Some Finitely Additive Probabilities and Decisions

Teddy Seidenfeld
Carnegie Mellon University

Based on joint work with



Mark Schervish **CMU**

Jay Kadane **CMU**

Rafael Stern **USP - Brazil**

Part 1: Some (Purely) Finitely Additive Probabilities for Logic
Using analysis by M. Amer (1985a,b)

1.1 Structural assumptions for finitely additive probabilities.

A finitely additive measure space, $<\Omega, \mathcal{B}, P>$,

where P is a finitely additive probability (see below)

defined on a field of sets $\mathcal{B}$,

with the sure-event Ω equal to a set of disjoint and mutually exhaustive possibilities, called "states": $\Omega = \{\omega_i: i \in I\}$.

A set B is *measurable* if $B \in \mathcal{B}$.

A numerical probability P is <u>finitely additive</u> [f.a.] if it satisfies:

- **(i) For each measurable set B, $0 \leq P(B) \leq 1$.**
- **(ii) $P(\Omega) = 1$.**
- **(iii) If E and F are disjoint measurable sets, with $G = E \cup F$, then $P(G) = P(E) + P(F)$.**

For probability P to be *countably additive* is to require also either of the following two axioms, which, given P is f.a., are equivalent provided that $\mathcal{B}$ is at least a field of sets. [See Billingsley, §2.]

(iv-a) Let $\{A_i : i = 1, \ldots\}$ be a sequence of measurable, pairwise disjoint sets, and their union, A, is measurable.

That is, assume $A_i \cap A_j = \emptyset$ if $i \neq j$, and $A = \cup_i A_i \in \mathcal{B}$.

P is *countably additive₁* [c.a.₁] if $P(A) = \sum_i P(A_i)$ for each such sequence.

(iv-b) Let $\{B_i : i = 1, \ldots\}$ be a decreasing (or, respectively increasing) sequence of measurable sets, and their limit, B, is measurable.

That is, $B_i \supseteq B_{i+1}$, and assume $B = \cap_i B_i$

(or respectively $B_i \subseteq B_{i+1}$ and $B = \cup_i B_i$ is measurable).

P is *countably additive₂* [c.a.₂] if $P(B) = \lim_i P(B_i)$ for each such sequence.

- **Countable additivity$_2$ is a continuity condition for probability.**

If P is finitely but not countably additive, it is *merely* finitely additive.

An extreme version of mere f.a. probability is a <u>purely</u> f.a. probability:

A f.a. probability P is called a *<u>purely</u>* finitely additive if, for each $\varepsilon > 0$, there exists a (measurable) denumerable partition of the sure event $\Pi = \{A_i : i = 1, \ldots\}$, such that $\sum_i P(A_i) < \varepsilon$.

And P is called *<u>strongly</u>* f.a. if $\sum_i P(A_i) = 0$.

1.2 Probabilities for infinite free Boolean algebras.

Amer's (1985*b*) Theorem 5 reports the following result:

Proposition 1: Let $\mathcal{A}$ be an infinite, free Boolean algebra.

There are no countably additive probabilities on $\mathcal{A}$.

Amer proves that each infinite, free Boolean algebra embeds mandatory discontinuities in the sense of continuity required by c.a.$_2$.

***Corollary*: Each f.a. probability on such a Boolean algebra is purely finitely additive. (See the appendix to these slides for details.)**

***Example*: The Lindenbaum-Tarski algebra $\mathcal{L}$ for sentential logic (with top $\top$ and bottom $\perp$) is a countable, free Boolean algebra with the denumerable set of sentence letters $\{S_i: i = 1, \ldots \}$ serving as a set of free generators.**

Let φ_i be either one of S_i or $\neg S_i$.

Regardless of which is chosen, the algebra compels: $\bigvee\{\varphi_i\} = \top$!!

Then, $\mathcal{L}$ supports no countably additive probability.

But $\mathcal{L}$ supports uncountably many merely f.a. probabilities.

Additional clarificatory remarks about this example.
See Theorem 5 of Amer (1985a).

There are $2^{\aleph_0}$ different, 2-valued, semantic models M for $\mathcal{L}$.

A model M provides a truth value (T or F) for each sentence letter S_i.

Each model M provides a 2-valued probability over all of $\mathcal{L}$:

For each $\lambda \in \mathcal{L}$, $P_M(\lambda) = 1$ if $M(\lambda) = T$, and

$P_M(\lambda) = 0$ if $M(\lambda) = F$.

Consider the sequence of φ_i where each is F under model M.

Then, for $k = 1, \ldots$, $P_M(\varphi_1 \vee \ldots \vee \varphi_k) = 0$, but $P_M(\bigvee\{\varphi_i\}) = P_M(\top) = 1$, which violates c.a.$_2$.

Thus, there are uncountably many (strongly) f.a. probabilities on $\mathcal{L}$.

***Summary of Part 1*: Some Boolean algebras compel mere f.a. by mandating failures of continuity in the sense of c.a.$_2$**

Part 2: Some Finitely Additive (Statistical) Decisions
or
How Bruno de Finetti might have channeled Abraham Wald

Based on our (2019) ***What Finite-Additivity Can Add to Decision Theory***

Organization of Part 2 of this presentation.

