essential Riemann integral) We define the *essential Riemann integral of* f to be the Riemann integral of any function which is Lebesgue equivalent (almost everywhere equal) to f.

To see that the essential Riemann integral is well-defined, we need to check that any two functions which are both Riemann integrable, and almost everywhere equal, have the same Riemann integral. This is a simple consequence of the fact that the Riemann integral is equal to the Lebesgue integral, in the case that both exist. We refer to the set of functions that are Lebesgue equivalent to some Riemann integrable function as RI^*.

The Pollard-Getchell integral is another simple variant of the Riemann integral. The following description is based on that given in [28] (Chapter 1 section 5).

Definition 4.3.3 (Pollard-Getchell integral) A real number M is the *Pollard-Getchell* integral of f if for every $\varepsilon > 0$ there exists some partition $\mathcal{P}_0$ such that $|\sum_T f - M| < \varepsilon$ whenever T is a tagged partition with the inclusion condition whose underlying partition is a refinement of $\mathcal{P}_0$.

Since a refinement of a partition with size δ has size at most δ, we can see that all Riemann integrable functions are Pollard-Getchell integrable. It is also clear to see that it is at least as strong as the improper Riemann integral. For any partition $\mathcal{P}_0$ there exists a gauge function g such that all refinements of $\mathcal{P}_0$ are smaller than g. This means that Pollard-Getchell integrable functions are Kurzweil-Henstock integrable and that their Pollard-Getchell integral is equal to their Kurzweil-Henstock integral.

4.4 Further gauge integrals

4.4.1 The McShane integral

The McShane integral is very similar to the Kurzweil-Henstock integral. It is in fact a restriction of that, designed so that if f is integrable then $|f|$ will be too. This will mean that it is equivalent to the Lebesgue integral. The definition was first given in [34], and it is discussed in both [23] and [40].

Definition 4.4.1 (McShane integral) A function f is *McShane integrable* and M is its *McShane integral* iff for every $\varepsilon > 0$, there exists a gauge function $\delta(x)$ such that

$$\left|\sum_{\mathcal{T}} f - M\right| < \varepsilon \tag{4.7}$$

holds for any tagged partition $\mathcal{T}$ smaller than $\delta(x)$.

The only difference between this and the definition of the KH-integral is that here the partitions do not need to have the inclusion condition. It is clear that the set of functions which are McShane integrable is a subset of the set of Kurzweil-Henstock integrable functions.

The effect of the lack of the inclusion condition is not immediately obvious, however it can be seen through the following remark.

Remark Suppose that we have two tagged partitions $\mathcal{T}_0$ and $\mathcal{T}_1$ both of which are smaller than some gauge function $\delta(x)$. We can find a tagged partition $\mathcal{T}$ which is also smaller than $\delta(x)$ and which is a refinement of both $\mathcal{T}_0$ and $\mathcal{T}_1$, and we can choose the tags of $\mathcal{T}$ so that each one is a tag of either $\mathcal{T}_0$ or $\mathcal{T}_1$. To be precise, we will require that if (t, I) belongs to $\mathcal{T}$, then there is some interval $I' \supset I$ so that (t, I') belongs to either $\mathcal{T}_0$ or $\mathcal{T}_1$.

The partition $\mathcal{T}$ can for example consist of all the sets from

$$\{X \cap Y : X \in P(\mathcal{T}_0), Y \in P(\mathcal{T}_1)\}$$

which are nondegenerate intervals, where $P(\mathcal{T})$ denotes the underlying partition

of $\mathcal{T}$. For each such interval, contained within an interval from $\mathcal{T}_0$ and one from $\mathcal{T}_1$, we can take the tag from one or the other as its tag. If we had to restrict ourselves to partitions with the inclusion condition, this would not be possible as we have no guarantee that either tag point would belong to the interval. We can use this to show that that McShane integrability is inherited by measurable subsets. From this it is clear that the McShane integrable functions are absolutely integrable. So both the McShane integrable functions and the Lebesgue integrable functions are subspaces of the Kurzweil-Henstock integrable functions, such that if f is f is integrable then so is $|f|$. The following theorem can therefore be proved by showing that the positive KH-integrable functions are both Lebesgue integrable and McShane integrable.

Proposition 4.4.3 *A function f is McShane integrable if and only if it is Lebesgue integrable, and in the case that they exist, both integrals are the same.*

The way in which the inclusion condition makes the behavior of Riemann sums regular for less well-behaved functions is shown by the following result, taken from [7]:

Proposition 4.4.4 *A function f is improper Lebesgue integrable if and only if there exists some finite set $A \subset \mathbb{I}$ and for every $\varepsilon > 0$ there exists some gauge $\delta(x)$ so that for every tagged partition $\mathcal{T}$ of $\mathbb{I}$ smaller than $\delta(x)$ and with the inclusion condition holding on A, we have*

$$|\sum_{\mathcal{T}} f - \text{KH}\int f| < \varepsilon.$$

Here "the inclusion condition holding on A" means that if (t, I) is a member of the tagged partition, $t \in A$ implies $t \in I$.

In fact the original definition of the Denjoy integral, bears resemblance to an improper Lebesgue integral. The Denjoy integral, equivalent to the Kurzweil-Henstock integral, is something of a generalization of improper integration. In effect this allows limits of the Lebesgue integral to be taken on a much wider

range of subsets than the improper integral.

4.4.2 The C-integral

The following construction was first given by Bongiorno in [1], although its defining quality was proved by Bongiorno, DiPiazza and Preiss in [6]. The descriptive definition of an equivalent integral was first given in [8]. Based on the usage in these papers we refer to it as the C-integral. It imposes a new restriction on the tagged partitions which we will accept in our Riemann sums.

Definition 4.4.5 With ε a positive real number, we say that a tagged partition $\mathcal{T}$ has the *ε-proximity condition* if

$$\sum_{(t,I)\in\mathcal{T}} \operatorname{dist}(t,I) < \frac{1}{\varepsilon}.$$

It is clear that partitions with the inclusion condition have the ε-proximity condition for all positive ε. Thus the C-integral is stronger than the McShane integral and weaker than the Kurzweil-Henstock integral. So it is unnecessary to say what the value of the C-integral is and we simply define what it means to be C-integrable.

Definition 4.4.6 A function f is C-integrable if for every $\varepsilon > 0$ there exists a gauge function $\delta(x)$ such that

$$|\sum_{\mathcal{T}} f - \text{KH}\int f| < \varepsilon$$

for all partitions $\mathcal{T}$ which are smaller than $\delta(x)$ and which have the ε-proximity condition.

Proposition 4.4.7 *A measurable function f is C-integrable iff it is the sum of a Lebesgue integrable function and a derivative.*

It is trivial that a Lebesgue integrable function is C-integrable, and straightforward to see why a derivative is C-integrable. Suppose that $f = F'$ for F some

differentiable function on $\mathbb{I}$. Given a point $x \in \mathbb{I}$, firstly we can find a value $\delta(x)$ such that for any interval $I = (a, b)$ in $(x - \delta(x), x + \delta(x))$, we have

$$F(b) - F(a) = \int_a^b f = F'(x) + e$$

where $|e|$ is small. But in addition to this, we can make sure that $e = \mathsf{o}(d(x, I))$. Summing over all intervals in the partition would give us exactly that the ε-proximity condition is required for the Riemann sums to converge.

The proof of the converse is more involved and is given in [6].

Remark As is suggested by the definition of the ε-proximity condition, given a gauge function we can find a partition of $\mathbb{I}$ smaller than that gauge such that

$$\sum \operatorname{dist}(t, I)$$

is as large as we want. Suppose that $\phi(x)$ is a decreasing function such that $\lim_{x \to 0^+} f(x) = +\infty$. We can make a (ε, ϕ)-condition on partition sizes by replacing the bound $\frac{1}{\varepsilon}$ in the definition by $\phi(x)$. In fact any integral defined using such a variant proximity condition would be equivalent to the C-integral, with the choice of ϕ affecting only the choice of the gauge functions.

Additional properties of the C-integral are given in [2] and in [3], which discuss its primitives and prove several convergence theorems.

Chapter 5

Random Riemann integrals

In this chapter we introduce the *random Riemann integral*. In the first section, we give the definition of the random Riemann sum of a function on a partition. This is a random variable in the space of Riemann sums of f. We show the easy result that if f is Lebesgue integrable these random variables converge in probability as the size of the partitions goes to 0. This property we will refer to as weak random Riemann integrability. In the second section we consider almost sure convergence of random Riemann sums. This will depend on the sequence of partitions taken, and is referred to as random Riemann integrability with respect to a sequence. We prove that if f is in L^p and the sequence of partition sizes is in ℓ^{p-1} then this holds. We go on to give a partial converse, as well as a counterexample that shows that this is not the exact necessary condition. The third section gives two classes of not Lebesgue integrable functions which are not weakly random Riemann integrable. The proof depends on a lemma which is proved slightly differently for either class. The result is suggestive that the set of weakly random Riemann integrable functions may be exactly L^1.

5.1 Basics

Let a partition $\mathcal{P}$ of $\mathbb{I}$ be given, and suppose it contains n intervals $(I_k)_{k=1}^n$. For each k let t_k be a random variable distributed uniformly on I_k, independently of all t_j, $j \neq k$. The partition

$$T_{\mathcal{P}} := \{(t_k, I_k) : 1 \leq k \leq n\}$$

is a random variable in the space of tagged partitions on $\mathbb{I}$ with the inclusion condition whose underlying partition is $\mathcal{P}$.

Definition 5.1.1 We define the *random Riemann sum* of f on $\mathcal{P}$ to be the Riemann sum of f on $T_{\mathcal{P}}$,

$$\Sigma_{\mathcal{P}} f := \sum_{k=1}^{n} |I_k| f(t_k),$$

which is a random variable in $\mathbb{R}$ if f is an almost everywhere finite Lebesgue measurable function.

It is clear that the random Riemann sum of f is linear in f, and that if $f \geq g$ almost everywhere then $\Sigma_{\mathcal{P}} f \geq \Sigma_{\mathcal{P}} g$ almost surely.

Definition 5.1.2 We say that f is *weakly random Riemann integrable* and that M is its *random Riemann integral* if for all $\varepsilon > 0$ there exists a δ such that

$$|\mathcal{P}| < \delta \implies \mathsf{P}(|\Sigma_{\mathcal{P}} f - M| > \varepsilon) < \varepsilon.$$

This is to say that we have convergence in probability of the random Riemann sum on $\mathcal{P}$ to M, as the size of $\mathcal{P}$ tends to 0.

Our first and easiest task is to show that the weak random Riemann integral extends the Riemann integral, in the following sense.

Lemma 5.1.3 *If f is a Riemann integrable function, then it is weakly random*

Riemann integrable, and then the random Riemann integral of f is equal to the Riemann integral of f.

Proof. This is obvious since $\Sigma_{\mathcal{P}} f$ is a random variable supported on the set of Riemann sums of f on all tagged partitions with the inclusion condition, whose underlying partition is $\mathcal{P}$. ■

We expect this lemma to hold for every reasonable construction of a random variable involving Riemann sums of f.

The following lemma will be used to prove that if we can approximate a function in L^1 by random Riemann integrable functions then that function too is random Riemann integrable. In later chapters we will see further constructions involving stochastic Riemann sums. Where we require their convergence in probability an analogous property will usually be immediate. However it does not necessarily follow in the case of almost sure convergence.

Lemma 5.1.4 *Let $\varepsilon > 0$ be given. There exists δ such that*

$$\int |f| < \delta \implies \mathsf{P}(|\Sigma_{\mathcal{P}} f| > \varepsilon) < \varepsilon$$

for any partition $\mathcal{P}$.

Proof. The expectation of $\Sigma_{\mathcal{P}} f$ is

$$\mathsf{E}(\Sigma_{\mathcal{P}} f) = \sum_{k=1}^{n} |I_k| \mathsf{E} f(t_k) = \sum_{k=1}^{n} \int_{I_k} f = \int f.$$

Since[(1)]

$$\begin{aligned} |\Sigma_{\mathcal{P}} f| &= |\Sigma_{\mathcal{P}} f^+ - \Sigma_{\mathcal{P}} f^-| \\ &\leq \Sigma_{\mathcal{P}} f^+ + \Sigma_{\mathcal{P}} f^- \\ &= \Sigma_{\mathcal{P}} |f| \end{aligned}$$

[(1)] The function f^+ is the positive part of f, $f^+(x) := \max(f(x), 0)$. Similarly the negative part of f is $f^- := \max(-f(x), 0)$.

we have that

$$|\mathsf{E}\Sigma_{\mathcal{P}} f| \leq \mathsf{E}\sum_{k=1}^{n}\int_{I_k}|f| = \int|f|.$$

Therefore by Chebyshev's inequality $\mathsf{P}(|\Sigma_{\mathcal{P}} f| > \varepsilon) < \frac{\int|f|}{\varepsilon}$ as required. ■

We can now show that the weak random Riemann integral extends the Lebesgue integral.

Theorem 5.1.5 *If f is Lebesgue integrable, then it is weakly random Riemann integrable, and the two integrals are equal.*

Proof. We can write f as the sum of two function

$$f = g + h$$

where g is Riemann integrable, and $\int|h|$ is less than $\frac{\varepsilon}{3}$, and (using Lemma 5.1.4) also so small that $\mathsf{P}(|\Sigma_{\mathcal{P}} h| > \frac{\varepsilon}{3}) < \frac{\varepsilon}{3}$ for any partition $\mathcal{P}$.

Since g is Riemann integrable there exists a δ such that if $|\mathcal{P}| < \delta$ then $|\Sigma_{\mathcal{P}} g - \int g| < \frac{\varepsilon}{3}$. Putting these together yields that

$$\begin{aligned}\mathsf{P}(|\Sigma_{\mathcal{P}} f - \int f| > \varepsilon) &\leq \mathsf{P}(|\Sigma_{\mathcal{P}} g - \int g| > \frac{\varepsilon}{3}) + \mathsf{P}(|\Sigma_{\mathcal{P}} h - \int h| > \frac{2\varepsilon}{3}\\ &\leq 0 + \mathsf{P}(|\Sigma_{\mathcal{P}} h| > \frac{\varepsilon}{3}\\ &\leq \frac{\varepsilon}{3} \leq \varepsilon\end{aligned}$$

for all partitions smaller than δ, as required. ■

5.2 The random Riemann integral

Suppose that $(\mathcal{P}_n)_{n\in\mathbb{N}}$ is a sequence of partitions. It will be useful to write $(|\mathcal{P}_n|)$ to denote the sequence of sizes of the partitions $(\mathcal{P}_n)$. Suppose that we take random Riemann sums on ¶$_n$ for each n, distributed exactly as before, and the distribution of each is independent of all others. If I is an interval of $\mathcal{P}_n$ we call its tag points $t_{I,n}$.

Definition 5.2.1 A function f is *strongly random Riemann integrable*, or just *random Riemann integrable* with respect to $(\mathcal{P}_n)_{n\in\mathbb{N}}$, where $(\mathcal{P}_n)$ is a sequence of partitions of $\mathbb{I}$, if the sequence of random variables $(\Sigma_{\mathcal{P}_n} f)$ converges almost surely.

Remark Suppose that $\Sigma_{\mathcal{P}_n} f$ converges almost surely to some real number. Then it converges in probability as well, with the same limit. So we do not study the value of the strong random Riemann integral, but only whether or not it converges.

The first result gives us a condition when we do have almost sure convergence. Its sufficiency was first proved in [24].

Theorem 5.2.3 *Suppose that $p > 1$. Let f be a measurable function in L^p and let $(\mathcal{P}_n)$ be a sequence of partitions with the sequence of partition sizes $(|\mathcal{P}_n|)$ in ℓ^{p-1}. Then f is strongly random Riemann integrable with respect to $(\mathcal{P}_n)$.*

Proof. First suppose that $1 < p \leq 2$. For each $n \in \mathbb{N}$, we define f_n as follows. On each interval of $\mathcal{P}_n$, f_n is constant and has the same integral as f on that interval. That is, if $x \in I \in \mathcal{P}_n$, then

$$f_n(x) := \frac{1}{|I|} \int_I f$$

Further define g_n by $g_n(x) := f(x) - f_n(x)$ for all $x \in \mathbb{I}$ and $n \in \mathbb{N}$.

It is easy to see that $\Sigma_{\mathcal{P}_n} f_n$ is constant and equal to $\int f$ with certainty for each $n \in \mathbb{N}$. We know that $\Sigma_{\mathcal{P}_n} f = \Sigma_{\mathcal{P}_n} f_n + \Sigma_{\mathcal{P}_n} g_n$, and so it is enough to show that $\Sigma_{\mathcal{P}_n} g_n \to 0$ almost surely.

We will use the fact that $\int |g_n|^p$ is bounded by a constant that does not depend on n. In fact each f_n is dominated by Mf, the Hardy-Littlewood maximal operator of f (see Definition 2.2.9). We know that $\int (Mf)^p$ is finite if $\int |f|^p$ is, and so by the triangle inequality in L^p, $\int |g_n|^p$ is at most

$$\left(\left(\int |f|^p \right)^{\frac{1}{p}} + \left(\int |f_n|^p \right)^{\frac{1}{p}} \right)^p \leq 2^p \int (Mf)^p =: c < \infty.$$

Write $X_{I,n} := |I| g_n(t_{I,n})$. Then $\Sigma_{\mathcal{P}_n} g_n = \sum_{I \in \mathcal{P}_n} X_{I,n}$. Note that

$$\mathsf{E} X_{I,n} = \int_I g_n = \int_I f - f_n = 0,$$

therefore also $\mathsf{E}(\Sigma_{\mathcal{P}_n} g_n) = 0$.

We use a lemma on sums of random variables with expectation 0.

Lemma 5.2.4 *If $\{X_j : 1 \leq j \leq n\}$ is a finite collection of independent random variables with expectation 0, and $1 < p \leq 2$, then*

$$\mathsf{E}\,(|\sum_{j=1}^{k} X_j\,|^p) \leq 2 \sum_{j=1}^{k} \mathsf{E}(|X_j|^p).$$

This is a consequence of a second lemma:

Lemma 5.2.5 *If x and y are real numbers, and $1 < p \leq 2$, then*

$$|x+y|^p \leq |x|^p + 2|y|^p + py|x|^{p-1}\,\mathrm{sign}(x).$$

Proof (Lemma 5.2.5). To prove this lemma it is sufficient to prove it for $x = 1$, in other words that

$$|1+y|^p \leq 1 + 2|y|^p + py \tag{5.1}$$

for all $y \in \mathbb{R}$ and all $1 < p \leq 2$. If $p = 2$ the inequality holds trivially, so assume $1 < p < 2$. Define

$$G(y) := \frac{|1+y|^p - 1 - py}{|y|^p}.$$

Straightforward calculation shows us that

$$G(-1) = p - 1$$

$$\lim_{y \to 0^-} G(y) = 0$$

$$\lim_{y \to 0^+} G(y) = 0$$

$$\lim_{y \to \infty} G(y) = 1$$

$$\lim_{y \to -\infty} G(y) = 1$$

and that the derivative of G is nonzero on $(-1, 0)$ and $(0, \infty)$. This means that $G(x)$ is bounded by $p - 1$ on $(-1, 0)$ and by 1 on $(0, \infty)$.

