FOUNDATIONS OF THE THEORY OF UNIFORM DISTRIBUTION

PETE L. CLARK

Introduction

The book *Uniform distribution of sequences* by Kuipers and Niederreiter, long out of print, has recently been made available again by Dover books.[1] I came across a copy at the Borders bookstore in San Francisco and decided to give it a try (the price, as they say, was right). It turned out to be full of interesting results, so I decided to take some notes on what seemed to me to be the parts most worthy of memory.

Uniform distribution – or, as it is now more often called, "equidistribution" – of various arithmetic, (and geometric, spectral,...) objects is a hot topic nowadays, and was the subject of a recent SMS-NATO conference in Montréal. It seems that there is not so much in the way of *foundational* material available online, and most contemporary treatments of equidistribution expose the foundations with such extreme brevity that mistakes can arise.[2] One notable exception is the 12 page article [**?**] based on Andrew Granville's opening lecture at the Montréal conference. Their article has the merits of (i) true brevity (it can be digested in one sitting leaving a full but not aching stomach) and (ii) having been written by two of the leading authorities of the day. In contrast [KN] possesses (ii), but – instead of (i) – (i'): in its 330 pages it gives a wonderfully careful and complete (especially if one reads the fine printed "notes") presentation of the foundations of the subject and some of the most important work up until about 1970. These notes, I hope, possess (i"): a treatment which is more systematic and complete than [**?**] but shorter than [KN] and presented in somewhat more modern language and style. Unfortunately I cannot claim (ii); *caveat emptor.*

A word on the organization and the style: in view of the fact that I recently taught an undergraduate analysis class that ended with a treatment of the Weierstrass approximation theorem, it was striking to me that the most basic results – and in particular, the nice application to the uniform distribution of the fractional parts of $n\alpha$ for irrational α – depend on nothing more advanced than this. So I consciously wrote §1 so as to be accessible to (presumably rather bright and inquisitive) students who have just taken such a course: in particular, I deliberately phrased everything so as to work in terms of Riemann rather than Lebesgue integrals. As

[1]In case you don't know, this is a company which publishes mathematical books at uniquely reasonable prices, the only catch being that they must have been written at least thirty or so years ago. For most of my mathematical life, their line of books seemed to be of interest primarily for amateurs, students and professionals in other fields, but in the last few years the number of books worthy of attention of research mathematicians seems to have spiked rather dramatically.

[2]In particular the definition of equidistribution given in [**?**, Chapter X] is incorrect.

I dug deeper into the theory, I learned that in fact the theory of equidistribution is closely linked to Riemann (and not Lebesgue!) integration. This connection is developed more fully than in most introductory treatments.

On the other hand, one of the charms of the subject is that it quite naturally draws upon a certain number of basic topics from measure theory, functional analysis, topology, groups and representation theory, and after §2 I have not shied away from presenting results in "proper generality." It would certainly be nice to balance out this rather Bourbakistic treament with a second part amassing examples of the sort of equidistributed sequences (and more general objects) one encounters in contemporary mathematics. Unfortunately I am by no means capable of doing so; fortunately the SMS-NATO proceedings address these disparate applications.

Some notation and terminology:

If $f,\ g : \mathbb{Z}^+ \to \mathbb{R}$, then by $f = O(g)$ we mean that there exists some C such that for all $n \in \mathbb{Z}^+$, $|f(n)| \leq C|g(n)|$. By $f = o(g)$ we mean that $g(n) \neq 0$ for sufficiently large n and $\lim_{n\to\infty} \frac{|f(n)|}{|g(n)|} = 0$). If f and g depend also on other quantities, then the "constant" C (resp. the convergence to zero) is *not* assumed to be uniform in these other quantities.[3]

$e(x) = e^{2\pi\sqrt{-1}x}$.

We write $\mathbb{E}$ to mean either $\mathbb{R}$ or $\mathbb{C}$ (n'importe quelle).

In §1, $\mathcal{C}$ denotes the $\mathbb{E}$-vector space of continuous functions $f : [0,1] \to \mathbb{E}$, an ($\mathbb{E}$)-Banach space for the supremum norm.

Following Bourbaki, we call a topological space for which each open cover admits a finite subcover quasi-compact; a compact space is a Hausdorff quasi-compact space.

1. Uniform distribution in $[0,1]$

1.1. The definition.

In this section, unless otherwise indicated all sequences $\{x_n\}_{n=1}^\infty$ will be sequences in the unit interval $I = [0,1]$. We say that $\mathbf{x}$ is **uniformly distributed** (often abbreviated **u.d.**) if for all $0 \leq a \leq b \leq 1$,

$$\lim_{N\to\infty} \frac{\#\{n \leq N \mid a \leq x_n \leq b\}}{N} = b - a. \tag{1}$$

This is a good starting point because it is concrete and reasonably perspicuous: we are requiring of a sequence that for every closed subinterval, the proportion of the first N elements lying in that subinterval should approach the length of the subinterval as N approaches ∞. On the other hand, some modern readers might see in its very concreteness a sort of puzzle, and try to figure out both "what it really means" – i.e., what are the fundamental properties of sequences in I, closed subintervals and their lengths that the definition purports to be relating, and how might the definition be extended to sequences in more general spaces? – and also

[3]Occasionally we may write something like $O_*(g)$, which is a rhetorical device: we are *emphasizing* the dependence on some auxiliary quantity $*$.

"whether it is strong enough" – e.g. should we perhaps be requiring a similar condition on more general subsets of I? Both of these questions will be addressed in due course.

There is a natural variant with, instead of a single infinite sequence $\mathbf{x}$, a double sequence of the following form: for each $n \in \mathbb{Z}^+$ we are given a finite sequence $\mathbf{x}_{n,i}$ in I of length ℓ_n (i.e., i ranges from 1 to ℓ_n). We can then define uniform distribution of $\mathbf{x}_{n,i}$ in the analogous way: for all $0 \leq a \leq b \leq 1$,

$$\lim_{n\to\infty} \frac{\#\{i \mid a \leq x_{n,i} \leq b\}}{\ell_n} = b - a.$$

This is closely related to the previous notion of uniform distribution but more general: given any infinite sequence $\mathbf{x}$, we can define a double sequence $\mathcal{D}(\mathbf{x})$ with $\ell_n = n$ by taking $x_{n,i} = x_i$ for $1 \leq i \leq n$, and then (tautologically) $\mathbf{x}$ is u.d. iff $\mathcal{D}(\mathbf{x})$ is. Note that for any given n the definition of u.d. of a double sequence does not depend upon the ordering of the terms of the finite sequence $x_{n,i}$, so an entirely equivalent notion is that of uniform distribution of a sequence S_n of finite *multisets*[4].

Exercise 1.1.1:
a) Suppose that S_n is a u.d. multiset sequence. Show that $\#S_n \to \infty$.[5]
b) Given a double sequence $x_{n,i}$, we can form an ordinary sequence $\mathcal{L}(x_{n,i})$ by lexicographic ordering: i.e., $x_{1,1}, \ldots, x_{1,\ell_1},\ x_{2,1},\ x_{2,\ell_2}, \ldots$. Show that, in general, the u.d. of the double sequence $x_{n,i}$ does not imply the u.d. of $\mathcal{L}(x_{n,i})$.
c) For a double sequence $x_{n,i}$, put $s_n = \sum_{i=1}^n \ell_i$. Suppose that $\ell_n = o(s_n)$ (e.g. $\ell_n = n$). Then $x_{n,i}$ u.d. implies $\mathcal{L}(x_{n,i})$ is u.d. Show moreover that this holds under all reorderings of the finite sequences $x_{n,i}$, so that we also get a criterion for passing from u.d. multiset sequences to u.d. sequences.

1.2. **First examples.** It is not quite trivial to show that *any* sequence is u.d. according to the definition. Perhaps the following is the simplest example.