2.1 Three dominance principles and finitely additive expectations – in increasing strength:

Uniform (bounded-away) dominance

Simple dominance

***Admissibility* (aka *Strict Dominance*)**

2.2 Finitely additive mixed strategies and Wald's (statistical) *Loss* functions.

- **An example involving a discontinuous, strictly proper scoring rule.**

Some results – assuming that *Loss* is bounded below:

Existence of a *Minimal, Complete Class of Bayes Decisions*

Existence of a *Minimax Strategy* and a *Worst-case prior*

Uniform dominance of never-Bayes decisions for bounded loss
– *generalized Rationalizability*

2.3 But, not all *priors* have Bayes-decisions (!)

2.1 Three dominance principles, in increasing order of logical strength

Fix a partition $\pi = \{\omega_1, \ldots, \omega_n, \ldots\}$, which might be infinite.

An *Act* is a function from π to a set of outcomes *O*.

Assume that outcomes may be compared by preference, at least within each ω.

	ω_1	ω_2	ω_3	…	ω_n	…
Act_1	$o_{1,1}$	$o_{1,2}$	$o_{1,3}$	…	$o_{1,n}$	…
Act_2	$o_{2,1}$	$o_{2,2}$	$o_{2,3}$	…	$o_{2,n}$	…

***Uniform dominance*:**

For each ω_i in π, outcome $o_{2,i}$ is strictly preferred to $o_{1,i}$ by at least $\varepsilon > 0$.

***Simple dominance*:**

For each ω_i in π, outcome $o_{2,i}$ is strictly preferred to outcome $o_{1,i}$.

***Admissibility* (Wald, 1950) – *Strict dominance* (Shimony, 1955):**

For each ω_i — $o_{2,i}$ is weakly preferred to $o_{1,i}$

and for some ω_j — $o_{2,j}$ is strictly preferred to $o_{1,j}$.

Then, by *dominance* applied with partition in π,: Act_2 is *strictly preferred* to Act_1.

de Finetti (1974):

A class {*X*} of real-valued variables defined on a <u>privileged partition</u> of *states*, Ω.

Let *P* be a (f.a) probability on Ω.

Denote by $\mathcal{E}_P(X)$ the (f.a.) expected value of variable *X* with respect to Ω.

***Preference* between pairs of variables based on finitely additive expectation:**

2. **obeys *Uniform Dominance* in Ω**
3. **but may fail *Simple Dominance* in Ω.**

Example₁ **– Let Ω be countably infinite Ω = {ω_1, ω_2, …}.**

Consider variables $X(\omega_n) = -1/n$, and the constant $Z(\omega_n) = 0$.

Let *P* be a (strongly) finitely additive probability $P(\{\omega\}) = 0$.

Then $\mathcal{E}_P(X) = 0 = \mathcal{E}_P(Z)$, so indifference between *X* and *Z*.

But *Z* simply dominates *X*.

	ω_1	ω_2	ω_3	…	ω_n	…
X	-1	-1/2	-1/3	…	-1/*n*	…
Z	0	0	0	…	0	…

Finitely additive mixed strategies: Making lemonade from lemons.

Example₂: Decision making under certainty: $\Omega = \{\omega\}$.

Consider the half-open interval of constant rewards, $\mathcal{X} = \{X: 0 \leq X < 1\}$.

Each *pure strategy X* is (uniformly) dominated.

Likewise, each countably additive *mixed strategy* P^σ over $\mathcal{X}$ has expectation < 1.

But let be *P* a f.a. *mixed strategy* over $\mathcal{X}$ where, for each $\varepsilon > 0$, $P[X > 1-\varepsilon] = 1$.

- **Then, $\mathcal{E}_P(X) = 1$.**

In f.a. jargon, P *agglutinates* X at the (missing) value 1.

2.2 Elementary Statistical Decision Theory in the fashion of A.Wald.

- **An agent has a set $\mathcal{A}$ of available (*pure strategy*) actions, and there is uncertainty over a set Θ of *parameters* or *states of Nature*. Θ forms a privileged partition.**
- **The agent suffers *loss* $L(\theta; a)$ if she chooses a and θ is the *state of Nature*.**
- **Sometimes the agent is allowed to choose action a using a probability measure (a mixed strategy) δ over $\mathcal{A}$, and (when there are no data) we replace *loss* $L(\theta; \cdot)$ by the *risk* $R(\theta; \delta) = \int_{\mathcal{A}} L(\theta; a)\, \delta(da)$.**

***Aside*: As usual, the probability measure $\delta_a(A) = I_A(a)$ for every $A \subseteq \mathcal{A}$ is equivalent to the pure strategy a.**

The agent wants to choose δ to minimize *Risk*: respect *dominance* in Θ.

A.Wald (1950): Respect *Admissibility* for *Risk* in Θ.

***Example 3a Brier Score* for two complementary events.**

$\Omega = \{B, B^c\}^2$ where B is also the indicator function I_B for some event B.

$\mathcal{A} = [0, 1]^2$. There are no data.

