A Subjective Foundation of Objective Probability

Luciano De Castro, *UIUC*
&
Nabil I. Al-Najjar, *MEDS*

RUD 2009

To obtain a copy of this paper and other related work, visit:
http://www.kellogg.northwestern.edu/faculty/alnajjar/htm/index.htm

Similarity as basis for probability judgements

Skip slide

In his classic 1937 article, de Finetti wondered:
"How should an insurance company evaluate the probability that an individual dies in a given year?"

De Finetti's view was:

- First choose a class of "similar" events
- use the frequency as base-line estimate of the probability
  - *e.g.,* : "death in a given year of an individual of the same age [...] and living in the same country."
- The choice of a class of "similar" events is subjective
  - ..maybe "not individuals of the same age and country, but those of the same profession and town, . . . etc, where one can find a sense of 'similarity' that is also plausible."

Three fundamental ideas

Skip slide

1. *Similarity is not optional:*
   probability judgements are founded on similarity
2. *Similarity is exchangeability*
   exchangeability *formalizes* the notion of similarity
3. *Similarity is subjective*
   Similarity is the decision maker's *model* or *theory* of the world.

   "we cannot repeat an experiment and look for a covering theory; we must have at least a partial theory before we know whether we have a repetition of the experiment."

Alternative title:

Similarity First!

Luciano De Castro, *UIUC*
&
Nabil I. Al-Najjar, *MEDS*

RUD 2009

De Finetti's ideas in action

Skip slide

- Sequence of experiments, each with outcome S
  - S is Polish (complete, separable metric space)
  - σ-algebra $\mathcal{S}$
- State space:

$$\Omega = S \times S \times \cdots$$

  - Generic element $\omega = (s_1, s_2, \ldots)$
  - Borel σ-algebra Σ on Ω
- Decision problem is static, but an inter-temporal interpretation may be suggestive.

Acts and permutations

Skip slide

- An act is any bounded, measurable function

$$f : \Omega \to \mathbb{R}$$

  - Interpret values of acts as utils
  - $\mathcal{F}$ set of all acts

- Π is the set of finite permutations π
- Given an act f and a permutation π, define $f \circ \pi$ by

$$f \circ \pi(s_1, s_2, \ldots) = f(s_{\pi(1)}, s_{\pi(2)}, \ldots)$$

Stylized example: *coin tosses*

Skip slide

- $S = \{H, T\}$; generic $\omega = \{H, T, T, H, \ldots\}$
- You might be interested in betting on the *i*th coin turning *Heads*
- Typical assumption:

  "the experiment is repeated under identical conditions"

- *"Absurd!"* says de Finetti !!
- If the experiments were really identical, then the coin should either always turns up H or always T

Exchangeability as model of similarity

Skip slide

- The fact that the outcome varies means that there are important factors that I do not understand or care to model
- But I *judge* the experiments as “similar”
- ..means: these factors affect experiments symmetrically

De Finetti’s exchangeability: for all $f \in \mathcal{F}, \pi \in \Pi$

$$f \sim f \circ \pi.$$

In the stylized coin example, H is equally likely in all experiments

$$\textit{Heads} \text{ in toss } i \quad \sim \quad \textit{Heads} \text{ in toss } j$$

De Finetti's Theorem

Skip slide

De Finetti's theorem: a preference that is

1. exchangeable *and*
2. subjective expected utility

.. has the representation

$$P(A) = \int_{\theta \in \Theta} P^{\theta}(A)\, d\mu.$$

1. Parameters: i.i.d. distributions P^{θ} indexed by $\theta \in \Theta \equiv \Delta(S)$
2. Bayesian belief μ over the parameters Θ

De Finetti's Theorem and its discontents

Skip slide

Contentment: Parameters formally capture the extent to which experiments are similar

- experiments are different enough to have different outcomes, but similar enough to share a common distribution P^θ

The discontents: de Finetti's theorem confounds similarity with the decision criterion. It simultaneously identifies

- the parameters Θ
- and that the decision maker has a prior about them:

$$P(A) = \int_{\theta \in \Theta} P^\theta(A) \; d\,\mu.$$

..but how about classical statisticians, ambiguity, etc?
Similarity is an overarching principle for all decision making and inference; Bayesianism is not.

