Generalizing the Arithmetic Mean

bharathk98

May 2020

0.1 Suggestion

Don't treat this as a research paper. Lightly skim.

0.2 Theories One Should Know

- Descriptive Set Theory
- Measure Theory
- Box-Counting Dimensions
- Folner Sequences

1 Purpose

- *I wish to create a measure that calculates arithmetic mean for functions where it's well-defined but calculates a new arithmetic mean for functions where it's not.*

I wish to find a measure (possibly new) that calculates a unique arithmetic mean for the largest class of functions and coincides with the arithmetic mean of other functions.

2 Problems With Other Definitions

2.1 Lebesgue Measure

Suppose $f : A \to B$ is an arbitrary measurable function, where A is a non-fixed arbitrary subset of the real numbers.

- The Lebesgue Measure of A or $\lambda(A)$ gives the average $\frac{1}{\lambda(A)}\int_A f(x)d\lambda$ and is undefined when $\lambda(A) = 0$.

2.2 Counting Measure

- The counting measure, μ, of the domain, A, is denoted $\mu(A)$ and $|A|$ is the cardinality of A such that

$$\mu(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ \infty & \text{if } A \text{ is infinite} \end{cases}$$

  and the average $\frac{1}{\mu(A)} \sum\limits_{x \in A} f(x)$ is undefined when $\mu(A) = \infty$.

- Hence when generalizing the arithmetic mean we consider functions whose domain has infinite points.

2.3 Hausdorff Measure

The Hausorff Measure seems like a proper solution but there lies an issue:
Suppose the domain has zero Lebesgue Measure and infinite elements.

- If we define the average on A using the Hausdorff Measure, we must rely on its dimension to get a defined mean.

- If the dimension is zero, we get the counting measure (this gives a defined mean when A is finite).

- If the dimension is one, we get the Lebesgue Measure (this gives a defined mean when A has a measure greater than zero).

- When A has zero Lebesgue Measure with infinite points there exists a dimension $d \in [0, 1]$ where $s < d$ give a measure of ∞ and $s > d$ give a measure of 0.

- If $d = s$ the Haursdorff Measure would be any number between zero and infinity and is therefore not unique.

2.4 Conditional Expectation

- When applying the conditional expectation for finding the average of f (in the measure-theoretic sense) one must find a distribution of A.

- If A has zero Lebesgue Measure and infinite elements, I assume there is multiple distributions to choose for defining the mean. Therefore we must apply the axiom of choice to get a unique mean.

I assume we are unable to come up with a choice function that gives a unique distribution for every A and in turn gives a unique measure for conditional expectation.

2.5 Properties Our Measure Should Have

If S is an arbitrary subset of A and A is an arbitrary subset of $\mathbb{R}$, we should define a probability measure $\mathcal{P}(S, A)$ which yields an arithmetic mean that coincides with other arithmetic means. The measure should have the following properties.

- When $S = A$, $\mathcal{P}(S, A) = \mathcal{P}(A, A) = 1$
- When $\lambda(A) > 0$, $\mathcal{P}(S, A) = \lambda(S)/\lambda(A)$
- When S is countable and A is uncountable, $\mathcal{P}(S, A) = 0$ (regardless if $\lambda(A) \geq 0$).
- When $\lambda(S), \lambda(A) = 0$, with S being nowhere dense and A being dense in an interval, then $\mathcal{P}(S, A) = 0$

3 First Attempt In Generalizing Arithmetic Mean

3.1 Defining Measure $\mathcal{P}$ (Try and Read The Whole Section)

If $S \subseteq A$:

- ℓ is the length of an interval
- $(I_j)_{j=1}^{m}$ and $(J_k)_{k=1}^{n}$, for $m, n \in \mathbb{N}$, are a sequence of open intervals where $\ell(I_1) = ... = \ell(I_n) = g \in \mathbb{R}^+$, $\ell(J_1) = ... = \ell(J_m) = g \in \mathbb{R}^+$ and the infimum of equations below are taken over all possible I_j and J_k

$$\mathcal{M}(g, S) = \begin{cases} g \cdot \inf\left\{m \in \mathbb{N} : S' \subseteq S, S' \text{ is countable}, \ (S \setminus S') \subseteq \bigcup_{k=1}^{m} I_m\right\} & A \text{ is uncountable} \\ g \cdot \inf\left\{m \in \mathbb{N} : S \subseteq \bigcup_{k=1}^{m} I_m\right\} & A \text{ is countable} \end{cases} \tag{1}$$

and

$$\mathcal{N}(S, A) = \lim_{g \to 0} \frac{\mathcal{M}(g, S)}{\mathcal{M}(g, A)} \tag{2}$$

then we add $\mathcal{O}(g, S)$ so if A is a interval with finite length, $\mathcal{O}(g, S)$ gives the Lebesgue Measure of S over the Lebesgue Measure of A.