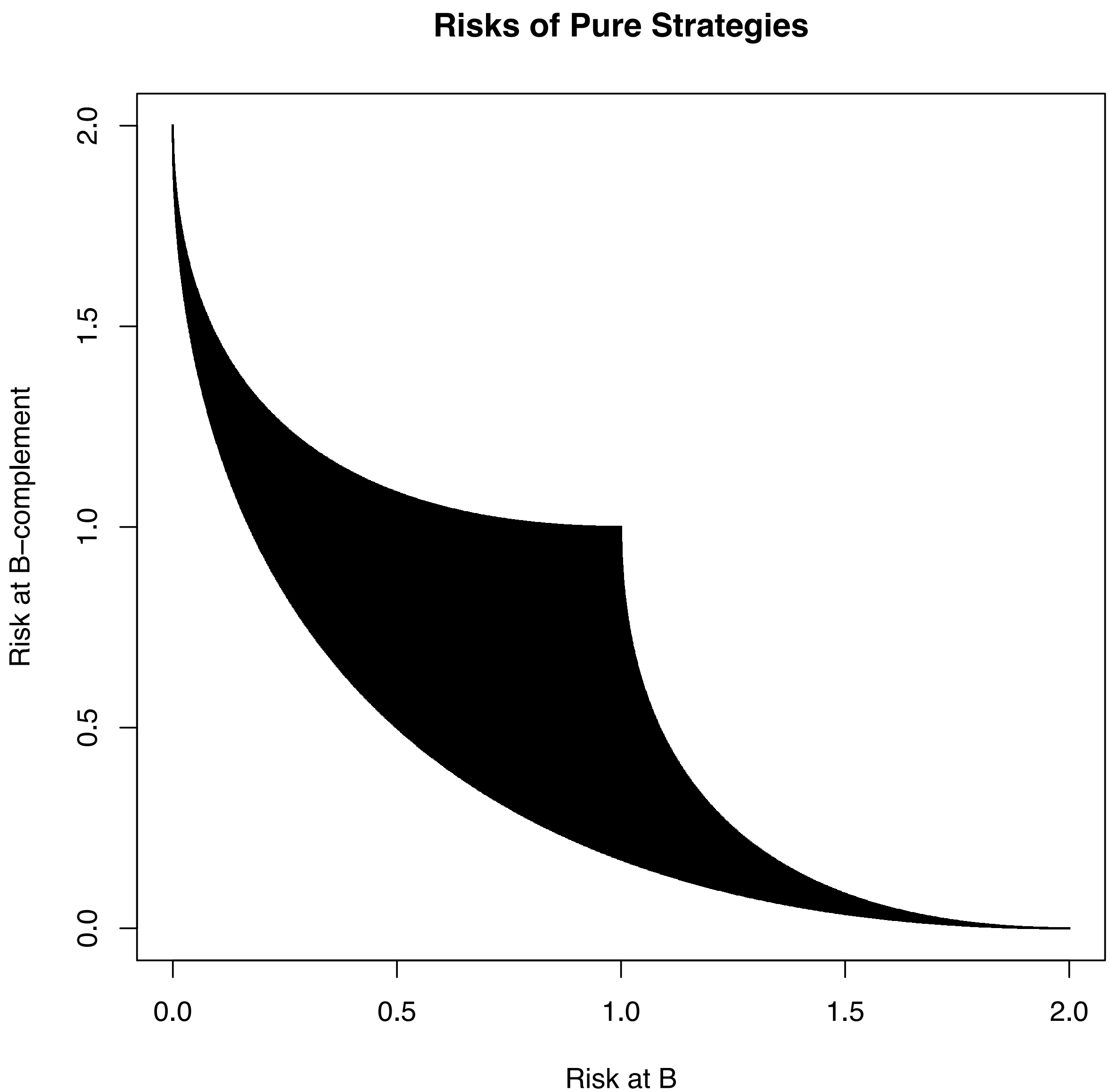
$L(\theta; (a_1, a_2)) = (I_B - a_1)^2 + (I_{Bc} - a_2)^2$

- **The only *admissible* actions are $\{(a_1, a_2): a_1 + a_2 = 1\}$,**

  which correspond to the lower boundary of the Risk set – see next slide.

- **Brier Score is a *strictly proper* scoring rule.**

  The Bayesian agent minimizes expected score <u>uniquely</u> by announcing her degrees of belief for (B, B^c): $a_1 = Prob(B)$ and $a_2 = Prob(B^c)$


Risks of Pure Strategies
Risk at B–complement
0.0
0.5
1.0
1.5
2.0
Risk at B
0.0
0.5
1.0
1.5
2.0

Example 3b: A discontinuous Brier Score.

$\Omega = \{B, B^c\}^2$ where B is also the indicator function I_B for some event B.

$\mathcal{A} = [0, 1]^2$. Again, there are no data.

$$L(\theta; (a_1, a_2)) \quad = \quad (I_B - a_1)^2 + (I_{Bc} - a_2)^2$$

$$+ (½) \times \begin{cases} (I_{[0,.5]}(a_1) + I_{(.5,1]}(a_2)) & \text{if } \theta = B \\ (I_{(.5,\, 1]}(a_1) + I_{[0,\, .5]}(a_2)) & \text{if } \theta = B^c \end{cases}$$