For $y < -1$, let us set for convenience $z := -y$, and define $G^*(z) := G(y)$. The function is given by

$$G^*(z) = \frac{(z-1)^p - 1 + pz}{z^p}$$

and we can see that as $z \to \infty$, this tends to 1. Assume that it attains a maximum for $1 < z$, then the location of this maximum would be given by

$$\frac{d}{dz} G^*(z) = pz^{(-1-p)} \left(z + z(z-1)^{p-1} + 1 - (z-1)^p - pz \right) = 0$$

or equivalently

$$(z-1)^p - 1 + pz = z + z(z-1)^{p-1}.$$

This means that the function $G^*(z)$ has the value

$$\frac{z + z(z-1)^{p-1}}{z^p} = \frac{1}{z^{p-1}} + \frac{(z-1)^{p-1}}{z^{p-1}}$$

at its maximum if there is one, and this is clearly less than 2 since both summands on the right hand side are less than 1. We have shown that $G(y) < 2$ for all y and all $1 < p < 2$ and this proves (5.1). ■

Proof (Lemma 5.2.4). We use induction, noting that the inequality is trivial for $n = 1$. Assume that for the collection X_j, $1 \leq j \leq k$ of independent random variables, each with expectation 0, we have

$$\mathsf{E}\left(\left|\sum_{j=1}^{k} X_j\right|^p\right) \leq 2\sum_{j=1}^{k} \mathsf{E}(|X_j|^p).$$

Then setting $x = \sum_{j=1}^{k} X_j$ and $y = X_{k+1}$ in Lemma 5.2.5, we obtain

$$\left|\sum_{j=1}^{k+1} X_j\right|^p \leq \left|\sum_{j=1}^{k} X_j\right|^p + 2|X_{k+1}|^p + pX_{k+1}\left|\sum_{j=1}^{k} X_j\right|^{p-1} \operatorname{sign}(\sum_{j=1}^{k} X_j).$$

Now if we take expectations of each term, we can use the fact that the last term has expectation 0. This is true because

$$\mathsf{E}\left(pX_{k+1}\left|\sum_{j=1}^{k} X_j\right|^{p-1} \operatorname{sign}(\sum_{j=1}^{k} X_j)\right) = \mathsf{E}X_{k+1} \cdot \mathsf{E}\left(p\left|\sum_{j=1}^{k} X_j\right|^{p-1} \operatorname{sign}(\sum_{j=1}^{k} X_j)\right)$$

by independence, and $\mathsf{E}X_{k+1}$ is 0. So using the induction step we obtain

$$\begin{aligned}
\mathsf{E}\left|\sum_{j=1}^{k+1} X_j\right|^p &\leq \mathsf{E}\left|\sum_{j=1}^{k} X_j\right|^p + 2\mathsf{E}|X_{k+1}|^p \\
&\leq 2\sum_{j=1}^{k} \mathsf{E}(|X_j|^p) + 2\mathsf{E}|X_{k+1}|^p \\
&= 2\sum_{j=1}^{k+1} \mathsf{E}(|X_j|^p)
\end{aligned}$$

as required. ■

By this lemma and since $\mathsf{E}X_{I,n} = 0$, we have the inequality

$$\mathsf{E}\left(\left|\sum_{I\in\mathcal{P}_n} X_{I,n}\right|^p\right) \leq 2\sum_{I\in\mathcal{P}_n} \mathsf{E}\left(|X_{I,n}|^p\right). \tag{5.2}$$

Now

$$\sum_{I\in\mathcal{P}_n} \mathsf{E}(|X_{I,n}|^p) \le |\mathcal{P}_n|^{p-1} \int |g_n|^p \tag{5.3}$$

which can be seen simply by expanding each component of the sum on the left hand side as $\mathsf{E}|g_n(t_{I,n})|^p|I|^p$. So by (5.2) and the choice of $(\mathcal{P}_n)$ with $(|\mathcal{P}_n|) \in \ell^{p-1}$, we have

$$\sum_{n=1}^{\infty} \mathsf{E}\left(\left|\sum_{I\in\mathcal{P}_n} X_{I,n}\right|^p\right) \le 2\sum_{n=1}^{\infty} |\mathcal{P}_n|^{p-1} \int |g_n|^p \le 2c\sum_{n=1}^{\infty} |\mathcal{P}_n|^{p-1} < +\infty.$$

An application of Markov's inequality tells us that

$$\sum_{n=1}^{\infty} \mathsf{P}\left(\left|\sum_{I\in\mathcal{P}_n} X_{I,n}\right| > \varepsilon\right) \le \frac{1}{\varepsilon^p}\sum_{n=1}^{\infty} \mathsf{E}\left(\left|\sum_{I\in\mathcal{P}_n} X_{I,n}\right|^p\right) < +\infty$$

for all positive ε. This property is called complete convergence of $\sum_I X_{I,n}$ to 0. By the Borel-Cantelli lemma it implies almost sure convergence to the same limit.

Now suppose that $p > 2$. For each $n \in \mathbb{N}$ we write f as a sum $f = f_n^* + f_n^{**}$. Here we define

$$f_n^* := \begin{cases} f & \text{if } |f| < |\mathcal{P}_n|^{\frac{-1}{p}}, \\ 0 & \text{otherwise} \end{cases}$$

and define $f_n^{**} := f - f_n^*$. Since f is in L^p, $\int f_n^{**} \to 0$ and $\int f_n^* \to \int f$ as $n \to \infty$.

We will use two lemmas. The first was proved by Evans and Humke in Theorem 2 of [18]. Their proof of the result for the random first-return Riemann sum applies equally to our random Riemann sum, which is identically distributed. See Chapter 7 for details of the random first-return Riemann sum.

Lemma 5.2.6 *Suppose that f is a measurable function on $\mathbb{I}$ with $|f| < M$. Then for any partition $\mathcal{P}$ and any $m \in \mathbb{N}$, we have*

$$\left|\mathsf{E}\left(\int f - \Sigma_{\mathcal{P}} f\right)^{2m}\right| \le c \cdot M^{2m}|\mathcal{P}|^m,$$

where c depends only on m.

Proof. We can express $\int f - \Sigma_{\mathcal{P}} f$ as the sum

$$\sum_{I_k \in \mathcal{P}} (\int_{I_k} f - |I_k| f(t_{I_k})).$$

Since $\mathsf{E}|I|f(t_I)$ is $\int_I f$ this is a sum of random variables with mean 0. Moreover each is independent of all the others. Let us write them as

$$F_k := \int_{I_k} f - |I_k| f(t_{I_k})$$

for $1 \leq k \leq n$. The expected value of the $2m^{th}$ power of their sum is given by

$$\mathsf{E}\left(\sum_k F_k\right)^{2m} = \mathsf{E} \sum_{\substack{j_1 \neq 1 \\ j_2 \neq 1 \\ \cdots \\ j_n \neq 1 \\ \sum_k j_k = 2m}} \prod_k F_k^{j_k}. \tag{5.4}$$

This is because if we take a product where one of the F_k is taken to the power 1, then the expectation of this will be 0:

$$\mathsf{E}\left(F_1 \prod_{k=2}^{n} F_k^{j_k}\right) = \mathsf{E}F_1 \mathsf{E} \prod_{k=2}^{n} F_k^{j_k} = 0.$$

We are going to show that the right hand side is less than

$$\mathsf{E}\left(\sum_k F_k^2\right)^m.$$

Let $Z_{n,m}$ be the set of indices given on the right hand side of 5.4, that is to say the family $\{(j_1, \ldots, j_n) : \sum_k j_k = 2m,\ j_k \neq 1 \, \forall k\}$. For $i \in \{0, 1, \ldots, \lfloor \frac{m}{3} \rfloor\}$ let $Z_{n,m,i}$ be the subset of $Z_{n,m}$ which has exactly $2i$ odd terms. We can show that

$$\binom{2i}{i} \# Z_{n,m,i} \leq \binom{m}{i} \binom{m-i}{i} \# Z_{n,m,0}. \tag{5.5}$$

This is a counting argument; we can form a mapping

$$\phi : \binom{2i}{i} \times Z_{n,m,i} \to Z_{n,m,0}$$

by considering the element of $\binom{2i}{i}$ to be a subset of the odd terms of $(j_1, \ldots, j_n)$. We increment all the odd terms in this subset by 1 and decrement all other odd terms by 1. We can extend the mapping to be

$$\phi^* \binom{2i}{i} \times Z_{n,m,i} \to \binom{m}{i} \times \binom{m-i}{i} \times Z_{n,m,0}$$

by setting the element of $\binom{m}{i}$ to be the subset of the nonzero terms which were increased and the element of $\binom{m-i}{i}$ to be the subset of those remaining terms which were decreased. (There are fewer than m nonzero terms.) This is an injection since we can map $(S, S', (j_1, \ldots, j_n))$ back into $\binom{2i}{i} \times Z_{n,m,i}$ by reversing the increments and decrements, and considering S to be a subset of $S \cup S'$. It is not a surjection as there are strictly fewer than m nonzero terms, and because if $j_k = 2$ for k in the set to be decremented, the image under the inverse map will not be in $Z_{n,m,i}$.

Now this implies that

$$\binom{2i}{i} \sum_{Z_{n,m,i}} \prod_k F_k^{j_k} \leq \binom{m}{i} \binom{m-i}{i} \sum_{Z_{n,m,0}} \prod_k F_k^{j_k}.$$

Here we used (5.5) as well as the fact that $2ab \leq a^2 + b^2$. The latter means that when we replaced $F_k^{j_k} F_l^{j_l}$ by alternately $F_k^{j_k+1} F_l^{j_l-1}$ and $F_k^{j_k-1} F_l^{j_l+1}$ we did not make the sum smaller.

So using the identity $\binom{m}{2i}\binom{2i}{i} = \binom{m}{i}\binom{m-i}{i}$,

$$\sum_{Z_{n,m}} \prod_k F_k^{j_k} \leq \sum_{i=0}^{\lfloor \frac{m}{3} \rfloor} \binom{m}{2i} \sum_{Z_{n,m,0}} \prod_k F_k^{j_k}$$

The right hand side is less than

$$2^m \sum_{Z_{n,m,0}} \prod_k F_k^{j_k}.$$

If we write $U_{n,m}$ for the sets of indices $(i_1, i_2, \ldots, i_n)$ where $\sum_{k=1}^n i_k = m$, the sum can also be expressed as

$$\sum_{U_{n,m}} \prod_k F_k^{2i_k}. \tag{5.6}$$

The terms of this sum correspond to a (small) subset of the terms of

$$\left(\sum_k F_k^2\right)^m = \sum_{\substack{(a_1,a_2,\ldots,a_m)\\ \in\{1,\ldots,n\}^m}} \prod_h F_{a_h}^2. \tag{5.7}$$

In fact, each term of (5.6) appears at least once and at most $m!$ times in (5.7). So since $2^m < m!$,

$$\sum_{Z_{n,m}} \prod_k F_k^{j_k} \le 2^m \sum_{Z_{n,m,0}} \prod_k F_k^{j_k} \le 2^m \left(\sum_k F_k^2\right)^m.$$

Now, each F_k is bounded by $2M|I_K|$ where M is the bound on $|f|$. Therefore

$$\left(\sum_k F_k^2\right)^m \le 2^{2m} M^{2m} \left(\sum_k |I_k|^2\right)^m.$$

It remains to see that

$$\sum_k |I_k|^2 \le |\mathcal{P}| \sum_k |I_k| = |\mathcal{P}|$$

and we have

$$\mathsf{E}\left(\sum_k F_k\right)^{2m} \le 2^{3m} M^{2m} |\mathcal{P}|^m$$

as required. ■

Now, choose an integer $m > \frac{p-1}{1-\frac{2}{p}}$. Since $|f_n^*| < |\mathcal{P}_n|^{\frac{-1}{p}}$, we have that

$$\mathsf{E}\left(\left|\int f_n^* - \Sigma_{\mathcal{P}_n} f_n^*\right|^{2m}\right) \leq c|\mathcal{P}_n|^{\frac{-2m}{p}}|\mathcal{P}_n|^m = c|\mathcal{P}_n|^{m(1-\frac{2}{p})}.$$

Since $(|\mathcal{P}_n|) \in \ell^{p-1}$, and by the choice of m, this expectation has finite sum over n. As before then, this means that $\Sigma_{\mathcal{P}_n} f_n^*$ converges completely and almost surely to $\lim_{n\to\infty} \int f_n^* = \int f$.

To deal with the random Riemann sums of f_n^{**}, we have a second lemma, an standard result about random variables.

Lemma 5.2.7 *Let $p > 1$ and suppose that X_k, $1 \leq k \leq n$, are independent random variables such that $\mathsf{E}|X_k|^p < +\infty$ for all k. Denote $\sum_{k=1}^n X_k$ by S. Then*

$$\mathsf{E}|S|^p \leq \max\left\{2^p \sum_k \mathsf{E}|X_k|^p, 2^{p^2}\left(\sum_k \mathsf{E}|X_k|\right)^p\right\}.$$

This result is given in [26]. We omit the proof.

Now, we apply this lemma to the random variables $X_{I,n}^{**}$, which are defined analogously to $X_{I,n}$ as the summands of the random Riemann sum, this time for the functions f_n^{**}. They satisfy the conditions of the lemma, and we find that

$$\mathsf{E}\left|(\Sigma_{\mathcal{P}_n} f_n^{**})^p\right| = \mathsf{E}\left|\sum_I X_{I,n}^{**}\right|^p \leq c \max\left(\sum_I \mathsf{E}|X_{I,n}^{**}|^p, \left(\sum_I \mathsf{E}|X_{I,n}^{**}|\right)^p\right),$$

where c is some constant. We wish to show that the left hand side is smaller than $c' \cdot |\mathcal{P}_n|^{p-1}$ for some other constant c'. Obviously we can do this by finding a bound of this form for each of the terms of the maximum on the right hand side.

Firstly we know that

$$\sum_I \mathsf{E}(|X_{I,n}^{**}|^p) \leq |\mathcal{P}_n|^{p-1} \int |f_n^{**}|^p,$$

which can be proved in the same way as (5.3). Since $f_n^{**} \to 0$ in L^p, $\int |f_n^{**}|^p$

is bounded over all n, and so we can take $c_1 := c \cdot \max\{|f_n^{**}|^p, n \in \mathbb{N}\}$ as the appropriate constant.

Now consider $(\sum_I \mathsf{E}|X_{I,n}^{**}|)^p$. We have that $\sum_I \mathsf{E}|X_{I,n}^{**}| \leq \int |f_n^{**}|$. Either $|f_n^{**}| = 0$ or $|f_n^{**}| > |\mathcal{P}_n|^{\frac{-1}{p}}$, and so

$$\int |f_n^{**}| \leq |\mathcal{P}_n|^{\frac{p-1}{p}} \int |f_n^{**}|^p.$$

Therefore

$$\left(\sum_I \mathsf{E}|X_n^{**}|\right)^p \leq |\mathcal{P}_n|^{p-1} \left(\int |f_n^{**}|^p\right)^p.$$

So we can take $c_2 := c \cdot (\max\{\int |f_n^{**}|^p, n \in \mathbb{N}\})^p$. Then if $c' := \max(c_1, c_2)$, we have

$$\mathsf{E}|(\Sigma_{\mathcal{P}_n} f^{**})^p| \leq c'|\mathcal{P}_n|^{p-1}.$$

This means that

$$\sum_{n=1}^{\infty} \mathsf{E}|(\Sigma_{\mathcal{P}_n} f^{**})^p| < +\infty,$$

and as before, this proves that $\Sigma_{\mathcal{P}_n} f_n^{**} \to 0$ almost surely as $n \to \infty$. ■

We have a partial converse to this theorem. This depends on f being non-negative, and on two further conditions, other than the conditions that $f \notin L^p$ and $(|\mathcal{P}_n|) \notin \ell^{p-1}$ for some p. The first of these is that we will only deal with sequences of partitions for which the ratio between the lengths of the largest and smallest intervals in each partition is bounded.

In fact if we seek a necessary and sufficient condition for f to be random Riemann integrable with respect to $(\mathcal{P}_n)_{n=1}^{\infty}$, we have to be content with a condition which bounds both the minimum and maximum size of intervals in $\mathcal{P}_n$. If we use the standard definition of the size of a partition, we can find functions which are random Riemann integrable with respect to $(\mathcal{P}_n)_{n\in\mathbb{N}}$, for $\mathcal{P}_n$ of any size, by choosing $\mathcal{P}_n$ with one large interval, in a region where f is well-behaved, and all other intervals of $\mathcal{P}_n$ much smaller.

On the other hand, we can find a sequence $(\mathcal{P}_n)_{n\in\mathbb{N}}$ where the *minimum* size of intervals in $\mathcal{P}_n$ is as small as we like, so that no non-constant f is random Riemann integrable with respect to $(\mathcal{P}_n)$, by taking one interval which is included in every $\mathcal{P}_n$, on which the distribution of $f(t_I)$ will not converge.

The second condition relates the growth of f to the sequence of partition sizes $(|\mathcal{P}_n|)_{n\in\mathbb{N}}$. It tries to avoid that for all $c > 0$, the measure of $\{x : f(x) > \frac{c}{|\mathcal{P}_n|}\}$ is small for most n. We can see both conditions in the following lemma.

Lemma 5.2.8 *Suppose that $p > 1$, that $d > 0$, and that f is a nonnegative function in L^1. Further suppose that $(\delta_n)_{n=1}^{\infty}$ is a sequence of positive numbers tending to 0 and for some $A > 0$, $B > 1$, $C > 0$, the set J of $n \in \mathbb{N}$ for which*

$$\lambda\left(\{x : f(x) > Z_n\}\right) > CZ_n^{-p} \text{ for some } Z_n \in \left(\frac{A}{\delta_n}, \frac{AB}{\delta_n}\right) \tag{5.8}$$

is such that $\sum_{n\in J} \delta_n^{p-1} = +\infty$. Let $(\mathcal{P}_n)$ be a sequence of partitions with $\delta_n < |I| < M\delta_n$ for all $I \in \mathcal{P}_n$ and all $n \in \mathbb{N}$, for some $M > 0$. Then there exists some $\varepsilon > 0$ with $|\Sigma_{\mathcal{P}_n} f - \int f| > \varepsilon$ for infinitely many n, almost surely.

To prove this we need a simple lemma about probabilities.

Lemma 5.2.9 *Suppose that $\{A_k : 1 \le k \le n\}$ is a finite collection of independent events, such that*

$$\sum_{k=1}^{n} \mathsf{P}(A_k) \le 1.$$

Then the probability of $\bigcup_{k=1}^{n} A_n$ is at least half of $\sum_{k=1}^{n} \mathsf{P}(A_k)$.

Proof (Lemma 5.2.9). Define S to be the sum of the probabilities of the events A_k,

$$S := \sum_{k=1}^{n} \mathsf{P}(A_k)$$

and further define

$$a_k := \frac{\mathsf{P}(A_k)}{S},\ \forall 1 \le k \le n.$$

Using independence, we obtain

$$\begin{aligned}\mathsf{P}\left(\bigcap_{k=1}^{n} A_k^C\right) &= \prod_{k=1}^{n}(1-\mathsf{P}(A_k)) \\ &= \prod_{k=1}^{n}(1-a_k S).\end{aligned}$$

Since $\sum_k a_k = 1$, we use the well known fact that if we fix the sum of several positive variables, their product is maximised when they are all equal.

$$\prod_{k=1}^{n}(1-a_k S) \leq (1-\frac{S}{n})^n$$

This last expression is monotonically increasing in n for $n \geq S$, and is bounded above by its limit e^{-S}, which is less than $1-\frac{S}{2}$ for $S \in [0,1]$. So

$$\mathsf{P}\left(\bigcap_{k=1}^{n} A_k^C\right) < 1-\frac{S}{2}$$

and

$$\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right) > \frac{S}{2}$$

as required. ■

Proof (Lemma 5.2.8). Assume without loss of generality that (5.8) holds for all $n \in \mathbb{N}$. Suppose that A, B, C and M are as in the statement of the lemma, and $(Z_n)_{n\in\mathbb{N}}$ is the sequence such that $Z_n \in \left(\frac{A}{\delta_n}, \frac{AB}{\delta_n}\right)$,

$$\lambda\left(\{x : f(x) > Z_n\}\right) > C Z_n^{-p},\ \forall n \in \mathbb{N}. \tag{5.9}$$

Also without loss of generality assume that $\int f < \frac{A}{4}$, by subtracting a Riemann integrable function from f if necessary. Trivially, the Riemann integrable function will be random Riemann integrable so we can neglect it. Since the Riemann integrable function is bounded, say by N, assume that $\frac{1}{\delta_n} > \frac{N}{2}$ for

all n. We can see that (5.9) still holds with $\frac{C}{2}$ in place of C since we will have $\frac{C}{2}Z_n^{-p} < C(Z_n^{-p} - N)$. By changing the values of C we keep the statement as given.

Now choose some $\varepsilon \in (4\int f, A)$. Fix $n \in \mathbb{N}$. For each $I \in \mathcal{P}_n$, let K_I be the set $\{x \in I : |I|f(x) > A\}$. We know that

$$K_I \supseteq \{x \in I : \delta_n f(x) > A\} \supseteq \{x : f(x) > Z_n\}$$

and so

$$\lambda\left(\bigcup_{I\in\mathcal{P}_n} K_I\right) \geq \lambda\{x \in \mathbb{I} : f(x) > Z_n\} \geq CZ_n^{-p}.$$

Since for each I such that $K_I \neq \emptyset$

$$\mathsf{P}\left(t_{I,n} \in K_I\right) = \frac{\lambda(K_I)}{|I|} \geq \frac{\lambda(K_I)}{M\delta_n}$$

where $t_{I,n}$ is the random tag point in the interval $I \in \mathcal{P}_n$, we have

$$\sum_{I\in\mathcal{P}_n} \mathsf{P}\left(t_I \in K_I\right) \geq \frac{\sum \lambda(K_I)}{M\delta_n} \geq \frac{CZ_n^{-p}}{M\delta_n} \geq \frac{C\delta_n^p}{M\delta_n A^p B^p} = C'\delta^{p-1}$$

where C' only depends on A, B, C and M.