Preorder and Monotonicity

Skip slide

Assumption 1: *$\succcurlyeq$ is reflexive and transitive.*
(may be incomplete)

Assumption 2: *If $f(\omega) \geq g(\omega)$ for all $\omega \in \Omega$, then $f \succcurlyeq g$.*

Assumption 3: *For every $x, y \in \mathbb{R}$,*

$$x > y \implies x \succ y.$$

Old Assumption: *For any $\alpha \in \mathbb{R}$, $f \succcurlyeq g$ implies*

$$f + \alpha \succcurlyeq g + \alpha.$$

Null sets

Skip slide

An event E is $\succcurlyeq$-null if any pair of acts that differ only on E are indifferent.

Assumption 4:

- $1_E \sim 0 \implies$ *E is null;*
- $\alpha 1_E \succcurlyeq \beta 1_E$ *for some* $\beta > \alpha \implies$ *E is null.*

Continuity

Skip slide

We require continuity with respect to:

$$f^n \to f \iff \left\{\omega \in \Omega : \lim_n f^n(\omega) \neq f(\omega)\right\} \succcurlyeq \textit{is}\text{-null}$$

Assumption 5:

- $f^n \to f$ *and* $g^n \to g$;
- $f^n \succcurlyeq g^n$ *for all* n
- $|f^n(\omega)| \leq b(\omega)$ *and* $|g^n(\omega)| \leq b(\omega)$, *for all* ω *and some* $b \in \mathcal{F}$.

Then $f \succcurlyeq g$.

Exchangeability

Skip slide

De Finetti's condition:

$$f \sim f \circ \pi$$

has little force without expected utility. In our more general setting, we need a stronger condition.

Definition

A preference $\succcurlyeq$ is *exchangeable* if for every act $f \in \mathcal{F}$ and permutations $\pi_1, \ldots, \pi_n$

$$f \sim \frac{f \circ \pi_1 + \cdots + f \circ \pi_n}{n}.$$

Stylized coin example:

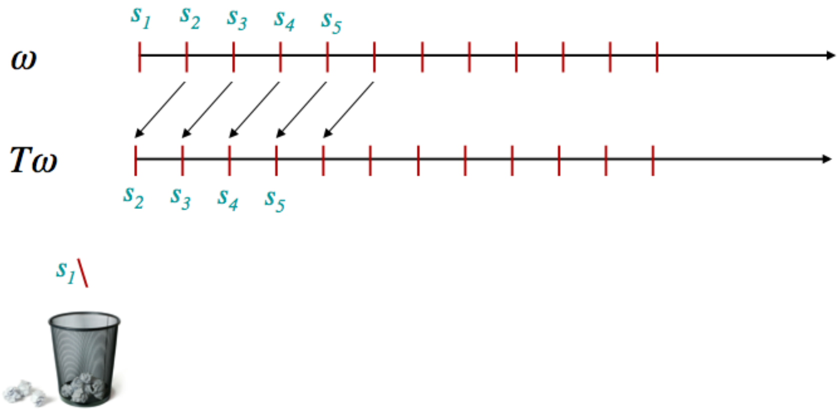
Heads in toss i $\sim$ average of *Heads* in tosses $1, \ldots, n$

To simplify the exposition, for the remainder of the paper:

restrict to the set of acts $\mathcal{F}_1$ that depend only on the first coordinate