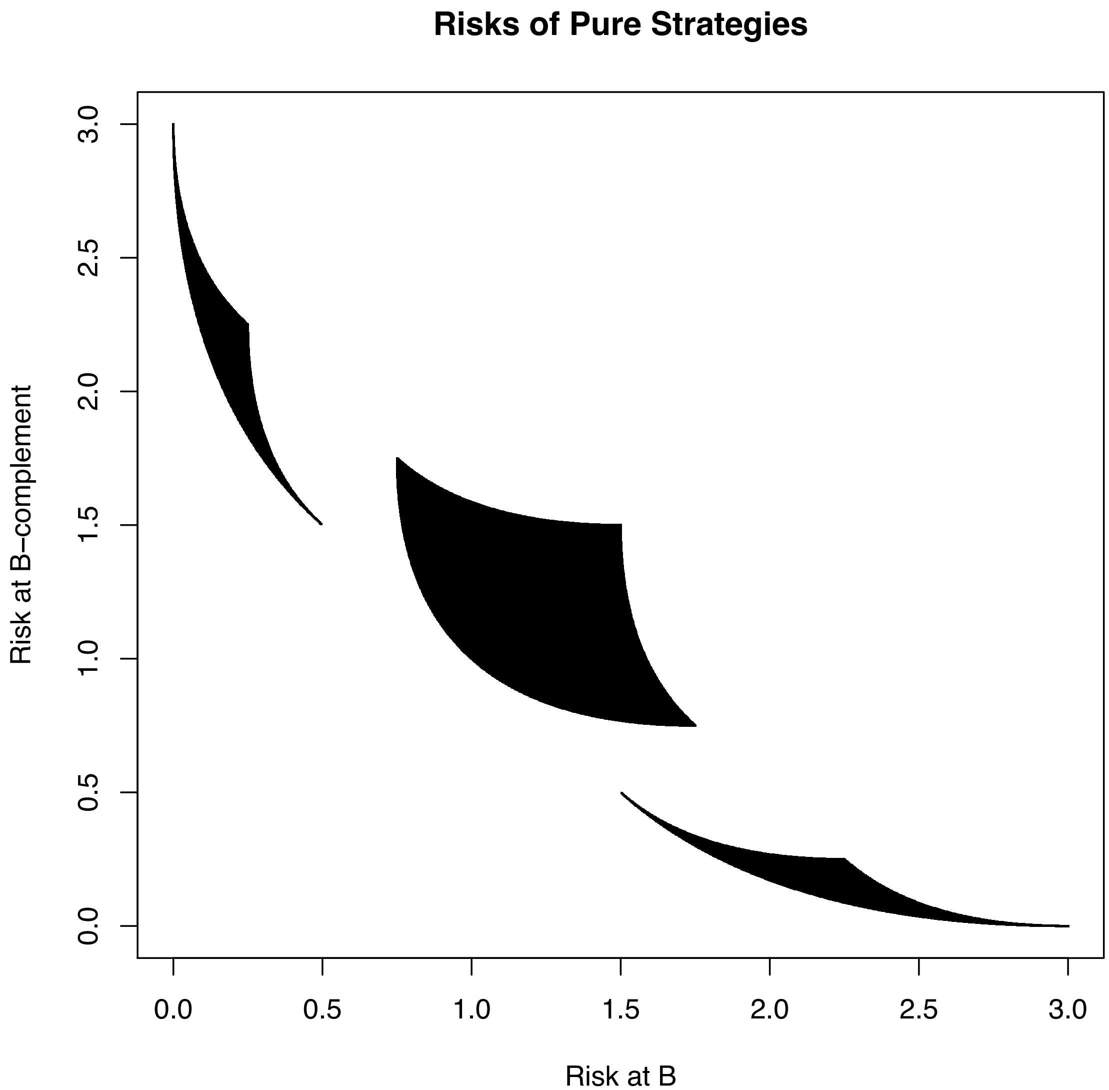
This *Loss* carries an added penalty when the forecast is on the wrong side of ½.

- **The only *admissible* actions are $\{(a_1, a_2): a_1 + a_2 = 1\}$.**
- **This discontinuous Brier Score is a *strictly proper* scoring rule.**

The Bayesian agent (uniquely) minimizes expected score by announcing her degrees of belief for (B, B^c): $a_1 = Prob(B)$ and $a_2 = Prob(B^c)$

but ...

$L(\theta; (a_1, a_2))$ is a point in a two-dimensional set $[0, 3]^2$.



Recall: The *admissible* options are on the lower boundary.

The shaded risk set has the properties that for pairs (p, 1-p):

Top **From (0, 3) down to but not including (.5, 1.5) are the points on the lower boundary, which correspond to $0 \leq p < .5$.**