Example 1.2.1: The multiset sequence $S_n = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ is u.d., since for all $0 \leq a \leq b \leq 1$, $\#S_n \cap [a,b] = n(b-a) + O(1) = n(b-a) + o(n)$. By Exercise 1, the associated concatenated sequence is u.d.:

$$0,\ 1,\ 0,\ \frac{1}{2},\ 1,\ 0,\ \frac{1}{3},\ \frac{2}{3},\ 1, \ldots$$

Observation 1. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two sequences such that the set*

$$\mathcal{D} = \mathcal{D}(\mathbf{x}, \mathbf{y}) = \{n \in \mathbb{Z}^+ \mid x_n \neq y_n\}$$

has density zero (i.e., $\lim_{N\to\infty} \frac{\#(\mathcal{D}\cap[1,N])}{N} = 0$*). Then* $\mathbf{x}$ *is u.d. iff* $\mathbf{y}$ *is.*[6]

Observation 2. *If* $\mathbf{x}$ *is u.d. in* I*, then it has dense image.*

[4]A multiset is the generalization of a set obtained by attaching a cardinal number to each element, its multiplicity. By a finite multiset we mean that there are finitely many distinct elements, each occurring with finite multiplicity.

[5]The cardinality of a multiset is defined in the obvious way, i.e., by taking the multiplicities into account. It is also true, however, that the number of distinct elements in S_n must tend to ∞.

[6]By an "observation" I mean an assertion whose statement is worth pointing out and whose truth the reader will have no trouble verifying unaided.

Observation 3. *The sets of u.d. sequences, of* non-*u.d. sequences, and of all sequences all have cardinality c.*

Whether a sequence has dense image does not depend upon the ordering of the terms. On the other hand, given a sequence with dense image, whether or not it is u.d. is *entirely* a matter of the ordering of the terms:

Theorem 4. *a) Any sequence has a rearrangement which is* not *u.d.*
b) If $\mathbf{x}$ *has dense image, it has a rearrangement which is u.d.*

Proof: a) If the sequence is not dense, *no* rearrangement is u.d. If it is dense, we can divide the index set $\mathbb{Z}^+$ into $I_1 = \mathbf{x}^{-1}([0,1/2])$ and $I_2 = \mathbf{x}^{-1}((1/2,1])$; the density implies that both of these sets are infinite. We can then rearrange the terms in such a way that $x_n \in [0,1/2]$ iff n is a multiple of 100 (say), and then the proportion lying in $[0,1/2]$ will be 99/100 instead of the desired 1/2.

b) Since $\mathbf{x}$ has dense image, a moment's thought shows that we may rearrange it so that $x_1 \in [0,\frac{1}{2}]$, $x_2 \in [\frac{1}{2},1]$, $x_3 \in [0,\frac{1}{3}]$, $x_4 \in [\frac{1}{3},\frac{2}{3}]$, and so on. This sequence is u.d. by the argument of Example 1.

Exercise 1.2.1: Show that if $\mathbf{x}$ is dense, we can in fact rearrange the terms such that for every $a < b$, in (1) we have that the lower limit is 0 and the upper limit is 1.

So at this point we have constructed "plenty" of u.d. sequences, all, however, of a rather contrived and uninteresting type.

1.3. **The Weyl Criterion.** As for e.g. irrationaity or transcendence, it is quite another matter if we ask about the uniform distribution of a pre-existing sequence.

Example 1.3.1: What about the sequence $\{n\sqrt{2}\}$ – here $\{x\} = x - \lfloor x \rfloor$ is the fractional part of the real number x? Even to show that for irrational α, $\{n\alpha\}$ is *dense* in the unit interval is a not entirely trivial result[7], due originally to Kronecker.

To answer this question, and many others, about u.d., the following is invaluable.

Theorem 5. *(Weyl Criterion) The sequence* $\mathbf{x} = \{x_n\}$ *is u.d. iff: for all* $0 \neq h \in \mathbb{Z}$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e(hx_n) = 0. \tag{2}$$

Remark: For $h = 0$ we have $\frac{1}{N}\sum_{n=1}^{N} e(hx_n) = 1$ for all sequences $\mathbf{x}$ and all N.

The ubiquitous first application is a refinement of Kronecker's theorem:

Corollary 6. *("The" equidistribution theorem) For* $\alpha \in \mathbb{R}$, *TFAE:*
(i) The sequence of fractional parts $\{n\alpha\}$ *is u.d.*
(ii) $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

[7] At least if we take a sufficiently elementary perpective. On the other hand a Lie theorist would probably regard it as trivial. More on this later.

Proof: That (i) $\implies$ (ii) is immediate from Observation 2. Conversely, the irrationality of α gives that for all $0 \neq h$, $e(h\alpha) \neq 1$, so

$$\left|\frac{1}{N}\sum_{n=1}^{N} e(hn\alpha)\right| = \left|\frac{1}{N} \cdot \frac{e(h(N+2)\alpha) - e(h\alpha)}{1 - e(h\alpha)}\right| \leq \frac{2}{N \cdot (1 - e(h\alpha))} \to 0.$$

The intriguing content of Weyl's criterion is that the problem of uniform distribution of a sequence is equivalent to one of cancellation in exponential sums, a topic whose importance in (analytic and algebraic) number theory and harmonic analysis could hardly be overstated. The idea here is that for all sequences $\mathbf{x}$ and all $h \in \mathbb{Z}$ we have that $|\frac{1}{N}\sum_{n=1}^{N} e(hx_n)| \leq 1$. This is the *trivial bound*, and it is sharp iff $h = 0$ or $x_1 = \ldots = x_N$. Otherwise the x_n's will be pointing in different directions, and like forces acting on us from all sides, will at least partially cancel each other out. Weyl's criterion asks simply for qualitative improvement: that the sum be $o_h(1)$ rather than just $O(1)$.[8]

Ideally, we would develop the theory in a way so as to make it completely "self-evident": in other words, we strive the illusion that we could have discovered all the main results ourselves. However, things will go much more smoothly if we allow ourselves a few well-chosen *hints*: i.e., every so often an oracle speaks to us to push us in the right direction. Here is a good example: it seems not to be at all obvious how to prove Weyl's criterion, until we receive the following

Hint 1. *For a subset $E \subset [0,1]$, consider $\chi_E : [0,1] \to \mathbb{R}$, the characteristic function of E. Then (1) says precisely that*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(x_n) = \int_0^1 f \tag{3}$$

holds with $f = \chi_{[a,b]}$.

Aha! This suggests the following important result.

Theorem 7. *(Fundamental Theorem) For a sequence $\mathbf{x} = \{x_n\}$ in $[0,1]$, the following are equivalent:*
a) (3) holds for all Riemann integrable $f : [0,1] \to \mathbb{E}$.
b) (3) holds for all continuous $f : [0,1] \to \mathbb{E}$.
c) $\mathbf{x}$ is uniformly distributed on $[0,1]$.

The following notation will be helpful in the proof:

Let $\mathcal{R}$ be the linear space of Riemann-integrable functions $f : [0,1] \to \mathbb{E}$;

$\mathcal{C}$ the linear space of continuous $\mathbb{E}$-valued functions on $[0,1]$;

$\mathcal{S}$ the $\mathbb{E}$-span of the characteristic functions of closed intervals ("step functions").

Proof: If (3) holds for all f in a subset S of $\mathcal{R}$, then by linearity it holds also for its $\mathbb{E}$-span $\langle S \rangle$. Therefore we must show that (3) holds for all $f \in \mathcal{R}$ iff it holds for all $f \in \mathcal{C}$ iff it holds for all $f \in \mathcal{S}$.

Since $\mathcal{C} \subset \mathcal{R}$, (a) implies (b).

[8] Note that the requirement that the sum approach 0 *uniformly* in h would be much stronger.