Let $q := \min(1, C'\delta_n^{p-1})$. We can choose a set of intervals $\mathcal{J} \subset \mathcal{P}_n$ such that $q/2 \leq \sum_{I\in\mathcal{J}} \mathsf{P}(t_I \in K_I) \leq 1$. Then, using Lemma 5.2.9, we know that

$$\mathsf{P}(\exists I \in \mathcal{J} : t_I \in K_I) > \frac{q}{4} = \min(\tfrac{1}{4}, c''\delta_n^{p-1}). \tag{5.10}$$

Now define

$$X_n := \sum_{I\in\mathcal{J}} f(t_I)|I| \qquad \text{and} \qquad Y_n := \Sigma_{\mathcal{P}_n} f - X_n = \sum_{I\notin\mathcal{J}} f(t_I)|I|.$$

From (5.10), and using the fact that f and therefore X_n is nonnegative, we

know that

$$\begin{aligned}\mathsf{P}(X_n > \varepsilon) &\geq \mathsf{P}(\exists I \in \mathcal{J} : |I| f(t_I) > \varepsilon) \\ &\geq \mathsf{P}(\exists I \in \mathcal{J} : t_I \in K_I) \\ &\geq \min(\tfrac{1}{4}, c'' \delta_n^{p-1}).\end{aligned}$$

As for Y_n,

$$\Big|Y_n - \int f\Big| \leq |Y_n| + \Big|\int f\Big|.$$

So,

$$\begin{aligned}\mathsf{E}\Big|Y_n - \int f\Big| &\leq \mathsf{E}Y_n + \mathsf{E}\Big(\int f\Big) \\ &\leq \sum_{I \notin \mathcal{J}} \int_I f + \int f \\ &\leq 2 \int f.\end{aligned}$$

Using Chebyshev's inequality, this tells us that

$$\mathsf{P}\left(\Big|Y_n - \int f\Big| > \frac{\varepsilon}{2}\right) \leq \frac{2 \int f}{\varepsilon/2} = \frac{4 \int f}{\varepsilon} =: e < 1,$$

where e does not depend on n, by choice of ε.

Now,

$$\begin{aligned}\mathsf{P}\left(\Big|\Sigma_{\mathcal{P}_n} f - \int f\Big| > \frac{\varepsilon}{2}\right) &\geq \mathsf{P}\left(X_n + Y_n > \int f + \frac{\varepsilon}{2}\right) \\ &\geq \mathsf{P}\left(X_n > \varepsilon \wedge Y_n > \int f - \frac{\varepsilon}{2}\right).\end{aligned}$$

The two events in the final line are independent, since X_n and Y_n depend only on the points from disjoint subsets of $\mathcal{P}_n$ (and the subsets in question are chosen

deterministically). So

$$\mathsf{P}\left(\left|\Sigma_{\mathcal{P}_n} f - \int f\right| \geq \frac{\varepsilon}{2}\right) \geq \mathsf{P}(X_n > \varepsilon)\mathsf{P}\left(Y_n > \int f - \frac{\varepsilon}{2}\right)$$
$$\geq \min(\tfrac{1-e}{4}, c'''\delta_n^{p-1}).$$

Since $\sum_{n\in\mathbb{N}} \delta_n^{p-1} = \infty$,

$$\sum_{n\in\mathbb{N}} \mathsf{P}\left(\left|\Sigma_{\mathcal{P}_n} f - \int f\right| > \frac{\varepsilon}{2}\right) = \infty$$

also, and therefore $\left|\Sigma_{\mathcal{P}_n} f - \int f\right| > \varepsilon/2$ holds for infinitely many values of n, almost surely. ■

We draw some consequences from this lemma, first giving a definition which expresses one of the restrictions on the partition sequence.

Definition 5.2.10 We define a *globally regular sequence of partitions*, in imperfect analogy with Pfeffer's regular partition (see [36]), as a sequence of partitions $(\mathcal{P}_n)_{n\in\mathbb{N}}$ such that

$$\frac{\max_{I\in\mathcal{P}_n} |I|}{\min_{I\in\mathcal{P}_n} |I|} < M$$

holds for some positive number M, and for every n.

Proposition 5.2.11 *For every nonnegative function which is not in L^{p-d}, where $p > 1$ and $d > 0$, there is some globally regular sequence of partitions $(\mathcal{P}_n)_{n\in\mathbb{N}}$ with sequence of sizes in ℓ^{p-1} so that $\Sigma_{\mathcal{P}_n} f$ does not converge almost surely.*

Proof. This is true because if $\int f^{p-d} = +\infty$, then there is a constant C and a sequence $M_n \to \infty$ such that

$$\lambda\{x : f(x) > M_n\} > CM_n^{-(p-\frac{\delta}{2})}$$

for all $n \in \mathbb{N}$.

We can define $(\delta_n)_{n\in\mathbb{N}}$ to be a sequence consisting only of elements from $\{\frac{1}{M_n} : n \in \mathbb{N}\}$. We take these in order, and repeat each enough times so that $\sum \delta_n^{p-1} < +\infty$ and $\sum \delta_n^{p-1-d} = +\infty$. Then f and (δ_n) will satisfy the conditions of Lemma 5.2.8.

Proposition 5.2.12 *If $p > 1$ and f is such that*

$$\{x : f(x) > z\} > cz^{\frac{1}{p}}$$

for all large enough z, and for some $c > 0$, and $(\mathcal{P}_n)_{n\in\mathbb{N}}$ is any globally regular sequence of partitions with $\sum_{n\in\mathbb{N}} |\mathcal{P}_n|^{p-1} = +\infty$, then $\Sigma_{\mathcal{P}_n} f$ does not converge almost surely.

Note that f is a function which just fails to be in L^p but which may be in weak L^p. (See Definition 2.1.1.)

Proof. This is obvious since for any value of δ_n we know that (5.8) holds. ■

We summarize the previous results, all corollaries of Lemma 5.2.8, by saying that if we assume nonnegativity of functions, global regularity of partition sequences, and the correct growth orders of both, then every function has a sequence of partitions for which its random Riemann sums do not converge strongly, and there exist functions which do not have convergence for any sequences of partitions. We should note that no attempt has been made to show either convergence or lack of convergence in any cases where the supremum of those p for which $f \in L^p$ is exactly equal to the infimum of p for which $(|\mathcal{P}_n|) \in \ell^{p-1}$, but that supremum and infimum are not both attained, either here or to the best of our knowledge anywhere else.

We arrived at Lemma 5.2.8 in attempting to prove the result we conjectured: that for any function in L^{p-d} and any globally regular sequence of partitions with sizes not in ℓ_p, strong convergence does not take place. In fact this is not true, as the following counterexample will show. Therefore, although the condition (5.8) may seem unnatural, it cannot be eliminated altogether.

Proposition 5.2.13 *For each $p > 1$ and each $d \in (0, p-1)$ there exists some function which is in L^1 but not in L^{p-d}, and a sequence $(\delta_n)_{n\in\mathbb{N}} \notin \ell^{p-1}$, such that for any sequence of partitions $(\mathcal{P}_n)_{n\in\mathbb{N}}$ where $\frac{\delta_n}{M} < |I| < M\delta_n$ for some $M > 1$ and all $I \in \mathcal{P}_n$, $n \in \mathbb{N}$, we have that f is strongly random Riemann integrable with respect to $(\mathcal{P}_n)$.*

Let p and d be given. Choose a natural number $k > 2p$. For all $n \in \mathbb{N}$, define

$$f_n := \begin{cases} 2^{\frac{k^n}{2}} & \text{on } (2^{-ak^{(n+1)}}, 2^{-ak^n}] \\ 0 & \text{elsewhere} \end{cases},$$

where $a := \frac{p-d}{2}$, and define $f := \sum_{n\in\mathbb{N}} f_n$. It is easy to see that $f \in L^1$ since

$$\int f_n < 2^{\frac{k^n}{2}} 2^{-\frac{(p-d)k^n}{2}} = b^{(k^n)}$$

where $b = 2^{\frac{1-(p-d)}{2}} < 1$, and that $f \notin L^{p-d}$ because

$$\int f^{p-d} = \sum_{n\in\mathbb{N}} \int f_n^{p-d} > \sum_{n\in\mathbb{N}} \frac{1}{2} 2^{-ak^n} 2^{\frac{p-d}{2}k^n} = \sum_{n\in\mathbb{N}} \frac{1}{2}$$

since the f_n have disjoint supports.

Now define $a_m := 2^{-k^m}$ and choose $(\delta_n)_{n\in\mathbb{N}}$ to be the sequence consisting of a_m repeated Q_m times for each m, taken in order, where $Q_m := 2^{(p-1)k^m}$. This means that

$$Q_m a_m^{p-1} = 2^{-(p-1)k^m} 2^{(p-1)k^m} = 1$$

and so $(\delta_n) \notin \ell^{p-1}$. Here we take a slight shortcut by defining Q_m to not necessarily be an integer. In practise we should take the nearest integer to Q_m as defined for the number of repetitions of a term a_m in the sequence. In fact it takes very little effort to check that this makes a difference which becomes vanishingly small very quickly in all subsequent calculations, and never affects convergence.

Now suppose that $M > 1$ is given, and that $(\mathcal{P}_n)$ is a sequence of partitions

with the sizes of all intervals satisfying the condition in the statement.

We will handle separately the leftmost interval of each partition $\mathcal{P}_n$ and all other intervals. For each n we call this leftmost interval J_n and its random tag point t_n. We refer to the two parts of the random Riemann sum by $X_n := |J_n|f(t_n)$ for the leftmost interval and $Y_n := \Sigma_{\mathcal{P}_n} f - X_n$ for the rest.

Fix some $m \in \mathbb{N}$, assuming that it is large enough for both $2^{-ak^{m+1}} < \frac{1}{M}2^{-k^m}$ and $\frac{1}{2}2^{-ak^{m-1}} > M2^{-k^m}$ to hold, and choose $n \in \mathbb{N}$ with $\delta_n = a_m$. The first inequality means that $[0, 2^{-ak^{m+1}}] \subseteq J_n$ since the length of J_n is at least $\frac{1}{M}2^{-k^m}$. The second means that since $2^{-ak^{m-1}} - 2^{-ak^m} > \frac{1}{2}2^{-ak^{m-1}} > |I|$ for all $I \in \mathcal{P}_n$, each interval of $\mathcal{P}_n$ except J_n contains at most one point from the set $\{2^{-ak^j} : j \in \mathbb{N}\}$.

We deal with the intervals to the right of J_n first. The last observation implies that on each of these intervals f is either constant or piecewise constant with a single jump discontinuity. In the former case $f(t_I)|I| = \int_I f$ with certainty, in the latter case $|f(t_I) - \frac{1}{|I|}\int_I f|$ is bounded by the size of the jump. So the total discrepancy between Y_n and the integral of f on $\mathbb{I}\backslash J_n$ is at most the maximal interval length, that is to say $M2^{-k^m}$, multiplied by the sum of the sizes of the jumps. Since f is monotone decreasing and $f(1) = 0$, that latter is the value of f at the rightmost point of J_n, or $2^{\frac{k^{m-1}}{2}}$.

We have shown that with certainty the difference between the random Riemann sum and the Lebesgue integral on the complement of J_n is bounded by $M2^{\frac{k^{m-1}}{2}-k^m}$ which of course tends to 0 for large m or large n.

Now we would like to show that $X_n - \int_{J_n} f \xrightarrow{a.s.} 0$. Since f is integrable, $\int_{J_n} f \to 0$ as $n \to \infty$. So it will be enough to show that $X_n \xrightarrow{a.s.} 0$. Fix some $\varepsilon > 0$, and consider the probability $\mathsf{P}(|X_n| > \varepsilon) = \mathsf{P}(X_n > \varepsilon)$. For $X_n = f(t_n)|J_n|$ to be greater than ε, we would need $f(t_n)$ to be at least $\frac{\varepsilon}{|J_n|} > \frac{1}{M}\varepsilon 2^{k^m}$. For large m this will be much bigger than $2^{\frac{k^m}{2}}$ and so in fact f will have to be at least $2^{\frac{k^{m+1}}{2}}$. The piece of $\mathbb{I}$ on which this holds has length $2^{-ak^{(m+1)}}$, therefore

the probability that $f(t_n) > \frac{\varepsilon}{|J_n|}$ is

$$\frac{2^{-ak^{(m+1)}}}{|J_n|} < M2^{-ak^{(m+1)}}2^{k^m}.$$

The sum of this probability over all values of n for which $\delta_n = a_m$ is

$$\begin{aligned} Q_m M2^{-ak^{(m+1)}}2^{k^m} &= M2^{(p-1)k^m}2^{-ak^{(m+1)}}2^{k^m} \\ &= M2^{k^m(p-1-(\frac{p-d}{2})k+1)} \\ &= M2^{k^m(p-(\frac{p-d}{2})k)}. \end{aligned}$$

Since $k > 2p$, the exponent is less than $k^m p(1-p+d) < 0$, and so the probability that $X_n > \varepsilon$ is summable over all $n \in \mathbb{N}$. As we have seen before, this is enough to prove that X_n, and therefore also $\Sigma_{\mathcal{P}_n} f - \int f$, converge to 0 almost surely.■

Remark The results above extend those results of Kahane in [29] and Pruss in [37]. They considered only the sequence of partitions

$$\mathcal{P}_n := \left\{ (\frac{j}{n}, \frac{j+1}{n}) : 0 \leq j < n \right\}$$

and proved that all functions in L^2 are strongly random Riemann integrable with respect to this sequence, and all functions which are not in L^2 are not. Theorem 5.2.3 implies a weaker version of the first part of this. In fact, it tells us that all functions in $L^{2+\varepsilon}$ are strongly random Riemann integrable with respect to this sequence. Proposition 5.2.13 shows that we cannot find the same generalization of the other direction. However, the results of this section may lead to a different generalization of the necessary condition for convergence.

5.3 The weak random Riemann integral

The main result of this section establishes that all the functions in two classes of not Lebesgue integrable functions are not weakly random Riemann integrable.

The first lemma simply allows us to find a collection of intervals contained in $\mathbb{I}$ on which the random Riemann sums do not converge weakly, in order to prove that it is not weakly random Riemann integrable on the whole of $\mathbb{I}$. In other words it is a Henstock lemma for the weak random Riemann integral, proving that integrability is inherited by subintervals.

Lemma 5.3.1 *If f is a function and A is a finite union of intervals in $\mathbb{I}$, then a partition $\mathcal{P}$ of A such that for all $M \in \mathbb{R}$,*

$$\mathsf{P}\left(\left|\Sigma_{\mathcal{P}} f - M)\right| > \varepsilon\right) > \varepsilon,$$

can be extended to a partition $\mathcal{P}'$ of $\mathbb{I}$ with $|\mathcal{P}'| \leq |\mathcal{P}|$ so that

$$\mathsf{P}(|\Sigma_{\mathcal{P}'}(f) - M| > \varepsilon) > \varepsilon$$

also holds for any $M \in \mathbb{R}$.

Proof. Let $\mathcal{P}$ be such a partition of A. We extend the partition arbitrarily to a partition $\mathcal{P}'$ of $\mathbb{I}$, of size at most $|\mathcal{P}|$. If for some M

$$\mathsf{P}(|\Sigma_{\mathcal{P}'}(f) - M| > \varepsilon) \leq \varepsilon$$

then since the distributions of $\Sigma_{\mathcal{P}} f$ and $\Sigma_{\mathcal{P}'\setminus\mathcal{P}} f$ are independent, we would have

$$\mathsf{P}\left(\left|\Sigma_{\mathcal{P}} f - M - \Sigma_{\mathcal{P}'\setminus\mathcal{P}} f\right| > \varepsilon | \mathcal{B}\left(\mathcal{P}'\setminus\mathcal{P}\right)\right) \leq \varepsilon$$

where $\mathcal{B}(\mathcal{P}'\setminus\mathcal{P})$ is the σ-algebra generated by $\Sigma_{\mathcal{P}'\setminus\mathcal{P}}(f|_{\mathbb{I}\setminus A})$.

This means that if x is an element of the measure space formed by projecting $\Omega_{\mathcal{P}'}$ onto $\mathcal{B}(\mathcal{P}'\setminus\mathcal{P})$, then

$$\mathsf{P}(|\Sigma_{\mathcal{P}}(f|_A) - M - c| > \varepsilon | x) < \varepsilon$$

and therefore by independence

$$\mathsf{P}(|\Sigma_{\mathcal{P}}(f|_A) - M - c| > \varepsilon) < \varepsilon$$

for some c which depends on x but not on M. Since c does not depend on M, by fixing x and taking arbitrary M, $M - c$ is arbitrary. This is a contradiction. ■

Corollary *If f is weakly random Riemann integrable on $\mathbb{I}$ then it is weakly random Riemann integrable on any subinterval of $\mathbb{I}$.* □

Proof. Lemma 5.3.1 shows that if f is not weakly random Riemann integrable on A then it is not weakly random Riemann integrable on $\mathbb{I}$. ■

Here we give a lemma on the first set of functions which we will prove not to be weakly random Riemann integrable, those which are measurable but not in $L^{1-\varepsilon}$ for some positive ε. For technical reasons we exclude here those which have large growth concentrated close to some point. They are also not weakly random Riemann integrable, but for them this will be proved using a different lemma.

Definition 5.3.3 A function f has the *left endpoint weak L^1 condition* on an interval (a, b), if f restricted to $(a, a + \varepsilon)$ is in weak L^1 for some $\varepsilon > 0$. The *right endpoint weak L^1 condition* is symmetric. We say that f has the *endpoint weak L^1 condition* on (a, b) if it has both the right and the left conditions.

Lemma 5.3.4 *Suppose that A_0 is an interval of $\mathbb{I}$, on which $\int_{A_0} |f|^{1-\varepsilon} = +\infty$ for some positive ε, and n is some positive integer. Suppose further that f has the endpoint weak L^1 condition on every subinterval of A_0.*[(2)] *Then we can find a positive constant M, and two disjoint subsets of A_0: a finite union of closed*

[(2)] The examiner, Prof. M. Laczkovich, pointed out that this condition would imply that f is weak L^1 on A_0. So in fact this lemma and the following one are redundant, as is case (i) in Theorem 5.3.8. Lemmas 5.3.6 and 5.3.7, and case (ii) of the Theorem are sufficient to prove that any function which is not weak L^1, is not weakly random Riemann integrable. We have left the obsolete material as is since this is the version which was approved by the examiners.

intervals $S := \bigcup_{i=1}^{k} I_i$, *each of length* $l > \frac{1}{M}$, *with*

$$\lambda(\{x \in S : |f(x)| > M\}) > \frac{l}{5n}$$

and

$$\lambda(\{x \in S : |f(x)| > \frac{3M}{4}\}) < \frac{2l}{n},$$

and an open interval A_1, *for which* $\int_{A_1} |f|^{1-\varepsilon} = +\infty$.

Lemma 5.3.5 *If* $A \subseteq \mathbb{I}$ *is a measurable set and* f *is Lebesgue measurable on* A *with* $\int_A |f|^{1-\varepsilon} = +\infty$ *for some positive* ε, *then given any two positive numbers* M_0 *and* r, *we can find* $M > M_0$ *such that*

$$M\lambda(\{x \in A : |f(x)| > M\}) > r \tag{5.11}$$

and

$$\lambda\left(\left\{x \in A : |f(x)| > \frac{3M}{4}\right\}\right) < 2\lambda(\{x \in A : |f(x)| > M\}). \tag{5.12}$$

Proof (Lemma 5.3.5). To show this, define

$$L(x) := \lambda(\{\theta \in A : |f(\theta)|^{1-\varepsilon/2} > x\}).$$

For any number $a > 0$ and for any y, we can find $x > y$ with

$$xL(x) > a \qquad \text{and} \qquad L(x) > \tfrac{1}{2}L(\tfrac{x}{2}). \tag{5.13}$$

Indeed, suppose that this were not the case; choose some $a > 0$ and some y such that for all $x > y$, either $xL(x) \leq a$ or $L(x) \leq \frac{1}{2}L(\frac{x}{2})$.

