Finitely additive measures and the first digit problem

by

Richard T. Bumby and Erik Ellentuck

March 02, 2006

The sources

The subject of this talk is a paper of the same title published in Fundamenta Mathematicae 65 (1969).

A misstatement of a theorem of the paper was submitted as Problem 5589 in the American Mathematical Monthly, and a counterexample to that statement by Dan Marcus, as well as a proof of the intended statement, was published in the February 1976 issue.

Milton Parnes used our methods in “On the measure of measurable sets of integers”, Acta Arithmetica 25 (1973).

Genesis

In those days, Rutgers gave large common final exams in all beginning course and used large spaces like the main gymnasium for this activity. All instructors were called together to proctor, but no individual was kept very busy. This led to Erik asking me if I could provide a number theoretical viewpoint on Benford's law. We met frequently after that and exchanged ideas. This paper was the result.

It also led to our meeting Roger Pinkham who had written on the subject earlier from a statistical viewpoint.

Equally likely integers

We want to describe the notion of "pick an integer at random", but traditional methods don't apply.

If a probability measure on the natural numbers assigns the same measure to each integer, that measure must be zero. Then the measure can't be countable additive, so traditional measure theory is not available.

The sets that we work with will be sets of **natural numbers**, but we will often be sloppy and say "integer".

An interesting alternative

Tarski showed that a **finitely additive** measure on an algebra of sets of integers containing all singletons can be extended to one defined on **the whole power set**. All measures that appear here will be finitely additive probability measures defined on all subsets.

Ultrafilters

A collection of nonempty sets that also contains the intersection of two sets in the collection and any **superset** of a set in the collection is called a **filter**. A **maximal** filter in the ordering by inclusion is called an **ultrafilter**. Since a set that meets all elements of a filter without containing any does not belong to the filter, but can added to it to create a larger filter, ultrafilters must contain any given set or its complement. The set of all sets containing a particular point (an integer in our case) is an ultrafilter, called a **principal** ultrafilter. There are many more, and it is these non-principal ultrafilters that are more interesting.

Integrals

If f is a **bounded** function on $\mathbb{N}$ and μ is a measure, $\int f \, d\mu$ can be defined. It is easily approximated, as in classical measure theory, by dividing the **range** at points c_k considering the sets where $f(n)$ lies in $[c_k, c_{k+1}]$. The sum of c_k times the μ-measure of the set of n with $f(n)$ in this interval is a lower bound on the integral, and replacing c_k by c_{k+1} gives an upper bound. These bounds are arbitrarily close, so every bounded function is integrable, and the integral is a linear functional. For the measure that tests membership in a non-principal ultrafilter $\mathcal{U}$, the integral of a function f gives something that is called a $\mathcal{U}$-limit of f.

A topology on the set of measures

The usual connection between measures and integrals holds and the weak* topology on the linear functionals gives a convenient topology on the space of all measures. Many interesting classes of measures will be closed in this topology.

Measurable sets for classes of measures

Instead of looking at individual measures, we will usually consider classes of all measures with some nice property. For a set $A \subseteq \mathbb{N}$ a class of measures $\mathcal{C}$, we can form the set of all $\mu(A)$ for $\mu \in \mathcal{C}$. This will be denoted $\mathcal{C}(A)$. Similarly, one can define $\mathcal{C}(f)$ as the set of values of integrals of the function f.

The sets on which all measures in $\mathcal{C}$ agree will be called measurable with respect to $\mathcal{C}$.

Convex classes

Many interesting classes of measures are **convex**. That is, if $\mu_0 \in \mathcal{C}$ and $\mu_1 \in \mathcal{C}$, and $0 \leq \alpha \leq 1$, then $\alpha\mu_0 + (1-\alpha)\mu_1 \in \mathcal{C}$.

For a convex class of measures, $\mathcal{C}$, the set $\mathcal{C}(A)$, or $\mathcal{C}(f)$ will always be an **interval**. This makes it easy to study measurability with respect to convex classes of measures.

The class of all measures

Let $\mathcal{M}$ denote the class of all measures. It is clearly a convex class. We have

$$\mathcal{M}(\emptyset) = 0 \quad \text{and} \quad \mathcal{M}(\mathbb{N}) = 1.$$

For all other sets A, $\mathcal{M}(A) = [0, 1]$ since we can use a measures that test whether a set belongs to an ultrafilter to obtain the extreme values.

The class of all non-atomic measures

Let $\mathcal{N}$ denote the class of all measures that give single points measure zero. Again, this is clearly a convex class. Finite sets have measure 0 and cofinite sets have measure 1. For all other sets A, $\mathcal{M}(A) = [0, 1]$ since we can use a measures that test whether a set belongs to a non-principal ultrafilter to obtain the extreme values.

Invariant measures

For $\mu \in \mathcal{N}$, $\mu(\mathbb{N}) = \mu(\mathbb{N}+1)$. A measure is said to be **translation invariant** if this is true for all sets. The class of all translation invariant measures is denoted $\mathcal{T}$. (We will show that it is nonempty.)