$$\mathcal{O}(g, S) = \begin{cases} g \cdot \inf\left\{n \in \mathbb{N} : S_j \subseteq S, \ \mathcal{N}(S_j, A) = 0, \ \bigcup_{j=1}^{\infty} S_j = S', \ (S \setminus S') \subseteq \bigcup_{k=1}^{n} J_k\right\} & A \text{ is uncountable} \\ g \cdot \inf\left\{n \in \mathbb{N} : S \subseteq \bigcup_{k=1}^{n} J_n\right\} & A \text{ is countable} \end{cases} \tag{3}$$

therefore the outer measure is

$$\mathcal{P}^*(S,A) = \lim_{g\to 0} \frac{\mathcal{O}(g,S)}{\mathcal{O}(g,A)} \tag{4}$$

And the inner measure is $\mathcal{P}_*(S,A) = 1 - \mathcal{P}^*(A \setminus S, A)$, meaning measure $\mathcal{P}(S,A)$ exists when $\mathcal{P}^*(S,A) = \mathcal{P}_*(S,A)$.

3.2 Defining $\mathcal{P}$-Average

When $\mathcal{P}(S,A)$ is defined, the $\mathcal{P}$-average of $f : A \to B$ or $\overline{f}$ is defined as follows:

Split $[a,b] = [\min(A), \max(A)]$ into sub-intervals using partitions x_i

$$a = x_0 \leq \cdots \leq x_i \leq \cdots \leq x_r = b$$

Here, if $1 \leq i \leq r$, then $[x_{i-1}, x_i]$ are sub-intervals of $[a,b]$.

For every i, choose a $v_i \in A \cap [x_{i-1}, x_i]$ and define $A_i = A \cap [x_{i-1}, x_i]$. Then define set P, such that $i \in P \subseteq \{1, ..., r \in \mathbb{N}\}$ when $A \cap [x_{i-1}, x_i] \neq \emptyset$, giving

$$\overline{f}(x) = \lim_{r\to\infty} \sum_{i \in P} f(v_i) \times \mathcal{P}(A_i, A) \tag{5}$$

(This looks tedious but one can use shortcuts to simplify the sum. If I write it out post it will be too long.)

3.3 Examples Not Well-Defined By Current Definitions of Mean But Defined By My Definition (of Mean)

3.4 First Example

Before defining function f, I will define sub-domains:

Suppose we set $G_0 = [0,1]$. We define $(G_0 + c)/d$ where if $c, d \in \mathbb{R}$, add every element in $x \in G_0$ by c and divide by d. Therefore, if we define $n \in \mathbb{Z}^+$, $c \in \{0, 2, 4\}$ and $d = 5$ then we extend G_0 to the recursion sequence G_{n+1}.

$$G_{n+1} = \frac{G_n}{5} \cup \frac{G_n + 2}{5} \cup \frac{G_n + 4}{5} \tag{6}$$

The first sub-domain, defined $\mathcal{G}$, is

$$\mathcal{G} = \bigcap_{n=0}^{\infty} G_n$$

The second sub-domain shall be $\mathcal{C}$ or the Cantor Set and the third sub-domain is $\mathbb{Q} \cap [0,1]$.

Therefore, we define $f : (\mathcal{C} \cup \mathcal{G} \cup \mathbb{Q}) \cap [0,1] \to \{1,2,3\}$:

$$f_2(x) = \begin{cases} 3 & x \in \mathcal{G} \setminus \mathbb{Q} \\ 2 & x \in \mathcal{C} \setminus (\mathcal{G} \cup \mathbb{Q}) \\ 1 & x \in \mathbb{Q} \cap [0,1] \end{cases}$$

3.5 Second Example

Consider function f where for it first sub-domain S_1,

$$S_1 = \left\{ \frac{1}{\sqrt{2s^2}} : s \in \mathbb{N} \right\}$$

And its second domain S_2

$$S_2 = \left\{ \frac{1}{2^s} + \frac{1}{2^t} : s, t \in \mathbb{N} \right\}$$

Therefore, $f_4 : S_1 \cup S_2 \to \{0,1\}$

$$f_4(x) = \begin{cases} 1 & x \in S_1 \\ 0 & x \in S_2 \end{cases}$$

3.6 Example Where $\mathcal{P}$ Is Undefined

Consider the following:

