Concept Paper

Defining the Most Generalized, Natural Extension of the Expected Value on Measurable Functions

Bharath Krishnan [*]

Concept Paper

Defining the Most Generalized, Natural Extension Of The Expected Value

Bharath Krishnan

Department of Mathematics, Indiana East University; bharathk98@gmail.com

Abstract: In this paper, we will extend the expected value of the function w.r.t the uniform probability measure on the Carathèodory extension to a larger class of functions, since the set of all functions with infinite or undefined expected values may form a prevalent subset of the set of all measurable functions. Before we get to the specific problem (or main question) of the paper, we will outline some preliminary definitions. We then will define a precise main question that will attempt to offer a unique solution and we'll offer a partial solution to the question. Along the way, we will ask a series of questions that will clarify our understanding of the paper.

Keywords: expected value; uniform measure; measure theory; prevalence; entropy; sample; linear; superlinear; choice function; Bernard's Paradox; pseudo-random

1. Background

I am an undergraduate from Indiana University despite being the age of a grad student. I should have graduated by now, but my obsession with research prevents me from moving forward. There is a chance that I might have a learning disability since writing isn't very easy for me.

As I've been in and out of college, I never got the chance to rigorously learn the subjects I'm researching. Most of what I learned was from Wikipedia, blogs and random research articles. I know little of what I read but I learn what I can from questions on math stack exchange.

What I truly want, however; is for someone to take my ideas and publish them.

I warn that the definitions may not be rigorous so try to go easy on me. (I recommend using programming such as Mathematica to understand these definitions).

2. Preliminaries

Suppose A is a set measurable in the Carathèodory sense, such for $n \in \mathbb{N}$, $A \subseteq \mathbb{R}^n$, and function $f : A \to \mathbb{R}$.

2.1. Motivation

It seems the set of all functions with infinite or undefined expected values (def. 1), using the uniform measure [1, p.32-37], may be a prevalent subset [2,3] of the set of all measurable functions, meaning "almost every" measurable function has infinite or undefined expected values. Furthermore, when the uniform measure of A, measurable in the Caratheodory sense, has zero or infinite volume (or undefined measure), there may be multiple, conflicting ways of defining a "natural" uniform measure on A.

Below I will attempt to define a question regarding an extension of the expected value (when it's undefined or infinite) which allows for a finite value instead.

Note the reason the question will be so long is there are plenty of "meaningless" extensions of the expected value (e.g. if the expected value is infinite or undefined we can just replace it with zero).

Therefore we must be more specific about what is meant by "meaningful" extension but there are some preliminary definitions we must clarify.

2.2. Preliminary Definitions

Definition 1 (Expected Value w.r.t the Uniform Probability Measure). *From an answer to a question in cross validated [4] , let* $\mathbf{X} \sim Uniform(A)$ *denote a uniform random variable on set* $A \subseteq \mathbb{R}^n$ *and* $p_{\mathbf{X}}$ *denote the probability density function from the radon-nikodym derivative [5, p.419-427] of the uniform probability measure on sets in the* σ*-algebra of Carathèodory measurable sets. If* $\mathbb{I}(x \in A)$ *denotes the indicator function on* $x \in A$:

$$\mathbb{I}(x \in A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

then the radon-nikodym derivative of uniform probability measure must have the form $\mathbb{I}(x \in A)/U'(A)$. *(Note* U' *is not the derivative of* U *in the sense of calculus but rather different from the uniform probability measure defined as* U.)

Therefore, using the law of the unconscious statistician, we should get

$$\mathbb{E}[f(\mathbf{X})] = \int_{\mathbb{R}^n} f(x) \cdot p_{\mathbf{X}}(x) \, d\mathbf{x} \tag{1}$$

$$= \int_{\mathbb{R}^n} f(\mathbf{x}) \cdot \frac{\mathbb{I}(x \in A)}{U'(A)} \, d\mathbf{x}$$

$$= \frac{1}{U'(A)} \int_A f(x) \, d\mathbf{x} \tag{P1}$$

$$= \mathbb{E}_{U'}[f(\mathbf{X})]$$

such the expected value is undefined when A *does not have a uniform probability distribution or* f *is not integrable w.r.t the measure* U'.

Definition 2 (Defining The Pre-Structure $\{F_r\}_{r \in \mathbb{N}}$**).** *Since there's a chance that* $\mathbf{X} \sim Uniform(A)$ *does not exist or* f *is not integrable w.r.t to* U', *using def. 1 we define a sequence of sets* $\{F_r\}_{r \in \mathbb{N}}$ *where:*

(1) $\bigcup_{r=1}^{\infty} F_r = A$
(2) *For all* $r \in \mathbb{N}$, $\mathbf{X}_r \sim Uniform(F_r)$ *exists (if* A *is countable infinite then for every* $r \in \mathbb{N}$, F_r *must be a finite set since* $\mathbf{X}_r$ *is a discrete uniform distribution of* F_r; *otherwise, if* A *is uncountable, then* $\mathbf{X}_r$ *is the normalized Lebesgue measure of* F_r *or some other uniform measure on* F_r *[6] where for every* $r \in \mathbb{N}$ *the Lebesgue measure or some uniform measure on* F_r *exists and is finite. [1, p.32-37].*
(3) *For all* $r \in \mathbb{N}$, $U'(F_r)$ *is positive and finite where* U' *is intrinsic. (For countably infinite* A, U' *is the counting measure, and* $U'(F_r)$ *is positive and finite since* F_r *is finite. For uncountable* A, U' *is the Lebesgue or radon-nikodym derivative on some uniform measures [6] where either of the measures on* F_r *are positive or finite.)*

then $\{F_r\}_{r \in \mathbb{N}}$ *is a* **pre-structure** *of* A, *since for every* $r \in \mathbb{N}$ *the sequence does not equal* A, *but "approaches" A as r increases.*

