Radically Elementary IP based on Extensive Measurement

T.Seidenfeld, M.J.Schervish, J.B.Kadane, R.Stern, and J.Cisewski – Carnegie Mellon University

Motivation for non-standard probability – three approaches and two old challenges

Approach #1 – Full Conditional Probability Distributions (Dubins, 1975) / Expected Utility
Approach #2 – Lexicographic Probability /Lex. Expected Value (e.g., Blume et al., 1991)
Approach #3 – Non-standard Probability/Expected Utility based on Extensive Measurement (Narens, 1974)

- **Challenge (1) Inference – satisfy symmetry of relevance/irrelevance relations (Relating to IP, see Cozman's research, 2015, …)**
  - **Subsidiary challenge – satisfy the *Sufficiency Principle*.**

  Defn.: Let $g(X) = Y$. Y is a *sufficient* summary for X with respect to quantity Z provided that $P(X \mid Y, Z) = P(X \mid Y)$, independent of Z.
- **Challenge (2) Decision theory – respect *admissibility* in finite partitions (Wald, 1950; Shimony, 1955)**

Continuing Example

$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$

	A	A^c
B	ω_1	ω_2
B^c	ω_3	ω_4

P(·) =

	A	A^c
B	.5	.5
B^c	0.	0.

Binary Decision Problem

	A	A^c
d_1	**1**	**0**
d_2	**0**	**1**

Approach #1
Full Conditional Probability Distributions

C-1 P(· | E) is well defined whenever E ≠ ∅
with P(E | E) = 1

C-2 P(FG | H) = P(F | GH)×P(G | H)
whenever GH ≠ ∅.

- *Asymmetry of the* P(· | ·) *relevance relation*

  $P(A) = P(A^c) = .50$ $P(B) = 1.0$ $P(B^c) = 0.0$

So, $P(A \mid B) = P(A^c \mid B) = .50$

Stipulate: $P(A \mid B^c) = .75 \neq P(A^c \mid B^c) = .25$.

Thus, $\{B, B^c\}$ is relevant to $\{A, A^c\}$.

	A	A^c
P(· | B)	.5	.5
P(· | B^c)	.75	.25

But $P(B) = P(B \mid A) = P(B \mid A^c) = 1$
and $P(B^c) = P(B^c \mid A) = P(B^c \mid A^c) = 0$

Thus, $\{A, A^c\}$ is irrelevant to $\{B, B^c\}$.

- Failure of *Sufficiency*

In the continuing example, Ω is a *sufficient* reduction of $\{B, B^c\}$ with respect to $\{A, A^c\}$ because
$P(B \mid \Omega \cap A) = P(B \mid \Omega) = P(B)$
and $P(B \mid \Omega \cap A^c) = P(B \mid \Omega) = P(B)$,
independent of $\{A, A^c\}$.
But $\{B, B^c\}$ is relevant to $\{A, A^c\}$.

- *Binary state decision problem*

Because $P(A) = P(A^c) = .50$,
d_1 and d_2 are indifferent, with expected value .50.

But, d_2 is inadmissible in the partition $\{B, B^c\}$.

	E(· | B)	E(· | B^c)
d_1	**.5**	**.75**
d_2	**.5**	**.25**

Conditional Expected Value

Neither challenge is met!

Approach #2
Lexicographic probability based on *tiers* from a full conditional probability.

Definitions (denumerable Ω):

- Atoms ω_i and ω_j belong to the same tier τ provided that $0 < P(\omega_i \mid \{\omega_i, \omega_j\}) < 1$.
- Tiers are well ordered: An atom belongs to a tier with a larger index if it is null relative one that belongs to a tier with lower index.
- The *lexicographic* (*countably additive*) *probability* $P^{\rightarrow} = < P_1, …., P_j, …>$, where $P_j = P(\cdot \mid \tau_j)$.
- The $P^{\rightarrow}$ *lex. prob. strict order* between two events is fixed by the first *tier* where they differ in probability.
- The $P^{\rightarrow}$ *lex. prob. expected utility strict order* between options is fixed by the first *tier* where they differ in expected utility

Continuing Example:

The full conditional prob P(· | ·) fixes 2 tiers:

$\tau_1 = \{\omega_1, \omega_2\}$ and $\tau_2 = \{\omega_3, \omega_4\}$.

$P_1(\omega) = P_1(\omega_2) = .50$: $P_1(\omega_3) = P_1(\omega_4) = 0.0$.

$P_2(\omega_1) = P_2(\omega_2) = 0.0$: $P_2(\omega_3)=.75$ and $P_2(\omega_4)=.25$.

So, A is strictly more $P^{\rightarrow}$ *lex.* probable than A^c.
Then, d_1 is strictly preferred to d_2,
under the $P^{\rightarrow}$ *lex. prob. expected utility strict order,*
A*dmissibility* is satisfied.

