

Conditional Independence, Imprecise Probabilities, Null-Events, and Graph-Theoretic Models

**Teddy Seidenfeld – Carnegie Mellon University
J.B.Kadane and M.J. Schervish (CMU)
Fabio Cozman –University of Sao Paulo**

Collaborated work with:

Outline

**Two problems relating to conditional probabilities and irrelevance.
And suggestions for redress.**

- 1. Competing senses of conditional irrelevance and independence within IP theory – indeterminate probability with sets of probabilities – matched with Decision theory for IP.**
- 2. Conditioning on (potentially) null events, i.e. when the conditioning event might carry 0 probability for some distributions in the IP set: the theory of full, conditional distributions (Dubins-Krauss theory).**

1. *Indeterminate or Imprecise Probability* [IP] theories provide one direction for relaxing traditional Bayesian theory in order to accommodate each of the follow demands:

- (i) Robust Bayesian inference – use a class of likelihood functions, or a class of priors, or both, e.g., inference with a class of exchangeable distributions; ε -contaminated models, etc..**
- (ii) Incomplete elicitation – use a class of distributions to constrain some aspects of uncertainty but not others. For example, de Finetti’s Fundamental Theorem can be interpreted this way.**

***Example 1:* Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, corresponding to the roll of a die. The *Bookie* announces *fair odds* (aka *previsions*) P for these four events:**

$$P(\{1\}) = 1/6; P(\{3,6\}) = 1/3; P(\{1,2,3\}) = P(\{1,2,4\}) = 1/2.$$

According to the rules of de Finetti’s betting game, the *Bookie* is obliged to take finite combinations of wagers with these fair odds – betting either way on the event or its complement. That fixes other fair odds for only 34 of the 64 events in the

algebra over Ω . For the remaining 15 event-pairs, coherence for fair odds on the remaining events compels the *Bookie* to a probability distribution within a (closed, convex) set, where the constraints on the Bookie's (*unspecified*) previsions are given by lower and upper probabilities $[\underline{P}(\bullet), \bar{P}(\bullet)]$.

For instance, though the prevision $P(\{4,6\}) = 1/3$ is determinate,

$$\begin{array}{l} \underline{P}(\{6\}) = 0 < \bar{P}(\{6\}) = 1/3; \\ \text{likewise } \underline{P}(\{4\}) = 0 < \bar{P}(\{4\}) = 1/3. \end{array}$$

(iii) Consensus in inference with more than one “expert.

It is widely understood that, e.g., standard pooling rules for amalgamating a collection of coherent distributions into a single distribution cannot preserve both unanimity in unconditional probabilities and unanimity in conditional judgments of relevance and irrelevance. But using the set of their opinions can do both – as we shall see shortly.

In this phase of our conversation, let us bracket concerns with null events and infinite partitions, and concentrate on simple IP theory with well defined conditional probabilities, and where conditional probabilities do not have 0 lower probabilities.

- *Our primary concern is how to define irrelevance.*

In the traditional, “precise” theory, the following identities are equivalent:

- $P(E,F) = P(E)P(F)$ *iff* $P(E|F) = P(E)$ *iff* $P(F|E) = P(F)$.

And that elementary fact, via Bayes’ theorem, shows how likelihood-concepts of relevance (e.g., *sufficiency, ancillarity*) relate to posterior-relevance.

The following IP concepts are presented using probability for events, i.e., expectations for indicator variables, but they generalize to expectations for bounded random variables.

Let $\underline{P}(E)$ and $\bar{P}(E)$ be the lower and upper probability for event E with respect to the set \mathcal{F} .

Likewise, let $\underline{P}(E | F)$ and $\bar{P}(E | F)$ be the lower and upper conditional probability for event E given event F , with respect to the set \mathcal{F} .

Weak Irrelevance of F to E :

$$\underline{P}(E) = \underline{P}(E | F) \quad \text{and} \quad \bar{P}(E) = \bar{P}(E | F)$$

This concept is easy to characterize with *one-sided* previsions for bets and called-off bets.

Alas, within IP theory, *Weak Irrelevance* is not a symmetric relation:

F may be weakly irrelevant to E but not conversely. (Cozman, 1999.)