$(b) \implies (c)$: Given a step function f and $\epsilon > 0$, there are continuous functions g_1 and g_2 such that $g_1 \leq f \leq g_2$ on I and $\int_0^1 (g_2 - g_1) < \epsilon$. Then:

$$\left(\int_0^1 f\right) - \epsilon \leq \left(\int_0^1 g_2\right) - \epsilon \leq \int_0^1 g_1 = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N g_1(x_n)$$
$$\leq \liminf_N \frac{1}{N}\sum_{n=1}^N f(x_i) \leq \limsup_N \frac{1}{N}\sum_{n=1}^N f(x_n) \leq \limsup_N \frac{1}{N}\sum_{n=1}^N g_2(x_n)$$
$$= \int_0^1 g_2 \leq \int_0^1 g_1 + \epsilon \leq \int_0^1 f + \epsilon.$$

$(c) \implies (a)$: if f is Riemann integrable, there exist two step functions $g_1,\ g_2$ with $g_1 \leq f \leq g_2$ on I and $\int (g_2 - g_1) \leq \epsilon$. The result follows by exactly the same inequalities as in the previous paragraph.

This result is "part one" of the foundation of the "Monte Carlo" approach to integration. It remains also to formalize and prove the fact that a randomly chosen sequence is u.d. ("part two", coming up) and to develop a *quantitative* theory leading to explicit upper bounds on the error ("part three", to be almost completely ignored in these notes).

Exercise 1.3.1: Check that the proof of Theorem 7 works to give analogous results for double sequences $\mathbf{x}_{n,i}$ and multiset sequences S_n.

We are now ready to prove the Weyl criterion: let us view $\mathcal{R}$ as a (complete) normed linear space, with $||f|| = \sup_{x\in[0,1]} |f(x)|$. For a given sequence $\mathbf{x}$, we can define $\mathbb{E}$-linear functionals on $\mathcal{R}$

$$F_N : f \mapsto \frac{1}{N}\sum_{n=1}^N f(x_i),\ G : f \mapsto \int_0^1 f.$$

These functionals have norm one: for all $f \in \mathcal{R}$, $|F_n(f)| \leq ||f||$ with equality for $f \equiv 1$ (and the same for G). It follows that if $F_N(f) \to G(f)$ holds for all f in some subset S of $\mathcal{R}$, then it holds also on the closed linear span $\overline{\langle S\rangle}$ of S. By the Weierstrass approximation theorem, we may take for instance the polynomial functions $f : [0,1] \to \mathbb{R}$, or indeed the sequence of monomials x^n. Just by separating real and imaginary parts, it follows immediately from this that the set of polynomials with $\mathbb{C}$-coefficients is dense in the space of $\mathbb{C}$-valued continuous functions. Composing with $e_1 = e^{2\pi i x}$, it follows that every continuous complex valued function $f : [0,1] \to \mathbb{C}$ with $f(0) = f(1)$ is a uniform limit of trigonometric polynomials – i.e., expressions of the form $\sum_{n\in\mathbb{Z}} \alpha_n e^{nix}$ with $\alpha_n = 0$ for all but finitely many n (Fejer's Theorem). This is almost the Weyl criterion, except that we get (3) for "periodic" step functions, i.e., f with $f(0) = f(1)$. However, if x is u.d. then $x_n = 1$ for only $o(n)$ terms of the sequence (apply the definition with $a = b = 1$), and this means that for an arbitrary step function f and $\epsilon > 0$, for all sufficiently large N there exists a periodic step function g such that $|F_N(f - g)|,\ |G(f - g)| < \epsilon$. Thus the Weyl criterion is necessary and sufficient for uniform distribution.

Exercise 1.3.2: Verify the analogue of the Weyl criterion for double sequences and multiset sequences.

Exercise 1.3.3: Show that if, in fact, (3) holds for all $h \in \mathbb{Z}^+$, then $\mathbf{x}$ is u.d.

Exercise 1.3.4: If $\mathbf{x}$ is u.d. and $\alpha \in \mathbb{Z} \setminus 0$, then $\alpha\mathbf{x}$ is u.d.

The proof of Weyl's criterion suggests that things will be equivalent but slightly easier if we consider instead uniform distribution on $S^1 = [0,1]/(0 \sim 1)$. For the remainder of this section – when the group structure of S^1 will not be used – we will stubbornly stick with $[0,1]$. However, in giving examples of $\mathbf{x} \in [0,1)$, we will allow sequences of real numbers, with the understanding that we are really speaking of the associates sequence $\{x_n\}$ of fractional parts (contained in $[0,1)$).

By calling Theorem 7 "fundamental" we are imparting a slightly nonstandard spin on matters; it is the Weyl criterion which is usually held to be paramount. We give the following justifications: (i) the Weyl criterion is readily deduced as a corollary of Theorem 7, but the converse is not true; (ii) Theorem 7 is any case also due to Weyl; and (iii) as we shall see, Theorem 7 generalizes nicely to any compact space, whereas the Weyl criterion has an analogue only on compact groups.

1.4. **Uniform distribution and Riemann integrability.**

Perhaps the reader noticed that the property of Riemann integrable functions used in the proof $c) \implies a)$ is in fact a characteristic property, by a famous theorem of Darboux. This leads us to suspect the following relation betweeen uniform distribution and Riemann integrability, proved by de Bruijn and Post in 1968, nearly fifty years after Weyl's Theorem 7.

Theorem 8. *Let $f : [0,1] \to \mathbb{R}$ be any function, and L a real number. TFAE:*
a) For all u.d. sequences $\mathbf{x}$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{n} f(x_n) = L$.
a') For all u.d. double sequences $x_{n,i}$, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{\ell_n} f(x_n) = L$.
b) f is Riemann integrable to L.

It seemed more fun[9] to try to figure out the proof for myself than track down the reference (apparently other proofs have subsequently appeared in the literature). So I don't know if the following proof is "new."

Proof: Theorem 7 and its variant for double sequences shows $b) \implies a)$ and $b) \implies a')$.

All of the Riemann sums intervening in the proof will be with respect to $\mathcal{P}_n$, the uniform partition of $[0,1]$ into n subintervals, for some n, so let us employ the simplified notation $R_n(f, x_{n,i})$ for $R(f, \mathcal{P}_n, x_{n,i})$. We will use the fact that a function f is Riemann integrable to L iff: for any sequence $x_{n,i}$ of taggings of $\mathcal{P}_n$ (i.e., $x_{n,i} \in [\frac{i-1}{n}, \frac{i}{n}]$) the Riemann sums $R_n(f, x_{n,i})$ approach L. (For instance, one can check that this implies equality of the upper and lower Darboux integrals.)

This remark makes the proof that a$'$) implies b) easy: indeed for any function g,

$$R_n(g, x_{n,i}) = \frac{1}{n}\sum_{i=1}^{n} g(x_i),$$

[9]For me, at least. Perhaps apologies are owed to the reader.

so it follows from Theorem 7 that any sequence of taggings $x_{n,i}$ of $\mathcal{P}_n$ is u.d. (note that this generalizes Example 1.2.1). So in a') we have assumed, in particular, the convergence of all sequences $R_n(f, x_{n,i})$ to a common value L, which gives the Riemann integrability of f to L.

The proof that a) implies b) is morally the same but with a few unpleasant technicalities: given any sequence of taggings $x_{n,i}$ as above, we can concatenate lexicographically to get a sequence $\mathbf{x}$ which, by the previous paragraph and Exercise 1.1.1, is u.d. We would like to argue that if a) holds for all these sequences, f is Riemann integrable to L. This is true, but not as obvious, since the sum involved is no longer a Riemann sum but rather (has a subsequence which is) a weighted average of Riemann sums. One thing one learns in undergraduate real analysis[10] is that it is much easier to show that a function is Darboux integrable, so let us do this instead. In other words, define

$$L_n(f) = \frac{1}{n}\sum_{i=1}^{n} \inf f|_{[\frac{i-1}{n}, \frac{i}{n}]},$$

$$U_n(f) = \frac{1}{n}\sum_{i=1}^{n} \sup f|_{[\frac{i-1}{n}, \frac{i}{n}]},$$

and it suffices to show that $\inf_n \Omega_n(f) = U_n(f) - L_n(f) = 0$.