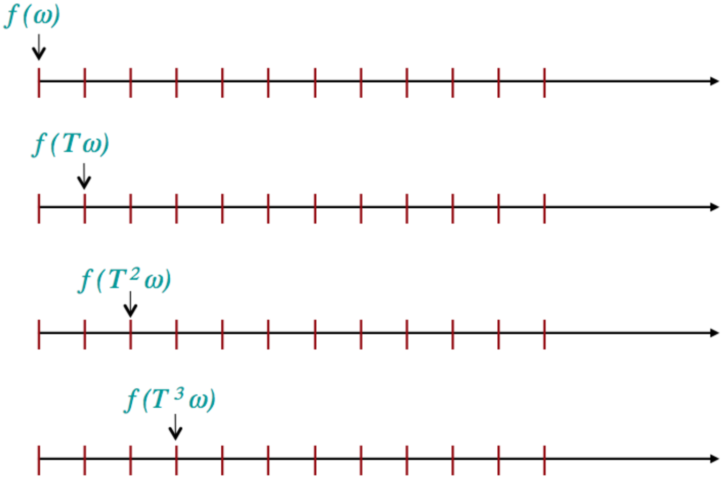
Shifting states

▸ Skip slide



Shifting acts

Skip slide



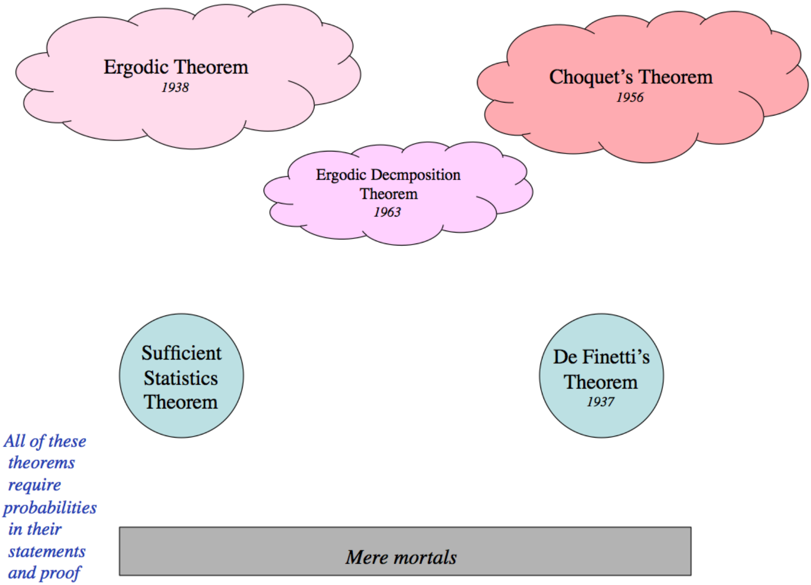
Subjective Ergodic Theorem

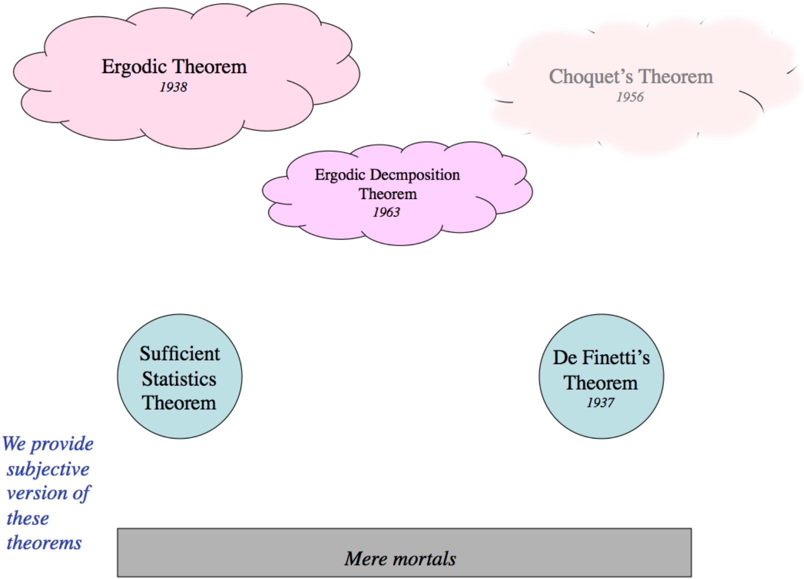
If $\succcurlyeq$ is exchangeable then for every act f, the act

$$f^{\star}(\omega) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^{j}\omega\right)$$

- is well-defined, except on a $\succcurlyeq$-null event, and
- $f^{\star} \sim f$.

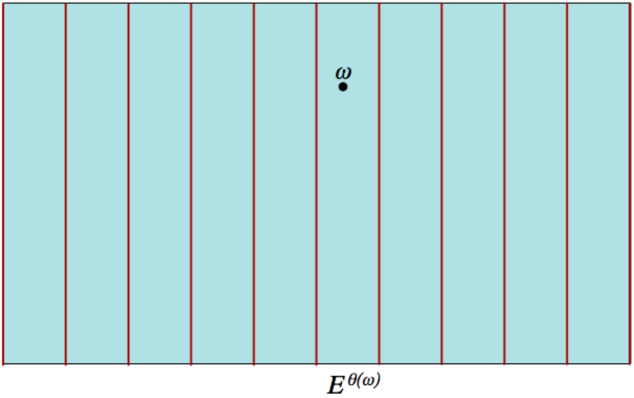
Note that $f^{\star}$, *when it exists,* is objective. Its existence is consequence of subjective assessment of similarity


Ergodic Theorem
1938
Choquet's Theorem
1956
Ergodic Decmposition
Theorem
1963
Sufficient
Statistics
Theorem
De Finetti's
Theorem
1937
All of these
theorems
require
probabilities
in their
statements
and proof
Mere mortals


