When Large also is Small

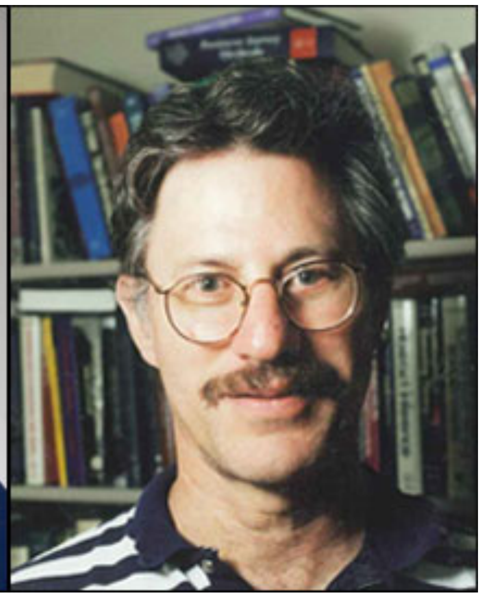
conflicts between Measure Theoretic and Topological senses of a <u>negligible</u> set

Teddy Seidenfeld

Collaboration: Jessica Cisewski, Jay Kadane, Mark Schervish, and Rafael Stern

Statistics Depts – Yale, Carnegie Mellon University (x 2), and U. Federal de São Carlos

Jessi



Jay

Mark

Rafael

<u>Outline</u>

1. **Review two strong-laws that rely on measure 0 as the sense of a *negligible* set: where the probability-1 law fails.**

2. **A topological sense of a *negligible* set – *meager* (or *1st category*) sets.**

3. **Oxtoby's (1957, 1980) results – where the two senses of *negligible* conflict.**

4. **A generalization of Oxtoby's (1957) result.**

5. **Some concluding thoughts on where these two formal perspectives on *negligible* sets do and do not play well together.**

1. Two philosophically significant, strong-laws that rely on measure 0 as the sense of a <u>negligible</u> set: where the law fails. [See, e.g. Schervish (1995).]

- **The *strong law of large numbers* for independent, identically distributed (*iid*) Bernoulli trials – connecting *chance* with limiting relative frequency.**

Let X be a Bernoulli variable sample space {0, 1}, with $P(X = 1) = p$, for $0 \leq p \leq 1$.

Let X_i ($i = 1, 2, \ldots$) be a denumerable sequence of Bernoulli variables,

with a common parameter $P(X_i = 1) = p$ and where trials are independent.

***Independence* is expressed as follows.**

$$\text{For each } n = 1, 2, \ldots, \text{ let } S_n = \sum_{i=1}^{n} X_n.$$

$$\text{Then } P(X_1 = x_1, \ldots, X_n = x_n) = p^{S_n} \times (1 - p)^{(n - S_n)}.$$

The *weak-law* of large numbers for *iid* Bernoulli trials:

$$\text{For each } \varepsilon > 0, \qquad \lim_{n\to\infty} P(|S_n/n - p| > \varepsilon) = 0.$$

The *strong-law* of large numbers for *iid* Bernoulli trials:

$$P(\lim_{n\to\infty} S_n/n = p) = 1.$$

If P is countably additive, the strong-law version entails the weak-law version.

Let < X, $\mathcal{B}$, P > be the countably additive *measure space* generated by all finite sequences of repeated, probabilistically independent [*iid*] flips of a "fair" coin.

Let 1 denote a "Heads" outcome and 0 a "Tails" outcome for each flip.

Then a point x of X is a denumerable sequence of 0s and 1s,

$$x = \langle x_1, x_2, \ldots \rangle, \text{ with each } x_n \in \{0, 1\} \text{ for } n = 1, 2, \ldots$$

and where $X_n(x) = x_n$ designates the outcome of the n^{th} flip of the fair coin.

$\underline{\mathcal{B}}$ is the Borel σ-algebra generated by *rectangular* events, those determined by specifying values for finitely many coordinates in Ω.