Definition 3 (Expected Value of F_r**).** *If* $\{F_r\}_{r \in \mathbb{N}}$ *is a pre-structure of* A *(def. 2), then for* $r \in \mathbb{N}$, *if*

$$\mathbb{E}_{U'}[f(\mathbf{X}_r)] = \frac{1}{U'(F_r)} \int_{F_r} f \, d\mathbf{x} \tag{2}$$

we then have that the expected value of the pre-structure could be described as $\mathbb{E}_{U'}[f(\mathbf{X}_r)] \to \mathbb{E}_{U'}^{\star}[f]$ *(def. 1) where:*

$$\forall(\epsilon > 0)\exists(N \in \mathbb{N})\forall(r \in \mathbb{N})(r \geq N \Rightarrow |\mathbb{E}_{U'}[f(\mathbf{X}_r)] - \mathbb{E}_{U'}^{\star}[f]| < \epsilon) \Rightarrow$$

$$\forall(\epsilon > 0)\exists(N \in \mathbb{N})\forall(r \in \mathbb{N})(r \geq N \Rightarrow |\frac{1}{U'(F_r)}\int_{F_r} f\,d\mathbf{x} - \mathbb{E}_{U'}^{\star}[f]| < \epsilon) \tag{3}$$

Definition 4 (Uniform ε coverings of F_r). *We define the uniform ε coverings of F_r as a group of pair-wise disjoint sets that covers F_r for every $r \in \mathbb{N}$, such the measure U' of each of the sets have the same value of $\varepsilon \in range(U')$ where $\varepsilon > \inf(range(U'))$ and the total sum for U' of the coverings is minimized. As a shortcut, if*

- *The element $t \in \mathbb{N}$*
- *The set $T \supset \mathbb{N}$ is arbitrary and uncountable.*

and set Ω is defined as:

$$\Omega = \begin{cases} \{1, \cdots, t\} & \text{if there are t ways of writing uniform } \varepsilon \text{ coverings of } F_r \\ \mathbb{N} & \text{if there are countably infinite ways of writing uniform } \varepsilon \text{ coverings of } F_r \\ T & \text{if there are uncountable ways of writing uniform } \varepsilon \text{ coverings of } F_r \end{cases} \tag{4}$$

then for every $\omega \in \Omega$, the set of uniform ε coverings is defined using the notation $\mathscr{U}(\epsilon, F_r, \omega)$ where ω "enumerates" all possible uniform ε coverings of F_r for every $r \in \mathbb{N}$.

Definition 5 (Sample of the uniform ε coverings of F_r). *The set of points, such for every $\varepsilon \in range(U')$ and $r \in \mathbb{N}$, we take a point from each of pair-wise disjoint set in the uniform ε coverings of F_r (def. 4). As a shortcut, if*

- *The element $k \in \mathbb{N}$*
- *The set $\mathcal{K} \supset \mathbb{N}$ is arbitrary and uncountable.*

and set Ψ_ω is defined as:

$$\Psi_\omega = \begin{cases} \{1, \cdots, k\} & \text{if there are k ways of writing the sample of uniform } \varepsilon \text{ coverings of } F_r \\ \mathbb{N} & \text{if there are countably infinite ways of writing the sample of uniform } \varepsilon \text{ coverings of } F_r \\ \mathcal{K} & \text{if there are uncountable ways of writing the sample of uniform } \varepsilon \text{ coverings of } F_r \end{cases} \tag{5}$$

then for every $\psi \in \Psi_\omega$, the set of all samples of the set of uniform ε coverings is defined using the notation $\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)$, where Ψ_ω "enumerates" all possible samples of $\mathscr{U}(\epsilon, F_r, \omega)$.

Definition 6 (Entropy on the sample of uniform coverings of F_r). *Since there are finitely many points in the sample of the uniform ε coverings of F_r (def. 5), we:*

(1) *Arrange the x-value of the points in the sample of uniform ε coverings from least to greatest. This is defined as:*

$$Ord(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))$$

(2) *Take the multi-set of the absolute differences between all consecutive pairs of elements in (1). This is defined as:*

$$\nabla Ord(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))$$

(3) *Normalize (2) into a probability distribution, where for multi-set X, we have $|X|$ as the cardinality of all elements in the multi-set, including repeated ones. This is defined as:*

$$\mathbb{P}(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)) = \{y / |\nabla Ord(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))| : y \in \nabla Ord(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))\}$$

(4) Take the *entropy* of (3), (for further reading, see [7, p.61-95]). This is defined as:

$$E(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)) = - \sum_{x \in \mathbb{P}(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))} x \log_2 x$$

where (4) is the entropy on the sample of uniform coverings of F_r.

Definition 7 (Sequence of sets converging uniformly to A). *For every $r \in \mathbb{N}$ (using def. 4, 5, and 6) if set A is finite:*

$$\lim_{\epsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} E(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)) \geq E(F_r)$$

and if set A is non-finite:

$$\lim_{\epsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} E(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)) = +\infty$$

*we then say pre-structure $\{F_r\}_{r \in \mathbb{N}}$ **converges uniformly** to A (or in shorter notation):*

$$F_r \overset{r \in \mathbb{N}}{\rightrightarrows} A \tag{6}$$

(Note we wish to define a uniform convergence of a sequence of sets to A since the definition is analogous to a uniform measure.)