***Extreme Example (with potentially null events):* $\Omega = \{\theta = .4, \theta = .6\} \times \{H, T\}$.**

$P(H | \theta = .6) = P(T | \theta = .4) = .6$. $\underline{P}(\theta = .4) = \underline{P}(\theta = .6) = 0$ and $\underline{P}(H) = \underline{P}(T) = .4$.

Then, $.4 = \underline{P}(H) \neq \underline{P}(H | \theta = .6) = .6$ – so θ is relevant to $\{H, T\}$.

But $0 = \underline{P}(\theta = .4) = \underline{P}(\theta = .6) = \underline{P}(\theta = .4 | H) = \underline{P}(\theta = .6 | H)$, etc. – so $\{H, T\}$ is irrelevant to θ .

More complicated examples avoid potentially null events entirely!

We can force a symmetric relation of weak independence:

Weak Independence of F and E:

E is weakly irrelevant to F, and F is weakly irrelevant to E.

But *Weak Independence* fails two *graphoid* \perp -independence properties, referred to as *Contraction* and *Intersection* for \perp (Cozman & Walley, 2005).

- **Contraction for \perp : If $(E \perp F)$ and $(E \perp G \mid F)$ then $(E \perp [F,G])$.**

When F and G are logically independent events,

- **Intersection for \perp : If $(E \perp F \mid G)$ and $(E \perp G \mid F)$ then $(E \perp [F,G])$.**

- **Without *Contraction*, e.g., the *d*-separation theorem is false for DAGs.**

- ***Intersection* is needed for Markov Random Fields.**

See E. San Martin (2006) connecting *Intersection* and *Basu's Theorem*.

There are two, progressively stronger senses of independence available within IP theory, which entail symmetric accounts of irrelevance and are worthwhile discussing.

Strong Independence

Events E and F are *strongly independent* with respect to the closed, convex set \wp if each extreme point P of \wp satisfies the usual factorization: $P(EF) = P(E)P(F)$.

Complete Independence

Events E and F are completely independent with respect to the set \wp if each point P of \wp satisfies the usual factorization: $P(EF) = P(E)P(F)$.

In the balance of this section, I want to argue for *Complete Independence* but the price it carries for decision theory is not trivial!

Decision Theory for IP.

For simplicity here, the following four restrictions are used throughout:

- **Act-state independence: no cases of “moral hazards” are considered – so basic dominance is valid.**
- **State-independent utility: no cases where the value of a prize depends upon the state in which it is received.**
- **There is a common cardinal utility for prizes – only uncertainties in beliefs are involved in group/individual decision problems: team play.**
- **Focus is on normal form (aka “strategic form”) decisions. However, there is no assumption that normal and extensive form decisions are equivalent.**

- A basic result why an IP decision needs a Bayes-model.