$\underline{\text{P}}$ is the countably additive *iid* product *fair-coin* probability that is determined by

$$\text{P}(X_n = 1) = 1/2 \quad (n = 1, 2, \ldots)$$

and where each finite sequence of length n is equally probable,

$$\text{P}(\text{X}_1 = x_1, \ldots, \text{X}_n = \text{x}_n) = 2^{-n}.$$

Let $L^{½}$ be the set of infinite sequences of 0*s* and 1*s* with limiting relative frequency ½

for each of the two digits: a set belonging to $\mathcal{B}$.

Specifically, let $S_n = \sum_{i=1}^{n} X_n$. Then $L^{½} = \{x: \lim_{n\to\infty} S_n/n = 1/2\}$.

- **The *strong-law of large numbers* asserts that $P(L^{½}) = 1$.**

What is excused with the *strong law*, what is assigned probability 0,

is the null set N (= $[L^{½}]^c$) consisting of

the complement to $L^{½}$ among all denumerable sequences of 0*s* and 1*s*.

- **It is an old story within Philosophy that the Strong Law of Large Numbers offers a probabilistic link between *chance* and *limiting relative frequency*.**

- **The Blackwell-Dubins (1962) strong-law for consensus among Bayesian investigators with increasing shared evidence.**

Let $<X, \mathcal{B}>$ be a measurable Borel product-space as follows.

Consider a sequence of sets X_i (i = 1, …) each with an associated σ-field $\mathcal{B}_i$.

The Cartesian product $X = X_1 \times \ldots$ of sequences $(x_1, \ldots) = x \in X$, for $x_i \in X_i$.

That is, each x_i is an atom of its algebra $\mathcal{B}_i$.

$\mathcal{B}$ be the σ-field generated by the measurable rectangles.

***Definition*: A *measurable rectangle* $(A_1 \times \ldots) = A \in \mathcal{B}$ is one where**

$A_i \in \mathcal{B}_i$ and $A_i = X_i$ for all but finitely many *i*.

Blackwell and Dubins (1962) consider the idealized setting where:

Two Bayesian agents consider a common product space and share evidence of the growing sequence of *histories* $\langle x_1, x_2, \ldots, x_k\rangle$.

Each has her/his own countably additive personal probability, with regular conditional probabilities for the future given the past.

- **Two measure spaces $\langle X, \mathcal{B}, P_1\rangle$ and $\langle X, \mathcal{B}, P_2\rangle$.**

- **Assume P_1 and P_2 agree on which events in $\mathcal{B}$ have probability 0.**

In order to index how much these two are in probabilistic disagreement, use the total-variation distance.

Define $\rho(\, P_1(\,\cdot \mid X_1{=}x_1, \ldots, X_n{=}x_n), P_2(\,\cdot \mid X_1{=}x_1, \ldots, X_n{=}x_n)\,) =$

$$\sup_{E \in \mathcal{B}} \mid P_1(E \mid X_1{=}x_1, \ldots, X_n{=}x_n) - P_2(E \mid X_1{=}x_1, \ldots, X_n{=}x_n) \mid .$$

The index ρ focuses on the greatest differences between the two agents' conditional probabilities.

The B-D (1962) strong-law about asymptotic consensus: For *i* = 1, 2

$$P_i\,[\, lim_{n\to\infty}\, \rho(\, P_1(\,\cdot \mid X_1{=}x_1, \ldots, X_n{=}x_n), P_2(\,\cdot \mid X_1{=}x_1, \ldots, X_n{=}x_n)\,) = 0\,] = 1.$$

- **Almost surely, increasing shared evidence creates consensus.**

2. A topological sense of a <u>negligible</u> set – meager (or 1st category) sets.

A <u>topology</u> $\mathcal{T}$ for a set *X* is a class of *open* subsets of *X* that

includes *X* and ∅

is closed under arbitrary unions and finite intersections.

The pair X = (*X*, $\mathcal{T}$) is called a *topological space*.