Definition 8 (Equivalent Sequences of Sets). *The pre-structures $\{F_r\}_{r \in \mathbb{N}}$ and $\{F'_j\}_{j \in \mathbb{N}}$ of A are **equivalent** if, from def. 3, $(\mathbb{E}_{U'}[f(\mathbf{X}_r)] \to \mathbb{E}_{U'}^\star[f]$ and $\mathbb{E}_{U'}[f(\mathbf{X}'_j)] \to \mathbb{E}_{U'}^{\star\star}[f]) :$*

$$\forall \left(f \in \mathbb{R}^A \right) (\mathbb{E}_{U'}^\star[f] = \mathbb{E}_{U'}^{\star\star}[f])$$

Definition 9 (Non-Equivalent Sequences of Sets). *The pre-structures $\{F_r\}_{r \in \mathbb{N}}$ and $\{F'_j\}_{j \in \mathbb{N}}$ of A are **non-equivalent** if, from def. 3, $(\mathbb{E}_{U'}[f(\mathbf{X}_r)] \to \mathbb{E}_{U'}^\star[f]$ and $\mathbb{E}_{U'}[f(\mathbf{X}'_j)] \to \mathbb{E}_{U'}^{\star\star}[f])$*

$$\exists \left(f \in \mathbb{R}^A \right) (\mathbb{E}_{U'}^\star[f] \neq \mathbb{E}_{U'}^{\star\star}[f])$$

Definition 10 (Sequence converging Sublinearly, Linearly, or Superlinearly to A compared to that of another Sequence). *Suppose pre-structures $\{F_r\}_{r \in \mathbb{N}}$ and $\{F'_j\}_{j \in \mathbb{N}}$ are non-equivalent and converge uniformly to A. Also, suppose for every $\epsilon \in range(U')$ where $\epsilon > \inf(range(U'))$ and $r \in \mathbb{N}$:*

(a) We take the cardinality of the sample of the uniform ϵ coverings of F_r (def. 5) divided by the smallest cardinality of the sample of the uniform ϵ coverings of F_j (def. 5), where the entropy on the sample of uniform coverings on F_j (def. 6) is larger than the entropy on the sample of uniform coverings on F_r (def. 6). In other words, if:

$$\overline{\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)} = \tag{7}$$
$$\inf \left\{ |\mathcal{S}(\mathscr{U}(\epsilon, F'_j, \omega'), \psi')| : j \in \mathbb{N}, \ \omega' \in \Omega, \ \psi' \in \Psi_\omega, \ E(\mathcal{S}(\mathscr{U}(\epsilon, F'_j, \omega'), \psi')) \geq E(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)) \right\}$$

then the ratio at the beginning of the paragraph is defined (using 7) as

$$\overline{\alpha}(\epsilon, r, \omega, \psi) = |\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))| / \overline{|\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)|} \tag{8}$$

(b) We take the cardinality of the sample of uniform ϵ covering of F_r (def. 5) divided by the largest cardinality of the sample of the uniform ϵ covering of F_j (def. 5), where the entropy on the sample of

uniform coverings on F_j (def. 6) is smaller then the entropy on the sample of uniform coverings on F_r (def. 6). In other words if:

$$\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi) = \tag{9}$$

$$\sup\left\{ |\mathcal{S}(\mathscr{U}(\epsilon, F_j', \omega'), \psi')| : j \in \mathbb{N}, \ \omega' \in \Omega, \ \psi' \in \Psi_\omega, \ E(\mathcal{S}(\mathscr{U}(\epsilon, F_j', \omega'), \psi')) \le E(\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)) \right\}$$

then the ratio at the beginning of the paragraph is defined (using 9) as

$$\underline{\alpha}(\epsilon, r, \omega, \psi) = |\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi))| \, / \, |\mathcal{S}(\mathscr{U}(\epsilon, F_r, \omega), \psi)| \tag{10}$$

(1)   *If using equations 8 and 10 we have that:*

$$\lim_{\varepsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \overline{\alpha}(\epsilon, r, \omega, \psi) = \lim_{\varepsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \underline{\alpha}(\epsilon, r, \omega, \psi) = 0$$

we say $\{F_r\}_{r \in \mathbb{N}}$ *converges uniformly to A at a **superlinear rate** to that of* $\{F_j'\}_{j \in \mathbb{N}}$.

(2)   *If using equations 8 and 10 we have that:*

$$0 < \lim_{\varepsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \overline{\alpha}(\epsilon, r, \omega, \psi) = \lim_{\varepsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \underline{\alpha}(\epsilon, r, \omega, \psi) < +\infty$$

we say $\{F_r\}_{r \in \mathbb{N}}$ *converges uniformly to A at a **linear rate** to that of* $\{F_j'\}_{j \in \mathbb{N}}$.

(3)   *If using equations 8 and 10 we have that:*

$$\lim_{\varepsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \overline{\alpha}(\epsilon, r, \omega, \psi) = \lim_{\varepsilon \to \inf(range(U'))} \sup_{r \in \mathbb{N}} \sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \underline{\alpha}(\epsilon, r, \omega, \psi) = +\infty$$

we say $\{F_r\}_{r \in \mathbb{N}}$ *converges uniformly to A at a **sublinear rate** to that of* $\{F_j'\}_{j \in \mathbb{N}}$.

I assume $\overline{\alpha}$ *and* $\underline{\alpha}$ *are always equal but I'm not sure how to prove this.*

2.3. Question on Preliminary Definitions

(1)   Are there "simpler" alternatives to either of the preliminary definitions? (Keep this in mind as we continue reading forward).

3. Main Question

Does there exist a unique extension (or a method that constructively defines a unique extension) of the expected value of f when the value's finite (def. 1), using the radon-nikodym derivative [5, p.419-427] of the uniform probability measure [1, p.32-37] on sets measurable in the Carathèodory sense, such we replace f with infinite or undefined expected values with f defined on a pre-structure where:

(1)   The expected value of f on each term of the pre-structure is finite
(2)   The pre-structure *converges uniformly* to A (def. 7)
(3)   The pre-structure *converges uniformly* to A at a *linear or superlinear* rate to that of other non-equivalent pre-structures of A (def. 9 & 10) which satisfy (1) and (2).
(4)   The generalized expected value [1] of f on a pre-structure satisfying (1), (2), and (3) (while the pre-structure *converges uniformly* to A) is finite (i.e. def. 3)
(5)   A choice function is defined which chooses a pre-structure satisfying (1), (2), (3), and (4) such the generalized expected value [1] is unique and finite for the largest subset of $\mathbb{R}^A$.

---

[1]   The result of algebraic manipulation on the expected value in def. 3 that is unique and finite for the largest subset of $\mathbb{R}^A$.