Ergodic Theorem
1938
Choquet's Theorem
1956
Ergodic Decmposition
Theorem
1963
Sufficient
Statistics
Theorem
De Finetti's
Theorem
1937
We provide
subjective
version of
these
theorems
Mere mortals

Ergodic Decomposition

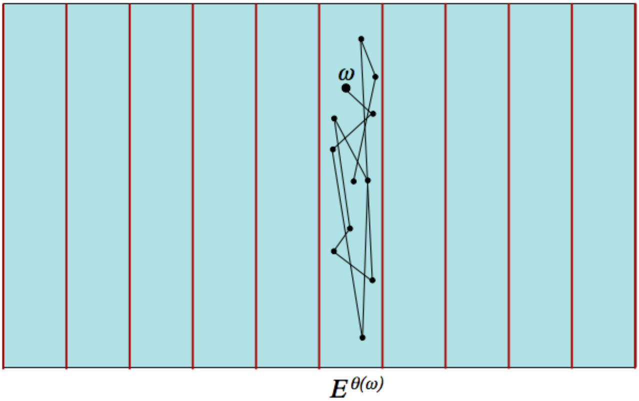
Skip slide



$$E^{\theta(\omega)} \equiv (f^\star)^{-1}(\omega) \equiv \{\text{ all states } \omega' \text{ such that } f^\star(\omega') = f^\star(\omega)\}$$

Ergodic Decomposition

Skip slide



Starting from ω, T defines an orbit that remains in $E^{\theta(\omega)}$

Subjective Ergodic Decomposition

There is a parametrization $\{(E^\theta, P^\theta)\}_{\theta\in\Theta}$ such that:

- the E^θ's partition Ω
- P^θ is the unique exchangeable distribution supported by E^θ:

$$P^\theta(E^\theta) = 1$$

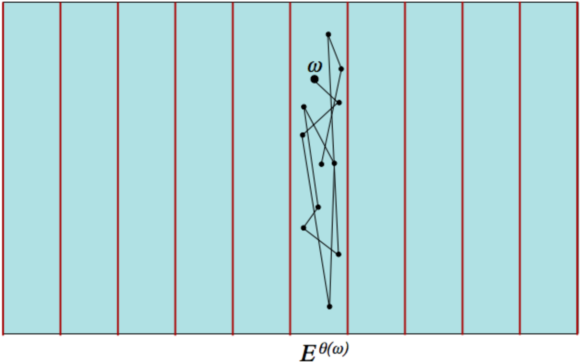
For every exchangeable preference $\succcurlyeq$ and parameter θ

- For P^θ-almost all $\omega \in E^\theta$:

$$f^\star(\omega) = \int_\Omega f \, dP^{\theta(\omega)}.$$

Ergodic Decomposition

Skip slide



Starting with ω, we obtain same answer if we:

- compute the 'objective' frequentist limit: $f^\star(\omega)$
- calculate the expected utility: $\int_\Omega f\, dP^{\theta(\omega)}$

Quick note on 'learning'

Skip slide

- Infinite data $\implies$ ω is observed $\implies$ θ can be learned
- But the model does not admit a mechanism to change the exchangeability relationship
- $\left\{\left(E^{\theta}, P^{\theta}\right)\right\}$ is the decision maker's window to the world! His way to organize information.
- The decision maker interprets the sequence

$$H, T, H, T, H, \ldots\ldots$$

as confirming that $\theta = 0.5$

Sufficient statistics: *Goal and summary*

Skip slide

We will show that the i.i.d. parametrization Θ is sufficient, in the sense of sufficient statistics, for the class of exchangeable preferences.

Sufficient statistics

Skip slide

- In statistics, a parameter is sufficient for a family of distributions if it contains all “relevant information” about that family
- Formally, a σ-algebra $\mathcal{Z}$ is sufficient for a family of distributions $\mathcal{P}$ if

  $$P(\cdot \mid \mathcal{Z})$$

  does not depend on $P \in \mathcal{P}$.
- Here we have preferences, not distributions, so this cannot make sense for us

Subjective Sufficient Statistics

Skip slide

- Define $\mathcal{F}_\Theta \subset \mathcal{F}$ to be the set of acts measurable with respect to $\{E^\theta\}_{\theta\in\Theta}$.
- Define the mapping

$$\Phi : \mathcal{F} \to \mathcal{F}_\Theta \qquad \textit{by} \qquad \Phi(f)(\omega) = \int_\Omega f \, dP^{\theta(\omega)}.$$

- We allow ambiguity *about* parameters, of course.
- However, a parameter *means* that, once its value is know, acts are evaluated using expected utility.