We claim that $2^k yL(2^k y) \leq \max\{a, yL(y)\}$ for all $k \geq 0$. This can be proved by induction, since if it does not hold for some k, then

$$2^k yL(2^k y) > \max\{a, yL(y)\} \geq a,$$

which implies $2^k yL(2^k y) \leq 2^{k-1} yL(2^{k-1} y)$, and therefore it does not hold for $k-1$ either. However it is trivially true for $k = 0$.

Since L is decreasing, we thus have

$$\forall z \geq y: \qquad zL(z) \leq 2\max\{a, yL(y)\},$$

and so $L(x) \in \mathsf{O}(1/x)$ as $n \to \infty$. This would mean that $|f|^{1-\varepsilon/2}$ is in L^p for all $p < 1$, which contradicts the fact that $|f|^{1-\varepsilon}$ is not in L^1.

Now set $y =: M_0^{1-\frac{\varepsilon}{2}}$ and a large enough that $ay^{\frac{\varepsilon}{2-\varepsilon}} > r$. Take $x > y$ with $xL(x) > a$ and $L(x) > \frac{1}{2}L(\frac{x}{2})$. This means that

$$\begin{aligned} x\lambda(\{\theta \in A : |f(\theta)|^{1-\varepsilon/2} > x\}) &> a \\ x^{\frac{1}{1-\varepsilon/2}} \cdot x^{-\frac{\varepsilon}{2-\varepsilon}} \lambda\left(\left\{\theta \in A : |f(\theta)| > x^{\frac{1}{1-\varepsilon/2}}\right\}\right) &> a \\ x^{\frac{1}{1-\varepsilon/2}} \lambda\left(\left\{\theta \in A : |f(\theta)| > x^{\frac{1}{1-\varepsilon/2}}\right\}\right) &> ax^{\frac{\varepsilon}{2-\varepsilon}} > r. \end{aligned}$$

Here we used that $\varepsilon < 2$, in order that $x^{\frac{\varepsilon}{2-\varepsilon}}$ should be greater than $y^{\frac{\varepsilon}{2-\varepsilon}}$, without loss of generality.

If we set $M := x^{1/(1-\varepsilon/2)}$, then $M > M_0$ and M satisfies (5.11). To check that M satisfies (5.12), we use the fact that $L(x) > \frac{1}{2}L(\frac{x}{2})$, and assuming without loss of generality that $\varepsilon < 1/2$,

$$2^{\varepsilon/(2-\varepsilon)} < 2^{1/3} < \frac{3}{2}$$

which means that

$$\begin{aligned}
\lambda\left(\left\{\theta \in A : |f(\theta)| > \frac{3M}{4}\right\}\right) &\leq \lambda\left(\left\{\theta \in A : |f(\theta)| > \frac{M}{2} \cdot 2^{\varepsilon/(2-\varepsilon)}\right\}\right) \\
&= \lambda\left(\left\{\theta \in A : |f(\theta)| > \frac{x^{\frac{1}{1-\varepsilon/2}}}{2} \cdot 2^{\varepsilon/(2-\varepsilon)}\right\}\right) \\
&= \lambda\left(\left\{\theta \in A : |f(\theta)| > \left(\frac{x}{2}\right)^{\frac{1}{1-\varepsilon/2}}\right\}\right) \\
&= L\left(\frac{x}{2}\right) \\
&\leq 2L(x) \\
&= 2\lambda(\{\theta \in A : |f(\theta)| > M\}). \qquad \blacksquare
\end{aligned}$$

Proof (Lemma 5.3.4). Write the interval A_0 as (a, b). Since f has the endpoint weak L^1 condition on A_0, there must exist some $\delta > 0$ such that

$$\int_{a+\delta}^{b-\delta} |f|^{1-\varepsilon} = +\infty.$$

Define $A := (a + \delta, b - \delta)$.

Choose M' so big that

$$\lambda(\{x \in A : |f(x)| > M'\}) < \frac{\delta}{2n}.$$

Using Lemma 5.3.5, we can find $M_1 > M'$ such that

$$M_1 \lambda(\{x \in A : |f(x)| > M_1\}) > 1 \tag{5.14}$$

and

$$\lambda(\{x \in A : |f(x)| > \frac{3M_1}{4}\}) < 2\lambda(\{x \in A : |f(x)| > M_1\}). \tag{5.15}$$

Write the set $\{x \in A : |f(x)| > M_1\} =: F_1$.

Suppose that either

$$\lambda(\{x \in (a+\delta-2n|F_1|, a+\delta) : |f(x)| > \frac{3M_1}{4}\}) \tag{5.16}$$

or

$$\lambda(\{x \in (b-\delta, b-\delta+2n|F_1|) : |f(x)| > \frac{3M_1}{4}\}) \tag{5.17}$$

is bigger than $|F_1|$ (which implies it is bigger than $\frac{1}{M_1}$). Then using Lemma 5.3.5 again, find $M_2 > 2M_1$ such that (5.14) and (5.15) hold for M_2.

We continue this process until we find an M_j for which both (5.14) and (5.15) hold, and for which (5.16) and (5.17) are smaller than $|F_j|$. We must eventually find such an M_j since otherwise either (5.16) or (5.17) is bigger than $\frac{1}{M_k}$ for arbitrarily large values of k, which would mean that the endpoint weak L^1 condition fails for one of $(a, a+\delta)$ or $(b-\delta, b)$. So set $M := M_j$ and $F_M := F_j$.

Choose some point c such that

$$\lambda(F_M \cap (-\infty, c)) = \lambda(F_M \cap (c, \infty)).$$

Call $A \cap (-\infty, c) =: A_L$ and $A \cap [c, \infty) =: A_R$. The union of A_L and A_R is A, and so at least one of

$$\int_{A_L} |f|^{1-\varepsilon} = +\infty \text{ or } \int_{A_R} |f|^{1-\varepsilon} = +\infty.$$

So choose A_1 to be one of A_L or A_R, on which $|f|^{1-\varepsilon}$ is not integrable, and call the other one A^*. Suppose without loss of generality that $A^* = A_R$. Let $l := 2n|F_M|$, noting that $l > \frac{2n}{M} > \frac{1}{M}$, and define the set S as the collection of intervals $\{(c+jl, c+(j+1)l] : 0 \leq j \leq k\}$, where $k \in \mathbb{N}$ is chosen so that $b-\delta \in (c+kl, c+(k+1)l]$.

Since

$$\begin{aligned} |F_M| &= \lambda(\{x \in A : |f(x)| > M\}) \\ &\leq \lambda(\{x \in A : |f(x)| > M'\}) \\ &\leq \frac{\delta}{2n}, \end{aligned}$$

S is contained in $A_0 \backslash A_1$, because $l = 2n|F_M| < \delta$ by choice of M'.

It remains to check two things. Firstly that

$$\lambda(\{x \in S : |f(x)| > M\}) = \frac{l}{4n} \geq \frac{l}{5n},$$

which is true by the choice of l, and because S contains half of F_M by measure. Secondly that

$$\lambda(\{x \in S : |f(x)| > \frac{3M}{4}\}) < \frac{2l}{n}$$

which follows from the facts that

$$\lambda(\{x \in A : |f(x)| > \frac{3M}{4}\}) \leq 2|F_M| = \frac{l}{n}$$

by (5.15) and the definition of l, and that

$$\lambda(\{x \in S \cap (b-\delta, b) : |f(x)| > \frac{3M}{4}\}) \leq |F_M| = \frac{l}{2n},$$

which follows from (5.16), (5.17) and the choice of M. ■

This was the first condition which implies nonconvergence of the weak random Riemann integral. Now we deal with the other class of functions; those which do not have the endpoint weak L^1 condition.

Lemma 5.3.6 *Suppose that f does not have the left endpoint weak L^1 condition on $\mathbb{I}$. Let ε_0 be a point in $(0,1]$, and n some positive integer. We can find $\varepsilon_1 \in (0, \varepsilon_0)$, a positive number M, and $S \subseteq [\varepsilon_1, \varepsilon_0)$, a disjoint union of intervals*

$\{I_i : 1 \leq i \leq k\}$, *each of length* $l > \frac{1}{M}$, *with*

$$\lambda(\{x \in S : |f(x)| > M\}) > \frac{l}{5n} \tag{5.18}$$

and

$$\lambda(\{x \in S : |f(x)| > \frac{3M}{4}\}) < \frac{2l}{n}. \tag{5.19}$$

Note that the conclusion of this lemma is the same as that of Lemma 5.3.4, with interval A_0 replaced by the interval $[0, \varepsilon_0)$ and the interval A_1 replaced by $[0, \varepsilon_1)$, on which the failure of the left endpoint weak L^1 condition is preserved just as the condition "f not in $L^{1-\varepsilon}$" is preserved in Lemma 5.3.4. As with Lemma 5.3.4, we start the proof by proving a lemma purely about the growth of the function f, analogous to Lemma 5.3.5 and with a similar proof.

Lemma 5.3.7 *Suppose that* f *is not in weak* L^1. *Then for all* $y > 0$ *we can find* $x \geq y$ *with*

$$x\lambda(\{z : |f(z)| > x\}) > 1$$

and

$$\lambda(\{z : |f(z)| > x\}) > \frac{1}{2}\lambda(\{z : |f(z)| > \frac{x}{2}\}).$$

Proof. Write $L(x) := \lambda(\{z : |f(z)| > x\})$. Suppose that the lemma is not true, and let y be such that for all $x > y$ either

$$xL(x) < 1, \text{ or } L(x) \leq \frac{1}{2}L(\frac{x}{2}). \tag{5.20}$$

By induction,

$$2^k yL(2^k y) \leq \max\{1, yL(y)\}$$

for all $k \geq 0$. The statement is trivially true for $k = 0$. Suppose it does not hold for k, then $2^k L(2^k y) > 1$ and therefore

$$2^k yL(2^k y) \leq 2^{k-1} yL(2^{k-1} y)$$

by (5.20) and so it does not hold for $k-1$ either. So in fact it holds for all k.

Since $L(x)$ is decreasing, we have

$$xL(x) \leq 2\max\{1, yL(y)\} \text{ for all } x \geq y.$$

This contradicts f not being in weak L_1. ■

Proof (Lemma 5.3.6). Choose M_0 large enough that

$$\lambda(\{z < \varepsilon_0 : |f(z)| > M_0\}) < \frac{\varepsilon_0}{20n}. \tag{5.21}$$

Since f restricted to $[0, \varepsilon_0)$ is not in weak L^1, we can use Lemma 5.3.7 to find $M > M_0$ so that

$$M\lambda(\{z < \varepsilon_0 : |f(z)| > M\}) > 1 \tag{5.22}$$

and

$$\lambda(\{z < \varepsilon_0 : |f(z)| > M\}) > \frac{1}{2}\lambda(\{z < \varepsilon_0 : |f(z)| > \frac{M}{2}\}) \tag{5.23}$$

Choose $0 < \varepsilon_1 < \frac{\varepsilon_0}{5}$ so that both these inequalities still hold with "$\varepsilon_1 < z < \varepsilon_0$" in place of "$z < \varepsilon_0$".

We define

$$l_0 := 4n\lambda(\{\varepsilon_1 < z < \varepsilon_0 : |f(z)| > M\}).$$

Since by (5.21), $l_0 < \frac{\varepsilon_0}{5}$, $4l_0 < \varepsilon_0 - \varepsilon_1$, by the choice of ε_1. Define

$$l := \frac{\varepsilon_0 - \varepsilon_1}{\left\lfloor \frac{\varepsilon_0 - \varepsilon_1}{l_0} \right\rfloor}$$

and observe that whereas l_0 divides a possibly fractional number of times into $\varepsilon_0 - \varepsilon_1$, l has an integer multiple, at least 4, equal to $\varepsilon_0 - \varepsilon_1$. So we let S be the partition of $[\varepsilon_1, \varepsilon_0)$ into intervals all of length l.

By the choice of l_0, and from (5.22), we have that $l_0 > \frac{4n}{M} > \frac{1}{M}$ and so $l > \frac{1}{M}$ also. Since l must be less than $\frac{5l_0}{4}$ and again by the definition of l_0, we

have (5.18). It remains to check (5.19), this follows from (5.23)

$$\begin{aligned}\lambda(\{x \in S : |f(x)| > \frac{3M}{4}\}) &< \lambda(\{x \in S : |f(x)| > \frac{M}{2}\}) \\ &< 2\lambda(\{x \in S : |f(x)| > M\}) \\ &= \frac{l_0}{2n} < \frac{2l}{n}\end{aligned}$$

since l is clearly at least as large as l_0. ■

Theorem 5.3.8 *If f is such that either*

(i) f is not in $L^{1-\varepsilon}$ for some $\varepsilon > 0$, or

(ii) f does not have the endpoint weak L^1 condition on some subinterval of $\mathbb{I}$,

then f is not weakly random Riemann integrable.[3]

Again we will use a basic lemma on finite collections of random events.

Lemma 5.3.9 *Suppose that $0 < a < b < 1$ and that $(E_k)_{k=1}^n$ is a finite collection of independent random events, with $a \leq \sum_{k=1}^n \mathsf{P}(E_K) \leq b$. The probability of the event "E_k holds for exactly one value of k between 1 and n" is at least $a(1-b)$.*

Proof (Lemma 5.3.9). The probability of the event "exactly one of E_k" can be expressed as:

$$\sum_{k=1}^{n} \mathsf{P}(E_k \cap \bigcap_{j \neq k} E_j^C).$$

[3] Case (ii) implies (i) – this leads to a simpler proof of a slightly stronger theorem. See the footnote attached to the statement of Lemma 5.3.4.

We obtain

$$\begin{aligned}\sum_{k=1}^{n} \mathsf{P}(E_k) \prod_{j\neq k}(1-\mathsf{P}(E_j)) &\geq \sum_{k=1}^{n} \mathsf{P}(E_k)(1-\sum_{j\neq k}\mathsf{P}(E_j)) \\ &\geq \sum_{k=1}^{n} \mathsf{P}(E_k)\left(1-\sum_{j}\mathsf{P}(E_j)\right) \\ &\geq a(1-b)\end{aligned}$$

where the first step is given by the independence of the E_k, the second is a basic inequality, and the rest is obvious. ■

Proof (Theorem 5.3.8). Suppose that both (i) and (ii) hold for f. Then we will prove the result for f using (ii). So we can assume that either (ii) holds, or (i) holds and (ii) does not hold for any subinterval. We refer to the second case as case (i').

If case (ii) holds, then we will show that f is not weakly random Riemann integrable on the subinterval in question, which proves that it is not weakly random Riemann integrable on $\mathbb{I}$ by Lemma 5.3.1. So assume without loss of generality that f does not have the left endpoint weak L^1 condition on $\mathbb{I}$.

In the first stage of the proof, we will construct a 'pseudopartition'[(4)], an infinite collection of non-overlapping intervals contained in $\mathbb{I}$. We do this by induction.

Let E_0 be the interval $[0,1)$. At the k^{th} stage, we apply Lemma 5.3.4 or Lemma 5.3.6 in case (i') or case (ii) respectively, setting the interval E_k to be either A_0 or $[0,\varepsilon_0)$, and using $n=k$.

We add the intervals of S given by the relevant lemma to our pseudopartition, and let E_{k+1} be the interval A_1 in case (i') or $[0,\varepsilon_1)$ in case (ii). We call the

[(4)] The term pseudopartition has been used in [5] and [12] to refer to a collection of infinitely many non-overlapping intervals which cover $\mathbb{I}$, where only finitely many intervals are to the right of any non-zero point. The purpose of a pseudopartition is that, given any desired partition size, we can take finitely many intervals of less than that size from the pseudopartition, and use them to form a partition. We extend this notation to refer to any infinite collection of intervals which partition $\mathbb{I}$, and from which we intend to form partitions, each containing some finite subset.

number M from the lemma in question, M_k.

This construction will yield a collection of disjoint intervals $\{J_{n,j} : n \in \mathbb{N}, 1 \leq j \leq r_n\}$, where $J_{n,j}$ is the j^{th} interval from those added at the n^{th} stage, and r_n is the number of intervals added at that stage; and a sequence of positive numbers $(M_k)_{k=1}^{\infty}$. From this point on we no longer need to distinguish between the cases (i') and (ii).

Define $m_n := |J_{n,j}|$, which does not depend on j. We know that $m_n > \frac{1}{M_n}$ for every n, and that

$$\lambda(\{x \in S_n : |f(x)| > M_n\}) > \frac{m_n}{5n}$$

and

$$\lambda(\{x \in S_n : |f(x)| > \frac{3M_n}{4}\}) < \frac{2m_n}{n},$$

where $S_n := \bigcup_{j=1}^{r_n} J_{n,j}$.

This means that if $t_{n,j}$ is chosen uniformly from $J_{n,j}$ then

$$\sum_{j=1}^{r_n} \mathsf{P}(|f(t_{n,j})| > M_n) > \frac{1}{5n}$$

and

$$\sum_{j=1}^{r_n} \mathsf{P}\left(|f(t_{n,j})\,| < \frac{3M_n}{4}\right) > 1 - \frac{2}{n}.$$

We also observe that $M_n|J_{n,j}| > 1$.

Let $\delta > 0$ be given and choose $N > 904$, big enough that $m_n < \delta$ for all $n > N$. This must be possible since $\bigcup_n \bigcup_j J_{n,j}$ is a disjoint union of intervals whose measure is at most 1.

Now choose $m > N$ so that

$$\frac{1}{904} \leq \sum_{n=N}^{m} \sum_{j} \frac{\lambda(\{x \in J_{n,j} : |f(x)| > M_n\})}{\lambda(J_{n,j})} < \frac{3}{904}.$$

Since the inner sum of the middle expression is between $\frac{1}{5n}$ and $\frac{2}{n} \leq \frac{2}{904}$

this is always possible. Let the partition $\mathcal{P}$ of $[0,1]$ consist of the intervals $\bigcup_{n=N}^{m}\bigcup_{j=1}^{r_n} J_{n,j}$ and any other intervals, each of length at most δ, whose non-overlapping union is the rest of $\mathbb{I}$. $\mathcal{P}$ has size at most δ.

We can write the random Riemann sum of $\mathcal{P}$ as

$$\Sigma_{\mathcal{P}}(f) = \sum_{n=N}^{m}\sum_{j=1}^{r_n} f(t_{n,j})m_n + \sum_{k=1}^{q} f(t_k^*)|J_k^*|$$

where $t_{n,j}$ are distributed uniformly on $J_{n,j}$, and t_k^* on J_k^*, the remaining intervals of $\mathcal{P}$.

In the following an event always refers to a measurable subset of the probability space on which the random Riemann sum is distributed. We estimate the probability of the event

$$\{\exists! q, l,\ N \le q \le m,\ 1 \le l \le r_q \text{ such that } |f(t_{q,l})| > M_q\}.$$

If we rename the events $\{|f(t_{n,j})| > M_n\}$, for $N \le n \le m$ and $1 \le j \le r_n$ as $\{B_k, 1 \le k \le s\}$, where $k = j + \sum_{i=N}^{n-1} r_i$ and $s := \sum_{i=N}^{m} r_i$, then we can describe this event as

$$\mathsf{P}(B_k \text{ for exactly one } k)$$

which by Lemma 5.3.9 is at least $\frac{901}{904^2}$ since by construction $\sum_{k=1}^{s}\mathsf{P}(B_k)$ is between $\frac{1}{904}$ and $\frac{3}{904}$.

Now also rename the events $\{|f(t_{n,j})| < \frac{3M}{4}\}$, $N \le n \le m$, $1 \le j \le r_n$, to be called C_k for $1 \le k \le s$, where k and s are as before. The event $\{\text{not } C_k \text{ for at least 2 values of } k\}$ has probability at most

$$\begin{aligned}\sum_{k=1}^{s}\sum_{\substack{k'=1\\k'\neq k}}^{s}(1-\mathsf{P}(C_k))(1-\mathsf{P}(C_{k'})) &\le \left[\sum_{k=1}^{s}(1-\mathsf{P}(C_k))\right]^2 \\ &\le \left(\frac{30}{904}\right)^2 = \frac{900}{904^2},\end{aligned}$$

since

$$\begin{aligned} 1 - \mathsf{P}(C_k) &= \mathsf{P}(|f(t_{n,j})| > \frac{3M_n}{4}) \\ &\leq 10\mathsf{P}(|f(t_{n,j})| > M_n) \\ &= 10\mathsf{P}(B_k) \end{aligned}$$

and $\sum_{k=1}^{s} \mathsf{P}(B_k) \leq \frac{3}{904}$. So $\mathsf{P}(\{B_k \text{ for exactly one } k, \text{ and } C_k \text{ for all other } k\})$, being no less than the difference between the probabilities of these two events, is at least $\frac{1}{904^2}$.