(6) Out of all the choice functions which satisfy (1), (2), (3), (4) and (5), we choose the choice function with the "simplest form", (meaning for a general pre-structure of A, when each choice function is fully expanded, we take the choice function with the fewest variables/numbers (excluding those with quantifiers) for which the variables are added and exponentiated by infinitesimal amounts and multiplied by the difference of one and infinitesimal amount.

How do we answer this question?

4. Informal Attempt to Answer Main Question

4.1. Choice Function

Suppose $\mathbb{S}'(A)$ is the set of all pre-structures of A which satisfy criteria (1) and (2) of the question in sec. 3 where the expected value of the pre-structures, as they converge uniformly to A is finite (def. 7), where the pre-structure $\{F''_r\}_{r \in \mathbb{N}} \in \mathbb{S}'(A)$ should be a sequence of sets that satisfy criteria (1), (2), (3), (4), and (5) of the question such that $\{F'_j\}_{j \in \mathbb{N}} \in \mathbb{S}''(A)$ represents all other non-equivalent pre-structures of A from $\{F''_r\}_{r \in \mathbb{N}}$. (Note I'm unable to find a choice function which satisfies criteria (6)).

Note from (a) and (b) (in def. 10), we represent the "smallest cardinality of the sample of the uniform ε covering of F_j (def. 5), where the entropy on the sample of uniform coverings on F_j (def. 6) is larger the entropy on the sample of uniform coverings on F_r (def. 6)" to be defined as follows:

$$\overline{\mathcal{S}(\mathcal{U}(\epsilon, F''_r, \omega), \psi)} = \tag{11}$$
$$\inf \left\{ |\mathcal{S}(\mathcal{U}(\epsilon, F'_j, \omega'), \psi')| : j \in \mathbb{N}, \ \omega' \in \Omega, \ \psi' \in \Psi_\omega, \ E(\mathcal{S}(\mathcal{U}(\epsilon, F'_j, \omega'), \psi')) \geq E(\mathcal{S}(\mathcal{U}(\epsilon, F_r, \omega), \psi)) \right\}$$

and we represent the "largest cardinality of the sample of the uniform ε covering of F_j (def. 5), where the entropy on the sample of uniform coverings on F_j (def. 6) is smaller then the modified entropy on the sample of uniform coverings on F_r (def. 6)" to be defined as follows:

$$\underline{\mathcal{S}(\mathcal{U}(\epsilon, F''_r, \omega), \psi)} = \tag{12}$$
$$\sup \left\{ |\mathcal{S}(\mathcal{U}(\epsilon, F'_j, \omega'), \psi')| : j \in \mathbb{N}, \ \omega' \in \Omega, \ \psi' \in \Psi_\omega, \ E(\mathcal{S}(\mathcal{U}(\epsilon, F'_j, \omega'), \psi')) \geq E(\mathcal{S}(\mathcal{U}(\epsilon, F_r, \omega), \psi)) \right\}$$

Therefore, using def. 5 and equations 11 and 12, if we take:

$$\sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \overline{\mathcal{S}(\mathcal{U}(\epsilon, F''_r, \omega), \psi)} = \overline{\mathcal{S}'}(\varepsilon, F''_r) = \overline{\mathcal{S}'} \tag{13}$$

$$\sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \underline{\mathcal{S}(\mathcal{U}(\epsilon, F''_r, \omega), \psi)} = \underline{\mathcal{S}'}(\varepsilon, F''_r) = \underline{\mathcal{S}'} \tag{14}$$

$$\sup_{\omega \in \Omega} \sup_{\psi \in \Psi_\omega} \mathcal{S}(\mathcal{U}(\epsilon, F''_r, \omega), \psi) = \mathcal{S}'(\varepsilon, F''_r) = \mathcal{S}' \tag{15}$$

and we take:

$$S(r) = \left(\sup(F''_{r+1}) - \sup(F''_r)\right)\left(\inf(F''_r) - \inf(F''_{r+1})\right) \big| \left(\inf(F''_r) - \inf(F''_{r+1})\right)\left(\sup(F''_{r+1}) - \sup(F''_r) - 1\right) \big| \tag{16}$$

such that

$$T(r) = \left(\sup(F''_{r+1})\inf(F''_r) - \sup(F''_r)\inf(F''_{r+1})\right)\left(\left(\inf(F''_r) - \inf(F''_{r+1})\right) - \left(\sup(F''_{r+1}) - \sup(F''_r)\right) - 1 \right) \tag{17}$$
$$\left(\inf(F''_r) - \inf(F''_{r+1})\right)\left(\sup(F''_{r+1}) - \sup(F''_r)\right)$$

then, using equations 13, 14, 15, 16 and 17, for the nearest integer function $[\times]$ of equations that replace $\times$ (and the absolute value function $||\times||$ of equations that replace $\times$) we want:

$$K(\varepsilon, F''_r) = ||1 - S(r)|| \left(\left|\left|\frac{|\mathcal{S}'| \left(1 + \left[\frac{|\mathcal{S}'|(|\mathcal{S}'|+2|\mathcal{S}'|)}{(|\mathcal{S}'|+|\mathcal{S}'|)(|\mathcal{S}'|+|\mathcal{S}'|+\overline{|\mathcal{S}'|})}\right]\right)\left(1 + \left[\underline{|\mathcal{S}'|}/|\mathcal{S}'|\right]\right)}{\left(1 + \left[|\mathcal{S}'|/\overline{|\mathcal{S}'|}\right]\right)\left(1 + \left[\underline{|\mathcal{S}'|}/\overline{|\mathcal{S}'|}\right]\right)} - |\mathcal{S}'|\right|\right| + |\mathcal{S}'| \right) - T(r) \qquad (18)$$

such, using equations 15 and 18, if set $S''(A) \subseteq S'(A)$, the choice function is the following set (when the set contains only one element):

$$C(A) = \qquad\qquad (19)$$

$$\sup \Bigg\{ S''(A) : S''(A) \subseteq S'(A),\ \exists (M \in \mathbb{N}) \forall (\varepsilon \in \text{range}(U')) \exists\, (j \in \mathbb{N}) \vee (r \in \mathbb{N}) \vee (\{F''_r\} \in \mathbb{S}''(A))$$

$$\left(\inf\left(\text{range}(U')\right) \leq \varepsilon \leq M,\ r \geq j \Rightarrow K(\varepsilon, F''_r) \leq \mathcal{S}'(\varepsilon, F''_r) \right) \Bigg\}$$