Subjective Sufficient Statistic Theorem

The i.i.d. parametrization Θ is *sufficient* for the set of exchangeable preferences:

For every f and g

$$f \succcurlyeq g \iff \Phi(f) \succcurlyeq \Phi(g).$$

- The parameter-based act associated with f:

$$F(\theta) \equiv \Phi(f)(\theta) \equiv \int_{\Omega} f \, dP^{\theta}.$$

- $\succcurlyeq$ induces a preference $\succcurlyeq$ on parameter-based acts:

$$f \succcurlyeq g \iff F \succcurlyeq G.$$

The Classic de Finetti's Theorem

An expected utility preference $\succcurlyeq$ is exchangeable if and only if there is a belief μ on Θ such that:

$$F \succcurlyeq G \iff \int_{\Theta} F \, d\mu \geq \int_{\Theta} G \, d\mu.$$

This confounds:

- similarity-based, statistical constructs:

$$P^{\theta}\text{'s}$$

- Bayesian criterion to resolve uncertainty about parameters

$$\mu$$

with no statistical or similarity interpretation

Inter-subjective consensus

Skip slide

Corollary

Let $\succcurlyeq$ be any exchangeable preference.

If all exchangeable Bayesians prefer f to g, then so would $\succcurlyeq$.

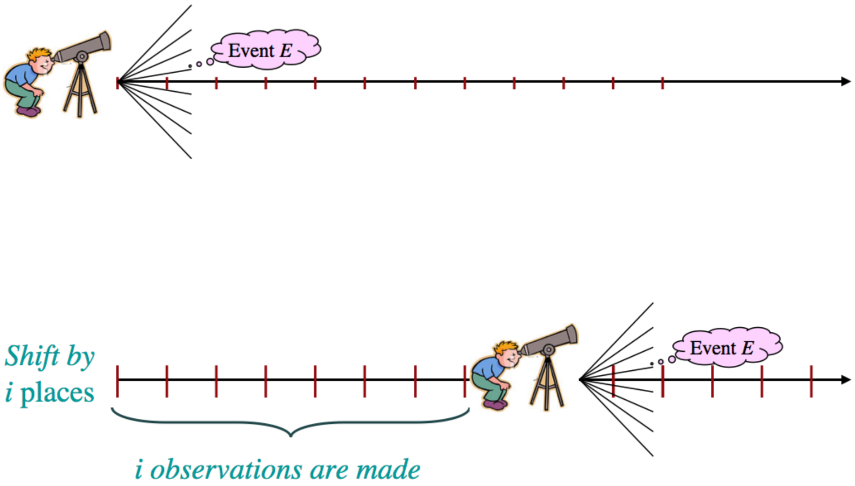
Ergodicity: *Goals*

Skip slide

We introduce learning-style conditions that imply subjective expected utility.

Ergodicity

Skip slide



Subjective Probabilities as Frequencies

Skip slide

- E is *invariant* if $T(E) = E$;
- $\succcurlyeq$ is *ergodic* if
  1. it is exchangeable, and
  2. E is invariant $\implies$ either E or E^c is $\succcurlyeq$-null.

Empirical distributions

Skip slide

- The *empirical distribution* at ω is the set-function which assigns to each event A the value

$$\nu(A, \omega) \equiv 1^{\star}_{A}(\omega),$$

if this value exists, and is not defined otherwise.

- this is just the empirical frequency of A along ω.

Subjective Probability from Frequencies

If $\succcurlyeq$ is ergodic, then:

- There is an event Ω' with $\succcurlyeq$-null complement, such that the empirical distribution $\nu(\cdot) = \nu(\cdot, \omega)$ is a well-defined probability distribution on $\mathcal{S}$ that is constant in $\omega \in \Omega'$; and
- $\succcurlyeq$ restricted to $\mathcal{F}_1$ is an expected utility preference with subjective probability ν.

Different derivations

Skip slide

Savage's theory:

normative axioms: P1, P2, P4, P6

+

exchangeability

$\Downarrow$

exchangeable subjective probability P

Our result:

Exchangeability & Ergodicity

$\Downarrow$

Expected utility: subjective belief = empirical measure ν

thus, P1, P2, P4, P6 !!

Statistical ambiguity: *Goal*

Skip slide

We define and characterize unambiguous events purely based on learning foundations.