A subset $Y \subseteq X$ is *<u>dense</u>* (in *X*) provided that,

Y has non-empty intersection with each (non-empty) open set in $\mathcal{T}$.

A subset $Y \subseteq X$ is *<u>nowhere dense</u>* (in *X*) provided that

for each (non-empty) open set *O*, there is a (non-empty) open $O' \subseteq O$

where $Y \cap O' = \varnothing$.

Topologically negligble (meager) and large (residual) sets

A set *M* is *meager* (or *1st Category*) *iff*

M is the denumerable union of nowhere dense sets.

A set R is *residual* (or *comeager*) *iff* $R = M^c$.

R is the complement of a meager set M.

3. *Oxtoby's (1957, 1980) results* – where the two senses of *negligible* conflict.

There are some evident similarities between

the measure theoretic sense of a *negligible* set – a P-null set

and

the topological sense of a *negligible* set – a meager set.

A trivial example:

If X is uncountable with P({x}) = 0 for each $x \in X$, and

the topology $\mathcal{T}$ on X has makes each point *nowhere dense* in X,

then a denumerable set of points is *negligible* in both senses simultaneously.

More significantly (Oxtoby, 1980, *T. 19.4*) establishes an important *duality*. Relative to Lebesgue measure and Euclidean topology on the real line –

***<u>Duality Theorem</u>*: Assume the *Continuum Hypothesis*.**

Let φ be a proposition involving only the concepts of:

measure 0 set, meager set, and pure set theory.

Let φ* be the proposition that results by interchanging

'measure 0' and 'meager' wherever these appear in φ.

Then, φ *if and only if* φ*.

However, this *duality* does not establish the same sets are judged negligible in both perspectives.

- ***Old News*: The real line can be decomposed into two complementary sets *N* and *M* where *N* has Lebesgue measure 0, and *M* is meager.**

Existence of a radically opposed decomposition of *negligible* sets is captured, more generally, by Oxtoby's [1980, p. 64] Theorem 16.5.

If the measure space $< X , \mathcal{B} , P>$, satisfies

- **P is nonatomic,**
- ***X* has a metrizable topology $\mathcal{T}$ with a base whose cardinality is less than the first weakly inaccessible,**
- **and, the σ-field $\mathcal{B}$ includes the Borel sets of $\mathcal{T}$,**

then *X* can be partitioned into a set of P-measure 0 and a meager set.

But are any of these problematic decompositions of practical significance?

Return to the setting of the *Law of Large Numbers*.

Let X_i (i = 1, 2, …) be a denumerable sequence of Bernoulli {0,1} variables.

Let < X, $\mathcal{B}$, P> be the *measure space* with $\mathcal{B}$ the Borel σ-algebra generated by all finite sequences of flips, and P is the *iid* "fair coin" measure on sequences.

Topologize this space using the product of the *discrete* topology on each X_i,

$$\mathcal{T}_i(X_i) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\} \quad \text{and then} \quad \mathcal{T}^\infty = \mathcal{T}_1 \times \mathcal{T}_2 \times \ldots .$$

Topology $\mathcal{T}^\infty$ is (homeomorphic to) the *Cantor Space*.

Let $L^{½}$ be the set of binary sequences with limiting relative frequency ½ for each of the two digits: a set belonging to $\mathcal{B}$.

Specifically, let $S_n = \sum_{i=1}^{n} X_n$ and then $L^{½} = \{x: \lim_{n\to\infty} S_n/n = 1/2\}$.

- **The *strong-law of large numbers* asserts that $P(L^{½}) = 1$.**
- **BUT (Oxtoby, 1957) the set $L^{1/2}$ is a meager set in the topology $\mathcal{T}^\infty$ (!!)**

4. A generalization of Oxtoby's (1957) result.

In our (2017) we show (*Theorem A1*) that the tension over rival senses of *negligble* generalizes in a dramatic way to sequences of random variables relative to a large class of infinite product topologies. A *Corollary* applies to Bernoulli sequences.