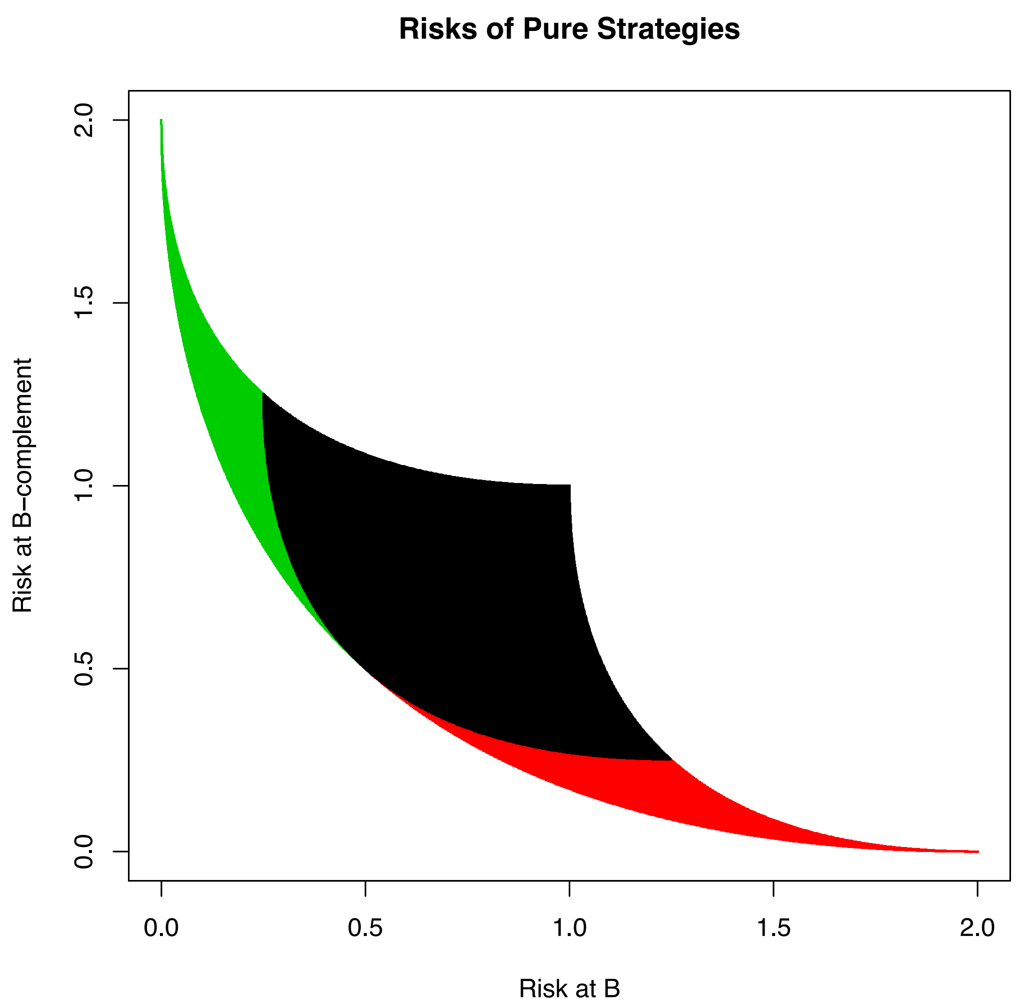
***Middl*e** **In the middle section, only the point (1, 1) is on the lower boundary, corresponding to p = .5.**

Bottom **From (but not including) (1.5, .5) to (3, 0) are the points on the lower boundary, which correspond to $.5 < p \leq 1$.**

So, points in the middle section (other than (1,1)) are *inadmissible* though some are not dominated by (1,1).

But those are dominated too, but only by other *inadmissible* options.

The discontinuous (strictly proper) Brier Score carves up the continuous Brier Score.



Some decision-theory results in the fashion of Wald (1950)

Definitions:

Call a subclass $\mathcal{C} \subseteq \mathcal{A}$ of available decisions *Complete*

if for each decision $\delta \notin \mathcal{C}$ there is $\delta_0 \in \mathcal{C}$ where δ_0 *dominates* δ

in the sense of *admissibility*.

Call a subclass $\mathcal{C} \subseteq \mathcal{A}$ of available decisions *Minimally Complete*

if $\mathcal{C}$ is complete and no proper subset of $\mathcal{C}$ is Complete.

- **If there exists a *Minimally Complete* class it consists of the *admissible* decisions.**
- **In *Example 3*$_b$ (discontinuous Brier), there is no *Minimally Complete* class.**

And using countably additive mixed strategies does not help this way.

BUT, augment the decision space by allowing (merely) f.a. mixed strategies.

Then, these f.a. mixed strategies fill in the missing lower boundary for *Risk*.

For example, consider f.a. mixed strategies δ_1 and δ_2 with the features that

$\forall \varepsilon > 0$,	$\delta_1\{ a_1 : .5 - \varepsilon < a_1 < .5 \} = 1$
and	$\delta_2\{ a_1 : .5 < a_1 < .5 + \varepsilon \} = 1$,
and	**where** $a_2 = 1 - a_1$.

Then $\quad R(\theta; \delta_1) = (.5, 1.5)$ **and** $R(\theta; \delta_2) = (1.5, .5)$

***Aside*: As $R(\theta; (.5,.5)) = (1,1)$, the 3 risk points $R(\theta; \delta_1)$, $R(\theta; (.5,.5))$ and $R(\theta; \delta_2)$ are colinear.**

Some results – assuming that *Loss* is bounded below:

Strategies are (f.a.) *mixed strategies*.

Strategy δ is *admissible* if there is no strategy δ′ such that

$$(\forall\theta)\quad R(\theta;\delta) \geq R(\theta;\delta') \quad \text{and} \quad (\exists\theta)\quad R(\theta;\delta) > R(\theta;\delta').$$

Strategy δ^0 is *Bayes* with respect to a f.a. "prior" probability λ on Θ if

$$\int_\Theta R(\theta;\delta^0)\,\lambda(d\theta) = \textit{infimum}_\delta \int_\Theta R(\theta;\delta)\,\lambda(d\theta).$$

Strategy δ^* is *minimax* provided that

$$\textit{supremum}_\theta\, R(\theta;\delta^*) = \textit{infimum}_\delta\, \textit{supremum}_\theta\, R(\theta;\delta)$$

Denote the *Bayes-risk* for δ *wrt* "prior" λ by $r(\lambda,\delta) = \int_\Theta R(\theta;\delta)\,\lambda(d\theta)$.

A "prior" λ^* on Θ is *least favorable* provided that

$$\textit{infimum}_\delta\, r(\lambda^*,\delta) = \textit{supremum}_\lambda\, \textit{infimum}_\delta\, r(\lambda;\delta).$$

- **Assume that the Loss function is bounded below, that decision rules are mixed strategies, and that "prior" probabilities are finitely additive.**

***Theorem*$_1$: The decision rules whose risks form the lower boundary of the risk function constitute a minimal complete class of admissible rules.**

Each admissible rule is a Bayes rule.