We have established that with probability greater than some positive constant, namely $\frac{1}{87216}$, $|f(t_{n,j})|$ will be greater than M_n for exactly one value of n and of j, and smaller than $|\frac{3M_n}{4}|$ for all other pairs of values.

For each $1 \leq k \leq s$, we divide the set $\left(-\frac{3M_n}{4}, \frac{3M_n}{4}\right]$, where n is such that $\sum_{i=N}^{n-1} r_i < k \leq \sum_{i=N}^{n} r_i$, into intervals of equal length

$$E_{k,j} := \left(-\frac{3M_n}{4} + \frac{3(j-1)M_n}{2z_k}, -\frac{3M_n}{4} + \frac{3jM_n}{2z_k}\right],$$

for $1 \leq j \leq z_k$. Here $z_k \in \mathbb{N}$ is chosen large enough that $\frac{3M_n}{2z_k} < \frac{1}{8sM_n}$.

This allows us to define $D_{\alpha_1,\alpha_2,\ldots,\alpha_s} := \bigcap_{k=1}^{s} \{f(t_k) \in E_{k,\alpha_k}\}$ for $1 \leq \alpha_k \leq z_k$ (where of course $t_k = t_{n,j}$ if $k = \sum_{i=1}^{n-1} r_i + j$, $1 \leq j \leq r_n$). Then we have

$$\bigsqcup_{k_1=1}^{z_1} \bigsqcup_{k_2=1}^{z_2} \cdots \bigsqcup_{k_s=1}^{z_s} D_{k_1,\ldots,k_s} = \bigcap_{k=1}^{s} C_k =: D.$$

Here $\bigsqcup$ denotes a disjoint union. We also define

$$F^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_s} := B_k \cap \bigcap_{\substack{i=1 \\ i \neq k}}^{s} \{f(t_i) \subset E_{i,\alpha_i}\}$$

for which

$$\begin{aligned} F^k &:= \{B_k \text{ and } C_i \text{ for } i \neq k\} \\ &= \bigsqcup_{\alpha_1=1}^{z_1} \cdots \bigsqcup_{\alpha_{k-1}=1}^{z_{k-1}} \bigsqcup_{\alpha_{k+1}=1}^{z_{k+1}} \cdots \bigsqcup_{\alpha_s=1}^{z_s} F^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_s} \end{aligned}$$

holds.

Further define for $1 \leq j \leq z_k$

$$T_{k,j} := \mathsf{P}(f(t_k) \in E_{k,j} | C_k) = \frac{\mathsf{P}(f(t_k) \in E_{k,j})}{\mathsf{P}(C_k)}$$

noting that $\sum_{j=1}^{z_k} T_{k,j} = 1$. We will define a measure-preserving mapping $\phi : \bigcup_{k=1}^{s} F^k \to D$ as follows. We divide each $F^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_s}$, where $1 \leq k \leq s$ and $1 \leq \alpha_i \leq z_i$ for all $1 \leq i \leq s$, $i \neq k$, into disjoint sets $\tilde{F}^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_k,\alpha_{k+1},\ldots,\alpha_s}$ for $1 \leq \alpha_k \leq z_k$. We do this arbitrarily, with the condition that the division is made in the ratio $T_{k,1} : T_{k,s} : \ldots : T_{k,z_k}$, i.e.

$$\mathsf{P}(\tilde{F}^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_k,\alpha_{k+1},\ldots,\alpha_s}) = T_{k,j} \cdot \mathsf{P}(F^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_s}).$$

Now we define $\phi \restriction_{\tilde{F}^k_{\alpha_1,\ldots,\alpha_s}}$ to be an arbitrary measure preserving bijection onto some subset of $D_{\alpha_1,\ldots,\alpha_s}$ which is disjoint from the images of $\tilde{F}^k_{\alpha_1,\ldots,\alpha_s}$ for all other k. We can do this provided that

$$\sum_{k=1}^{s} \mathsf{P}(\tilde{F}^k_{\alpha_1,\ldots,\alpha_s}) \leq \mathsf{P}(D_{\alpha_1,\ldots,\alpha_s}) \tag{5.24}$$

But

$$\begin{aligned} \mathsf{P}(\tilde{F}^k_{\alpha_1,\ldots,\alpha_{k-1},j,\alpha_{k+1},\ldots,\alpha_s}) &= \frac{\mathsf{P}(f(x) \in E_{k,j})}{\mathsf{P}(C_k)} \mathsf{P}(F^k_{\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_s}) \\ &= \frac{\mathsf{P}(f(x) \in E_{k,j})}{\mathsf{P}(C_k)} \mathsf{P}(B_k) \prod_{\substack{i=1 \\ i \neq k}}^{s} \mathsf{P}(f(t_i) \in E_{i,\alpha_i}) \\ &= \frac{\mathsf{P}(B_k)}{\mathsf{P}(C_k)} \mathsf{P}(D_{\alpha_1,\ldots,\alpha_s}) \end{aligned}$$

and this implies (5.24) since

$$\sum_{k=1}^{s} \frac{\mathsf{P}(B_k)}{\mathsf{P}(C_k)} \leq \frac{\sum_{k=1}^{s} P(B_k)}{1 - \frac{30}{904}} \leq \frac{3}{874} \leq 1.$$

So we can construct ϕ as desired.

Note that if $x \in \tilde{F}^k_{\alpha_1,\ldots,\alpha_s}$, where $k = j + \sum_{i=N}^{n-1} r_i$, $1 \leq j \leq r_n$ then

$$|f(t_{n,j})| \cdot |J_{n,j}| \restriction_{\{x\}} > M_n \cdot |J_{n,j}|$$

where "$\restriction_{\{x\}}$" means that the random variable is evaluated for the point x of the probability space; whereas

$$|f(t_{n,j})| \cdot |J_{n,j}| \restriction_{\{\phi(x)\}} < \frac{3M_n}{4} \cdot |J_{n,j}|.$$

Since $M_n|J_{n,j}| > 1$, we have that

$$\left|\left(f(t_{n,j})|J_{n,j}| \restriction_{\{x\}}\right) - \left(f(t_{n,j})|J_{n,j}| \restriction_{\{\phi(x)\}}\right)\right| > \frac{1}{4}.$$

Furthermore, for all other pairs n, j it must be the case that $f(t_{n,j})$ is in the same interval $E_{k,j}$ for both x and $\phi(x)$ and so

$$\left|\left(f(t_{n,j})|J_{n,j}| \restriction_{\{x\}}\right) - \left(f(t_{n,j})|J_{n,j}| \restriction_{\{\phi(x)\}}\right)\right| \leq |J_{n,j}| \frac{1}{8sm_n} = \frac{1}{8s}.$$

Hence

$$\left|\Sigma^*_{\mathcal{P}}(f) \restriction_{\{x\}} - \Sigma^*_{\mathcal{P}}(f) \restriction_{\{\phi(x)\}}\right| \geq \frac{1}{4} - (s-1)\frac{1}{8s} \geq \frac{1}{8},$$

where $\Sigma^*_{\mathcal{P}}(f) = \sum_{n,j} f(t_{n,j})|J_{n,j}| = \Sigma_{\mathcal{P}}(f) - \sum_{k=1}^{q} f(t^*_k)|J^*_k|$.

So, given an interval of width less that $\frac{1}{8}$, $\Sigma^*_{\mathcal{P}}(f)$ lies in that interval in at most one of the cases x, $\phi(x)$. In particular, since ϕ is defined everywhere on

$$\bigsqcup_{k=1}^{s} F^k = \{B_k \text{ for exactly one value of } k, \text{ and } C_k \text{ for all other values}\},$$

which has probability at least $\frac{1}{904^2}$, and since the image of ϕ has the same probability, there can be no Z for which

$$\mathsf{P}(|\Sigma^*_{\mathcal{P}}(f) - Z| < \frac{1}{16}) < \frac{1}{904^2}$$

holds.

By Lemma 5.3.1, if this holds for $\Sigma^*_{\mathcal{P}}(f)$, the random Riemann sum on $\bigcup_{n,j} J_{n,j}$, it also holds for the random Riemann sum on $\mathcal{P}$, $\Sigma_{\mathcal{P}}(f)$. Since $|\mathcal{P}| < \delta$ and δ is arbitrary, this proves that f is not random Riemann integrable. ■

Chapter 6

Mycielski approximation and Riemann sums

6.1 Notation

Definition 6.1.1 Suppose we have a set S consisting of k distinct points from the unit interval $\mathbb{I}$,

$$S := \{x_i : 1 \leq i \leq k\}.$$

For each i, $1 \leq i \leq k$, define I_i^S to be the interval consisting of those points of $\mathbb{I}$ which are closer to x_i than to any other point of S. The division of $\mathbb{I}$ into such intervals is known as a *Voronoi tesselation*, and the sets I_i^S are called *Voronoi cells*. They cover the whole of $\mathbb{I}$ except for $k-1$ points which are tied for closest point. We can form the tagged partition

$$\mathcal{T}^S := \{(x_i, I_i^S) : 1 \leq i \leq k\}.$$

This tagged partition has the inclusion condition. The *nearest neighbour Riemann sum* is the Riemann sum on this tagged partition.

Now suppose that we have a random variable $\overline{X}$, which is distributed on Ω, the space of sequences of points in $\mathbb{I}$. The elements of $\overline{X}$ are $(X_k)_{k\in\mathbb{N}}$, each of which is distributed uniformly on $\mathbb{I}$, and is independent of all the others. In what follows we shall consider only that subset Ω^*, of full measure in Ω, for which no two of the X_k are equal, and such that their image set $\{X_n : n \in \mathbb{N}\}$ is dense in $\mathbb{I}$.

Set S_n to be the set $\{X_k : 1 \leq k \leq n\}$, and construct the intervals $I_i^{S_n}$ for $1 \leq i \leq n$ and the tagged partition $\mathcal{T}^{S_n}$. We will refer to the former as I_i^n and the latter as $\mathcal{T}^n$. Both are functions of the random variable $\overline{X}$ and so are random variables in their own right. In particular, $\mathcal{T}^n$ is a random variable in the space of tagged partitions. Unlike the random variables distributed on this space that we have seen in the previous chapter, its underlying partition is also a random variable.

Definition 6.1.2 Now we take some function f and we construct for each n, the Riemann sum random variable

$$\tilde{\Sigma}_n(f) := \sum_{\mathcal{T}^n} f = \sum_{i=1}^{n} f(X_i)|I_i^n|.$$

We call this the *Mycielski random Riemann sum.* We also define the step function on the same random tagged partition

$$f_n(x) := f(X_i) \text{ if } x \in I_i^n$$

setting it as the left or right limit arbitrarily at those points which do not belong to any I_i^n. We refer to this random variable as the *Mycielski step function of* f.

The reason for defining this step function is so that we can distinguish between convergence of the *Mycielski random Riemann sum* $\tilde{\Sigma}_n(f)$ to $\int f$ and convergence in L^1 of f_n to f. Of course the latter implies the former.

These definitions are a special case of a general construction which requires only a measure on a metric space. This was suggested in [35] and is discussed

at length in [20]. In fact we could modify the definition of the intervals I_i^n, by using some other metric to form the Voronoi tesselation, and we could use some other measure than Lebesgue measure, both for the distribution of the points in Ω and in place of $|I_i^n|$ in the Riemann sum.[(1)] Fremlin gives in [20] several pairs of metric and measure whose behaviour is considerably more pathological than that which we study here.

6.2 Basic results

Proposition 6.2.1 *For every f and all $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that*

$$\mathsf{P}(\int |f - f_n| > \varepsilon) < \varepsilon$$

for all $n > n_0$, where f_n is the n^{th} Mycielski step function of f.

This is proved in [35] as Theorem 1. The proof is identical to that of Theorem 5.1.5.

Suppose that we refer to the index of the nearest point to x among $S_n = \{X_k : 1 \leq k \leq n\}$ as $G_n(x)$. In other words x is closer to $X_{G_n(x)}$ than to X_j for any $j \neq G_n(x)$, and x is contained within $I^n_{G_n(x)}$. This defines, for each $x \in \mathbb{I}$, a sequence of points $(X_{G_n(x)})_{n=1}^{\infty}$. The limit of this sequence is x, and it consists of elements of (X_n), each repeated 1 or more times. If we remove these repetitions, and assuming that $x \notin \{X_k : k \in \mathbb{N}\}$, we have a subsequence of (X_n), call it $(A_n(x))_{n\in\mathbb{N}}$. We can describe this as the points of (X_n) which are closer to x than any previous term of (X_n).

We define the functions

$$f_n^{\circ}(x) := f(A_n(x))$$

[(1)] We could also use two different measures, one for the distribution of the random points, and one in place of length in the Riemann sums. This variant, which is different to that above only when the second measure is not absolutely continuous with respect to the first, has not to our knowledge been considered before.

and examine their convergence to f. We call the function $f_n^\circ(x)$ the *Mycielski pseudo-step function.*[2] As we will see this function can also be called the n^{th} *first-return recovery function of* f (depending on the trajectory (X_n)). This name will be explained in the next chapter.

Proposition 6.2.2 *If f is Lebesgue integrable, the Césaro sum of f_n°,*

$$F_n^\circ(x) := \frac{1}{n}\sum_{k=1}^{n} f_n^\circ(x)$$

converges almost surely in L^1 to f.

This is Theorem 2 in [35].

Proposition 6.2.3 *There exists a bounded function f and a set $A \subset \mathbb{I}$ of positive measure such that $f_n^\circ(x)$ does not converge to $f(x)$ for every x in A.*

This proposition will be proved in the next chapter, in the context of the first-return recovery of f. Section 7.1 makes clear the identity between these functions and their definition in the language of first-return points, and Section 7.3 gives the construction of f and discusses counterexamples to this proposition in general.

Corollary *There exists some bounded function f such that the Mycielski step functions f_n do not converge to f for any x in a set of positive measure.* □

6.3 Almost sure convergence in L^1

We can summarise the results of the previous section by saying that the Mycielski step functions converge in probability to f both pointwise and in L^1, but do not almost surely converge pointwise. This section examines almost sure L^1 convergence.

[2] By analogy with "pseudo-partition": there is a decomposition of $\mathbb{I}$ into countably many intervals together with a set of isolated points, and the function is constant on each of the intervals.

Proposition 6.3.1 *If f is in L^p for some $p > 2$, then then Mycielski step functions f_n converge in L^1 to f almost surely.*

This was proved by Fremlin as Theorem 7I of [20]. It is also worth noting that Theorem 3A and its Corollary 3B, together with Proposition 7A, all in the same work, give a more straightforward proof of the same statement, in the case that f is in L^∞.

The remainder of this chapter is devoted to some remarks which may suggest a way towards showing that L^1 convergence of f_n to f holds for some larger class of functions. These are inspired by the author's attempts to prove his conjecture as follows.

Conjecture 6.3.2 *If f is in $L^1 \log(L)$, then the Mycielski step functions f_n converge in L^1 to f almost surely.*

Remark We can make two simplifications to the definition of the Fremlin-Mycielski integral. Firstly we can replace the "nearest neighbour random sum" with the "*leftmost neighbour Riemann sum*". This means that if we order the points $\{X_k : 1 \leq k \leq n\}$ as $\{X_{\sigma(k)} : 1 \leq k \leq n\}$, where $i < j$ implies $X_{\sigma(i)} < X_{\sigma(j)}$, then the intervals of the partition are given by $[0, X_{\sigma(1)}]$, $(X_{\sigma(k-1)}, X_{\sigma(k)}]$ for $1 < k < n$ and $(X_{\sigma(n-1)}, 1]$. The nearest neighbour Riemann sum is the arithmetic mean of the leftmost neighbour random sum and the rightmost neighbour Riemann sum, therefore it converges almost surely if they both do.

Remark Secondly, it is slightly simpler to look at either "nearest neighbours" or "leftmost (rightmost) neighbours" not on $\mathbb{I}$ but on $\mathbb{S}$, the circle obtained by identifying the endpoints of $\mathbb{I}$. So if x_i is the smallest among the points, and x_j is the largest, we let the left endpoint of the interval I_i be $\frac{x_i - x_j + 1}{2}$ in the case of nearest neighbours, x_i in the case of leftmost neighbours, or x_j in the case of rightmost neighbours, rather than 0. This means that we can treat all points in a similar manner, rather than the smallest and largest among the tag points differently from all other points. Since the size of the set of points which

are moved from one partition interval to another by this change, tends to zero for large k, this does not affect the limiting behavior of the Riemann sums. We now have a construction which is translation-invariant.

Remark Darling gave in [15] a method permitting the exact determination of the expected value of the lengths of the partition intervals, or of the sum of any function of their lengths.

Theorem 6.3.6 *Suppose that $(X_i)_{i=1}^n$ are n points distributed uniformly and independently of one another on $\mathbb{I}$. Reorder them as $(X_{j_i})_{i=1}^n$ where*

$$X_{j_1} \leq X_{j_2} \leq \cdots \leq X_{j_n}$$

or without loss of generality

$$X_{j_1} < X_{j_2} < \cdots < X_{j_n}$$

since any equality holds only with probability 0. Let $\{l_i : 1 \leq i \leq n-1\}$ be defined by $l_i = X_{i+1} - X_i$, the length of the intervals in the "leftmost neighbour" partition.

Then if h is a measurable function defined on $[0,1]$,

$$\mathsf{E}(\sum_{i=1}^{n-1} h(l_i)) = \frac{1}{n(n-1)} \int_0^1 h(t)(1-t)^n dt.$$

An attempt to apply this directly to prove the conjecture above fails as follows: If we defined the "length" of each interval in the n^{th} Mycielski random Riemann sum to be $\frac{1}{n}$, than the Riemann sum would reduce to the average of n points taken uniformly and independently from the distribution of $f_{\#}\lambda$. This is the so-called 'pullback' measure of the Lebesgue measure by f, and is defined by

$$f_{\#}\lambda(A) := \lambda(f^{-1}[A])$$

where A is a (Borel) measurable subset of $\mathbb{R}$. By the strong law of large numbers

this converges almost surely to $\mathsf{E} f_{\#}\lambda = \int f$. Obviously this would differ from the actual Riemann sum, however if the sum of the defects $\sum_I \left||I| - \frac{1}{n}\right|$, taken over all intervals I were to become vanishingly small for large n, this would indicate that the difference between the Riemann sum and the simple average also becomes small for bounded f.

However, the expected number of intervals of length greater than $\frac{1}{n}$ tends to $\frac{n}{e}$, and their total length has expectation tending to $\frac{2}{e}$. Therefore the total difference between the actual lengths and $\frac{1}{n}$, given by

$$\sum_{|I|>\frac{1}{n}} \left(|I| - \frac{1}{n}\right) + \sum_{|I|<\frac{1}{n}} \left(\frac{1}{n} - |I|\right) = 2 \sum_{|I|>\frac{1}{n}} \left(|I| - \frac{1}{n}\right)$$
$$\to 2\left(\frac{2}{e} - \frac{n}{e}\frac{1}{n}\right)$$

remains close to $\frac{2}{e}$ for arbitrarily large n.

Remark This problem has some similarities to the question which was answered in Chapter 8. Both are part of a class of problems which we can express like this: *For what functions f, and what sequences h_n of functions mapping $\mathbb{I}$ into itself, converging in some sense to* $\mathrm{id}_{\mathbb{I}}$*, do we have convergence (pointwise a.e., in measure, in L^1, etc) of $f \circ h_n$ to f?* (In fact, many questions on the convergence of Riemann sums can be formulated like this.)

As expressed this is extremely general. In fact, it seems clear that we must choose the sequence (h_n) to be some very natural sequence, or else taken from an extremely restricted space of functions, to have any chance of $f \circ h_n$ converging at all.

In the present case the functions h_n are neither bijective nor measure-preserving, both of which should make the link between convergence of h_n and of $f \circ h_n$ stronger. However, it may be possible to use the regularity of h_n, which is monotone and constant on an open ball around nearly every point, and the fact that $|h_n(x) - x|$ is a monotone sequence for all $x \in \mathbb{I}$.