More generally, we can take any function $g : \mathbb{N} \longrightarrow \mathbb{N}$ and say that a measure μ is g-invariant if $\mu\big(g^{-1}(A)\big) = \mu(A)$ for all A, where $g^{-1}(A)$ denotes the complete inverse image of A. The class of g invariant measures will be denoted $\mathcal{I}_g$.

Constructing invariant measures

If f is any bounded function, $\mathfrak{I}_g(f)$ can be estimated using

$$f^{(m)}(n) = \frac{1}{m} \sum_{k=0}^{m-1} f\big(g^k(n)\big)$$

(This is only really suitable for g that are one to one. The general case will be more similar to our later discussion of scale invariant measures.)

A theorem

Theorem 1. $\mathcal{I}_g$ *is non-empty and*

$$\mathcal{I}_g(f) = [\liminf_m \operatorname{glb}_n f^{(m)}(n), \limsup_m \operatorname{lub}_n f^{(m)}(n)]$$

Also,

$$\limsup_m \operatorname{lub}_n f^{(m)}(n) = \operatorname{glb}_m \operatorname{lub}_n f^{(m)}(n)$$

and similarly for the lower bound. Furthermore, if $\mathcal{I}_g \subseteq \mathcal{N}$, *the upper bound is equal to* $\operatorname{glb}_m \limsup_n f^{(m)}(n)$.

Proof of the theorem

The later versions of the upper bound are smaller than the earlier ones. To prove the theorem, measures attaining the larger upper bound will be constructed while the smaller one will be shown to be an upper bound.

The construction step

Here is a construction of a measure with

$$\mu(f) = \lim_m \sup \operatorname{lub}_n f^{(m)}(n).$$

Let $\mu_n^{(m)} = k/m$ where k is the number of members of A in $\langle n, g(n), \ldots, g^{m-1}(n) \rangle$. Then $f^{(m)}(n)$ is the integral of f with respect to this measure.

Construction continued

Choose a sequence of m with $m \to \infty$ realizing this limit. Choose your favorite sequence of $\epsilon_m \to 0$. For each m choose n depending on m with $f^{(m)}(n) > \mathrm{lub}_n\, f^{(m)}(n) - \epsilon_m$. Form the $\mathcal{U}$-limit of these $f^{(m)}(n)$ over some ultrafilter. This limit is in $\mathcal{I}_g$ because $\left| f^{(m)}\left(n \right) - f^{(m)}\left(g(n) \right) \right| < 1/m$.

Bounds on measures

If $\mu \in I_g$, $\int f \, d\mu = \int f(g) \, d\mu$, which shows that the integral is also the same as $\int f^{(m)} \, d\mu$. This shows that $\int f \, d\mu \leq \text{lub}_n f^{(m)}(n)$ for every m. This leads to the alternate expression as upper bound.

The proof by Marcus

When the equality of limits for $g(n) = n + 1$ was presented as a problem in the **Monthly** a simple direct proof was given. It demonstrated that $\text{lub}_n\, f^{(m)}(n)$ is essentially a decreasing function of m so it has a limit. This also shows why it was necessary to add the word **bounded** to the statement.

Examples

For the class $\mathcal{T}$:

Any set that contains exactly one of each pair $\{2n, 2n+1\}$ has measure $1/2$.

If $\alpha > 1$ and $\beta > 0$ are real numbers, the **generalized arithmetic progression** $\lfloor n\alpha + \beta \rfloor$ has measure $1/\alpha$.

The set of SF squarefree numbers has $\mathcal{T}(SF) = [0, 6/\pi^2]$.

Sparse sets of real numbers

Some terminology used in the paper:

A set $A \subseteq \mathbb{R}^+$ is called **discrete** if

$$\Psi_A(n) = \#\left\{ A \cap [n, n+1) \right\}$$

is finite for all n, and A is **sparse** if $\Psi_A(n)$ is bounded.

Distortions

For each $\mu \in \mathcal{M}$, $\phi(A) = \int \Psi_A \, d\mu$ is a set function on sparse sets extending μ. If $\mu \in \mathcal{T}$, ϕ will be invariant under maps f with $|f(x) - x|$ bounded. Such maps are called **distortions**. Restricting any distortion invariant ϕ to subsets of $\mathbb{N}$ (and normalizing) gives a measure in $\mathcal{T}$ with $\phi(A) = \int \Psi_A \, d\mu$.

Measure inducing sets

Measures in $\mathcal{T}$ are identified with normalized distortion invariant set function on the family of sparse subsets of $\mathbb{R}^+$.

Other subsets of $\mathbb{R}^+$ of interest are the **measure inducing** sets S which are unions of intervals S_i such that sets of integers, one from each S_i have measure zero for all measures in $\mathcal{T}$. Equivalently, there is a function f with $\lim_{t\to\infty} f(t)/t = 0$ such that each interval of length t meets at most $f(t)$ of the S_i.