First we must convince ourselves that $\Omega_n(f) < \infty$ for all n, i.e., that f is bounded. If not, assume WLOG that f is unbounded above, so that for each n there is some point y_n with $f(y_n) \geq 2^n$. Let x_n be any u.d. sequence, so by assumption $\frac{1}{N}\sum_{n=1}^{n} f(x_n) \to L$; in particular

$$\sum_{n=1}^{N^2} f(x_n) = O(N^2). \tag{4}$$

If we then modify the sequence x_n by taking $x_{n^2} = y_n$, then since we have changed it only on a set of density zero it remains u.d. But clearly (4) does not hold on our modified sequence, a contradiction.

Now fix $\epsilon > 0$ and let us choose two double sequences $x'_{n,i}$, $x''_{n,i}$ (in linear order) so that the corresponding Riemann sums $R_n(f, x'_n)$ and $R_n(f, x''_n)$ are extremely close to $L_n(f)$ and $U_n(f)$. We cannot, in general, make them equal because f need not attain its subinterval infima and suprema; however, for each n and $1 \leq i \leq n$, we can choose the tagging point $x'_{n,i}$ (resp. $x''_{n,i}$) to be within ϵ of the infimum (resp. supremum) of f on $[\frac{i-1}{n}, \frac{i}{n}]$, and then we will have

$$R_n(f, x'_n) \leq L_n(f) + \epsilon,$$

$$R_n(f, x''_n) \geq U_n(f) - \epsilon.$$

By assumption, the sequence $\frac{1}{N}\sum_{n=1}^{N} f(x''_n) - f(x'_n)$ converges to zero, hence so does the subsequence obtained by restricting to N of the form $1 + \ldots + M = \frac{M(M+1)}{2}$. So for all sufficiently large M we have

$$\epsilon > \frac{1}{1 + \ldots + M}\sum_{n=1}^{1+\ldots+M} f(x''_n) - f(x'_n) = \sum_{n=1}^{M} \frac{n}{1 + \ldots + M}\left(R_n(f, x''_n) - R_n(f, x'_n)\right)$$

[10] One learns it especially well by *teaching* real analysis.

$$\geq \sum_{n=1}^{M} \frac{n}{M(M+1)/2}(\Omega_n(f) - 2\epsilon).$$

If $\Omega_n(f)$ were greater than 3ϵ for all n, the final expression would exceed ϵ, a contradiction. Since ϵ was arbitrary, this shows that $\inf_n \Omega_n(f) = 0$ so f is Riemann integrable (to L, clearly). This completes the proof.

To my knowledge this result is only of "philosophical" importance, but it nevertheless seems intriguing: it suggests that, as soon as we can formulate a notion of uniform distribution in an abstract setting, we will then get at least a candidate for the notion of an "abstract Riemann integrable function."

1.5. **"Linearity" properties of u.d. sequences.** Let $\mathbf{x}$, $\mathbf{y}$ be sequences (taken (mod 1) as usual), and $\alpha \in \mathbb{R}^{\times}$.

Nonexample 1.5.1: Take $x_n = n\sqrt{2}$, $y_n = -n\sqrt{2}$. Then $\mathbf{x}$ and $\mathbf{y}$ are u.d., but $\mathbf{x} + \mathbf{y}$ is not.

Nonexample 1.5.2: Take $x_n = n\sqrt{2}$, $\alpha = \frac{1}{\sqrt{2}}$. Then $\mathbf{x}$ is u.d. but $\alpha\mathbf{x}$ is not.

Thus the set of all sequences of real numbers which are u.d. (mod 1) is not an $\mathbb{R}$-subspace of the space of all real sequences. Nevertheless there is something to the idea, as the following two results show:

Proposition 9. *If $\mathbf{x}$ is u.d. and $\mathbf{y}$ (mod 1) is convergent, then $\mathbf{x} + \mathbf{y}$ is u.d.*

Proof: Let us first do the special case in which $y_n = y$ is a constant sequence. From the right viewpoint (i.e., that of Haar measure on S^1) the result is obvious, but it is no trouble to give a direct proof: note that as $x \mapsto f(x)$ ranges through all continuous functions on S^1, so does $x \mapsto f(y+x)$, and apply Theorem 7. To tackle the general case we may now perform a translation and hence assume $\mathbf{y} \to 0$. Let $f : S^1 \to \mathbb{E}$ be continuous; since S^1 is compact, f is uniformly continuous. So given $\epsilon > 0$, for all $N \geq N_0$, $|f(x_n + y_n) - f(x_n)| < \epsilon$ and $|\sum_{n=1}^{N} f(x_n) - \int_0^1 f| < \epsilon$, so

$$|\frac{1}{N}\sum_{n=1}^{N} f(x_n + y_n) - \int_0^1 f| \leq |\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f| + |\frac{1}{N}\sum_{n=1}^{N} f(x_n + y_n) - f(x_n)|$$

$$\leq \epsilon + \frac{N_0||f||}{N} + \left(\frac{N - N_0}{N}\right)\epsilon \to 0$$

as $N \to \infty$, completing the proof.

Theorem 10. *(Weyl) If $\mathbf{x}$ is a sequence of distinct* integers*, then the set of real numbers α such that $\alpha\mathbf{x}$ is* not *u.d. has measure zero.*

Proof: It is easy to see that it suffices to look at $\alpha \in [0, 1)$. Now, for fixed $0 \neq h \in \mathbb{Z}$, define

$$S_h(N, \alpha) = \frac{1}{N}\sum_{n=1}^{N} e(hx_n\alpha).$$

Then

$$|S_h(N, \alpha)|^2 = S_h(N, \alpha)\overline{S_h(N, \alpha)} = \frac{1}{N^2}\sum_{m,n=1}^{N} e(h(x_m - x_n)\alpha),$$

so

$$\int_0^1 |S_h(N,\alpha)|^2 d\alpha = \frac{1}{N^2}\sum_{m,n=1}^{N}\int_0^1 e(h(x_m - x_n)\alpha)d\alpha = \frac{1}{N}.$$

So

$$\int_0^1 \sum_{N=1}^{\infty} |S_h(N^2),\alpha)|^2 d\alpha \leq \sum_{N=1}^{\infty}\int_0^1 |S_h(N^2,\alpha)|^2 d\alpha = \sum_N \frac{1}{N^2} < \infty,$$

where the first inequality holds by Fatou's Lemma. It follows that $\sum_N |S_h(N^2,\alpha)|^2$ is finite for almost all $\alpha \in [0,1]$ and *a fortiori* that $\lim_{N\to\infty} S_h(N^2,\alpha) = 0$ for almost all α. For fixed N, let m be the positive integer such that $N \in [m^2,(m+1)^2)$. Trivial estimates give:

$$|S_h(N,\alpha)| \leq |S_h(m^2,\alpha)| + \frac{2m}{N} \leq |S_h(m^2,\alpha)| + \frac{2}{\sqrt{N}}.$$

We therefore get that $S_h(N,\alpha) \to 0$ for almost all α. Since the nonzero integers form a countable set, we are done by Weyl's criterion.

Remark: Here some of Lebesgue's integration theory has snuck in. It is not clear that this is necessary – one can certainly define "measure zero" without defining "measure" – and I would be inclined to doubt that Weyl's 1914 proof uses Fatou's Lemma. Perhaps someone can suggest a more elementary argument.