4.2. Questions on Choice Function

Suppose for $k \in \mathbb{N}$, $C^k(A)$ represents the k-th iteration of the choice function of A, e.g. $C^3(A) = C(C(C(A)))$, where the infinite iteration of $C(A)$ (if it exists) is $\lim_{k \to \infty} C^k(A) = C^\infty(A)$. Therefore, when taking the following:

$$C'(A) = \begin{cases} C(A) & \text{if } C(A) \text{ contains one element} \\ C^j(A) & \text{if } j \in \mathbb{N}, \text{ such for all } k \geq j,\ C^k(A) \text{ contains one element} \\ C^\infty(A) & \text{if it exists, and } C^\infty(A) \text{ contains one element} \end{cases}$$

(1)   What unique pre-structure would $C'(A)$ contain (if it exists) for:

  (a) $\mathbb{Z}$ where if $\{F_r\}_{r\in\mathbb{N}} \in C'(\mathbb{Z})$, we want $\{F_r\}_{r\in\mathbb{N}} = \{\{m \in \mathbb{Z} : -r \leq m \leq r\}\}_{r\in\mathbb{N}}$
  (b) $\mathbb{Q}$ where if $\{F_r\}_{r\in\mathbb{N}} \in C'(\mathbb{Q})$, we want $\{F_r\}_{r\in\mathbb{N}} = \{\{s/t! : s \in \mathbb{Z},\ t \in \mathbb{N},\ -rt! \leq s \leq rt!\}\}_{r\in\mathbb{N}}$
  (c) $\mathbb{R}$ where we're not sure what $\{F_r\}_{r\in\mathbb{N}} \in C'(\mathbb{R})$ would be in this case. What would $\{F_r\}_{r\in\mathbb{N}}$ be if it's unique?

(2)   There are a total of 113 variables in the choice function C (excluding quantifiers). Is there a choice function with fewer variables that answers criteria (1), (2), (3), (4) & (5) of the question in sec. 3 for a "larger" subset of $\mathbb{R}^A$? (This might be impossible to answer since such a solution cannot be shown with prevalence or shyness [2,3]). Therefore, we need a more precise version of "size" with some examples (as shown in this answer [8]) being the following:

  (a)   Fractal Dimension notions
  (b)   Kolmogorov Entropy
  (c)   Baire Category and Porosity

4.3. Generalized Expected Values

Using the choice function in section 4.1, if the image of f under A is $f[A] := \{f(x) : x \in A\}$, we take $\{F''_r\} \in C(f[A])$ and then take the pre-image under f of F_r (which is defined as $f^{-1}[F''_r]$) where:

$$f^{-1}\left[F''_r\right] \overset{r\in\mathbb{N}}{\rightrightarrows} A$$

However, the expected value of $f^{-1}\left[F_r''\right]$ (def. 3) may be infinite (e.g. unbounded f). Hence, for every $r \in \mathbb{N}$, we take $\left\{\left\{F_{t_r}'''\right\}_{t_r \in \mathbb{N}}\right\}_{r \in \mathbb{N}} \in C(f^{-1}\left[F_r''\right])$ where:

$$\forall(r \in \mathbb{N})\left(F_{t_r}''' \overset{t_r \in \mathbb{N}}{\rightrightarrows} F_r''\right)$$

Thus, if there exists a unique and finite $\ddot{\mathbb{E}}[f]$ where:

$$\forall(\epsilon > 0)\exists(N \in \mathbb{N})\forall(r \in \mathbb{N})\forall(t_r \in \mathbb{N})\vee\left(\left\{\left\{F_{t_r}'''\right\}_{t_r \in \mathbb{N}}\right\}_{r \in \mathbb{N}} \in C\left(f^{-1}\left[F_r''\right]\right)\right) \qquad (20)$$

$$\left(r \geq N, t_r \geq N \Rightarrow \frac{1}{U'\left(F_{t_r}'''\right)}\int_{F_{t_r}'''} f\, d\mathbf{x} - \ddot{\mathbb{E}}[f] < \epsilon\right)$$

Then $\ddot{\mathbb{E}}[f]$ is the **generalized expected value w.r.t choice function** C, which answers criteria (1), (2), (3), (4), (perhaps (5)) of the question in sec. 3; however, there is still a chance that the equation 20 fails to give an unique $\ddot{\mathbb{E}}[f]$. Hence; if $k \in \mathbb{N}$, we take the k-th iteration of the choice function C in 19, such there exists a $j \in \mathbb{N}$, where for all $k \geq j$, the new expected value $\mathbb{E}^{\dagger}[f]$ (or the **generalized expected value w.r.t finitely iterated** C) is unique and finite.

　　Hence, if the k-th iteration of C is represent as $C^{[k]}$ (where e.g. $C^3(f[A]) = C(C(C(f[A])))$), we want a unique $\mathbb{E}^{\dagger}[f]$ where:

$$\forall(\epsilon > 0)\exists(N \in \mathbb{N})\forall(r \in \mathbb{N})\forall(t_r \in \mathbb{N})\exists(j \in \mathbb{N})\forall(k \in \mathbb{N})\left(k \geq j \Rightarrow \right. \qquad (21)$$

$$\left. \forall\left(\left\{\left\{F_{t_r}'''\right\}_{t_r \in \mathbb{N}}\right\}_{r \in \mathbb{N}} \in C^{[j]}\left(f^{-1}\left[F_r''\right]\right)\right)\left(r \geq N, t_r \geq N \Rightarrow \frac{1}{U'\left(F_{t_r}'''\right)}\int_{F_{t_r}'''} f\, d\mathbf{x} - \mathbb{E}^{\dagger}[f] < \epsilon\right)\right)$$