***Theorem*$_2$: There exists a minimax decision rule and a corresponding least-favorable prior.**

Each minimax rule is Bayes *wrt* each least-favorable prior.

***Theorem*$_3$ (*Rationalizability for infinite games*):**

Assume that the loss function is bounded above and below.

Suppose that δ_0 is a decision rule that is *<u>not</u>* Bayes for any prior,

i.e., δ_0 is not *E-admissible* against the vacuous prior.

Then there is decision rule δ_1 and $\varepsilon > 0$ such that

$$(\forall\theta) \quad R(\theta; \delta_0) > R(\theta; \delta_1) + \varepsilon.$$

That is, then there is a rival δ_1 that *uniformly dominates* δ_0 in Risk.

3. But not all *priors* have Bayes-decisions (!)

One of the challenges associated with (merely) f.a. expectations is that the order of integration matters – *Fubini's Theorem* has restricted validity.

So, even though the *Risk* function has a closed (lower) boundary composed of Bayes decisions, it does *not* follow that for an arbitrary "prior" λ on Θ there is a Bayes decision δ^0 *wrt* λ, where

$$\int_\Theta R(\theta; \delta^0)\, \lambda(d\theta) = \textit{infimum}_\delta \int_\Theta R(\theta; \delta)\, \lambda(d\theta).$$

***Example*₄:**

Parameter space, Θ = (0, 1)

$\mathcal{A}$ is the set of all non-empty open subintervals of (0, 1).

That is, $\mathcal{A} = \{ (x, y):\ 0 \leq x < y \leq 1\}$.

Denote by *Len*[(*x*, *y*)] = *y* - *x*, the length of interval *a*.

The *Loss* function reflects a goal of *anti-estimation* for θ:

$$L(\theta;\ a) \;=\; I_a(\theta)/Len[a] \;+\; (1-Len[a])/10$$

Consider a (strongly) f.a. "prior" $\lambda^{\#}$ on Θ where, for each *y* > 0,

$$\lambda^{\#}\{\theta: 0 < \theta < y] = 1.$$

In f.a. jargon, $\lambda^{\#}$ *agglutinates* its mass at the (missing) $\theta = 0$.

The *Bayes risk* with respect to $\lambda^{\#}$ satisfies, for each $n = 1, 2, \ldots$,

$$r(\lambda^{\#}; (1/n, 1)) = \int_{\Theta} R(\theta; (1/n, 1)\, \lambda^{\#}(d\theta)$$

$$= \int_{\Theta} [I_a(\theta)/Len[a] + (1-Len[a])/10]\, \lambda^{\#}(d\theta)$$

$$= \int_{\Theta} [I_{(1/n,\,1)}(\theta) \,/\, (n-1)/n] \;+\; (1/n)/10]\, \lambda^{\#}(d\theta)$$

$$= 0 + 1/10n = 1/10n.$$

Hence, $\quad$ $infimum_{\delta}\, r(\lambda^{\#}; \delta) = 0.$

- **But, there is no decision rule $\delta^{\#}$ with *Bayes risk* $r(\lambda^{\#}; \delta^{\#}) = 0$.**

To see this, note that, by indirect reasoning:

If $r(\lambda^{\#}; \delta) = 0$, then – from the 2nd term in the *Loss* function – $\mathcal{E}_{\lambda\#}[Len(\delta)] = 1$.

But then, because of the order of integration, with the "prior" $\lambda^{\#}(d\theta)$ integration on the outside – from the 1st term in the *Risk* function –

$$\int_{\Theta} (I_{\delta}(\theta)/Len[\delta])\, \lambda^{\#}(d\theta) \;>\; 0.$$

SUMMARY – Part 2

We have reviewed how the use of some merely f.a. mixed strategies convert

the failure of simple dominance – a *lemon*,

into the closure of the lower-boundary for a (bounded-below) statistical *Loss function*, understood in the fashion of A. Wald – *lemonade*!

It follows that there exist:

- **a *Minimal Complete Class* of *Admissible* decisions, each of which is Bayes with respect to some (f.a.) "prior";**
- **a Minimax rule and Worst-case "prior" for which the Minimax is Bayes;**

and

- **a generalized *Rationalizability result* where each never-Bayes decision is uniformly dominated by some alternative (mixed strategy) decision.**

BUT – not every "prior" has its Bayes rule.

Some References

Amer, M. A. (1985a) Classification of Boolean algebras of logic and probabilities defined on them by classical models. *Zeitschr f. math. Logik und Grundlagen d. Math*. 31: 509-515.

Amer, M.A. (1985b) Extension of relatively σ-additive probabilities on Boolean algebras of logic. *J. Symbol Logic* 50: 589-596.

Billingsley, P. (1986) *Probability and Measure* (2nd ed.). John Wiley & Sons: New York

de Finetti, B. (1974) *Theory of Probability* (vol. 1). John Wiley & Sons: New York.

Schervish, M., Seidenfeld, T., Stern, R. and Kadane, J. (2019) What finite-additivity can add to decision theory. *Stat. Methods and Applications* 29: 237-263.

Wald, A. (1950) *Statistical Decision Functions*. John Wiley & Sons: New York.

Appendix

- **Corollary**: Let $\mathcal{A}$ be an infinite, free Boolean algebra. Each finitely additive probability P on $\mathcal{A}$ is purely finitely additive.