Another theorem

Theorem 2. *If S is measure inducing, $\alpha > 0$, and $\mu \in \mathcal{T}$, then* $\mu(S \cap \mathbb{N}) = \alpha\mu(S \cap \alpha\mathbb{N})$.

Scale invariance

It follows from Theorem 2 that $\mu(\mathbb{N}) = \alpha\mu(\alpha\mathbb{N})$. Then, the set function defined by $\mu'(A) = \alpha\mu(\alpha A)$ also belongs to $\mathcal{T}$. If $\mu' = \mu$, then μ is said to be α-**scale invariant**. The class of all such measures will be denoted $\mathcal{S}_\alpha$.

One can also consider the class $\mathcal{S} = \bigcap_\alpha \mathcal{S}_\alpha$.

Construction of scale invariant measures

Take any $\mu \in \mathcal{T}$ and let

$$\mu^{(n)}(A) = \frac{1}{n}\sum_{k=0}^{n-1} \alpha^k \mu(\alpha^k A).$$

Then, let $\nu(A) = \mathcal{U}-\lim \mu^{(n)}(A)$. One sees that $\nu \in \mathcal{S}_\alpha$ and if $\mu \in \mathcal{S}_\beta$, then $\nu \in \mathcal{S}_\alpha \cap \mathcal{S}_\beta$. Furthermore, if α and β are multiplicatively independent, and if

$$\mu(\delta A) \leq \mu(A) \quad \text{for all} \quad \delta \geq 1 \qquad (*)$$

then $\nu \in \mathcal{S}$.

Connection with asymptotic density

Consideration of asymptotic density leads to a class of measures $\mathcal{A}$ for which

$$\int f \, d\mu \leq \operatorname*{lub}_{n} \frac{1}{n} \sum_{k=1}^{n} f(k)$$

All measures in $\mathcal{A}$ satisfy $(*)$, but we were unable to decide the converse.

Connections with Benford's law

The connection is expressed in

Theorem 3. *Let U be an arc on the unit circle in $\mathbb{C}$ and*

$$S = \left\{ x \in \mathbb{R}^+ : e(\log_\beta x) \in U \right\}.$$

If $\theta \in \mathbb{R}^+$ and $\log_\beta \theta$ is irrational, and $\mu \in \mathbb{S}_\theta$, then $\mu(S \cap \mathbb{N}) = \lambda(U)$ where λ is the Lebesgue measure on the circle normalized to give the whole circle measure 1.

Logarithmic density

(This was not in the original paper.)

In theorem 1, translation invariant measures were constructed as limits of the atomic measures $\mu_n^{(m)}$. If we started with measures for which the measure of individual integers k in the support was proportional to $1/k$, we would obtain measures extending the familiar notion of **logarithmic density**. I haven't checked all details, but such measures appear to belong to $\mathbb{S}$, i.e., they are α-scale invariant for all α

The measure of measurable sets

This deals with the work of Milton Parnes. A set $A \subseteq \mathbb{N}$ may be encode by the real number $\sum \chi_A(n)2^{-n}$, so a set of such sets may be viewed as a subset of $\mathbb{R}$. In particular, the sets measurable with respect to $\mathcal{I}_g$ may be considered from this point of view forming a set $J_g \subseteq \mathbb{R}$.

Theorem. *J_g has Lebesgue measure zero.*

A special collection of a measurable sets

If k and L are integers with $L > 2$ and

$$0 \leq \frac{k-1}{L} < J_g(A) < \frac{k}{L} \leq 1$$

then **measurability** gives that all $\mu_{LM}^{(n)}(A)$ for sufficiently large M lie between $(k-1)/L$ and k/L (this use of the multiples of a base number of iterations was also part of the work of Marcus).

A bound on the measure

Using this for $n = 1$ already shows that the Lebesgue measure of the set of real numbers corresponding the set of all such A is bounded by

$$p = 2^{-LM} \sum_{(k-1)M}^{kM} \binom{LM}{i} < 1.$$

Conclusion of proof

Now, take $n = g_{LM}(1)$. The value of the measure gives a condition that is **independently** satisfied with probability p.

Continuing in this fashion, the sets with these measures give a set of real numbers with measure bounded by p^k for all k, so that the measure is zero.

A similar argument is possible to deal with sets of measure 0 or 1. All together, there are countably many choices of pairs (k, L) that cover all measures, so the theorem follows.

The answer of a question of Buck

Parnes applies this to a question raised by Buck in Amer. J. Math. 68 (1946). A class of **outer measures** giving the correct densities of arithmetic progressions was studied, and the measurability question was raised for its collection of measurable sets.

We have seen that the $\mathcal{T}$-measurable sets include all arithmetic progressions, with the correct density. Studies based on outer measure are subsumed under our theory, so Buck's measurable sets are all $\mathcal{T}$-measurable. Since the latter class gives a set of Lebesgue measure zero, so does Buck's class.