If this still does not give a unique and finite expected value, we then take the **most generalized expected value w.r.t an infinitely iterated** C i.e. $\mathbb{E}^{\ddagger}[f]$ where if the *infinite iteration* of C is stated as $\lim_{k \to \infty} C^{[k]}(f[A]) = C^{\infty}(f[A])$, we then take:

$$\forall(\epsilon > 0)\exists(N \in \mathbb{N})\forall(r \in \mathbb{N})\forall(t_r \in \mathbb{N}) \qquad (22)$$

$$\forall\left(\left\{\left\{F_{t_r}'''\right\}_{t_r \in \mathbb{N}}\right\}_{r \in \mathbb{N}} \in C^{\infty}\left(f^{-1}\left[F_r''\right]\right)\right)\left(r \geq N, t_r \geq N \Rightarrow \frac{1}{U'\left(F_{t_r}'''\right)}\int_{F_{t_r}'''} f\, d\mathbf{x} - \mathbb{E}^{\ddagger}[f] < \epsilon\right)$$

　　However, the averages $\ddot{\mathbb{E}}[f]$, $\mathbb{E}^{\dagger}[f]$, and $\mathbb{E}^{\ddagger}[f]$ in equations 20, 21, and 22 (respectively) should only be attempted for functions where *the expected value is infinite or undefined* or for **worst-case functions**—poorly behaved $f : A \to \mathbb{R}$ (where for $n \in \mathbb{N}$, $A \subseteq \mathbb{R}^n$, where f is a function) defined on infinite points covering an infinite expanse of space. For example:

(1) For a worst-case f defined on countably infinite A (e.g. countably infinite "pseudo-random points" non-uniformly scattered across the real plane), one might typically use $\ddot{\mathbb{E}}[f]$ (20) (since countable sets may need just one iteration of C for the generalized expected value to be unique); otherwise, one may use $\mathbb{E}^{\dagger}[f]$ (21) for finite iterations of C.

(2) For a worst-case f defined on uncountable A, we may have to use $\mathbb{E}^{\ddagger}[f]$ (22) as the function is so difficult to analyze. We can imagine this function as an uncountable number of "pseudo-random" points non-uniformly generated on a subset of the real plane (see 5.1 for a visualization.)

Note that no matter how generalized and "meaningful" the extension of an expected value is, there will always be an f where the expected value does not exist.

4.4. Questions Regarding The Answer

(1) Using prevalence and shyness [2,3], can we say the set of f where $\ddot{\mathbb{E}}[f]$ (20), $\mathbb{E}^{\dagger}[f]$ (21), or $\mathbb{E}^{\ddagger}[f]$ (22) are unique and finite form either a **prevalent** or *neither prevalent nor shy* subset of $\mathbb{R}^A$. (If the subset is *prevalent*, this implies the generalized expected values are unique and finite for a "large" subset of $\mathbb{R}^A$; however, if the subset is *neither prevalent nor shy* we need a more precise definition of "size" which takes "an exact probability that the expected values are unique & finite"—some examples (as shown in this answer [8]) being:

    (a)   Fractal Dimension notions
    (b)   Kolmogorov Entropy
    (c)   Baire Category and Porosity

(2) Can either $\ddot{\mathbb{E}}[f]$ (20), $\mathbb{E}^{\dagger}[f]$ (21), or $\mathbb{E}^{\ddagger}[f]$ (22) (when A is the set of all Liouville numbers & $f = \mathrm{id}_A$) give a finite value on the Liouville numbers? What would the value be?

(3) Similar to how def. 11 approximates the expected value in def. 1, how do approximate $\ddot{\mathbb{E}}[f]$ (20), $\mathbb{E}^{\dagger}[f]$ (21), and $\mathbb{E}^{\ddagger}[f]$ (22)?

(4) Can we use programming to estimate $\ddot{\mathbb{E}}[f]$ (20), $\mathbb{E}^{\dagger}[f]$ (21), and $\mathbb{E}^{\ddagger}[f]$ (22) (if a unique and finite results exists)?

4.5. Application

(1) In Quanta magazine [9], Wood writes on Feynman Path Integrals: "No known mathematical procedure can *meaningfully average*[2] an infinite number of objects covering an infinite expanse of space in general. The path integral is more of a physics philosophy than an exact mathematical recipe."—despite Wood's statement, mathematicians Bottazzi E. and Eskew M. [10] found a constructive solution to the statement using integrals defined on filters over families of finite sets; however, the solution was not unique as one has to choose a value in a partially ordered ring of infinite and infinitesimal elements. In addition, although there were ways of preventing the use of the axiom of choice (within their integral), the axiom was still required for certain cases.

    (a)   Perhaps, if Botazzi's and Eskew's Filter integral [10] is not enough to solve Wood's statement, could we replace the path integral with expected values $\ddot{\mathbb{E}}[f]$ (20), $\mathbb{E}^{\dagger}[f]$ (21), and $\mathbb{E}^{\ddagger}[f]$ (22)? (See, again, sec. 5.1 for a visualization.)

(2) As stated in sec. 2.1, "when the uniform measure of A, measurable in the Caratheodory sense, has zero or infinite volume (or undefined measure), there may be multiple, conflicting ways of defining a "natural" uniform measure on A." This is an example of Bertand's Paradox which shows, "the principle of indifference (that allows equal probability among all possible outcomes when no other information is given) may not produce definite, well-defined results for probabilities if applied uncritically when the domain of possibilities (i.e. sample space) is infinite [11].