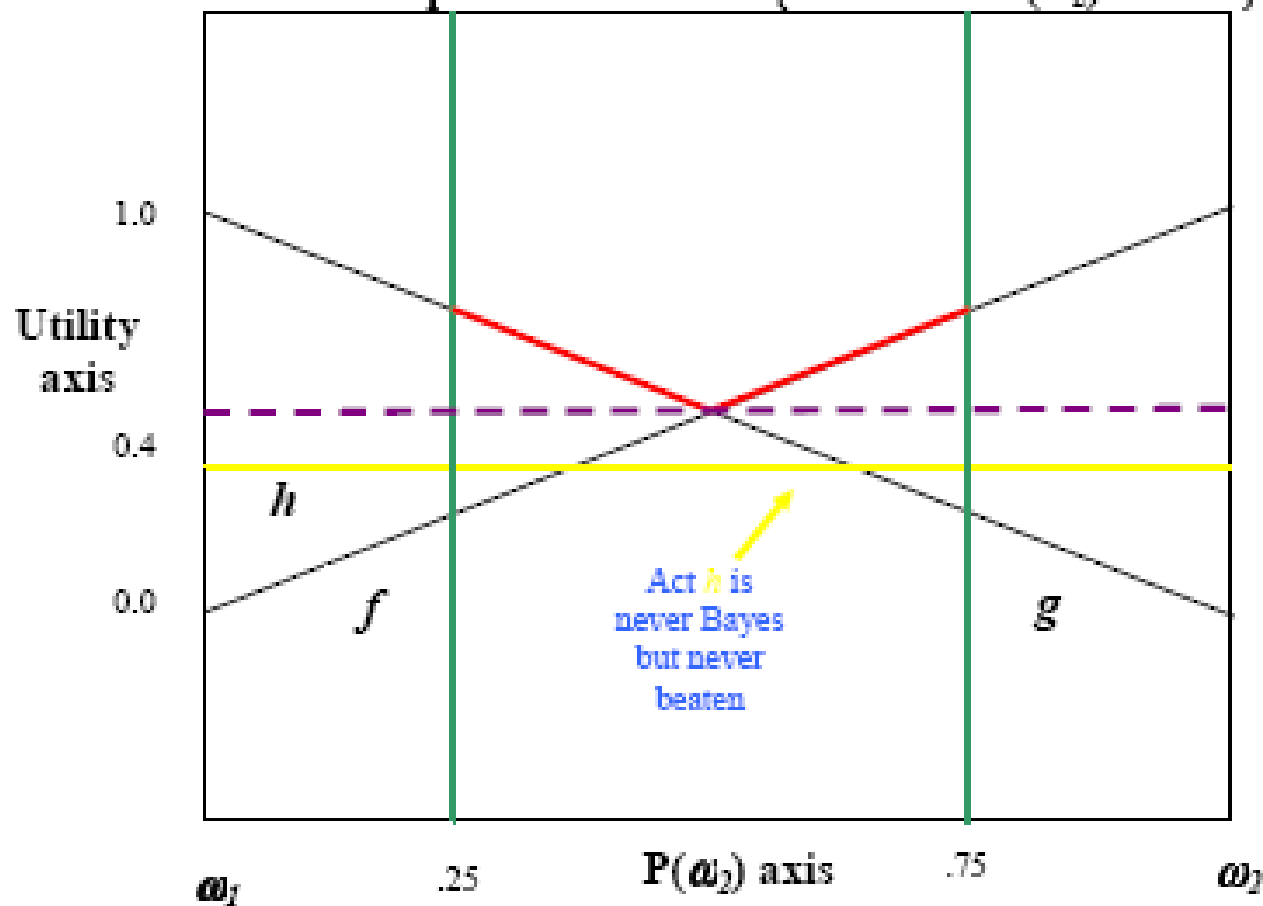
General framework: Given a (closed) set O of feasible options, a choice function C identifies the set A of acceptable options $C[O] = A$, for a non-empty subset $A \subseteq O$.

Definition: Option $o \in O$ has a local Bayes model P if o maximizes the P -expected utility over the options in O .

Theorem 1 (Pearce, 1984 for finite state spaces): If an option $o \in O$ fails to have a local Bayes model then it is uniformly, strictly dominated by a finite mixture of options already available from O .

In this sense, options that do not have even a local Bayes model suffer de Finetti's penalty – being uniformly strictly dominated by a mixed option – within the decision at hand: a strengthened *Complete Class* Theorem. Using a result due to Kindler (1983) it applies also to infinite state spaces.

Consider a decision problem on 2 states $\{\omega_1, \omega_2\}$, with 3 options $\{f, g, h\}$, and a convex set of probabilities $\mathcal{P} = \{P: .25 \leq P(\omega_2) \leq .75\}$.



Example 1

Contrast three decision rules for extending Expected Utility [EU] theory when probability – but not cardinal utility – is indeterminate. Decisions involve uncertainty, represented by a set of probabilities \wp .

- ***\wp -Maximin* (Gilboa-Schmeidler, 1989) – Maximize Minimum expectations over \wp . This decision rule is summarized by a binary preference relation that satisfies postulates for *Ordering* but not *Independence*. It uses a closed, convex set \wp .**

In Example 1 only h is *\wp -Maximin* from $\{f, g, h\}$.