Using similar methods, one can show:

Theorem 11. *(*[KN, Cor. 1.4.3]*) Let* $\mathbf{x}$ *be a sequence of real numbers such that* $\inf_{m\neq n} \lambda_m - \lambda_n > 0$. *Then* $\alpha\mathbf{x}$ *is u.d.* (mod 1) *for almost every real number* α.

On the other hand, there is the following result of Dress:

Theorem 12. *(Dress) If* $\mathbf{x}$ *is a nondecreasing sequence with* $x_n = o(\log n)$, *then for* no *real* α *is* $\alpha\mathbf{x}$ *u.d.*

I was not able to find in [KN] an answer to the following

Question 1. *If* $\mathbf{x}$ *is u.d., then is it true that for almost every* α, $\alpha\mathbf{x}$ *is u.d.?*

1.6. **The difference lemma.**

Theorem 13. *(Fejér) Let* $\mathbf{x}$ *be a sequence of real numbers such that* $(\Delta\mathbf{x})_n := x_{n+1} - x_n$ *is monotone. If, further,*

$$\lim_{n\to\infty} (\Delta\mathbf{x})_n = 0, \ \lim_{n\to\infty} n(\Delta\mathbf{x})_n = \infty,$$

then $\mathbf{x}$ *is u.d.*

Corollary 14. *(Fejér) Let* $f : [1,\infty) \to \mathbb{R}$, *differentiable for sufficiently large* x. *If* $f'(x)$ *tends monotonically to* 0 *as* $x \to \infty$ *and* $\lim_{x\to\infty} x|f'(x)| = \infty$, *then* $\{f(n)\}$ *is u.d.*

Theorem 15. *(Van der Corput Difference Theorem) If for each* $h \in \mathbb{Z}^+$, $\{x_{n+h} - x_n\}$ *is u.d., then* $\mathbf{x}$ *is u.d.*

As a consequence, we get what is perhaps the single most interesting "classical" result.

Theorem 16. *Let $P(x) = a_n x^n + \ldots + a_0 \in \mathbb{R}[x]$ be a degree n polynomial with real coefficients. TFAE:*
a) There exists $i > 0$ such that a_i is irrational.
b) $\{P(n)\}$ is u.d.

Proofs to be added, perhaps by someone else.

1.7. **Discrepancy.** There is a rich *quantitative* theory of u.d.; it is this part of the theory that has the clearest connections with "classical number theory" and especially with Diophantine approximation. We will not even attempt to survey this aspect of the theory, except to give one basic definition and one remarkable theorem.

Definition: The *discrepancy* of a finite sequence $x_1, \ldots, x_n$ (still in $[0,1]$) is defined to be

$$D(x_1, \ldots, x_n) = \sup_{0 \leq a \leq b \leq 1} |\frac{\#\{i \,||a \leq x_i \leq b\}}{n} - (b-a)|.$$

Given an infinite sequence $\mathbf{x}$ we define for each n, $D_n(\mathbf{x}) = D(x_1, \ldots, x_n)$.

Proposition 17. *(*[KN, Thm 2.1.1]*) A sequence $\mathbf{x}$ is u.d. iff $\lim_{n\to\infty} D_n(\mathbf{x}) = 0$.*

The fact that $D_n(\mathbf{x}) \to 0$ implies u.d. is obvious; the other direction is not, and indeed expresses the fact that the uniformity over all subintervals $[a,b]$ of the convergence of the expression in (1) is automatic. (Given the compactness of $[0,1]$, this is not too surprising, however.)

We can use the discrepancy to quantify the idea that some sequences are more uniformly distributed than others.

One ought not even to define the discrepancy without citing the following result:

Theorem 18. *(Erdos-Turan) For any sequence $\mathbf{x}$ and any positive integers N and m we have*

$$D_N(\mathbf{x}) \leq \frac{6}{m+1} + \frac{4}{\pi} \sum_{h=1}^{m} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e(hx_n) \right|.$$

This is a spectacular sharpening of the sufficiency of Weyl's criterion for u.d. (Conversely, given the necessity of Weyl's criterion, we recover Proposition 17.)

We will refer the reader to Chapter 2 of [KN] for the development of this theory, and in particular §2.5 which discusses applications to **Monte Carlo** integration.

Interlude: Where to go from here?

At this point we have presented what we understand to be a fairly complete picture of the classical qualitative theory of uniform distribution. In some sense it is parallel to the elementary theory of convergence of infinite series: our basic problem is to resolve a dichotomy (convergent/divergent) or (u.d./not u.d.) We have some general criteria that work at least for certain important classes of examples; nevertheless, when presented with a specific, and not overly simple, example, we may be left to our own devices and not able to resolve the question. For instance, it is not known whether $\{e^n\}$ is uniformly distributed $\pmod 1$.

Where do we go from here?

There are first of all some obvious questions:

Question 2. *Are* most *sequences* $\mathbf{x}$ *u.d. mod* 1*?*

We must, of course, define what we mean by "most." Apart from the (often trivial, as here) sense of cardinality, there most common interpretations across mathematics are in the sense of measure and in the sense of Baire category. Both will be considered.

Question 3. *How do we extend the notion of uniform distribution to sequences in more general spaces?*

If by a more general space we meant only the hypercubes $[0,1]^d$ or the tori $(S^1)^d$, then it is fairly clear how one would try to do it (say in $[0,1]^d$): we define uniform distribution in terms of volumes of rectangular subboxes. One finds without any trouble that the fundamental theorem holds verbatim. To find the analogue of the Weyl criterion it helps to have had some exposure to Fourier analysis on $(S^1)^d$: in place of the scalar exponential functions we use all functions of the form $v = (x^1, \ldots, x^d) \mapsto e(v \cdot h)$, where $h = (h^1, \ldots, h^d) \in \mathbb{Z}^d$, i.e., the characters of $(S^1)^n$.

If we want to talk about uniform distribution on, say, a compact real (or p-adic!) manifold, it is clear that some further data is needed: e.g. if φ is a homeomorphism of $[0,1]$ – then the u.d. of a sequence $\mathbf{x}$ obviously does not imply that of $\varphi(\mathbf{x})$.

Question 4. *What about sequences which are asymptotically distributed according to a non-linear distribution function?*

In other words, if $g : [0,1] \to [0,1]$ is a monotone function with $g(0) = 0$ and $g(1) = 1$, we could define a sequence $\mathbf{x}$ as being asymptotically g-distributed if for all $0 \leq a \leq b \leq 1$ we have

$$\lim_{N \to \infty} \frac{\#\{n \leq N \mid a \leq x_n \leq b\}}{N} = g(b) - g(a);$$

of course we recover uniform distribution by taking $g(x) = x$. In most modern number-theoretic applications (e.g. random matrices, spectral graph theory) the distribution function of interest is *not* $g(x) = x$.

If we were to continue our push to keep things as elementary as possible, we could develop a theory of g-distributed sequences by using, in place of the Riemann integral, the Riemann-Stieltjes integral dg. But here (for me at least) the elementary approach loses its charm: rather than fussing about what happens at the points of discontinuity of g, why not just work with any measure μ on $[0,1]$ which is absolutely continuous with respect to Lebesgue measure and thus recover g as the Radon-Nikodym derivative $d\mu/dx$? It seems that Question (3) is virtually begging for us to consider a measure space (X, μ), and once we do this, studying g-distribution on $[0,1]$ becomes the special case of $([0,1], gdx)$.

2. Uniform distribution in a compact measure space

2.1. **First steps.** So let us start anew and try to build a theory of uniform distribution in a measure space $(X, \mathcal{A}, \mu)$ – here $\mathcal{A}$ is a σ-algebra which is the "domain" of μ. We would like to work in as much generality as possible, and what we require of our theory is that it at least recovers the theory of §1 in the special case of Lebesgue measure on $[0, 1]$. Moreover it would be nice to have an analogue of the Weyl criterion and/or the "fundamental" Theorem 7. This turns out to be less straightforward (and hence more interesting) than one might think.