Using sec. 4.2, perhaps if we take:

$$C'(A) = \begin{cases} C(A) & \text{if } C(A) \text{ contains one element} \\ C^j(A) & \text{if } j \in \mathbb{N}, \text{ such for all } k \geq j, C^k(A) \text{ contains one element} \\ C^\infty(A) & \text{if it exists, and } C^\infty(A) \text{ contains one element} \end{cases}$$

---

[2]   Meaningful Average—The measure inside the average is canonical when the measure is normalized as a uniform probability measure [1, p. 32-37]

then for $\{F_r\}_{r\in\mathbb{N}} \in C'(A)$, if $S \subseteq A$ if we have the following:

$$\forall(\epsilon > 0)\exists(N \in \mathbb{N})\forall(r \in \mathbb{N})\left(r \geq N \Rightarrow \frac{U'(S \cap F_r)}{U'(F_r)} - \mathscr{U}(S) < \epsilon\right) \tag{23}$$

Then $\mathscr{U}(S)$ might serve as a solution to Bertand's Paradox (unless there is a simpler solution to the main question in sec. 3).

(a)   How do we apply $\mathscr{U}(S)$ (or a simpler solution) to the usual example which demonstrates the Bertand's Paradox as follows: for an equilateral triangle (inscribed in a circle), suppose a chord of the circle is chosen at random—what is the probability that the chord is longer than a side of the triangle? [12] (According to Bertand's Paradox there are three arguments which correctly use the principle of indifference yet give different solutions to this problem [12]:

   (i)   The "random endpoints" method: Choose two random points on the circumference of the circle and draw the chord joining them. To calculate the probability in question imagine the triangle rotated so its vertex coincides with one of the chord endpoints. Observe that if the other chord endpoint lies on the arc between the endpoints of the triangle side opposite the first point, the chord is longer than a side of the triangle. The length of the arc is one-third of the circumference of the circle, therefore the probability that a random chord is longer than a side of the inscribed triangle is $1/3$.

   (ii)   The "random radial point" method: Choose a radius of the circle, choose a point on the radius, and construct the chord through this point and perpendicular to the radius. To calculate the probability in question imagine the triangle rotated so a side is perpendicular to the radius. The chord is longer than a side of the triangle if the chosen point is nearer the center of the circle than the point where the side of the triangle intersects the radius. The side of the triangle bisects the radius, therefore the probability a random chord is longer than a side of the inscribed triangle is $1/2$.

   (iii)   The "random midpoint" method: Choose a point anywhere within the circle and construct a chord with the chosen point as its midpoint. The chord is longer than a side of the inscribed triangle if the chosen point falls within a concentric circle of radius $1/2$ the radius of the larger circle. The area of the smaller circle is one-fourth the area of the larger circle, therefore the probability a random chord is longer than a side of the inscribed triangle is $1/4$.

5. Glossary

5.1. Example of Case (2) of Worst Case Functions

(If the explanation below is difficult to understand, see this visualization to accompany the explanation [13], then when changing the sliders each time, wait a couple of seconds for the graph to load.)

We wish to create a function that appears to be a "pseudo-randomly" distributed but has infinite points that are non-uniform (i.e. does not have complete spatial randomness [14]) in the sub-space of $\mathbb{R}^2$, where the expected value or integral of the function w.r.t uniform probability measure [1][p.32-37] is non-obvious (i.e. not the center of the space the function covers nor the area of that space).

Suppose for real numbers x_1, x_2, y_1 and y_2, we generate an uncountable number of "nearly pseudo-random" points that are non-uniform in the subspace $[x_1, x_2] \times [y_1, y_2] \subseteq \mathbb{R}^2$.

We define the function as $f : [x_1, x_2] \to [y_1, y_2]$.

Now suppose $b \in \{2, 3, \cdots, 10\}$ where the base-b expansion of real numbers, in interval $[x_1, x_2]$, have infinite decimals that approach x from the right side so when $x_1 = x_2$ we get $f(x_1) = f(x_2)$.

Furthermore, for $\mathbb{N} \cup \{0\} = \mathbb{N}_0$, if $r \in \mathbb{N}_0$ and $\text{digit}_b : \mathbb{R} \times \mathbb{Z} \to \{0, 1, \cdots, b-1\}$ is a function where $\text{digit}_b(x, r)$ takes the digit in the b^r-th decimal fraction of the base-b expansion of x (e.g. $\text{digit}_{10}(1.789, 2) = 8$), then $\{g_r{'}\}_{r \in \mathbb{N}_0}$ is a sequence of functions such that $g_r{'} : \mathbb{N}_0 \to \mathbb{N}_0$ is defined to be:

$$g_r'(x) = \left[\frac{10}{b} \sin(rx) + \frac{10}{b} \right] \tag{24}$$

then for some large $k \in \mathbb{N}$ and $x_1, x_2 \in \mathbb{R}$, the intermediate function (before f) or $f_1 : [x_1, x_2] \to \mathbb{R}$ is defined to be

$$f_1(x) = \left| \left(\sum_{r=0}^{\infty} g_{r+1}' \left(\sum_{p=r}^{r+k} \text{digit}_b(x, p) \right) \middle/ b^r \right) - 10 \right| =$$

$$\left| \left(\left(\sum_{r=0}^{\infty} \left[\frac{10}{b} \sin \left((r+1) \left(\sum_{p=r}^{r+k} \text{digit}_b(x, p) \right) \right) + \frac{10}{b} \right] \right) \middle/ b^r \right) - 10 \right|$$

where the points in f_1 are "almost pseudo-randomly" and non-uniformly distributed on $[x_1, x_2] \times [0, 10]$. What we did was convert every digit of the base-b expansion of x to a pseudo-random number that is non-equally likely to be an integer, including and also in-between, 0 and $(10 \cdot 10^s)/b$. Furthermore, we make the function truly "appear pseudo-random", by adding the b^r-th decimal fraction with the next k decimal fractions; however, we also want to control the end-points of $[0, 10^{s+1}]$ such if $y_1, y_2 \in \mathbb{R}$, we convert $[x_1, x_2] \times [0, 10]$ to $[x_1, x_2] \times [y_1, y_2]$ by manipulating equation 25 to get:

$$f(x) = y_2 - \frac{y_2 - y_1}{10} f_1(x)$$

$$y_2 - \left(\frac{y_2 - y_1}{10} \right) \left| \left(\left(\sum_{r=0}^{\infty} \left[\frac{10}{b} \sin \left((r+1) \left(\sum_{p=r}^{r+k} \text{digit}_b(x, p) \right) \right) + \frac{10}{b} \right] \right) \middle/ b^r \right) - 10 \right|$$

such that the larger k is, the more pseudo-random the distribution of points in f in the space $[x_1, x_2] \times [y_1, y_2]$ but unlike most distributions of these points, f is uncountable.