Let χ be a set with topology $\mathcal{T}$ and Borel σ-field, $\mathcal{B}$, i.e., the σ-field generated by the open sets in $\mathcal{T}$. Let χ^∞ be the countable product set with the product topology $\mathcal{T}^\infty$ and product σ-field, $\mathcal{B}^\infty$, which is also the Borel σ-field for the product topology (because it is a countable product).

Let <Ω, $\mathcal{A}$, P> be a probability space.

Relate these two spaces with a sequence of random quantities

$\{X_n\}_{n=1}^{\infty}$, where, for each *n*, $X_n: \Omega \rightarrow \chi$ is ($\mathcal{A}$ and $\mathcal{B}$) measurable.

Define $X: \Omega \rightarrow \chi^\infty$ by $X(\omega) = \langle X_1(\omega), X_2(\omega), \ldots\rangle$.

Let $S_X = X(\Omega)$ be the image of X, i.e., the set of sample paths of X.

We denote elements of S_X as $y = <y_1, y_2, \ldots>$.

As S_X is a subset of χ^∞ we endow S_X with the subspace topology.

We require a degree of *logical independence* between the X_n's.

In particular, we need the sequence $\{X_n\}_{n=1}^{\infty}$ to be capable of moving to various places in χ^∞ regardless of where it has been so far.

We express this as *Condition 1*, below, in terms of the interior of a set.

- **The *interior* of a set B is the union of all open subsets of B.**

***Condition 1*: For each j, let $B_j \in \mathcal{B}$ be a set with nonempty interior B_j^o.**

Require that, for each n, for each $x = \langle x_1, \ldots, x_n \rangle \in \langle X_1, \ldots, X_n \rangle(\Omega)$,

and for each j, there exists a positive integer $c(n, j, x)$ such that

$$\langle X_1, \ldots, X_n, X_{n+c(n,j,x)} \rangle^{-1}(\{x\} \times B_j^o) \neq \emptyset.$$

***Condition* 1 asserts that, no matter where the sequence of random variables has been up to time n, there is a finite time, $c(n, j, x)$, after which it is possible that the sequence reaches the set B_j^o.**

For each sample path $y \in S_X$, define $\tau_0(y) = 0$, and for $j > 0$, define

$$\tau_j(y) = \begin{cases} \min\{n > \tau_{j-1}(y): y_n \in B_j\}, & \text{if the minimum is finite,} \\ \infty & \text{if not.} \end{cases}$$

Let $B = \{y \in S_X : \tau_j(y) < \infty \text{ for all j}\},$

And let $A = S_X \setminus B = B^c \cap S_X.$

- ***A* is the set of sample paths each of which *fails* to visit at least one of the B_j sets, in the order specified.**

***Aside*: Because we do not require that the sets B_j are nested, it is possible that the sequence reaches B_k for all $k > j$ without ever reaching B_j.**

- ***Theorem*: *A* is a meager set. (!!)**

The following *Corollary* generalizes Oxtoby's (1957) result that the *Strong Law* for *iid* Bernoulli variables provides a *measure 1* set that is *meager*.

As before, let $X_i = \{0,1\}$, $i = 1, 2, \ldots$, be a sequence of Bernoulli {0, 1} variables.

Let < X, $\mathcal{B}$, P> be the *measure space* with B the Borel σ-algebra generated by all finite sequences of flips, and P is the *iid* "fair coin" measure on sequences.

Topologize the measurable space < X, $\mathcal{B}$ > using the product of the *discrete* topology on each X_i, $\mathcal{T}_i(X_i) = \{\varnothing, \{0\}, \{1\}, \{0,1\}\}$ and then $\mathcal{T}^{\infty}(X) = \mathcal{T}_1 \times \mathcal{T}_2 \times \ldots$.

Let $L^{½}$ be the set of binary sequences with limiting relative frequency ½ for each of the two digits: a set belonging to $\mathcal{B}$.