We establish the result for a subalgebra $\mathcal{A}_\Gamma \subseteq \mathcal{A}$ generated by $\Gamma = \{\gamma_1, \ldots\}$, a denumerable subset of generators of $\mathcal{A}$. Without loss of generality, we use the Lindenbaum-Tarksi algebra $\mathcal{L}$ of sentential logic for this subalgebra. That is, up to isomorphism, $\mathcal{L}$ is the free Boolean algebra with countably many generators. Next, we summarize relevant details of $\mathcal{L}$.

Let **L** be the first order sentential language with denumerably many proposition letters, *p*, which are the atomic propositions of **L**, $\mathcal{P} = \{p_1, p_2, \ldots\}$. For convenience, let '**&**', '**v**' and '¬' be the sentential operators in **L**, whose semantics are respectively the usual truth functions 'and', 'or', and 'not'. Let ***WFF*** be the denumerable set of well formed formulas in **L**, which is the syntactic, recursive closure of the atomic propositions under the sentential operators.

Let ≡ denote (semantic) logical equivalence, an equivalence relation between pairs of well formed formulas in **L**.

For $s \in$ **WFF**, let $\bar{s}$ be the equivalence class of its logically equivalent well formed formulas.

$\mathcal{L}$ is the Lindenbaum-Tarski algebra over **WFF**/≡.

$\mathcal{L}$ is a countable Boolean algebra, where, for $s, t \in$ **WFF**

the algebraic join $\bar{s} \vee \bar{t} = \overline{s \mathbf{v} t}$,

the algebraic meet $\bar{s} \wedge \bar{t} = \overline{s \,\&\, t}$

the algebraic complement $\bar{s}' = \overline{\neg s}$.

For convenience, denote $\mathbf{T}$ = equivalence class of tautologies**,**

and $\perp$ = equivalence class of contradictions**.**

Define the (transitive) partial order $\preccurlyeq$ on $\mathcal{L}$ by $\bar{s} \preccurlyeq \bar{t}$ if *s* (semantically) entails *t*.

Note that $\preccurlyeq$ is a strict partial order; that is, if $\bar{s} \preccurlyeq \bar{t}$ and $\bar{t} \preccurlyeq \bar{s}$ then $\bar{s} = \bar{t}$.

Denote by $\bar{s} \prec \bar{t}$ the asymmetric, transitive relation, $\bar{s} \preccurlyeq \bar{t}$ and $\bar{s} \neq \bar{t}$.

$\mathbf{T}$ is the maximal element and $\perp$ is the minimal element of this strict partial order.

That is, $\perp \prec \mathbf{T}$ and if $\perp \neq \bar{s} \neq \mathbf{T}$ then $\perp \prec \bar{s} \prec \mathbf{T}$.

When neither $\bar{s} \preccurlyeq \bar{t}$ nor $\bar{t} \preccurlyeq \bar{s}$, say that $\bar{s}$ and $\bar{t}$ are *independent*.

Observe that $\mathcal{L}$ is atomless: That is, consider $s \in$ **WFF** where $\bar{s} \neq \perp$. Let t = s & p, where 'p' does not appear among the atomic propositions in s. Then $\perp \prec \bar{t} \prec \bar{s}$. So $\bar{s}$ is not an atom of $\mathcal{L}$.

$\mathcal{L}$ is (up to isomorphism) the countable, atomless Boolean algebra. Because $\mathcal{L}$ is a countable Boolean algebra, it is not a Boolean σ-algebra. (See Sikorski [1969, p. 66, (E)].) So we have to be careful about the existence of infinitary joins and infinitary meets within $\mathcal{L}$. That is, an infinitary join is the least upper bound under $\preccurlyeq$ and the infinitary meet is the greatest lower bound under $\preccurlyeq$ of a (countable) set of elements of the Boolean algebra. These need not exist in $\mathcal{L}$.

For $\bar{S} = \{\bar{s}_i \in \mathcal{L}: i = 1, 2, \ldots\}$, say that $\bar{t} \in \mathcal{L}$ is the *infinitary join* of $\bar{S}$, written $\bar{t} = \bigvee \bar{S}$, provided that,

for each $\bar{s}_i \in \bar{S}$, $\bar{s}_i \preccurlyeq \bar{t}$, and

if also there exists $\overline{t'} \in \mathcal{L}$ where, for each $\bar{s}_i \in \bar{S}$, $\bar{s}_i \preccurlyeq \overline{t'}$, then $\bar{t} \preccurlyeq \bar{t}'$.

The infinitary meet of $\bar{S}$ is defined similarly.

A (finitely additive) probability P on $\mathcal{L}$ satisfies:

(i) $0 \leq P(\bar{s}) \leq 1$

(ii) $P(\mathbf{T}) = 1$, $P(\bot) = 0$

(iii) $P(\bar{s} \vee \bar{t}) = P(\bar{s}) + P(\bar{t})$ whenever $\bar{s} \wedge \bar{t} = \bot$.

Definition: P is countably additive$_1$ on $\mathcal{L}$ provided that, whenever $\bar{S} = \{\bar{s}_i \in \mathcal{L}: i = 1, 2, \ldots\}$ is a denumerable partition, i.e., satisfying

(i) $\bar{s}_i \wedge \bar{s}_j = \bot$ whenever $i \neq j$, and

(ii) where the infinitary join $\bar{t} = \bigvee \bar{S}$ exists,

then $P(\bar{t}) = \sum_i P(\bar{s}_i)$.