But, **$P^{\rightarrow}$ has an asymmetric irrelevance relation**.

$P^{\rightarrow}(A \mid B) = P_1(A) = .50 = P^{\rightarrow}(A^c \mid B) = P_1(A^c)$

$P^{\rightarrow}(A \mid B^c) = P_2(A) = .75 > P^{\rightarrow}(A^c \mid B^c) = P_2(A^c) = .25$.

Hence, $\{B, B^c\}$ is $P^{\rightarrow}$-relevant to $\{A, A^c\}$

But $P^{\rightarrow}(B \mid A) = <P_1(B \mid A), P_2(B \mid A) > = < 1, 0 >$
$= <P_1(B \mid A^c), P_2(B \mid A^c) > = P^{\rightarrow}(B \mid A^c)$.

Therefore, $\{A, A^c\}$ is $P^{\rightarrow}$-irrelevant to $\{B, B^c\}$.

$P^{\rightarrow}$ fails the Sufficiency principle.

Under $P^{\rightarrow}$, Ω is a *sufficient* reduction of $\{B, B^c\}$ with respect to $\{A, A^c\}$ because
$P^{\rightarrow}(B \mid \Omega \cap A) = P^{\rightarrow}(B \mid \Omega) = P^{\rightarrow}(B)$ and $P^{\rightarrow}(B \mid \Omega \cap A^c) = P^{\rightarrow}(B \mid \Omega) = P^{\rightarrow}(B)$, independent of $\{A, A^c\}$.

Approach #3
Radically elementary – non-standard probability based on extensive measurement.

Here is a non-standard probability *P to replace the full conditional P(· | ·) and the lexicographic $P^{\rightarrow}$

*P(·) =

	A	A^c
B	(1−ε)/2	(1−ε)/2
B^c	**3ε/4**	**ε/4**

Inference Problem

$*P(A^c) = (2\text{-}\varepsilon)/4 < *P(A) = (2+\varepsilon)/4 >$
$*P(A \mid B) = 1/2 < *P(A \mid B^c) = 3/4$.

So $\{B, B^c\}$ is *P-relevant to $\{A, A^c\}$.

$*P(B^c) = \varepsilon < P(B) = 1\text{-}\varepsilon <$
$*P(B \mid A) = (2\text{-}2\varepsilon)/(2+\varepsilon) <$
$*P(B \mid A^c) = (2\text{-}2\varepsilon)/(2\text{-}\varepsilon)$.

So $\{A, A^c\}$ is *P-relevant to $\{B, B^c\}$.
Thus, A and B are *P-negatively correlated.

***P has a symmetric relevance relation.**

Decision Problem

The *P expected value of $d_1 = (2+\varepsilon)/4$
The *P expected value of $d_2 = (2\text{-}\varepsilon)/4$,
which is less than the *P-expected value of d_2.

***P satisfies Admissibility.**

***P meets both challenges!**

Representation (and elicitation) for radically elementary, non-standard IP theory

Theorem (Narens, 1974): Given a *Radically Elementary Closed Extensive Structure* [***RECES***] { ≽, ⊕ } with domain $\mathcal{D}$ there exists a non-standard $*\mathfrak{R}^+$-valued function $*\boldsymbol{g}: \mathcal{D} \rightarrow *\mathfrak{R}^+$ where $*\boldsymbol{g}(d_1) \geq *\boldsymbol{g}(d_2)$ *iff* $d_1 \succcurlyeq d_2$ and $*\boldsymbol{g}(d_1 \oplus d_2) = *\boldsymbol{g}(d_1) + *\boldsymbol{g}(d_2)$. [See our Abstract for definitions.]

Corollary₁: Apply this Theorem to an agent's preferences over the domain of non-negative vectors on the finite set $\Omega = \{\omega_1, …, \omega_k\}$. The function ***g** induces a unique, non-standard regular probability, *P, defined on Ω, where P*(E) = 0 *iff* E = ∅.
Thus, *P(· | ·) is a non-standard, full-conditional probability on Ω satisfying Dubins' two conditions, **C-1** and **C-2**, above.

Corollary$_2$: Modify Axiom$_1$ for a ***RECES*** so that ≻ is a strict partial order. Follow the construction in [SSK, 1990] to create an *IP-representation of {≻, ⊕ } with a convex set of regular, non-standard *P-probabilities. Because each *P in the IP-set is regular, the corresponding convex set of conditional, non-standard probabilities, *P(· | E) is well defined for each E ≠ ∅.

Conjecture: As in [SSK, 2010] modify Axiom$_1$ for a ***RECES*** so that the decision maker provides a *coherent choice function*.

- Does this theory characterize **all** IP-sets of regular, non-standard probabilities on Ω so that no two have the same choice function?