- ***Maximality* (Sen 1977/Walley 1991) – admissible choices are those that are undominated in expectations over $P \in \wp$ – a convex, usually closed set – by any single alternative choice. It uses pairwise comparisons for expectation dominance. This decision rule is given by binary comparisons, but does not support *Ordering*.**

In Example 1 each of $f, g,$ and h is *Maximal* from $\{f, g, h\}$

- ***E*-admissibility (Levi, 1980 / Savage, 1954!) – admissible choices have Bayes models, i.e., they maximize *EU* for some probability in the (*convex*) set \wp . This decision rule does not reduce to pairwise comparisons.**

In Example 1 only *f* and *g* are *E*-admissible from $\{f, g, h\}$.

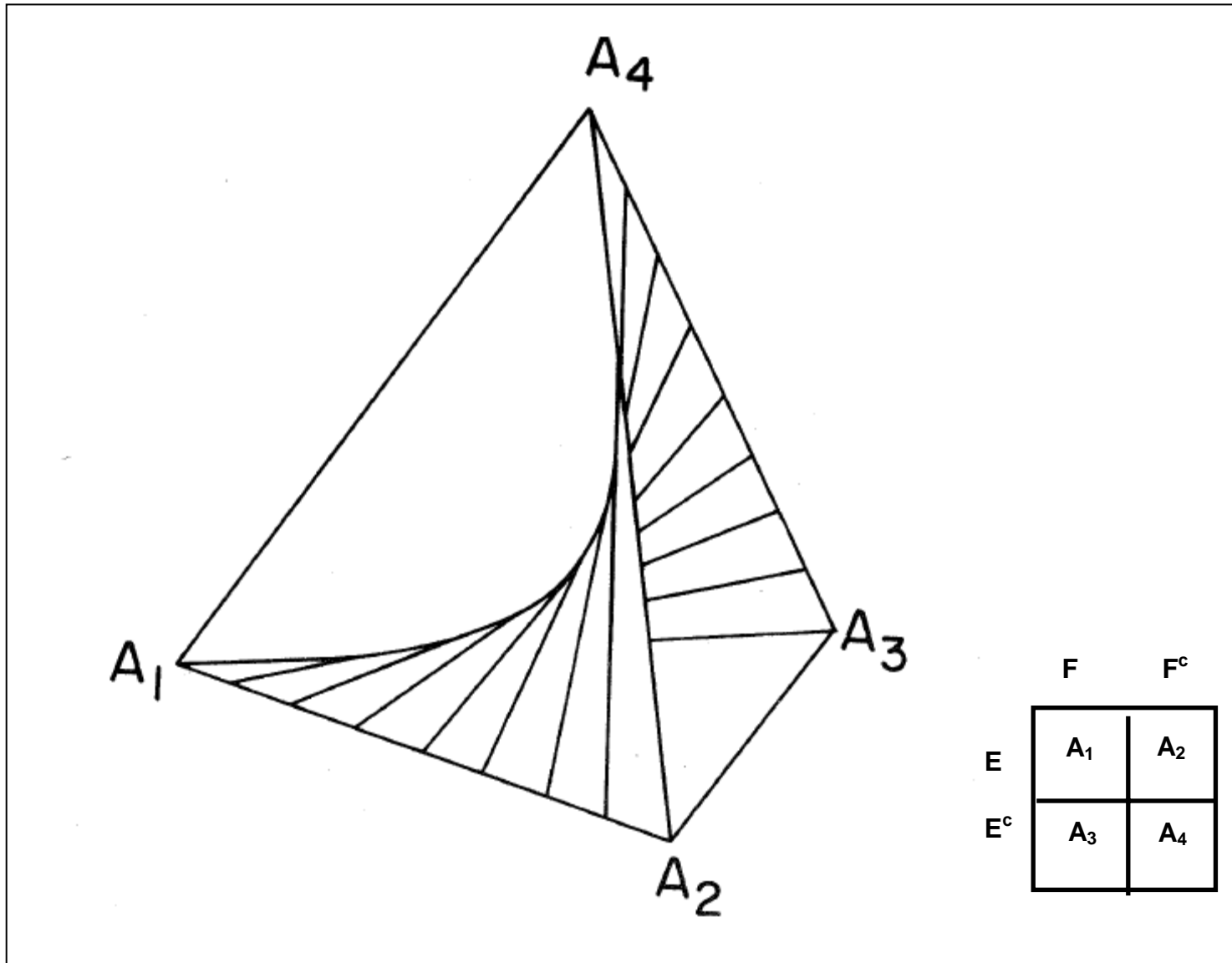
- **Each rule has *EU* Theory as a special case when probability is determinate, when \wp is comprised by a single probability distribution.**
- **And each rule avoids sure-loss – no “Book” is possible in betting, when \wp is convex and closed, and the option set is convex as well – closed under mixed strategies.**
- **The three rules reflect a progression, where each rule relaxes more of the ordering assumption than does its predecessors.**