To fix ideas, let us start with the case of a regular Borel measure μ on a compact space X. In particular, let us always take the σ-algebra in question to be the Borel algebra, i.e., the one generated by the open subsets. (Afterwards, we can think about how to extend to a locally compact space.) Let us write $\mathcal{B} = \mathcal{B}(X)$ for the Banach algebra of $\mathbb{R}$-valued Borel measurable functions, with the supremum norm. Write $\mathcal{C} = \mathcal{C}(X)$ for the closed subalgebra of continuous functions.

As a first attempt, given a sequence $\mathbf{x}$ in X, suppose we define it to be u.d. if for each Borel subset A of X,

$$\lim_{N\to\infty} \frac{\#\{1 \leq i \leq N \mid x_i \in A\}}{N} = \mu(A). \tag{5}$$

First of all, since $\limsup_N$ of the left-hand expression is clearly at most one, we must evidently be thinking that μ is a *probability* measure, i.e., $\mu(X) = 1$. Let us make this assumption.

Unfortunately this does not work: the given condition does not hold for all Borel subsets of $[0, 1]$. For instance it would tell us that the density of the set of n for which x_n is rational is zero, but we saw that there are uniformly distributed sequences consisting entirely of rational numbers.

In (3) the limit is only required to exist when A is a closed subinterval of $[0, 1]$, but it is less than clear what the analogue of a closed interval in our arbitrary compact space X should be, so our next guess is to require that (5) hold for all closed subsets A. Does this work?

No again. There exist nowhere dense closed subsets C of $[0, 1]$ with any given Lebesgue measure $0 < \lambda < 1$ – e.g., consider a set which is constructed like the Cantor set except instead of removing the middle $\frac{1}{3}$ from each line segment we remove a centered line segment whose length is α percent of the total length for a suitably chosen $\alpha < \frac{1}{3}$. Then the complement U of C is dense so there exists a uniformly distributed sequence with image contained in U, so that taking $A = C$ in (5) the left-hand side is zero and the right hand side is λ. The same example, of course, shows that we cannot require (5) to hold for all open sets.[11] What then is the nice property of closed intervals that made the theory work?

Hint 2. *Define a subset A of X to be a* Jordan set *if it is a Borel set with $\mu(\partial A) = 0$. (Recall $\partial A = \overline{A} \cap \overline{X \setminus A}$.)*

[11]Unfortunately this is the definition suggested by Iwaniec and Kowalski.

Exercise X.X: a) Show that the Jordan sets form an algebra of sets, which we will denote $\mathcal{J}$, and call the Jordan algebra.[12] Show that $\mathcal{J}$ is, in general, not a σ-algebra.
b) When $X = [0,1]$, show that the algebra generated by the closed subintervals is a proper subalgebra of the Jordan algebra.
c) When $X = [0,1]$, show that Jordan sets are precisely the sets whose characteristic functions are Riemann-integrable.

Exercise X.X: Show that every nonempty open subset U contains a nonempty Jordan open subset.

2.2. The Fundamental Theorem.

Theorem 19. *(Abstract Fundamental Theorem) For a sequence* $\mathbf{x}$ *in* X*, TFAE:*
a) For all $f \in \mathcal{C}$*,*

$$\frac{1}{N}\sum_{n=1}^{N} f(x_n) \to \int_X f d\mu. \tag{6}$$

b) We have

$$\frac{\#\{1 \leq n \leq N \mid x_n \in A\}}{N} \to \mu(A) \tag{7}$$

for all Jordan sets A*.*
c) Equation (7) holds for all closed Jordan sets.

It is therefore reasonable to define a sequence $\mathbf{x}$ in X to be μ-**equidistributed** if it satisfies the equivalent conditions of Theorem 19.

Remark: Note that Theorem 19 is not just an abstraction the case of $X = [0,1]$, as even in that classical case it tells us that u.d. on closed subintervals implies u.d. on all Jordan sets.

The proof of Theorem 19 cannot quite consist of reshashing the proof of Theorem 7 because in a general compact space it is not immediately clear where to come by nontrivial examples of Jordan sets. Rather, we must exploit Urysohn's Lemma. But first a straightforward preliminary result:

Proposition 20. *Suppose* $\mathcal{S}$ *is a subset of the Borel functions* $\mathcal{B}(X)$ *whose closed linear span contains* $\mathcal{C}$*, the continuous functions. Let* $\mathbf{x}$ *be any sequence in* X*. Then if (6) holds for all* $f \in \mathcal{S}$*, it holds for all* $f \in \mathcal{C}$*.*

Proof: Define

$$\overline{F} : f \mapsto= \limsup_N \frac{1}{N}\sum_{n=1}^{N} f(x_n),$$

$$\underline{F} : f \mapsto= \liminf_N \frac{1}{N}\sum_{n=1}^{N} f(x_n),$$

and $G : f \mapsto \int_X f d\mu$. Then $\overline{F}$, $\underline{F}$, and G are all linear functionals on $\mathcal{B}$ of norm 1, hence in particular are continuous. Therefore the assumed equality $\underline{F} = \overline{F} = G$ on

[12] The Jordan in question here is Camille Jordan, not to be confused with the Jordan algebras named after the mathematical physicist (and brownshirt) Pascuale Jordan.

$\mathcal{S}$ implies the same equality on the closed linear span, which was to be shown.

Now let us prove Theorem 19. Of course $b) \implies c)$.

Let us show $a) \implies b)$: if we can show that for all Jordan sets A and all $\epsilon > 0$, there exist continuous functions g_1 and g_2 on X such that $g_1 \leq \chi_A \leq g_2$ and $\int_X (g_2 - g_1) d\mu < \epsilon$, then the same two lines which proved $(b) \implies (c)$ in Theorem 7 will work here. Now, using the fact that A is a Jordan set and the regularity of the measure, we can find a closed subset $C \subset A^0$ with $\mu(A^0 \setminus C) < \epsilon/2$ and an open subset $D \supset \overline{A}$ with $\mu(D - \overline{A}) < \epsilon/2$. Next recall that as a compact space X is completely regular, so that disjoint closed sets may be separated by continuous functions. In particular there is a continuous function $g_1 : X \to [0,1]$ which is identically equal to 1 on C and identically 0 on $X \setminus A^0$, and another continuous function $g_2 : X \to [0,1]$ which is identically 1 on $\overline{A}$ and identically zero on $X \setminus D$. This works, since

$$\int_X (g_2 - g_1) d\mu - \int_{D \setminus C} (g_2 - g_1) d\mu \leq \mu(D \setminus C) = \mu(D \setminus A) + \mu(\partial A) + \mu(A^0 \setminus C) < \epsilon.$$

We now show that $c) \implies a)$. Note that in asuming c) we are equivalently assuming that (6) holds for all characteristic functions of closed Jordan sets. By Proposition 2.2, it is enough to show that, say $\mathcal{W}$, the closed linear span of the characteristic functions of closed Jordan sets, contains $\mathcal{C}$. Since $\mathcal{W}$ is a linear subspace containing the constant functions, it is harmless to rescale and thus assume that $f(X) \subset [0,1)$. Consider the closed sets

$$M_\alpha = f^{-1}([\alpha, \infty)).$$

Since $\partial M_\alpha \subset f^{-1}(\alpha)$, M_α is a Jordan set for all but at most countably many α.[13] Fix $\epsilon > 0$. There exists a finite sequence

$$0 = \alpha_a < \alpha_1 < \ldots < \alpha_n = 1$$

such that for all i, $\alpha_{i+1} - \alpha_i < \epsilon$ and M_{α_i} is a Jordan set. Take an $x \in X$. There exists some k such that $\alpha_k \leq f(x) < \alpha_{k+1}$. Then

$$\sum_{i=0}^{k-1} (\alpha_{i+1} - \alpha_i) \chi_{M_{\alpha_i}}(x) - f(x)| = |\sum_{i=0}^{k} (\alpha_{i+1} - \alpha_i) - f(x)| = \alpha_{k+1} - f(x) \leq \epsilon.$$

Thus we have shown that

$$||\sum_{i=0}^{n-1} (\alpha_{i+1} - \alpha_i) \chi_{M_{\alpha_i}} - f|| \leq \epsilon,$$

i.e., that f is in $\mathcal{W}$. This completes the proof of the theorem.