5.2. Question Regarding Case (2) of The Worst Case Function

Let us be more specific, suppose for the function in equation 25 of sec. 5.1, we have:

- $b = 3$
- $[x_1, x_2] \times [y_1, y_2] = [0, 1] \times [0, 1]$
- $k = 100$

(one could try simpler parameters); what is the expected value using either $\mathbb{E}^{\dagger}[f]$ (21), or $\mathbb{E}^{\ddagger}[f]$ (22) if the answer is finite and unique?

What about for f in general (i.e. in terms of b, x_1, x_2, y_1, y_2 and k)?

(Note if $x_1, y_1 \to -\infty$ and $x_2, y_2 \to \infty$, then the function is an explicit example of the function that Wood [3] describes in Quanta Magazine)

---

[3]   Wood wrote on Feynman Path Integrals: "No known mathematical procedure can *meaningfully average* [2] an infinite number of objects covering an infinite expanse of space in general. The path integral is more of a physics philosophy than an exact mathematical recipe."

5.3. Approximating the Expected Value

Definition 11 (Approximating the Expected Value). *In practice, the computation of this expected value may be complicated if the set A is complicated. If analytic integration does not give a closed-form solution then a general and relatively simple way to compute the expected value (up to high accuracy) is with importance sampling. To do this, we produce values* $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_M \sim$ *IID g for some density function g with support* $A \subseteq support(g) \subseteq \mathbb{R}^n$ *(hopefully with support fairly close to A) and we use the estimator:*

$$\hat{\mu}_M \equiv \frac{\sum_{i=1}^{M} \mathbb{I}(\mathbf{X}_i \in A) \cdot f(\mathbf{X}_i)/g(\mathbf{X}_i)}{\sum_{i=1}^{M} \mathbb{I}(\mathbf{X}_i \in A)/g(\mathbf{X}_i)} \tag{25}$$

From the law of large numbers, we can establish that $\mathbb{E}[f(\mathbf{X})] = \lim_{M \to \infty} \hat{\mu}_M$ *so if we take M to be large then we should get a reasonably good computation of the expected value of interest.*

Note importance sampling requires three things:

(1) *We need to know when point x is in set A or not*
(2) *We need to be able to generate points from a density g that is on a support that covers A but is not too much bigger than A*
(3) *We have to be able to compute $f(x)$ and $g(x)$ for each point $x \in A$*

References

1. T., L.; E., R. The maximum entropy of a metric space. https://arxiv.org/pdf/1908.11184.pdf.
2. Ott, W.; Yorke, J.A. Prevelance. *Bulletin of the American Mathematical Society* **2005**, *42*, 263–290. https://www.ams.org/journals/bull/2005-42-03/S0273-0979-05-01060-8/S0273-0979-05-01060-8.pdf.
3. Hunt, B.R. Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces **1992**. https://arxiv.org/abs/math/9210220, doi:10.48550/ARXIV.MATH/9210220.
4. (https://stats.stackexchange.com/users/173082/ben), B. In statistics how does one find the mean of a function w.r.t the uniform probability measure? Cross Validated, [https://stats.stackexchange.com/q/602939]. https://stats.stackexchange.com/q/602939 (version: 2023-01-24).
5. B., P. 3 ed.; John Wiley & Sons: New York, 1995; pp. 419–427. https://www.colorado.edu/amath/sites/default/files/attached-files/billingsley.pdf.
6. (https://mathoverflow.net/users/46214/mark mcclure), M.M. Integral over the Cantor set Hausdorff dimension. MathOverflow, [https://mathoverflow.net/q/235609]. https://mathoverflow.net/q/235609 (version: 2016-04-07).
7. M., G. 2 ed.; Springer New York: New York [America];, 2011; pp. 61–95. https://ee.stanford.edu/~gray/it.pdf, doi:https://doi.org/10.1007/978-1-4419-7970-4.
8. (https://math.stackexchange.com/users/13130/dave-l renfro), D.L.R. Proof that neither "almost none" nor "almost all" functions which are Lebesgue measurable are non-integrable. Mathematics Stack Exchange, [https://math.stackexchange.com/q/4623168]. https://math.stackexchange.com/q/4623168 (version: 2023-01-21).
9. C., W. Mathematicians Prove 2D Version of Quantum Gravity Really Works. *Quanta Magazine*. https://www.quantamagazine.org/mathematicians-prove-2d-version-of-quantum-gravity-really-works-20210617.
10. E., B.; M., E. Integration with Filters. https://arxiv.org/pdf/2004.09103.pdf.
11. Shackel, N. Bertrand's Paradox and the Principle of Indifference. *Philosophy of Science* **2007**, *74*, 150–175. https://orca.cardiff.ac.uk/id/eprint/3803/1/Shackel%20Bertrand's%20paradox%205.pdf, doi:10.1086/519028.
12. Drory, A. Failure and Uses of Jaynes' Principle of Transformation Groups. *Foundations of Physics* **2015**, *45*, 439–460. https://arxiv.org/pdf/1503.09072.pdf, doi:10.1007/s10701-015-9876-7.
13. B., K. Visualization of Uncountable Number of Psuedo-random Points Generate on Subset of the Real Plane, 2023. https://www.wolframcloud.com/obj/4e78f594-1578-402a-a163-ebb16319ada2.

14.   Maimon O., R.L. 2 ed.; Springer New York: New York [America];, 2010; pp. 851–852. https://doi.org/10.1007/978-0-387-09823-4, doi:https://doi.org/10.1007/978-0-387-09823-4.