If $f : \left\{ \frac{1}{s} : s \in \mathbb{N} \right\} \to \{0,1\}$ and

$$f(x) = \begin{cases} 1 & x \in \left\{ \frac{1}{2s+1} : s \in \mathbb{N} \right\} \\ 0 & x \in \left\{ \frac{1}{2s} : s \in \mathbb{N} \right\} \end{cases}$$

Suppose $S_1 = \left\{ \frac{1}{2s+1} : s \in \mathbb{N} \right\}$ and $S_2 \in \left\{ \frac{1}{2s} : s \in \mathbb{N} \right\}$

Using $\mathcal{P}^*$, I approximated (and must prove using box-counting dimensions) $\mathcal{P}^*(S_1, S_1 \cup S_2) \geq 1/\sqrt{2}$ and $\mathcal{P}^*(S_2, S_1 \cup S_2) \geq 1/\sqrt{2}$. Hence, $\mathcal{P}_*(S_1, S_1 \cup S_2) = 1 - \mathcal{P}^*(S_2, S_1 \cup S_2) \leq 1 - 1/\sqrt{2}$ but $1/\sqrt{2} \neq 1 - 1/\sqrt{2}$. Hence with this case my measure is undefined.

Instead we need a measure that gives the same value as $\mathcal{P}(S, A)$ when $\mathcal{P}$ defined but gives different value when $\mathcal{P}(S, A)$ is undefined.

3.7 Defining Density

Suppose $A \subseteq \mathbb{R}$ and $S \subseteq A$.

And suppose $F_1, ..., F_\omega$ are a sequence of sets denoted $\{F_n\}$ where $\mathcal{P}(F_1, A), ..., \mathcal{P}(F_\omega, A)$ are defined, $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_\omega$ and $\lim\limits_{\omega\to\infty} \bigcup\limits_{n=1}^{\omega} F_n = A$

We write density of S as

$$d\left(S, \{F_n\}\right) = \lim_{\omega\to\infty} \mathcal{P}\left(S \cap F_\omega, F_\omega\right)$$

The problem is there are several ways to write $\{F_n\}$. For $A = \mathbb{Q} \cap [0,1]$, $\{F_n\}$ can be written as:

$$\{F_n\} = \left\{\frac{m}{n} : m, n \in \mathbb{N}, m \leq n \leq \omega\right\}$$

Since $\left\{\frac{m}{n} : m, n \in \mathbb{N}\right\} = \mathbb{Q}$

or

$$\{F_n\} = \left\{\frac{m}{n^2} : m, n \in \mathbb{N}, m \leq n^2 \leq \omega\right\}$$

Since $\left\{\frac{m}{n^2} : m, n \in \mathbb{N}\right\} = \mathbb{Q}$

Moreover, when A is countable and dense in an interval and S is countable and almost nowhere dense, one can come up $\{F_n\}$ where the density of S is non-zero. (This is non-intuitive.)

For instance, consider $A = \left(\mathbb{Q} \cup \left\{\frac{1}{\ln(n)} : n \in \mathbb{N}\right\}\right) \cap [0,1]$ and $S = \left\{\frac{1}{\ln n} : n \in \mathbb{N}\right\} \cap [0,1]$. One can come up with $\{F_n\} = \left\{\frac{m}{n} : m, n \in \mathbb{N}, n \neq 0, m \leq n \leq \omega\right\} \cup \left\{\frac{1}{\ln(n)} : n \in \mathbb{N}, 0 < \ln(n) \leq \omega\right\}$ which makes $d\left(S, \{F_n\}\right)$ (I assume this would be simple to prove but would like to check using mathematical proof)

Therefore, we must define a new density, where if A is dense in an interval, and S is almost nowhere dense, then S shall always have zero density.

3.8 Defining New Density

To get a unique $\{F_n\}$, we must define a choice function that can be applied to arbitrary A and coincide with $\mathcal{P}(S, A)$ when $\mathcal{P}(S, A)$ is defined.

I'm don't know how this can be done but here is my attempt:

3.8.1 Attempt

Note the density without a choice function is simply:

$$d(S, \{F_n\}) = \lim_{\epsilon \to 0} \lim_{\omega \to \infty} \mathcal{P}\left(S \cap F_\omega, F_\omega\right)$$

[I assume] we can define a choice function for any arbitrary A by doing the following:

If σ is the standard deviation of a set such that

- X is a non-fixed arbitrary set and $\mathcal{P}$-measurable
- $\overline{X}$ is the $\mathcal{P}$-average of X (Section 3.2)
- X^2 squaring all elements in X
- $\sigma\left\{X\right\} = \sqrt{\overline{X^2} - \overline{X}^2}$