Statistical ambiguity: *Summary*

▸ *Skip slide*

Assume our weak assumptions

Exchangeability

(but no substantive ambiguity-flavor assumptions: eg incompleteness, ambiguity aversion..)

$$\Downarrow$$

(a) Subjective set of priors driven by learning conditions

(b) expected utility on statistically unambiguous events

(c) no implications about the attitude towards ambiguity
(no pessimism, fear of malevolent nature..)

Statistical ambiguity

Skip slide

Assume:

- finite outcome space S
- focus on the set of acts $\mathcal{F}_1$ that depend only on the first coordinate.

Definition

Given a family of subsets $\mathcal{C} \subset 2^S$, a partial probability ν on $\mathcal{C}$ is a function $\nu : \mathcal{C} \to [0, 1]$ such that there is a probability distribution on S that agrees with ν on $\mathcal{C}$.

Statistical ambiguity

Skip slide

Definition

An event $A \subset S$ is $\succcurlyeq$-*statistically unambiguous* if there is $\Omega' \in \Sigma$ with $\succcurlyeq$-null complement, such that the frequency of A, $\nu(A) = \nu(A, \omega)$, is constant in $\omega \in \Omega'$.

The crucial part of the definition is the requirement that $\nu(A, \omega)$ is independent of ω off a $\succcurlyeq$-null set.

Characterization

Skip slide

Theorem (Statistical Ambiguity)

Assume S is finite. For any exchangeable preference $\succcurlyeq$

- *The set of* $\succcurlyeq$*-statistically unambiguous events* $\mathcal{C} \subset 2^S$ *is a* λ*-system, i.e., a family of sets closed under complements and disjoint unions;*
- *The empirical measure* $\nu(\cdot)$ *is a partial probability on* $\mathcal{C}$*; and*
- *For every* $\mathcal{C}$*-measurable acts* $f, g \in \mathcal{F}_1$*:*

$$f \succcurlyeq g \iff \int f \, d\nu \geq \int g \, d\nu.$$

$\mathcal{C}$ is only a λ-system.

Concluding remarks

Skip slide

- De Finetti understood the distinction between
  1. *Defining* subjective probability, and
  2. *Formulating* probability judgements
- Current conception of subjective probability emphasizes the first, and largely ignores the second
- Exchangeability, as formal model of similarity, is the part of the preference on which probability judgements are based
- This paper does what de Finetti's theorem cannot do:

  separate similarity judgements from the decision criterion

Moving forward

Skip slide

- An exchangeability relationship is the decision maker's *theory*, of what constitutes similar experiments
- ..but so far we have nothing to say about theory choice, theory change...
- Given our results, we can potentially talk about
  - Normative criteria that guide theory or model choice
  - Model uncertainty
  - Integrating classical and Bayesian methods

The End!!

Skip slide

Epstein-Seo, Part 1

Skip slide

Epstein-Seo. Part 1, strengthen:

$$f \sim f \circ \pi$$

to:

$$\alpha f + (1-\alpha) f \circ \pi \sim f \qquad (*)$$

Under our general assumptions, our condition

$$f \sim \frac{f \circ \pi_1 + \cdots + f \circ \pi_n}{n}.$$

neither implies nor is it implied by (*)

More important is the difference between our results

Epstein-Seo, Part 1

Skip slide

Assume our weak assumptions plus exchangeability

Epstein-Seo, Part 1:

Gilboa-Schmeidler axioms (completeness, C-indep., ambiguity aversion)

$\Downarrow$

(1) Gilboa-Schmeidler set of priors is exchangeable;
(b) pessimistic MEU criterion given this set

Our approach:

Various learning conditions

$\Downarrow$

(a) Subjective set of priors driven by learning conditions;
(b) no implications about the attitude towards statistical ambiguity

Epstein-Seo, Part 2

Skip slide

Epstein-Seo second model maintains:

$$f \sim f \circ \pi$$

but allow:

$$\alpha f + (1-\alpha) f \circ \pi \succ f \tag{$*$}$$

Now, parameters are *sets of measures*; the representation is a probability distribution over sets of measures

Examples of parameters:

- the set of all Dirac measures on Ω is a possible parameter
- so is the set of all independent distributions
- the set of all independent distributions with marginals belonging to $\{.2, .6\}$
- or to $\{.43, .91\}$, ... etc.

Epstein-Seo, Part 2

Skip slide

- Important contribution because it shows the sort of anomalies that might arise when exchangeability is weakened.
- A parametrization where individual parameters are sets, including the set of all sample paths, seems far removed from the intuitive idea of parameters as useful devices to summarize information.
- It is difficult to imagine how statistical inference can proceed on this basis.