- **The *strong-law of large numbers* asserts that $P(L^{½}) = 1$.**
- **BUT (Oxtoby, 1957) the set $L^{1/2}$ is a meager set in the topology $\mathcal{T}^{\infty}$**

Now, consider the set of sequences:

OM = {*x*: the observed relative frequency of 1 oscillates *maximally*}

Specifically, for each $x = < x_1, x_2, \ldots > \in OM$,

$$\textit{lim.inf.} \sum_{j=1}^{n} x_j/n = 0 \text{ and } \textit{lim sup.} \sum_{j=1}^{n} x_j/n = 1.$$

OM is a $\mathcal{B}$-measurable set.

The complement to OM, $OM^c = L^{<0,1>}$, is the measurable set of binary sequences whose observed relative frequencies *fail* to oscillate maximally.

$$L^{<0,1>} = \{x: \textit{lim.inf.} \sum_{j=1}^{n} x_j/n > 0 \text{ or } \textit{lim sup.} \sum_{j=1}^{n} x_j/n < 1\}.$$

- ***Corollary*: $L^{<0,1>}$ is a *meager* set in $\mathcal{T}^{\infty}$.** See also Calude and Zamfirescu (1999).

***Challenge*: What stochastic process P treats $L^{<0,1>}$ as a P-null event?**

The conflict between the two senses of *negligible* runs deeper still.
Build a hierarchy of events by considering the *sojourn time*s for relative frequencies, and then relative frequencies of frequencies, *etc*. .

Let the sequence of Bernoulli outcomes $x = <x_1, x_2, \ldots >$ count as the sequence of 0^{th} *tier* events – the sequence of 0s and 1s.

- **Define the 1^{st} tier event $F^{1}_{[.2,.4]}$ as occurring whenever**

the relative frequency of 1 in the sequence x falls in the interval [.2, .4].

Even though OM is a residual set of sequences, the subset of OM for which the relative frequency of $F^{1}_{[.2,.4]}$ *fails* to *oscillate maximally* is a *meager* set.

2^{nd} tier events are defined by intervals of frequencies of 1^{st} tier events.

Since the countable union of *meager* sets is *meager*:

- **The set of sequences that have relative frequencies of events that oscillate maximally at each countable tier is residual!**

5. Some concluding thoughts on where the two formal perspectives on *negligible* sets do, and do not play well together.

***Q*: What roles can these two different senses of negligible play together?**

***Tentative Answer*:**

Use a topological sense of "negligible" for sets that are *not* within the domain of the measure – where probability does not apply.

Example:

Regarding Blackwell-Dubins asymptotic ρ-consensus among Bayesian agents who share evidence,

use topology to investigate the size of the community for which the shared evidence creates asymptotic merging.

Or, as convergence is a topological notion,

use a different topology than the one induced by sup-norm, ρ, to define asymptotic merging.

But do not let the measure and the topology compete over the same family of sets as to which are *negligible*.

That way lies conflict!

Selected References

Blackwell, D. and Dubins, L. (1962) ***Merging of opinions with increasing information*****. Ann. Math. Stat. 33: 882-887.**

Calude, C. and Zamfirescu, T. (1999) ***Most numbers obey no probability laws*****. Publ. Math. Debrecen 54: 619-623.**

Cisewski, J., Kadane, J.B., Schervish, M.J., Seidenfeld, T. and Stern, R (2017) ***Standards for Modest Bayesian Credences*****. Forthcoming, Phil. Sci.**

Oxtoby, J.C. (1957) ***The Banach-Mazur Game and Banach Category Theorem*****, in Dresher, M., Tucker, A. W., & Wolfe, P.** ***Contributions to the Theory of Games*** **(Vol. 3). Princeton University Press: pp. 159-163**

Oxtoby, J.C. (1980) ***Measure and Category*** **(2nd edition). Springer-Verlag.**

Schervish, M.J. (1995) ***Theory of Statistics*****. Springer-Verlag.**