Therefore, if $\mathcal{F}(A)$ represents all sequences of sets with the same properties as $\{F_n\}$ but written in different forms, with $\{H_n\}, \{G_n\} \in \mathcal{F}(A)$ then:

$$\begin{aligned}\beta = \Bigg\{ &\{G_n\} : \forall\left(\{H_n\}\right) \forall\left(\omega_2\right) \exists\left(\omega_1\right) \left(\lim_{\omega_2 \to \infty} \frac{\sigma\left\{G_{\omega_1}\right\}}{\sigma\left\{H_{\omega_2}\right\}} \leq 1 \right), \\ &\forall(H_n) \forall\left(\omega_2\right) \exists(\omega_1) \left(\lim_{\omega_2 \to \infty} \frac{\mathcal{P}\left(G_{\omega_1}, G_{\omega_1} \cup H_{\omega_2}\right)}{\mathcal{P}\left(H_{w_2}, G_{w_1} \cup H_{w_2}\right)} = \right. \\ &\left. \inf\left\{ |1 - s| : s = \lim_{\omega_2 \to \infty} \frac{\mathcal{P}\left(G_{\omega_3}, G_{\omega_3} \cup H_{\omega_2}\right)}{\mathcal{P}\left(H_{\omega_2}, G_{\omega_3} \cup H_{\omega_2}\right)}, s < 1 \right\} \right) \Bigg\}\end{aligned}$$

What this means is we want to find all $\{G_n\}$ where for all $\{H_n\}$ and ω_2 there exists an ω_1 such that

$$\frac{\sigma\left\{G_{\omega_1}\right\}}{\sigma\left\{H_{\omega_2}\right\}} \leq 1$$

and for all $\{H_n\}$ and ω_2 there exists an ω_1 such that if

$$s = \lim_{\omega_2 \to \infty} \frac{\mathcal{P}\left(G_{\omega_1}, G_{\omega_1} \cup H_{\omega_2}\right)}{\mathcal{P}\left(H_{\omega_2}, G_{\omega_1} \cup H_{\omega_2}\right)}$$

and $s < 1$ then s should be as close to 1 as possible.

Hence, with $\{F_n\} \in \beta$, we get the density:

$$\mathcal{D}(S, \{F_n\}) = \lim_{\omega \to \infty} \mathcal{P}\left(S \cap F_\omega, F_\omega\right)$$

I assume this would give us a unique value.

3.9 Defining The Average

To define the average, first split $[a, b] = [\min(A), \max(A)]$ into sub-intervals using partitions x_i

$$a = x_0 \leq \cdots \leq x_i \leq \cdots \leq x_r = b$$

Here, if $1 \leq i \leq r$, then $[x_{i-1}, x_i]$ are sub-intervals of $[a, b]$.

For every i, choose a $v_i \in A \cap [x_{i-1}, x_i]$ and define $A_i = A \cap [x_{i-1}, x_i]$. Next we define set P, such that $i \in P \subseteq \{1, ..., r \in \mathbb{N}\}$ when $A \cap [x_{i-1}, x_i] \neq \emptyset$. This gives

$$\lim_{r \to \infty} \sum_{i \in P} f(v_i) \times \mathcal{D}(A_i, A) \tag{7}$$

(This may look tedious but we can use shortcuts to simplify the sum. If I write it out the post will be too long.)

4 Examples

Here are section where one must apply the density and average:

4.1 Third Example

Suppose we want to define f_3 such that if $\mathcal{C}$ is the Cantor Set, the first sub-domain of f_3 is:

$$S_n = \bigcup_{n=0}^{\infty} \bigcup_{k=0}^{2^n-1} \left\{3^{-n}(x+k) : x \in \mathcal{C}\right\}$$

If $\mathcal{G}$ is the first sub-domain in second example (section 3.4), the second sub-domain of f_3 is:

$$T_n = \left(\bigcup_{n=0}^{\infty} \bigcup_{k=0}^{3^n-1} \left\{5^{-n}(x+k) : x \in \mathcal{G}\right\} \right) \setminus S_n$$

Therefore $f_3 : S_n \cup T_n \to \{1, 2\}$

$$f_3(x) = \begin{cases} 2 & x \in S_n \\ 1 & x \in T_n \end{cases}$$

4.2 Fourth Example

Consider function $f_4(x)$ where $f_4 : \mathbb{Q} \cap [0, 1] \to \{0, 1\}$

$$f_4(x) = \begin{cases} 1 & x \in \left(\mathbb{Q} \setminus \left\{\frac{s}{2t+1} : s, t \in \mathbb{N}\right\}\right) \cap [0, 1] \\ 0 & x \in \left\{\frac{s}{2t+1} : s, t \in \mathbb{N}\right\} \cap [0, 1] \end{cases}$$