Proof of the Corollary: Let P be a finitely additive probability on $\mathcal{L}$. Let $\varepsilon > 0$. We show there exists a denumerable partition $\Psi = \{\psi_1, \psi_2, \ldots\}$ in $\mathcal{L}$ with $\sum_i P(\psi_i) < \varepsilon$.

Let $\Gamma = \{\gamma_1, \ldots\}$ be the set of the sentential generators of $\mathcal{L}$: the set of (equivalence classes of the) atomic propositions.

Choose integer k that satisfies, $(1+\varepsilon)/\varepsilon < 2^k$; equivalently, $\frac{1/2^k}{1-1/2^k} < \varepsilon$.

For $j = 1, 2, \ldots$, define successive (disjoint) blocks b_j containing $j \times k$ many generators from Γ.

That is, $b_j = \{ \gamma_{\frac{k(j-1)j}{2}+1}, \ldots, \gamma_{\frac{k(j+1)j}{2}} \}$.

Specifically, $b_1 = \{\gamma_1, \ldots, \gamma_k\}$, $b_2 = \{\gamma_{k+1}, \ldots, \gamma_{3k}\}$, $b_3 = \{\gamma_{3k+1}, \ldots, \gamma_{6k}\}$, etc.

The set of blocks partitions the set of generators in Γ.

Each block, b_j, generates $2^{j \times k}$ many Boolean elements β_m^j ($m = 1, \ldots, 2^{j \times k}$) of $\mathcal{A}_\Gamma$ of the form

$$\beta_m^j = \delta_{\frac{k(j-1)j}{2}+1} \wedge \ldots \wedge \delta_{\frac{k(j+1)j}{2}}$$

where $\delta_i = \gamma_i$ or $\delta_i = \gamma_i{}'$.

Note that, since the algebra $\mathcal{A}_\Gamma$ is free, each β_m^j satisfies: $\perp \prec \beta_m^j \prec \mathbf{T}$.

Trivially, for each block b_j, if $\beta_m^j \neq \beta_n^j$ then $\beta_m^j \wedge \beta_n^j = \perp$.

Equally evident, for each block b_j, $\mathbf{T} = \bigvee\{\beta_m^j : m = 1, \ldots, 2^{j \times k}\}$.

Because the generators are independent, the Boolean elements β_m^j β_n^k from different blocks b_j and b_k are also independent, i.e., neither $\beta_m^j \preccurlyeq \beta_n^k$ nor $\beta_n^k \preccurlyeq \beta_m^j$.

As P is finitely additive, then for each block b_j ($j = 1, 2, \ldots$), there exists (at least) one Boolean element β_m^j with $P(\beta_m^j) \leq 1/2^{j \times k}$. For ease of notation, denote this element of $\mathcal{A}_\Gamma$ as β_j.

Define elements ψ_j of $\mathcal{A}_\Gamma$ as follows:

for $j = 1$, $\psi_1 = \beta_1$; and for $j \geq 2$, $\psi_j = \beta_j \wedge \psi_1' \wedge \ldots \wedge \psi_{j-1}'$

and let $\Psi = \{\psi_j : j = 1, \ldots\}$.

Claim: Ψ is a partition:

(i) $\psi_j \wedge \psi_k = \mathbf{0}$ whenever $j \neq k$. (So, also Ψ is an anti-chain.)

(ii) $\mathbf{T} = \bigvee\Psi$

Proof: (i) Immediate from the definition of the ψ_j.

(ii) Trivially, $\psi_j \preccurlyeq \mathbf{T}$. Next we show $\mathbf{T}$ is the least upper bound for Ψ.

By a simple induction, for each n = 1, 2, …, $\psi_1 \vee \ldots \vee \psi_n = \beta_1 \vee \ldots \vee \beta_n$. So, for each n, $\psi_1 \vee \ldots \vee \psi_n$ and $\beta_1 \vee \ldots \vee \beta_n$ share the same upper bounds in $\mathcal{L}$. Argue indirectly. Let $\bar{s} \prec$ **T** and suppose $\bar{s}$ is an upper bound for Ψ. Then (β_1 v … v β_{k-1}) entails $\bar{s}$. Let γ_k be an atomic proposition not appearing in s. So $\beta_k \npreccurlyeq \bar{s}$, i.e., there is a truth assignment where **t**(β_k) = T and **t**($\bar{s}$) = F. Since the atomic propositions have independent truth assignments, there is a semantic model where, also, **t**(β_1 v … v β_{k-1}) = F and **t**($\bar{s}$) = F. (If not, i.e., if **t**(β_1 v … v β_{k-1}) = F entails **t**(s) = T, then (β_1 v … v β_{k-1})' entails $\bar{s}$. And then, as (β_1 v … v β_{k-1}) entails $\bar{s}$, **T** $\preccurlyeq \bar{s}$.) Thus, (β_1 v … v β_k) $\npreccurlyeq \bar{s}$. Therefore, $\mathbf{T} = \bigvee\Psi$, and the *claim* is verified.

It is evident that $P(\psi_j) \leq P(\beta_j)$. So, $P(\psi_j) \leq 1/2^{j \times k}$.

Then $\sum_j P(\psi_j) \leq \sum_j 1/2^{j \times k} = \frac{1/2^k}{1 - 1/2^k} < \varepsilon$, which establishes that P is purely finitely additive.◊**Corollary**

Note: What drives this result is the fact that $\mathbf{T} = \bigvee\Psi$. Were $\mathcal{A}$ a σ-algebra then $\bigvee\Psi = (\psi_1 \vee \ldots \vee \psi_n \vee \ldots) \prec \mathbf{T}$.