Chapter 7

First-return integrals

7.1 First-return recoverability

In this chapter we refer to a *trajectory* as a sequence of points from $\mathbb{I}$ which is dense in $\mathbb{I}$. Suppose that such a trajectory $\mathbf{t} = (t_n)_{n\in\mathbb{N}}$ is given.

Definition 7.1.1 If A is a subinterval of $\mathbb{I}$, the *first-return point* of A with respect to $\mathbf{t}$ is the point $r_{\mathbf{t}}(A) := t_n$, where

$$n = \min\{k \in \mathbb{N} : t_k \in A\}.$$

In other words t_n is the first point of the sequence $\mathbf{t}$ to belong to A. Since $[\mathbf{t}]$, defined as $\{t_n : n \in \mathbb{N}\}$ is dense in $\mathbb{I}$ such a point always exists.

Definition 7.1.2 Let $\mathbf{t}$ be a trajectory and x be a point of $\mathbb{I}$. The *first-return route* of $\mathbf{t}$ to x is the sequence defined inductively by

$$y_1^x(\mathbf{t}) := r_{\mathbf{t}}(\mathbb{I}) = t_1$$

and for all $n \in \mathbb{N}$

$$y_{n+1}^x(\mathbf{t}) := r_{\mathbf{t}}(B(x, |x - y_n^x(\mathbf{t})|)),$$

except in the case where $y_n^x(\mathbf{t}) = x$, in which case the above definition will not

work and we simply let $y^x_{n+1}(\mathbf{t}) := x$ as well.

So the first-return route is a subsequence of $\mathbf{t}$ which has the same initial element as $\mathbf{t}$ and such that each subsequent term is the first element of $\mathbf{t}$ to be strictly closer to x than the previous term, except in the case where $x \in [\mathbf{t}]$, in which case the first-return route is eventually constant and equal to x. The first-return route obviously converges to x in all cases.

Definition 7.1.3 Now, suppose that f is a function $\mathbb{I} \mapsto \mathbb{R}$ (not necessarily measurable) and that x is a member of $\mathbb{I}$. We say that f is *first-return recoverable with respect to* $\mathbf{t}$ *at* x if

$$f(y^x_n(\mathbf{t})) \to f(x) \text{ as } n \to \infty.$$

We say that f is (almost) everywhere first-return recoverable w.r.t. $\mathbf{t}$ if it is first-return recoverable at x for (almost) every $x \in \mathbb{I}$. Alternatively we say that $\mathbf{t}$ (a.e.) recovers f.

Remark Here we see that the functions f°_n defined in the previous chapter, for a fixed sequence $\overline{X} = \mathbf{t}$, may be defined by

$$f^\circ_n(x) = f(y^{x,\mathbf{t}}_n),$$

and that to say that they converge a.e. pointwise to f is the same as saying that f is almost everywhere first-return recoverable. This explains the name "first-return recovery functions". While the pointwise limits of the two sequences of functions, the Mycielski step functions and the first-return recovery functions, are identical, the natural context for discussing the Mycielski step functions is with the construction based on Voronoi regions in the previous chapter, and the first-return recovery functions are related to first-return integration, and so belong in this chapter.

It is clear that continuous functions are everywhere first-return recoverable with respect to any trajectory. On the other hand, a discontinuous function

must, for any discontinuity point, have a trajectory w.r.t. to which it is not recoverable at that point. We are first interested in those functions for which there exists at least one trajectory with respect to which they are everywhere first-return recoverable.

Remark For a function to be (almost) everywhere first-return recoverable with respect to some trajectory is equivalent to saying that for that trajectory the Mycielski step functions converge pointwise (almost everywhere) to f. Equivalently the Mycielski pseudo-step functions converge pointwise (almost everywhere) to f. The first characterization immediately yields one direction of the following proposition.

Proposition 7.1.6 *A function f is everywhere first-return recoverable w.r.t. some trajectory* $\mathbf{t}$ *iff it is Baire one.*

This was proved in [14]. As pointed out in the previous remark, the very simple proof of one direction can be deduced from the characterization of the limit in terms of step functions. The proof that a Baire one function must be everywhere first-return recoverable is longer and is given in [14]

Theorem 7.1.7 *A function is almost everywhere first-return recoverable w.r.t some trajectory* $\mathbf{t}$ *iff it is measurable.*

Proof. As with the previous proposition one direction, namely that "a.e. first-return recoverable" implies "measurable", is easy and follows from the characterization of a.e. first-return recoverable functions as the almost everywhere limit of step functions

The opposite direction is more complicated. A proof is given in [17] (Theorem 2.3). The proof relies on the fact, given below as Proposition 7.2.3, that a first-return integrable function must be almost everywhere first-return recoverable. Since, for example, all bounded Lebesgue measurable functions are first-return integrable, so they are first-return recoverable. But recoverability is a topological property which is unchanged by a homeomorphism of the image

space, for example left composition by arctan. Thus we can associate all measurable functions to some bounded measurable function, which shows that they all must be a.e. first-return recoverable. ■

The paper [13] gives a number of properties of functions closely related to (everywhere) first-return recoverability. These address the fact that, in order to recover a function with isolated points on its graph, we need to include the abscissae of these points in our trajectory. We can tighten the notion of recoverability to *first-return approachability* if the graph of our function does not have isolated points. This simply means that at each stage of the construction of the first-return route to x we take the next point of the sequence which is closer to x than the previous elements of the route, *and which is not x itself.* Something similar applies for points which are isolated from the right or the left, or points which are not isolated but which are separated from the set of continuity points. The paper proves one general theorem and all the characterizations of functions for which slightly stronger notions than recoverability hold follow from this. These characterizations are all of subsets of the class of Baire one functions whose graphs satisfy in addition various topological properties.

7.2 Integrating sequences

We now consider the question of first-return integration. This is a procedure based on Riemann sums, which reflects the values of the function on the trajectory in question.

Definition 7.2.1 The *first-return integral* of a function f on a measurable set A w.r.t. a trajectory $\mathbf{t}$ exists and is equal to M iff for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left| \sum_{I \in \mathcal{P}} f(r_{\mathbf{t}}(I)) |I \cap A| - M \right| < \varepsilon$$

for all partitions $\mathcal{P}$ with $|\mathcal{P}| < \delta$.

The first thing that we can see about this definition is that it only depends on the restriction of f to the points of $\mathbf{t}$. Therefore the first-return integral alone is extremely weak and it is possible to find examples where its value in no way resembles that of any reasonable integral. We will strengthen the first-return integral in three ways; by requiring it to integrate a function on all measurable subsets of $\mathbb{I}$, by requiring the trajectory to be one which also recovers f, or in the next section by looking at integrals which are the same for almost every trajectory, in a sense which will be made clear. The first idea, to look at integrals on all subsets, motivates the following definition. It does not strengthen the first-return integral by very much since of course it is still sensitive to changing the function on a countable set.

Definition 7.2.2 If f is Lebesgue integrable and the first-return integral of f w.r.t. $\mathbf{t}$ is equal to the Lebesgue integral on every measurable subset of $\mathbb{I}$, we say that $\mathbf{t}$ is a *integrating sequence* for f, or that $\mathbf{t}$ *integrates* f.

Proposition 7.2.3 *If f is in L^1, then every trajectory which recovers f almost everywhere is an integrating sequence for f.*

This proposition comes from [17]. However, a better proof can be deduced from part 3Bc of [19]. Here it is proved that if $F \subset \mathbb{I}$ is a closed set of positive measure, no trajectory has both of the two following properties:

1. The first-return route to x includes infinitely many points not in F, for almost every x (the set Σ' in section 3B of [19]). This is the same as saying that $\mathbf{t}$ does not recover f at almost all of the points of F.

2. The first-return integral of χ_F with respect to the trajectory is defined and equal to $\int_A \chi_F = \lambda(A \cap F)$ for every measurable set A (the set Σ_{χ_F} in section 3B of [19]). That is to say, the trajectory integrates χ_F.

So we can rephrase the statement that the set of trajectories satisfying both of these is empty as the following lemma.

Lemma 7.2.4 *If f is the characteristic function of a measurable set F, a trajectory which integrates f also recovers f almost everywhere.*

While F in [19] is a closed set, this is not necessary for this statement to be proved.

Proof. Suppose that $\mathbf{t}$ integrates f but that there is a set of positive measure A on which it does not recover f. Assume without loss of generality that $A \cap F$ has positive measure (if not, exchange F with its complement). For every $x \in A \cap F$ and every δ there exists some $\delta' < \delta$ such that $r_{\mathbf{t}}(B(x, \delta'))$ is not in F.

Fix some $\varepsilon > 0$. We can choose $\delta_0 > 0$ so that any first-return Riemann sum of f on a partition smaller than δ_0 differs from the Lebesgue integral by no more than $\frac{\varepsilon}{3}$. Now for each $x \in F \cap A$ consider the family of balls with center x, and with

- radius smaller than $\frac{\delta_0}{2}$.
- the first-return point of the ball not in F.

For every x in $A \cap F$ there exist arbitrarily small balls containing x with these properties.

Using the Vitali covering lemma, we can find a disjoint set of balls $B(x, \delta(x))$ with $x \in F \cap A$, with these two properties, whose union includes all but at most $\frac{1}{3}$ of the measure of $F \cap A$. Hence there is also a finite subset of this collection of balls, whose union includes all but at most $\frac{1}{2}$ of the measure of $F \cap A$. Call this subset D and its union

$$\mathscr{D} := \bigcup_{I \in D} I.$$

Now, if we take a partition containing all the intervals of D, and any other intervals with length less than δ_0 which cover $\mathbb{I} \backslash \mathscr{D}$, the size of the partition will be less than δ_0. The first-return integral of f on $\mathscr{D}$ with respect to $\mathbf{t}$ will be 0, and since

$$\int_{\mathscr{D}} f = \lambda(\mathscr{D} \cap F) \geq \frac{1}{2}\lambda(A \cap F),$$

it will differ from the Lebesgue integral on $\mathscr{D}$ by at least $\frac{1}{2}\lambda(A \cap F)$. But ε was arbitrary, so by choosing ε less than $\frac{1}{2}\lambda(A \cap F)$ we have a contradiction. ■

Proof (Proposition 7.2.3). Now assume that f has an integrating trajectory $\mathbf{t}$ and assume that this trajectory does not recover f at a set A of positive measure. Let A^+ be the subset of A where

$$\limsup_{n\in\mathbb{N}} f(y_n^x((t))) \geq f(x).$$

Without loss of generality we assume this to have positive measure. Then so does the set A_ε, defined by

$$A_\varepsilon := \{x \in A : \limsup f(y_n^x(\mathbf{t})) > f(x) + \varepsilon\}$$

for some $\varepsilon > 0$.

We can apply the same construction as in the proof of the preceding lemma, using the set A_ε in place of $A \cap F$, except that we add a third condition to the balls around x, that the radius, call it r_x, be chosen such that

$$\int_{x-r_x}^{x+r_x} f < 2r_x \left(f(x) + \frac{\varepsilon}{2}\right).$$

By the Lebesgue density theorem we can always do this, if necessary replacing A_ε by a subset of full measure. Now the first-return integral of f on the collection of balls will be more than

$$\sum_x 2r_x(f(x) + \varepsilon)$$

and the Lebesgue integral will be less than

$$\sum_x 2r_x \left(f(x) + \frac{\varepsilon}{2}\right)$$

Since $\sum_x r_x$ is bounded from below independently of the choice of ε we are done. ■

A partial converse holds only for bounded functions.

Proposition 7.2.5 *If f is bounded and measurable, then every integrating sequence for f recovers f almost everywhere.*

This lemma was also proved in [17].

The relationship between integrating trajectories and trajectories which recover f was exploited in [17] to show that every measurable function is recoverable. We mentioned this is the preceding section, where we stated but did not prove Theorem 7.1.7. In the remainder of this subsection we will show that every L^1 function has an integrating trajectory, and thus obtain that result as well as one further consequence.

Proposition 7.2.6 *If f is in L^1 then f has an integrating sequence.*

Remark In [17] it is proved that for f Lebesgue integrable, it is enough for the first-return integral of f w.r.t. to some trajectory to be equal to the Lebesgue integral on all closed intervals, for the same equality to hold on all measurable sets. Therefore, it is claimed that the proof in [11] that every L^1 function has a trajectory which yields the Lebesgue integral on every interval, is sufficient to show that every L^1 function has an integrating trajectory. In fact this is not explicitly proved anywhere in [11]. For the purposes of clarity we outline the three separate statements which prove the theorem in [17].

- If f is Lebesgue integrable, then there exists a trajectory such that the first-return integral of f on $\mathbb{I}$ with respect to that trajectory exists and is equal to $\int f$. This is the main theorem of [11].

- If f is a measurable function, $A \subset \mathbb{I}$ is an interval, then for any trajectory w.r.t. which the first-return integral of f on $\mathbb{I}$ is equal to the Lebesgue integral on $\mathbb{I}$, the two integrals will also be equal on A. This is essentially a Henstock lemma for the first-return integral. It is proved as Lemma 2.1 of [9].

- If the first-return integral of f with respect to $\mathbf{t}$ is equal to the Lebesgue integral on all closed intervals, then the two are also equal on all measurable sets. This is proved in [17] as Lemma 2.3.

Propositions 7.2.6 and 7.2.5 have the obvious corollary that every bounded function is a.e. recoverable w.r.t. some trajectory. We can say more; as remarked already a transformation of f which preserves the topology generated on $\mathbb{I}$ by its preimages preserves recoverability. By combining this argument with Proposition 7.2.3 we obtain the following statement, which was not given in [17].

Proposition 7.2.8 *For every Lebesgue integrable function there is a trajectory which integrates f and recovers f almost everywhere.*

Proof. Define $g : \mathbb{I} \mapsto (-\frac{\pi}{2}, \frac{\pi}{2})$ by $\tan(g(x)) = f(x)$. By Proposition 7.2.6 there exists some trajectory $\mathbf{t}$ which integrates g. By Proposition 7.2.5 this trajectory also recovers f almost everywhere. Since, given a sequence $(x_n)_{n\in\mathbb{N}}$, $g(x_n)$ converges to $g(x)$ exactly if $f(x_n)$ converges to $f(x)$, this same trajectory also recovers f. Then by Proposition 7.2.3, $\mathbf{t}$ also integrates f. ■

7.2.1 Functions not in L^1

We now look at the possibility that the first-return integral can yield the KH-integral of a function which is not necessarily Lebesgue integrable. In this section the term KH-function refers to a (Lebesgue equivalence class containing a) measurable function which is KH-integrable but not Lebesgue integrable. In this context we use the expression "integrating sequence" to refer to a trajectory with respect to which the first-return integral of f on $[a, b]$ is equal to $\text{KH}\int_a^b f$, for all $0 \leq a < b \leq 1$. We cannot ask for the first-return integral to match the KH-integral on all measurable subsets of $\mathbb{I}$ since the KH-integral is not in general defined on all of these.

Remark The results of this section show that these two definitions of integrating sequence can be reconciled. If we define an integrating sequence as one for

which the first-return integral is equal to the KH-integral on all subintervals of $\mathbb{I}$, then by Lemma 2.1 of [9] and Lemma 2.3 of [17] as mentioned above, where f is in L^1 this is equivalent to the first-return integral being equal to the Lebesgue integral on all measurable subsets of $\mathbb{I}$.

Bongiorno in [5] and [4] first constructed an example of a KH-function which does not have any integrating sequence. His example was adapted by Darji and Evans in [12], who gave a much larger class of KH-functions with this property. Bongiorno also asked whether there existed an KH-function for which the first-return integral does equal the KH-integral.

Since it is obviously trivial to find a function and a single sequence which integrates it, he placed two conditions on the integrating sequence. Firstly that it should also recover the function, and less formally that the sequence should be considered "natural". Section 2 of [19] is such a construction.

7.3 Almost sure first-return integration

We now turn to the question of which functions are integrated by almost every sequence, in the sense of the natural probability measure on the space of sequences previously considered.

As in Chapter 6, Ω is the space of sequences of points in $\mathbb{I}$, and we consider the natural probability on Ω. Since the support set of a trajectory must be dense in $\mathbb{I}$, we will consider only the conegligible subset Ω^* of Ω consisting of trajectories.

Definition 7.3.1 A function f is *almost surely first-return integrable* if almost every trajectory is an integrating trajectory for f.

It will also be useful to consider the analogous property for first-return recovery.

Definition 7.3.2 A function f is *almost surely, almost everywhere recoverable* if f is recovered almost everywhere by almost every trajectory.

These two definitions, taken with Lemma 7.2.3, divide the set of Lebesgue integrable functions into 3 classes[1]:

(I) f is a.s. first-return integrable (and therefore a.s. a.e. recoverable).

(II) f is a.s. a.e. first-return recoverable but not a.s. first-return integrable.

(III) f is not a.s. a.e. recoverable.

Of course, the Zero-One law applies to both first-return recovery and the first-return integral, so a function which is not almost surely recoverable almost everywhere has some positive set on which the first-return recovery sequence almost surely does not converge to the value of f. Similarly for the first-return Riemann sums, a function which is not almost surely first-return integrable, is almost surely not first-return integrable.

Class (I) obviously includes RI*, those functions Lebesgue equivalent to a Riemann integrable function. It is notable that no other functions have been shown to belong to (I), and that no functions are known to belong to (II)[2].

Fremlin's construction in 3B of [19] gives an example of a function in class (III). In fact this example is bounded, the characteristic function of a closed set. So no condition on the size of f can ensure that it is in classes (I) or (II). In fact, Fremlin's example is suggestive that a large family of Lebesgue integrable and indeed bounded functions are in class (III). We rephrase his proof to provide a sufficient condition for membership of class (III).

Lemma 7.3.3 *Suppose that A, B are sets of positive measure $A \subseteq B \subseteq \mathbb{I}$ so that for all $x \in A$ there exists a nonnegative, monotone decreasing sequence D_n^x, such that $\sum_{n=1}^{\infty} D_n^x = +\infty$ and*

$$1 - \lambda\left(B|B\left(x, e^{-n-\sqrt{2\log\log n}}\right)\right) > D_n^x$$

[1] If Lemma 7.2.3 can be extended to integrating sequences for KH-functions, that this decomposition would apply to all measurable functions (defining non-KH-integrable functions to be not almost surely first-return integrable). However we cannot exclude the (unlikely) possibility that there exist KH-functions which are almost surely first-return integrable but not almost surely almost everywhere recoverable.

[2] The functions given by Bongiorno and by Darji and Evans are almost surely almost everywhere recoverable, and are not almost surely first-return integrable. They do not appear here because they are not in L^1.

for all but finitely many values of n. Then χ_B is not a.s. a.e. recoverable.

Proof. Suppose that $x \in A$. We will show that the probability that the first-return route of $\mathbf{t} \in \Omega^*$ to x recovers $f(x)$ is 0. Using Fubini's theorem on $\lambda \restriction_A \times \mathsf{P}$, where P is the probability measure on Ω, this is enough to show that almost every trajectory does not recover f at any point of a set of full measure in A.

First we estimate from below the speed of convergence of the first-return route to x. In the case where $x = 0$ (or the symmetric case $x = 1$), we reason as follows. Assuming without loss of generality that $0 \notin [\mathbf{t}]$, the first-return route of $\mathbf{t}$ to 0, call it $y_n^0(\mathbf{t})$, consists of a point A_1 distributed uniformly on $[0, 1]$, followed by a point A_2 distributed uniformly on $[0, A_1]$ (and independently of A_1), and so on. Writing Y_n^0 for the random variable whose value is $y_n^0(\mathbf{t})$, we have

$$Y_n^0 = \prod_{i=1}^{n} U_n$$

where U_n are i.i.d. uniformly on $\mathbb{I}$.

Therefore $\log y_n^0 = -\sum_{i=1}^{n} E_n$, where E_n has an exponential distribution with mean 1. Write $E_n = 1 + X_n$, where X_n are i.i.d. with mean 0 and variance 1. Khinchin's law[3] tells us that $\sum_{i=1}^{n} X_n > \sqrt{2n \log\log n}$ for only finitely many values of n, almost surely. So $\left|Y_n^0 - 0\right| = Y_n^0 < e^{-n-\sqrt{2n \log\log n}}$ only finitely often almost surely.

To show a similar bound for a general point $x \in (0, 1)$, we first define $d_x = \min(x, 1 - x)$ and let N_x be the smallest natural number such that Y_n^x is contained in $B(x, d_x)$. Now consider the σ-algebra on $B(x, d_x)$ generated by the open balls with center x. Since we are ultimately interested only in $|Y_n^x - x|$, we see that we do not lose any information by considering the distribution of Y_n^x restricted to this σ-algebra.