Exercise XX: Let $\mu_1,\ \mu_2$ be distinct measures. Show that it is not possible for a sequence be u.d. w.r.t. both μ_1 and μ_2. (Hint: Riesz representation theorem.)

Define the *support* $\text{supp}(\mu)$ of a measure μ to be the set of points $x \in X$ such that $\mu(U) > 0$ for every open neighborhood U of x. It is a closed set.

[13]We are using here the fact that the sum of an uncountable set of non-negative real numbers can only be finite if all but at most countably many of these numbers are zero.

Exercise XX: Show that if $\mathbf{x}$ is u.d. in X, it is dense in $\mathrm{supp}(\mu)$.

2.3. **Metric theorems.** There are two major shortcomings in our development thus far. In contrast to the classical case, we do not yet have a Weyl criterion, and we do not yet know that there are any u.d. sequences at all!

Let us begin with the first point. We still have the Stone-Weierstrass theorem: in other words, given any algebra $\mathcal{A}$ of continuous functions containing the constant functions and separating points of X, its uniform closure will be all of $\mathcal{C}$, so by Proposition 2.2 we can test for u.d. using $\mathcal{A}$: i.e., a sequence in u.d. iff (6) holds for all $f \in \mathcal{A}$. It ought to be fairly clear that the existence of such a nice algebra as the algebra of trigonometric polynomials is closely linked to the group structure of S^1, and since we are not assuming that X is the underlying space of a topological group, we are not going to get anything so explicit. We will not even in general be able to find a countable subset $\mathcal{S}$ of $\mathcal{C}$ whose closed span is all of $\mathcal{C}$. Indeed, there is the following general fact:

Theorem 21. *For a compact space X, TFAE:*
a) X has a countable basis of open sets ("second countable").
b) $\mathcal{C}$ is the closed linear span of a countable subset (is "separable").
c) X is metrizable.

To see a space that satisfies none of these hypotheses, let X be any compact space consisting of more than a single point, and take $\tilde{X} = X^{\mathbb{R}}$ with the product topology. Then Tychonoff's theorem ensures that $\tilde{X}$ is compact, but it is easy to see that a) does not hold, because for any given open subset U of $\tilde{X}$, the set of $\alpha \in \mathbb{R}$ such that the projection of U onto the αth copy of X is proper in X is finite (by definition of the product topology), so arbitrary unions formed from a countable collection of open sets project surjectively onto all but countably many $\alpha \in \mathbb{R}$, whereas obviously for any $\alpha \in \mathbb{R}$ there exists an open set U_α which does not have this property.

It is also true that in a general compact space there may not exist any u.d. sequences at all. That is to say, in a general compact space, the weak sequential closure of the convex hull of the Dirac measures need not be equal to the entire set of probability measures on X. See ??? for a counterexample.

Thus to make progress we must assume that X has a countable basis.[14] The good news is that once we assume this we also get to assume the existence of a metric, which makes life much easier. As a very important example, we can demonstrate the existence of u.d. sequences in (X, μ) by proving a much stronger result. To set the stage, consider the space $X^\infty = X^{\mathbb{Z}^+}$ of all sequences in X. It is again a compact space, and carries a natural measure μ_∞, characterized by the following: for any subset Y such that $\pi_n(Y) = X$ for all but finitely many n, we have

$$\mu_\infty(Y) = \prod_{n \in \mathbb{Z}^+} \mu(\pi_n(Y)).$$

[14] I don't know about you, but the compact spaces I meet in my daily life all have countable bases, so it distresses me very little to make this extra assumption.

Let $\mathcal{U} \subset X^\infty$ be the subset of u.d. sequences. What is its measure? Note that since the property of membership in $\mathcal{U}$ is unaffected by modifying a sequence in any finite number of terms, the $0-1$ Law of Kolmogorov asserts that either (μ_∞)-almost every sequence is u.d. or almost every sequence is *not* u.d.[15] The following result ends the suspense:

Theorem 22. *We have in fact* $\mu_\infty(\mathcal{U}) = 1$.

Proof: The key is the following

Claim: For any given $f \in \mathcal{B}$, (6) holds for μ_∞-almost every $\mathbf{x} \in X^\infty$.

For if so, the result follows from the assumed separability of $\mathcal{C}$, Proposition 2.2, and the fact that a countable union of sets of measure zero has measure zero.

Let us prove the claim. Certainly it holds for f a constant function, so by replacing f by $f - \int_X f$, we may assume that $\int_f X d\mu = 0$. Consider the function F_N on X^∞ defined by

$$\mathbf{x} \mapsto \frac{1}{N}\sum_{n=1}^{N} f(x_n).$$

Then

$$\int_{X^\infty} F_N^2 d\mu_\infty = \frac{1}{N^2}\sum_{n=1}^{N}\int_{X^\infty} f^2(x_n)d\mu_\infty + \frac{2}{N}\sum_{1\leq i<j\leq N}\int_{X^\infty} f(x_i)f(x_j)d\mu_\infty = \frac{1}{N}\int_X f^2 d\mu.$$

So

$$\sum_{m=1}^{\infty}\int_{X^\infty}(F_{m^2})^2 d\mu_\infty = \left(\int_X f^2 d\mu\right)\sum_{m=1}^{\infty}\frac{1}{m^2} < \infty.$$

It follows from the dominated convergence theorem that $\lim_{m\to\infty} F_{m^2} = 0$ μ_∞-almost everywhere. For each $N \geq 1$, choose m such that $m^2 \leq N < (m+1)^2$. Then

$$|F_N| = |\frac{m^2}{N}F_{m^2} + \frac{1}{N}(f(x_{m^2+1}) + \ldots + f(x_N))| \leq |F_{m^2}| + \frac{2}{m}||f||.$$

Since the last expression tends to 0 μ_∞-almost everywhere as $N \to \infty$, we are done.

2.4. **Connections with ergodic theory.**

An attractive alternate proof of Theorem 22 can be obtained using the rudiments of ergodic theory.

Namely, let $(Y, \mathcal{A}, \mu)$ be a measure space and $T : U \to U$ be a measure-preserving transformation: for all $A \in \mathcal{A}$, $\mu(T^{-1}(A)) = \mu(A)$. A measure-preserving T is called *ergodic* if $T^{-1}(A) = A$ implies $\mu(A) \in \{0, 1\}$.

Proposition 23. *Let* (X, μ) *be a probability space,* $Y = (X^\infty, \mu_\infty)$ *the product space, and let* $T : Y \to Y$ *be the* ***shift operator****:*

$$\mathbf{x} = \{x_n\}_{n=1}^\infty \mapsto T(\mathbf{x}) = \{x_{n+1}\}_{n=1}^\infty.$$

Then T *is an ergodic transformation.*

[15] Or, technically, that $\mathcal{U}$ is not μ_∞-measurable.

Theorem 24. *(Birkhoff pointwise ergodic theorem) Let $(Y, \mathcal{A}, \mu)$ be a probability space, and let $T : Y \to Y$ be a measure-preserving transformation, and $g \in L^1(\mu)$.*
a) There exists a unique $\psi_g \in L^1(\mu)$ such that we have (for μ-.a.e. $y \in Y$)

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n y) = \psi_g(y).$$

b) If T is ergodic, then for μ-a.e. $y \in Y$ we have

$$\psi_g(y) = \int_Y g d\mu.$$

In particular, the function $y \mapsto \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n y)$ is μ-a.e. constant.

Now apply Proposition and Theorem 24 with $Y = (X^\infty, \mu_\infty)$, $T : Y \to Y$ the shift operator, f any element of $\mathcal{C} = \mathcal{C}(X, \mathbb{E})$ and $g = f \circ \pi_1 : Y \to \mathbb{E}$. The conclusion is precisely that for μ_∞-a.e. $\mathbf{x}$ in X^∞ we have

$$\frac{1}{N} \sum_{n=0}^{N-1} f(n) \to \int_X f d\mu,$$

which was the essence of the proof of 22. (The full proof follows, as before, by applying the result to each member of a countable basis of $\mathcal{C}$ and noting that the union of a countable collection of sets of measure zero has measure zero.)

Example: Consider the simplest nontrivial case $X = \{\pm 1\}$ with $\mu(\{-1\}) = \mu(\{1\}) = \frac{1}{2}$. A sequence x_n is μ-equidistributed if, asymptotically, half of its values are $+1$ and half are -1. Or, applying the Fundamental Theorem to the function $f : 1 \mapsto 1,\ -1 \mapsto -1$, one sees that if $\mathbf{x}$ is μ-equidistributed then the "expected value" converges to zero. Thus in this case Theorem 22 is none other than the **Strong Law of Large Numbers**.

Notwithstanding the name, as probabilists well know, stronger statements are indeed possible:

Theorem 25. *(Kolmogorov Law of the Iterated Logarithm) For any $f \in \mathcal{B}$, we have for μ_∞-a.e. $\mathbf{x}$*

$$\frac{1}{N} \sum_{n=1}^{N} f(x_n) - \int_X f d\mu = O\left(\sqrt{\frac{\log\log N}{N}}\right).$$

Exercise X.X (Thanks to R. Varley and J. Manning): Let $\alpha \in \mathbb{R}$, and consider the transformation $T : [0, 1) \to [0, 1)$ given by $x \mapsto \{x + \alpha\}$.
a) Show that T is ergodic iff $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.
b) Apply Birkhoff's Pointwise Ergodic Theorem (Theorem 24) to deduce the following result.

Theorem 26. *For $\alpha \in \mathbb{R}$, TFAE:*
(i) For any Lebesgue measurable subset E of $[0, 1)$ and ***almost every*** *$x \in [0, 1)$ we have*

$$\lim_{N\to\infty} \frac{\#\{n \le N \mid \{x + n\alpha\} \in E\}}{N} \to \mu(E). \tag{8}$$

(ii) $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Compare: According to Corollary 6 plus Proposition 9 (or just Theorem 16) together with Remark X.X, if α is irrational then (8) holds for every $x \in [0, 1)$ provided that E is a Jordan set. Thus, compared to Corollary 6, Theorem 26 has both a weaker hypothesis and a weaker conclusion.

2.5. **Equidistribution and Baire category.**

We can also ask if $\mathcal{U} = \mathcal{U}(\mu)$ is large in some topological sense. If we consider the following three facts –

(a) A sequence $\mathbf{x}$ in X can be u.d. w.r.t. at most one measure.
(b) Given any probability measure μ on X, there are, by Theorem 22 "plenty" of μ-u.d. sequences.
(c) The space of measures on X is infinite-dimensional if X is infinite –

then it is implausible that the subset $\mathcal{U}(\mu)$ is **residual** in X^∞, i.e., that its complement is of the first category:[16] certainly this could hold for at most one measure μ, so we would have to believe that X carries a "distinguished" measure, which for a general compact space is not the case.[17]

Theorem 27. *a) The set $\mathcal{U}(\mu)$ is dense in X^∞.*
b) If $\#X > 1$, $\mathcal{U}(\mu)$ is of the first category in X^∞.

2.6. **Equidistribution and rearrangement.**

Recall that in the case of $X = [0, 1]$, a sequence $\mathbf{x}$ can be rearranged to be equidistributed iff it has dense image. This is seen not to be true in general:

Example XX: If $X = \{\pm 1\}$, the sequence $1, \ -1, \ -1, \ -1, \ldots$ is not u.d., despite having dense (and even surjective) image.

We can, so to speak, isolate the problem: given a space (X, μ), let us say a point $x \in X$ is a **bad point** if all of the following hold: x is an isolated point of X, $X \neq \{x\}$, and $\mu(\{x\}) > 0$.

Theorem 28. *TFAE:*
a) X has no bad points.
b) Every sequence $\mathbf{x}$ with dense image has a u.d. rearrangement.

Proof: If X has a bad point x, we can construct a non-u.d. sequence with dense image as follows: take $x_1 = x$, and let $x_2, \ x_3, \ldots$ be any dense sequence in the (separable, compact) subspace $X \setminus \{x\}$. Conversely...

The general criterion for rearrangement is as follows:

Theorem 29. *A sequence $\mathbf{x}$ in X admits a u.d. rearrangement iff every open neighborhood of every point of* $\operatorname{supp}(\mu)$ *contains infinitely many terms of* $\mathbf{x}$.

---

[16]Recall that a subspace is *of the first category* if it can be expressed as a countable union of nowhere dense sets. A theorem of Baire asserts that no compact Hausdorff space is itself of the first category.

[17]E.g. there is no measure on $[0, 1]$ which is invariant under all homemorphisms.

2.7. **Riemann integration in compact metrizable spaces.** We are taught that one of the merits of the Lebesgue integral over the Riemann integral is its abstraction: the latter can be defined for functions on a general measure space, while the former is limited to functions defined on $\mathbb{R}^n$ or on very nice subsets thereof.

But this is not true. For instance, in studying functions on $\mathbb{Q}_p$ one sometimes encounters a limit of finite sums looking very much like a Riemann integral. If X is a metrizable compact space endowed with a probability measure μ, then comparing Theorems 7, 8 and 19, it seems plausible to define a μ-Riemann integrable function f to be a Borel function f such that (6) holds for all u.d. sequences $\mathbf{x}$ in X. At least this definition includes the continuous functions. In fact one has the following generalization of a well-known theorem of Lebesgue:

Theorem 30. *Let $f : X \to \mathbb{R}$ be any function. TFAE:*
a) There exists a number L such that for all μ-u.d. $\mathbf{x}$, $\frac{1}{N}\sum_{n=1}^{N} f(x_n) \to L$.
b) f is bounded and continuous μ-a.e.

By choosing a metric d on X, one can indeed define the Riemann integral of a function in a way that is almost a direct generalization of the usual case $X = [0,1]$. Namely, we define a Riemann sum $R(f, \mathcal{P}, x_i)$, where $\mathcal{P}$ is a finite collection $Y_1, \ldots, Y_n$ of Jordan sets[18] whose union is X, whose pairwise intersections have measure zero, and where $x_i \in Y_i$ is a sample point, to be $\sum_{i=1}^{n} f(x_i)\mu(Y_i)$. Define the *mesh* of a partition to be the largest diameter of any Y_i. We then define a function to be Riemann integrable to L if for every $\epsilon > 0$, there exists $\delta > 0$ such that for all tagged partitions $(\mathcal{P}, x_i)$ of mesh at most δ, $|R(f, \mathcal{P}, x_i) - L| < \epsilon$. One finds that the functions which are Riemann integrable in this sense are exactly the functions satisfying the equivalent conditions of the previous theorem. In particular one gets a version of Theorem 19 which mirrors Theorem 7 in allowing us to test for u.d. using the Riemann-integrable functions.

References

[KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences.*

[18]It would also suffice to work with the algebra of sets generated by the closed Jordan balls, since all but countably many of the closed balls with any given center are Jordan sets.