So write $W_n^x := |Y_n^x - x|$, for $n \geq N$, considered as a random variable on this σ-algebra. There is a natural measure isomorphism from Lebesgue

[3] See for example chapter 8 of [26].

measure restricted to the "centered" σ-algebra on $B(x, d_x)$ to Lebesgue measure on $[0, 2d_x]$ with the usual Borel σ-algebra. Moreover, this isomorphism maps the joint distribution of $(W_n^x)_{n=N}^{\infty}$ onto the joint distribution of $(\frac{1}{2}Y_n^0)_{n=M}^{\infty}$, where M is the first natural number such that $Y_M^0 \in [0, 2d_x]$. Further it is easy to see that N and M are identically distributed and almost surely finite.

This means that for all x in $\mathbb{I}$, we have that the slightly weaker bound

$$|Y_n^x - x| < 2e^{-n-\sqrt{2n \log\log n}}$$

holds no more than finitely many times almost surely.

The remainder of the proof is straightforward; suppose that for some x and some n and m

$$2e^{-m-\sqrt{2m \log\log m}} \leq W_n^x < 2e^{-(m-1)-\sqrt{2(m-1)\log\log(m-1)}}.$$

Since $B(x, e^{-m-\sqrt{2m \log\log m}})$ contains a proportion of its measure of at least $\frac{D_m^x}{2}$ of the complement of B, $B(x, W_n^x)$ similarly contains a proportion of no less than

$$\frac{D_m^x}{2} \frac{e^{-m-\sqrt{2m \log\log m}}}{e^{-(m-1)-\sqrt{2(m-1)\log\log(m-1)}}},$$

say no less that $\frac{D_m^x}{3}$ for all but a few m. So the probability that Y_{n+1}^x will not be contained in B is at least $\frac{D_m^x}{3}$. Since for all but finitely many n, $m \leq n$ and so $D_m(x) > D_n(x)$, we have that with probability 1, $Y_n^x \notin B$ infinitely often.

Because $x \in A \subseteq B$, this means that $\chi_B(Y_n^x)$ does not converge to $\chi_B(x)$ almost surely. ■

It is not difficult to put this together with Lemma 2.2.11 to prove that a characteristic function of the desired type does exist.

Proposition 7.3.4 *There exists a bounded measurable function which is not almost surely almost everywhere recoverable.*

Proof. Let a_n be some decreasing sequence with limit 0 and $\sum_{n\in\mathbb{N}} a_n = +\infty$.

Lemma 2.2.11 proves the existence of a set $H \subset \mathbb{I}$ such that for all $x \in H$, and for all but finitely many $n \in \mathbb{N}$ we have

$$1 - \lambda(H|B(x,\delta)) > a_n$$

for all $\delta > e^{-n-\sqrt{2n \log\log n}}$. This is clearly enough to imply the condition of Lemma 7.3.3, taking $A = B = H$. So χ_H is not a.s. a.e. recoverable. ■

In fact this construction implies that Class (III) is residual in L^∞.

Proposition 7.3.5 *The set of functions which are almost surely almost everywhere recoverable is nowhere dense in L^∞.*

Proof. Let $f := \chi_H$ be a function given by 7.3.4. If a is a small positive number then $a \cdot f$ is equally in Class (III). Take some number r smaller than $\frac{a}{2}$ and consider the ball in L^∞ around $a \cdot f$ of radius r. Every function in this ball is larger than $a - r > \frac{a}{2}$ on H and smaller than $r < \frac{a}{2}$ on $\mathbb{I}\backslash H$. It is impossible for the limit of the function on a sequence of points in the second set to be equal to the value of the function at point from the first set. Therefore at the set of points where f is not recovered almost surely, neither will be any function in this ball.

Now consider an open set E in L^∞. Assume that there exists some function g in E which is almost surely almost everywhere recoverable, otherwise we are done. Choose a small enough that $B(g, 2a) \subset E$. Then choose r as indicated above, this will mean that $B(g + a \cdot f, r)$ is also contained in E.

It remains to check that the sum of an almost surely almost everywhere recoverable function and a not almost surely almost everywhere recoverable function is not almost surely almost everywhere recoverable. But this is easy to see since all the relevant limits are additive in the two functions. ■

It is not possible for the set of almost sure almost everywhere recoverable functions to be nowhere dense in L^1, because the Riemann integrable functions are dense.

7.3.1 Connections to the random Riemann integral

Consider the distribution of trajectories given by the probability measure on Ω. If we fix some index $k \in \mathbb{N}$, then the distribution of t_k will naturally be uniform on $\mathbb{I}$. Furthermore the distribution of t_k conditional on it belonging to some measurable subset of $\mathbb{I}$ is likewise uniform on that subset. This allows us to see that the first-return point of some interval I, $r_{\mathbf{t}}(I)$, is distributed uniformly on I, and that if I, J are two intervals, $|I \cap J| = 0$, then $r_{\mathbf{t}}(I)$ and $r_{\mathbf{t}}(J)$ are independent.

This allows us to see that if we fix some partition $\mathcal{P}$, then the distribution of the tag points, and hence also of the Riemann sums as a function of $\mathbf{t} \in \Omega$, is identical to that given in the definition of the random Riemann integral in Section 5.1.

So some results concerning the random Riemann integral give us information about almost sure first-return integrability. In particular, the convergence of the weak random Riemann integral for all functions in L^1 is a necessary condition for the convergence of the first-return integral almost surely.

However, we must bear in mind that unlike the random Riemann integral, the distributions of the Riemann sums of the first-return integral on two different partitions are not independent. Since the results on the strong random Riemann integral in Chapter 5 all depend on the Borel-Cantelli lemma, we see that the positive results also hold for the dependent case, whereas the negative results do not.

So the strong convergence of the random Riemann integral (to the Lebesgue integral) implies that almost surely the first-return Riemann sums also converge to the integral, if we take the same sequence of partitions. However, for a function to be almost surely first-return integrable, we need almost sure convergence given any sequence of partitions with size tending to 0. In fact there are no positive results which do not rely on some condition on the speed of convergence of the partition sizes. So there is no function outside of RI* which we can show is in Class (I).

On the other hand the negative results for the random Riemann integral in Chapter 5 do not hold. So although we can show that for some functions and some sequences of partitions there is no strong convergence on the random Riemann integral, we can not show that the first-return Riemann sums do not converge on the same sequences of partitions, which would exclude such functions from being almost surely first-return integrable.

However those functions for which the random Riemann integral does not converge *weakly*, we know are not first-return integrable with probability 1. The following proposition is a direct corollary of Theorem 5.3.8.

Proposition 7.3.6 *If f is not in $L^{1-\varepsilon}$ for some $\varepsilon > 0$* or *f fails to have the endpoint weak L^1 condition on some subinterval of $\mathbb{I}$, then f is not almost surely first-return integrable.*

7.4 The Kieffer-Stanojević integral

We conclude this chapter by mentioning another probabilistic Riemann integral. The reason for placing it in this chapter is that it is a restriction of the first-return Riemann integral. In fact the distribution of the first-return Riemann sums on a sequence of partitions, each a refinement of the previous, is exactly that of the Kieffer-Stanojević integration procedure. The fact that all of L^1 is Kieffer-Stanojević integrable, but much of it is not almost surely first-return integrable shows that we cannot approximate first-return Riemann sums on a general sequence of partitions by those on a sequence consisting of successive refinements.

The Kieffer-Stanojević integral was suggested in [30]. As before, a sequence of partitions is given and a corresponding sequence of tagged partitions (with the inclusion condition) is defined as a random variable. Let $(\mathcal{P}_i)_{i\in\mathbb{N}}$ be a sequence of partitions, where each partition is a refinement of the previous one. This means that each interval is contained within an interval of the previous partition.

For $k \geq 2$, write ψ_k for the containment mapping from $\mathcal{P}_k$ to $\mathcal{P}_{k-1}$. This

function maps each interval to the unique one of the previous partition of which it is a subset. The first tagged partition is defined by $\mathcal{T}_1 := \{(t_I, I) : I \in \mathcal{P}_1\}$, where t_I is a random variable distributed on I uniformly and independently of all other t_I. Now suppose that $\mathcal{T}_{k-1}$ has been defined. Let $t_{k-1} : \mathcal{P}_{k-1} \to \mathbb{I}$ be the mapping which sends each interval in $\mathcal{T}_{k-1}$ to its tag. For those intervals $I \in \mathcal{P}_k$ such that $t_{k-1}(\psi_k(I)) \in I$, the tag of I in $\mathcal{T}_k$ is $t_{k-1}(\psi_k(I))$. In other words those intervals which contain a tag point of the previous partition, have that point as their tag. For all other intervals, the tag is chosen uniformly in that interval and independently of all previous choices.

We refer to the n^{th} tagged partition as $\mathcal{T}_n^{KS}$, and to the Riemann sum of f on that partition as $\Sigma_n^{KS} f$. The Riemann sums are distributed identically to the first-return Riemann sums on the same sequence of partitions.

Definition 7.4.1 (Kieffer-Stanojević integral) A function f is *Kieffer-Stanojević integrable* and M is its *Kieffer-Stanojević integral* in case $\Sigma_n^{KS} f$ converges to M almost surely.

Theorem 7.4.2 *If f is in L^1 then it is Kieffer-Stanojević integrable and its Kieffer-Stanojević integral is $\int f$.*

Proof. The proof of almost sure convergence was first given in [30]. It follows from the Reverse Martingale Convergence Theorem, originally proved in [16]. We simply show that the Kieffer-Stanojević random sums form a reverse martingale, and from this it follows that convergence in probability implies convergence almost surely to the same limit. Convergence in probability to the integral of f can be shown by the argument from Theorem 5.1.5.

To show that $\Sigma_n^{KS} f$ is a reverse martingale, we show that

$$\mathsf{E}(\Sigma_{n-1}^{KS}(f) \upharpoonright \Sigma_n^{KS} f) = \Sigma_n^{KS} f.$$

We have that

$$\Sigma_n^{KS} f := \sum_{k=1}^{a_n} f(t_{n,k}) |I_{n,k}| \tag{7.1}$$

and

$$\Sigma_{n-1}^{KS}(f) := \sum_{k=1}^{a_{n-1}} f(t_{n-1,k})|I_{n-1,k}| \tag{7.2}$$

where a_n is the number of elements in $\mathcal{T}_n^{KS}$ and $(I_{n,k}, t_{n,k})$ is the k^{th} element, numbered from left to right.

We know that all intervals of the underlying partitions are chosen deterministically. Each choice of a tag point is independent of all others, except in the case of the tag points of two intervals, one of which contains the other.

Therefore each term in (7.2), of the form $f(t_{n-1,j})|I_{n-1,j}|$, is independent of all terms in (7.1), except those $1 \leq k \leq a_n$ for which $I_{n,k} \subseteq I_{n-1,j}$. This means that

$$\mathsf{E}\left(f(t_{n-1,j})\,|I_{n-1,j}| \upharpoonright \Sigma_n^{KS} f\right) = \mathsf{E}\left(f(t_{n-1,j})\,|I_{n-1,j}| \upharpoonright \sum_{k:I_{n,k}\subseteq I_{n-1,j}} f(t_{n,k})|I_{n,k}|\right)$$

and

$$\mathsf{E}\left(\Sigma_{n-1}^{KS}(f) \upharpoonright \Sigma_n^{KS} f\right) = \sum_{j=1}^{a_{n-1}} \mathsf{E}\left(f(t_{n-1,j})\,|I_{n-1,j}| \upharpoonright \sum_{k:I_{n,k}\subseteq I_{n-1,j}} f(t_{n,k})|I_{n,k}|\right)$$

So it suffices to prove that

$$\mathsf{E}\left(f(t_{n-1,j})|I_{n-1,j}| \upharpoonright \sum_{k:I_{n,k}\subseteq I_{n-1,j}} f(t_{n,k})|I_{n,k}|\right) = \mathsf{E}\left(\sum_{k:I_{n,k}\subseteq I_{n-1,j}} f(t_{n,k})|I_{n,k}|\right). \tag{7.3}$$

Now $t_{n-1,j}$ is one of $\{t_{n,k} : I_{n,k} \subseteq I_{n-1,j}\}$. The conditional probability of it being $t_{n,i}$ for some $i \in A := \{k : I_{n,k} \subseteq I_{n-1,j}\}$, conditioned on $\sum_{k:I_{k,n}\subseteq I_{n-1,j}} f(t_{n,k})|I_{n,k}|$ (or equivalently on $\{t_{n,k} : k \in A\}$) is given by

$$\mathsf{P}(t_{n-1,j} \in I_{n,i}) = \frac{|I_{n,i}|}{|I_{n-1,j}|}.$$

So the left hand side of (7.3) is equal to

$$\mathsf{E}|I_{n-1,j}| \sum_{k\in A} \frac{|I_{n,k}|}{|I_{n-1,j}|} f(t_{n,k}) = \mathsf{E} \sum_{k\in A} |I_{n,k}|\, f(t_{n,k})$$

as required. ■

Chapter 8

Measure preserving transformations

This chapter deals with permutations of the unit interval, and their effect of composing them with measurable functions defined on that interval. Except for the final section which was added later, it is joint work with T. Nishiura and was adapted from the paper [25].

8.1 Preliminaries

In this chapter we consider a partition of $[0,1)$ to consist only of half-open intervals of the form $[a,b)$. The theory could work equally well with a general partition of $\mathbb{I}$ but this would introduce unnecessary technicalities.

Definition 8.1.1 A function $\varphi\colon [0,1] \to [0,1]$ is called a *permutation* of a partition Π of $[0,1)$ if φ restricted to I is a translation for each $I \in \Pi$, and $\varphi(1) = 1$.

Clearly a permutation is bijective and has the property that the Lebesgue measures of $\varphi(E)$ and $\varphi^{-1}(E)$ are equal to the Lebesgue measure of E for every Lebesgue measurable set $E \subset [0,1]$. The collection of all permutations will be denoted by $\mathscr{P}$.

The collection of all functions $h\colon [0,1] \to [0,1]$ that are almost everywhere limits of sequences in $\mathscr{P}$ will be denoted by $\mathscr{H}$. This is the same as the collection of functions that are limit in measure of sequences in $\mathscr{P}$.[1]

The collection of partitions of $[0,1)$ can be used to define another collection $\mathscr{H}'$. A function $h\colon [0,1] \to [0,1]$ is in $\mathscr{H}'$ if, for each sequence $(\Pi_n)_{n=1}^{\infty}$ of partitions of $[0,1)$ such that $|\Pi_n| \to 0$ as $n \to \infty$, there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ where φ_n is a permutation of Π_n such that φ_n converges almost everywhere to h. Obviously, $\mathscr{H}' \subset \mathscr{H}$. It will be shown in Section 8.3 that $\mathscr{H} = \mathscr{H}'$ and that $h \in \mathscr{H}$ if and only if h is λ-measurable and $\lambda = h_{\#}\lambda$. As mentioned in chapter 6, this notation denotes a measure defined by

$$f_{\#}\mu(A) := \mu(f^{-1}(A))$$

where μ is a measure and f is a μ-measurable function.

8.2 Factorization problems

The statement of the proposed factorization problem is the following: Find a minimal[2] class $\mathscr{G}$ of functions $g\colon [0,1] \to \mathbb{R}$ such that for each Lebesgue measurable function $f\colon [0,1] \to \mathbb{R}$ there is an h in $\mathscr{H}$ and a g in $\mathscr{G}$ such that the composition $g \circ h\colon [0,1] \to \mathbb{R}$ is Lebesgue equivalent to f. Observe that any collection of functions which includes one representative from each Lebesgue equivalence class of measurable functions, for example the set $\mathcal{B}_2$ of Baire class 2 functions, solves the factorization problem if we drop the condition that the class be minimal.

We can reformulate the question as follows: Find a minimal class $\mathscr{G}$ of functions $g\colon [0,1] \to [0,1]$ having the property that each Borel measurable function $f\colon [0,1] \to (0,1)$ has corresponding functions $h\colon [0,1] \to [0,1]$ in $\mathscr{H}$ and g in $\mathscr{G}$

[1] If a function is the almost everywhere limit of a sequence of functions then it is also their limit in measure. If a function is the limit in measure of some sequence of functions then it is the almost everywhere limit of a subsequence.

[2] As usual, we mean minimal in the sense of the partial order $\subset$. Of course, a minimal set need not be unique.

such that the composition $g \circ h$ is λ-equivalent to f.

Note that the Lebesgue measurability of $f\colon [0,1] \to \mathbb{R}$ has been replaced with Borel measurability of $f\colon [0,1] \to (0,1)$ and that $g\colon [0,1] \to \mathbb{R}$ has been replaced by $g\colon [0,1] \to [0,1]$. Clearly there is no loss of generality in making these replacements. We shall show that the class $\mathscr{G}$ of upper continuous, nondecreasing functions, with $g(0) = \lim_{h\to 0^+} g(h)$ will solve the factorization problem.

8.3 Characterization of $\mathscr{H}$ and $\mathscr{H}'$

The two collections $\mathscr{H}$ and $\mathscr{H}'$ were introduced to define the factorization problem. We have already pointed out that$\mathscr{H}' \subset \mathscr{H}$. We have the following chacterization.

Theorem 8.3.1 *$\mathscr{H}' = \mathscr{H}$, and $h \in \mathscr{H}$ if and only if h is a Lebesgue measurable function such that $\lambda = h_{\#}\lambda$.*

A Lebesgue measurable function $h\colon [0,1] \to [0,1]$ is said to be *measure preserving*[3] if $h_{\#}\lambda = \lambda$. Note that $\lambda(h^{-1}(1)) = 0$.

The next 2 propositions provide the proof of the theorem.

Proposition 8.3.2 *If $h \in \mathscr{H}$, then h is measure preserving.*

Proof. Let φ_n be a sequence of permutations converging in measure to h, and let ε_n be such that $\lambda(E_n) < \varepsilon_n$, where $E_n = \{\, x\colon |\varphi_n(x) - h(x)| \geq \varepsilon_n \,\}$, with $\varepsilon_n \to 0$ as $n \to \infty$. For closed sets K in $[0,1]$ let K_n be the ε_n-neighborhood of K. As ${\varphi_n}^{-1}[K] \subset h^{-1}[K_n] \cup E_n$, it follows that $\lambda(K) \leq \lambda(h^{-1}(K_n)) + \varepsilon_n$, whence $\lambda(K) \leq \lambda(h^{-1}(K))$. Consequently, $\lambda(U) \leq \lambda(h^{-1}(U))$ for every open set U in $[0,1]$. It now follows that $\lambda(K) = \lambda(h^{-1}(K))$ since $\lambda(h^{-1}([0,1])) \leq 1$.■

It remains to prove that if h is measure preserving then $h \in \mathscr{H}'$. The proof is a "pigeonhole" argument. That is, a partition Π of pigeons are to be assigned

[3] Here, h need not be bijective; in ergodic theory, "measure preserving" requires that h be bijective and both h and h^{-1} be measurable and measure preserving in our sense .

to a partition Π' of pigeonholes under certain rules φ. The final step of the proof of the characterization will depend on the following modified pigeonhole lemma.

Lemma 8.3.3 *Let $\Pi' = \{ I'_i \colon i = 1, 2, \ldots, n \}$ be a partition of $[0,1)$ and let $\{ K_i \colon i = 1, 2, \ldots, n \}$ be a disjoint collection of compact sets of $[0,1)$ such that $\lambda(I'_i) > \lambda(K_i)$ for each i. For all $\varepsilon > 0$, then there is a $\delta > 0$ such that for each partition $\Pi = \{ I_j \colon j = 1, 2, \ldots, m \}$ with $|\Pi| < \delta$ there is a permutation $\varphi \colon [0,1] \to [0,1]$ of Π such that, for each i,*

$$\lambda(I'_i) - 2|\Pi| > \lambda(\varphi(H_i)) > \lambda(\varphi(K_i)) - \varepsilon/n,$$

where $H_i := \bigcup_j \{ I_j \colon \varphi(I_j) \subset I'_i$ and $I_j \cap K_i \neq \emptyset \}$.

We should interpret the first inequality in the conclusion as saying that the intervals of the partition which intersect K_i approximate it well in measure, and the second inequality as saying that most of the points in these intervals are mapped to I'_i.

Proof. Let $\gamma > 0$ be small enough that $0 < 3\gamma < \lambda(I'_i) - \lambda(K_i)$ and $\lambda(U_i \setminus K_i) < \varepsilon/n$ for each i, where U_i is the γ-neighborhood of K_i, and such that 3γ is less than the minimum of the distances between distinct K_i's.

With $W(i, \Pi) = \bigcup \{ I_j \in \Pi \colon I_j \cap K_i \neq \emptyset \}$, observe that $\lambda(I'_i) - \lambda(W(i,\Pi)) + \lambda\big(W(i,\Pi) \setminus K_i\big) = \lambda(I'_i) - \lambda(K_i) > 3\gamma$. As $\lambda\big(W(i,\Pi) \setminus K_i\big) \to 0$ as $|\Pi| \to 0$, there is a δ such that $0 < \delta < \gamma$ and such that $\lambda\big(W(i,\Pi) \setminus K_i\big) < \gamma$ whenever $|\Pi| < \delta$. So, if $|\Pi| < \delta$, then

$$\lambda(I'_i) - 2|\Pi| > \textstyle\sum_j \{ \lambda(I_j) \colon I_j \in \Pi,\ I_j \cap K_i \neq \emptyset \}. \tag{8.1}$$

Let us construct the required φ. Let $|\Pi| < \delta$. Note that no interval I_j intersects more than one of the K_i. We separate the intervals I_j into the classes B_i, which consists of those intervals which intersect K_i, $1 \leq i \leq n$, and C, those intervals which do not intersect any K_i. (We distinguish between B_i and

$W(i,\Pi) = \bigcup_{B_i}$.) Denote $\sum\{\,\lambda(I_j)\colon I_j \in B_i\,\}$ by l_i and index the collection C as $J_1, J_2, \ldots, J_q$. Then

$$\textstyle\sum_{i=1}^{n} l_i + \sum_{k=1}^{q} \lambda(J_k) = 1.$$

Assume that the I_i' are numbered in increasing order from left to right, and call a_i the right endpoint of I_i'.

Let us describe the first step of the construction of φ. As $l_1 < a_1 - |\Pi|$ by (8.1), there is an m_1 such that $1 < m_1$ and $\sum_{k=1}^{m_1-1} \lambda(J_k) \le a_1 - l_1 < \sum_{k=1}^{m_1} \lambda(J_k)$. Define $a_1' = l_1 + \sum_{k=1}^{m_1} \lambda(J_k)$. Let φ be a permutation of $W(1,\Pi)$ onto $[0, l_1)$ and $\{\,J_k\colon k \le m_1\,\}$ onto $[l_1, a_1')$ with $a_1 \in \varphi(J_{m_1})$.

We now repeat this procedure for B_2. We have $a_1 < a_1' < a_1' + l_2 < a_2 - |\Pi|$, whence $[a_1', a_1' + l_2) \subset [a_1, a_2)$, and there is an m_2 such that $m_1 < m_2$ and $\sum_{k=m_1+1}^{m_2-1} \lambda(J_k) \le a_2 - (a_1' + l_2) < \sum_{k=m_1+1}^{m_2} \lambda(J_k)$. Define $a_2' = \sum_{i=1}^{2} l_i + \sum_{k=1}^{m_2} \lambda(J_k)$. Let φ be a permutation of B_2 onto $[a_1', a_1' + l_2)$ and $\{\,J_k\colon m_1 < k \le m_2\,\}$ onto $[a_1' + l_2, a_2')$ with $a_2 \in \varphi(J_{m_2})$.

This process continues up to the n^{th} stage, where $a_{n-1} \in \varphi(J_{m_{n-1}})$ and $a_{n-1}' = \sum_{i=1}^{n-1} l_i + \sum_{k=1}^{m_{n-1}} \lambda(J_k)$ satisfies $a_{n-1} < a_{n-1}' < a_{n-1}' + l_n < 1 - |\Pi|$. The remainder of the construction of φ is left to the reader.

As $W(i,\Pi) \setminus K_i \subset U_i \setminus K_i$, the construction is completed. ■

Proposition 8.3.4 *If $h\colon [0,1] \to [0,1]$ is measure preserving, then $h \in \mathscr{H}'$.*

Proof. Let $h\colon [0,1] \to [0,1]$ be measure preserving and for each m let Π_m' be a partition of $[0,1)$ such that $|\Pi_m'| < 2^{-m}$. Denote by n_m the number of intervals in Π_m'. Then $\{\,h^{-1}(I_{m,i}')\colon I_{m,i}' \in \Pi_m'\,\}$ and $h^{-1}(1)$ form a decomposition of $[0,1]$. For each $I_{m,i}'$, let $K_{m,i}$ be a compact subset of $h^{-1}(I_{m,i}')$ such that $\lambda\big(h^{-1}(I_{m,i}') \setminus K_{m,i}\big) < (n_m 2^m)^{-1}$.

For each Π_m' and $\varepsilon = 2^{-m}$, let δ_m be δ as provided by the modified pigeonhole lemma. We may assume $\delta_m > \delta_{m+1}$. For each m let

$$D_m = \textstyle\bigcup_{m \le m'} \bigcup\{\, h^{-1}(I_{m',i}') \setminus K_{m',i}\colon I_{m',i}' \in \Pi_{m'}'\,\}.$$

Note $D_{m'} \subset D_m$ whenever $m' \geq m$, and $\lambda(D_m) \leq 2^{-(m-1)}$. Also, if $x \notin D_m$ and $m' > m$, then x is in some $K_{m',i}$, whence $h(x) \in I'_{m',i} \in \Pi'_{m'}$.

Suppose that Π_k is a sequence of partitions such that $|\Pi_k| \to 0$ as $k \to 0$. Let k_m be the least k such that $|\Pi_{k'}| < \delta_m$ whenever $k' > k$. Observe that k_m is nondecreasing and converges to $+\infty$. If $k \leq k_1$, then let φ_k be the identity function. If $k_m < k \leq k_{m+1}$, then let φ_k be as given by the modified pigeonhole lemma for the partitions Π_k and Π'_m. Clearly φ_k, $k = 1, 2, \ldots$, is a well defined sequence.

The constructed sequence φ_k will converge almost everywhere to h. Indeed, let $\varepsilon > 0$ and let m be such that $2^{-(m-1)} < \varepsilon$. Suppose $x \notin D_m$ and $m' \geq m$. Let m'' and k be such that $k_{m'} = k_{m''} < k_{m''+1}$ and $k_{m''+1} \geq k > k_{m''}$. There is an $I_{k,j}$ in Π_k such that $x \in I_{k,j} \cap K_{m'',i}$ for some i. By the modified pigeonhole lemma, $\varphi_k(I_{k,j}) \subset I'_{m'',i} \in \Pi'_{m''}$. Hence $h(x)$ and $\varphi_k(x)$ are in the same $I'_{m'',i}$. As $\delta(\Pi'_{m''}) < 2^{-m''}$, $|\varphi_k(x) - h(x)| < 2^{-m''}$ whenever $k_{m''+1} \geq k > k_{m''}$. We infer from this that φ_k converges to h except on a subset of D_m. Hence the set in which φ_k does not converge to h has measure less than ε. ■

It now follows that the collection $\mathscr{G}$ of all nondecreasing, upper continuous functions $g\colon [0,1] \to [0,1]$ with $g(0) = \lim_{h \to 0^+} g(h)$ fulfills the requirements of the factorization. Indeed, for a Borel measurable function $f\colon [0,1] \to [0,1]$ let $g\colon [0,1] \to [0,1]$ be its nondecreasing, upper continuous distribution function (see the Remark at the end of the next section) that satisfies

$$g_{\#}\lambda([0,y]) = f_{\#}\lambda([0,y]), \quad y \in [0,1].$$

A measure-preserving h exists so that $f = g \circ h$ almost everywhere. To see that $\mathscr{G}$ is a minimal class, it is enough to observe that if two nondecreasing, upper continuous functions g_1 and g_2 are different, then they either differ only at 0, or on a set of positive measure. We have excluded the possibility that they differ only at 0 since the functions are left-continuous at 0. Hence there exists some

r such that

$$\lambda(\{\, x : g_1(x) > r \,\}) \neq \lambda(\{\, x : g_2(x) > r \,\}).$$

So there can be no measure-preserving function h such that $g_1 h = g_2$ almost everywhere. Consequently the following theorem yields a positive solution of the factorization problem.

Theorem 8.3.5 *The collection $\mathscr{G}$ of functions $g\colon [0,1] \to [0,1]$ that are nondecreasing and upper continuous is a minimal class having the property that each Borel measurable function $f\colon [0,1] \to (0,1)$ has corresponding functions $h\colon [0,1] \to [0,1]$ in $\mathscr{H}$ and g in $\mathscr{G}$ such that the composition gh is Lebesgue equivalent to f.*

Remark The construction of a nondecreasing function, called a *distribution function* or *monotone rearrangement*, was known to Hardy and Littlewood for measurable functions defined on the open interval $(0,1)$. That is, for each real-valued measurable function f on $(0,1)$, there corresponds a nondecreasing real-valued function g on $(0,1)$ that is upper continuous such that $\lambda(f^{-1}((0,y])) = \lambda(g^{-1}((0,y])))$ for every y (see [27, pages 91–92], [41, pages 29–30], and [38, page 272]). As $\mathbb{R}$ and $(0,1)$ are order isomorphic, there is no loss in assuming f and g map into $(0,1)$. Indeed, we infer from their result that each Lebesgue measurable function $f\colon [0,1] \to (0,1)$ corresponds to a nondecreasing, upper continuous distribution function $g\colon [0,1] \to [0,1]$ with $g^{-1}[\{0,1\}] \subset \{0,1\}$. Simply restrict f to the open interval $(0,1)$ and adjust the resulting Hardy-Littlewood distribution function that is defined on $(0,1)$ to the closed interval $[0,1]$ in the obvious way.

The factorization problem is posed in the context of almost everywhere convergence of a sequence of permutations to the function h. The following question remains.

Question. If h is measure preserving then does there exist a Lebesgue equivalent H such that H is the everywhere convergent limit of a sequence of permutations?

Clearly, such an H would be a Baire class 2 function.

8.4 Application to random Riemann sums

The motivation for defining the classes $\mathscr{H}$ and $\mathscr{H}'$ and for looking for the minimal class $\mathscr{G}$ that solves the factorization problem, comes from the study of random Riemann sums. If T is a tagged partition of $\mathbb{I}$, and h is some permutation of its underlying partition, than the Riemann sums of fh on T and f on $h^*(T)$ are the same, for any measurable f. Here $h^*(T)$ is defined by

$$h^*(T) := \{(h(t), h(I)) : (t, I) \in T\}.$$

This trivial observation can be extended using Theorem 8.3.5, in order that we can establish convergence of constructions similar to the random Riemann integral on some class of functions, just by examining a much more limited class.

Suppose that we have some sequence of tagged partitions, either deterministic or stochastic, and that we form for each function in L^1 the corresponding sequence of Riemann sums. Suppose further that this construction is linear and continuous considered as a function on L^1. In other words, we can disregard discrepancies of sufficiently small L^1 norm. Assume that the sequence of partitions has size tending to 0, either with certainty or almost surely. Then the following theorem suggests that we can exchange limits between Riemann sums of *rearrangements* of a function by permutations, and rearrangements of the Riemann sums on the same partitions (which in many cases might have no effect on the distribution of Riemann sums).

Theorem 8.4.1 *If f is a function in L^1, and $(h_n)_{n=1}^{\infty}$ is a sequence of measure-preserving bijections of $\mathbb{I}$ such that $h_n \to h$ a.e., then $f \circ h_n$ converges to $f \circ h$ in L^1.*

Proof. First of all it is enough to prove this for a sequence h_n which converges a.e. to the identity on $\mathbb{I}$. This is because we can replace convergence in L^1 for

Lebesgue measure for sequences tending to h with convergence in L^1 for the measure $\lambda_{\#}h$ for sequences tending to id.

Secondly by a common argument the result for f in L^1 will follow from the result for simple functions, which are dense in L^1. Therefore it will be enough to prove the statement for characteristic functions of measurable sets.

So suppose that $A \subset \mathbb{I}$ is a measurable set and that (h_n) is a sequence of bijective functions from $\mathbb{I}$ to $\mathbb{I}$, with the property that $\lambda(D) = \lambda(h_n(D))$ for all $n \in \mathbb{N}$, and that for almost every $x \in \mathbb{I}$ we have $h_n(x) \to x$. Let $\varepsilon > 0$ be given.

Choose some open set $E \supset A$ with $\lambda(E \backslash A) < \varepsilon$. If $x \in E$ and $x_n \to x$ then $f(x_n) = 1$ for all but finitely many values of n. Since $h_n \xrightarrow{a.e.} \mathrm{id}$, then $h_n(x) \to x$ for almost every $x \in E$. This means that there is some $n \in \mathbb{N}$ and some set $E' \subset E$ with $\lambda(E \backslash E') < \varepsilon$ such that $h_n(x) \in E$ for all x in E'.

So the measure of $\{x \in E' : h_n(x) \in A\}$ is at least $\lambda(E') - \varepsilon$, or $\lambda(E) - 2\varepsilon$, and the measure of $\{x \in E' \cap A : h_n(x) \in A\}$ is at least $\lambda(A) - 2\varepsilon$.

Therefore

$$\{x \in A : h_n(x) \notin A\} < 2\varepsilon$$

for all $n \geq n_0$ and so, since the image of that set has the same measure,

$$\int (f - f \circ h_n)^+ < 2\varepsilon.$$

An identical argument shows that $\int (f - f \circ h_n)^- < 2\varepsilon$ and so

$$\int |f - f \circ h_n| < 4\varepsilon$$

for all large enough n as required. ■

In many cases, including for example the weak random Riemann integral, the conditions required for this theorem to be useful are met. So we could prove that the weak random Riemann integral converges for all functions in L^1, provided that we know that it converges for all *decreasing* functions in L^1. Of course, this is not necessary to prove that the weak random Riemann integral

converges. However, the technique may allow further results to be proved for related integration procedures, such as those of chapters 7 and 6.[4]

Note this theorem is not directly applicable to questions of almost sure convergence as in the strong random Riemann integral. This is because L^1 convergence of $f \circ h_n$ to f does not imply almost sure convergence of the Riemann sums on the former to those on the latter. For this to be true we would have to obtain bounds on the speed of convergence in L^1 of rearrangements of f to f. For a variety of reasons this does not seem to be straightforward.

[4] The so-called distribution functions, which are shown in the previous sections, to be a minimal class of functions allowing rearrangement to all other functions under maps in $\mathscr{H}'$, were first constructed by Hardy and Littlewood in their paper [27]. As mentioned in Chapter 2, their interest in such functions was because they maximise certain functionals, such as the L^1 norm of the maximal operator of f. It seems at least plausible that it is possible to obtain using Hardy and Littlewood's work, similar results to those that we here suggest could follow from Theorem 8.4.1. In other words, if there exist counterexamples to some convergence property of Riemann sums, then there must be monotone counterexamples; and this is true because monotone functions can be rearranged to all other functions, or equivalently because functions with larger maximal operator have convergence properties at least as bad for these sums.

Bibliography

[1] B. Bongiorno, *Un nuovo integrale per il problema della primitive*, Le Matemiche **51** (1996), 299–313.

[2] ———, *On the C-integral*, Proceedings of the AMS Special Session on Nonabsolute Integration (2000), Available at `http://www.emis.de/proceedings/Toronto2000/`.

[3] ———, *On the minimal solution of the problem of primitives*, Journal of Mathematical Analysis and Applications **251** (2000), no. 2, 479 – 487.

[4] ———, *Notes on the first-return integral*, Real Analysis Exchange conference proceedings (2007), 45–47.

[5] ———, *On the first-return integrals*, J. Math. Anal. Appl. **333** (2007), 112–116.

[6] B. Bongiorno, L. Di Piazza, and D. Preiss, *A constructive minimal integral which includes Lebesgue integrable functions and derivatives*, J. London Math. Soc. **62** (2000), 117–126.

[7] D. Bongiorno, *Riemann-type definition of the improper integrals*, Czechoslovak Mathematical Journal **54** (2004), 717–725.

[8] A. Bruckner, R. Fleissner, and J. Foran, *The minimal integral which includes Lebesgue integrable functions and derivatives*, Colloq. Math. **50** (1986), 289–293.

[9] M. Csörnyei, U. B. Darji, M. J. Evans, and P. D. Humke, *First-return integrals*, J. Math. Anal. Appl. **305(2)** (2005), 546–559.

[10] M. Csörnyei, J. Grahl, and T. C. O'Neil, *Points of middle density in the real line*, submitted.

[11] U. B. Darji and M. J. Evans, *A first return examination of the Lebesgue integral*, Real Analysis Exchange **27** (2001/2002), 573–582.

[12] ———, *Functions not first-return integrable*, Journal of Mathematical Analysis and Applications **347** (2008), 381–390.

[13] U. B. Darji, M. J. Evans, C. Freiling, and R. J. O'Malley, *Fine properties of Baire one functions*, Fundamentae Mathematicae **155** (1998), 177–188.

[14] U. B. Darji, M. J. Evans, and R. J. O'Malley, *A first return characterization for Baire one functions*, Real Analysis Exchange **19(2)** (1993), 510–515.

[15] D. A. Darling, *On a class of problems related to the random division of an interval*, The Annals of Mathematical Statistics **24** (1953), 239–253.

[16] J. L. Doob, *Regularity properties of certain families of chance variables*, Transactions of the American Mathematical Society **47(3)** (1940), 455–486.

[17] M. J. Evans and P. D. Humke, *Almost everywhere first-return recovery*, Bulletin of the Polish Academy of Sciences, Mathematics **52(2)** (2004), 185–195.

[18] ———, *Almost every sequence integrates*, Acta Math. Hungar. **117** (2007), 35–39.

[19] D. H. Fremlin, *Notes on first-return integration*, available at `http://www.essex.ac.uk/maths/staff/fremlin/preprints.htm`, Version of 10.11.07.

[20] ———, *Problem GO*, available at `http://www.essex.ac.uk/maths/staff/fremlin/problems.htm`, Version of 14.10.10.

[21] ———, *Measure theory*, vol. 1, Torres Fremlin, 2000.

[22] ———, *Measure theory*, second ed., vol. 2, Torres Fremlin, 2010.

[23] R. A. Gordon, *The integrals of Lebesgue, Denjoy, Perron, and Henstock*, American Mathematical Society, 1994.

[24] J. Grahl, *A random approach to the Lebesgue integral*, J. Math. Anal. Appl. **340** (2008), 358–365.

[25] J. Grahl and T. Nishiura, *A factorization problem*, Real Analysis Exchange, accepted.

[26] A. Gut, *Probability: A graduate course*, Springer, 2005.

[27] G. H. Hardy and J. E. Littlewood, *A maximal theorem with function-theoretic applications*, Acta Math. **54** (1930), 81–116.

[28] R. Henstock, *Theory of integration*, Butterworths, London, 1963.

[29] C. S. Kahane, *Evaluating Lebesgue integrals as the limits of Riemann sums*, Math. Japonica **38** (1993), 1073–1076.

[30] J. C. Kieffer and Č. V. Stanojević, *The Lebesgue integral as the almost sure limit of random Riemann sums*, Proceedings of the American Mathematical Society **85** (1982), 389–392.

[31] V. I. Kolyada, *On the metric Darboux property*, Analysis Math. **9** (1983), 291–312, in Russian.

[32] O. Kurka, *Notes on exceptional densities*, preprint.

[33] ———, *Optimal quality of exceptional points of the Lebesgue density theorem*, preprint.

[34] E. J. McShane, *A unified theory of integration*, American Mathematical Monthly **80** (1973), 349–359.

[35] J. Mycielski, *Learning theorems*, unpublished note.

[36] W. F. Pfeffer, *The Riemann approach to integration*, Cambridge Univ. Press, 1993.

[37] A. R. Pruss, *Randomly sampled Riemann sums and complete convergence in the law of large numbers for a case without identical distribution*, Proceedings of the American Mathematical Society **124** (1996), 919–929.

[38] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton University Press, 1970.

[39] A. Szenes, *Exceptional points for Lebesgue's density theorem in the real line*, preprint.

[40] L. Peng Yee and P. Výborný, *The integral: An easy approach after Kurzweil and Henstock*, Cambridge University Press, 2000.

[41] A. Zygmund, *Trigonometric Series*, second ed., Cambridge University Press, Cambridge, 1977, (Combined volumes I and II).