The third decision rule – admissible options from O are those that have a Bayes-model against the set \wp – is different from the other two in that it does not require any special features of \wp , which may even be a disconnected set.

For an axiomatization of this decision rule for arbitrary sets of probabilities, see Seidenfeld, Schervish, Kadane (2007).

This is important when we consider how IP theory captures judgments of irrelevant and independence.

Recall, the surface of independence between two events E and F is not convex and has only distinguished convex regions. So, if our IP decision theory requires convexity of the set \wp , it will fail to preserve the set of distributions each of which makes E and F independent.



The following example illustrates how care is needed for sound reasoning even when complete independence obtains

Example 2: Dilation for pivotal variables

Suppose A is an *uncertain* event in the added sense of “uncertainty” that comes with a set of probabilities \wp .

That is $P_*(A) = .25$ and $P^*(A) = .75$.

Let $\{H,T\}$ indicate the flip of a fair coin whose outcomes are completely independent of A . That is, $P(A,H) = P(A)/2$ for each $P \in \wp$.

As A and H are completely independent, and as \wp is a closed, convex set, also A and H are strongly and weakly independent.

Define event E by, $E = \{(A,H) , (A^c,T)\}$.

	H	T
A	E	E^c
A^c	E^c	E

It follows, simply, that $P(E) = .5$ for each $P \in \wp$.

Thus, E is a *pivotal variable* with respect to the uncertain event A.

Then $.50 = P_*(E | A) = P_*(E) = P^*(E) = P^*(E | A) = .50,$

and $.25 = P_*(A | E) = P_*(A) < P^*(A) = P^*(A | E) = .75$

so A and E are weakly irrelevant – and even strongly irrelevant as

$.125 = P_*(E,A) = P_*(E)P_*(A),$ etc.

But $.25 = P_*(E | H) < P_*(E) = .5 = P^*(E) < P^*(E | H) = .75,$

and $.25 = P_*(H | E) < P_*(H) = .5 = P^*(H) < P^*(H | E) = .75$

so that H and E are *weakly relevant* in both directions, and thus strongly and completely relevant.

Note that upon learning the outcome of the coin flip {H, T} the determinate probability for E, $P(E) = .5,$ dilates to the interval $[.25, .75]$

$.25 = P_*(E | H) = P_*(E | T)$ and $P^*(E | H) = P^*(E | T) = .75$

likewise learning the outcome of the pivotal {E, E^c} dilates probability for the coin

flip $.25 = P_*(H | E) = P_*(H | E^c)$ and $P^*(H | E) = P^*(H | E^c) = .75.$

- As demonstrated in Cozman and Walley (2005) whenever the Intersection property fails for independence, then dilation ensues.**

Next, we consider the decision-theoretic difference between

***Strong Independence* – where extreme points of convex \wp factorize,**

and *Complete Independence* – where all points of \wp factorize.

Let us consider a group of N-many coherent “experts” who are forming a *team* with respect to some common goal, e.g., deciding on the allocation of a medical treatment for a subject.

- We assume that they share a common utility vis-à-vis the medical outcomes for the patient.**
- However, they have different judgments about how the patient will respond to rival treatments.**

Their cooperative decision making as a team requires an agreement on what features of their different beliefs are to be preserved in their collective actions.

What features of their shared beliefs and values will be reflected in their determination of acceptable options as a team?

***Proposal:* Preserve unanimity of unacceptable options.**

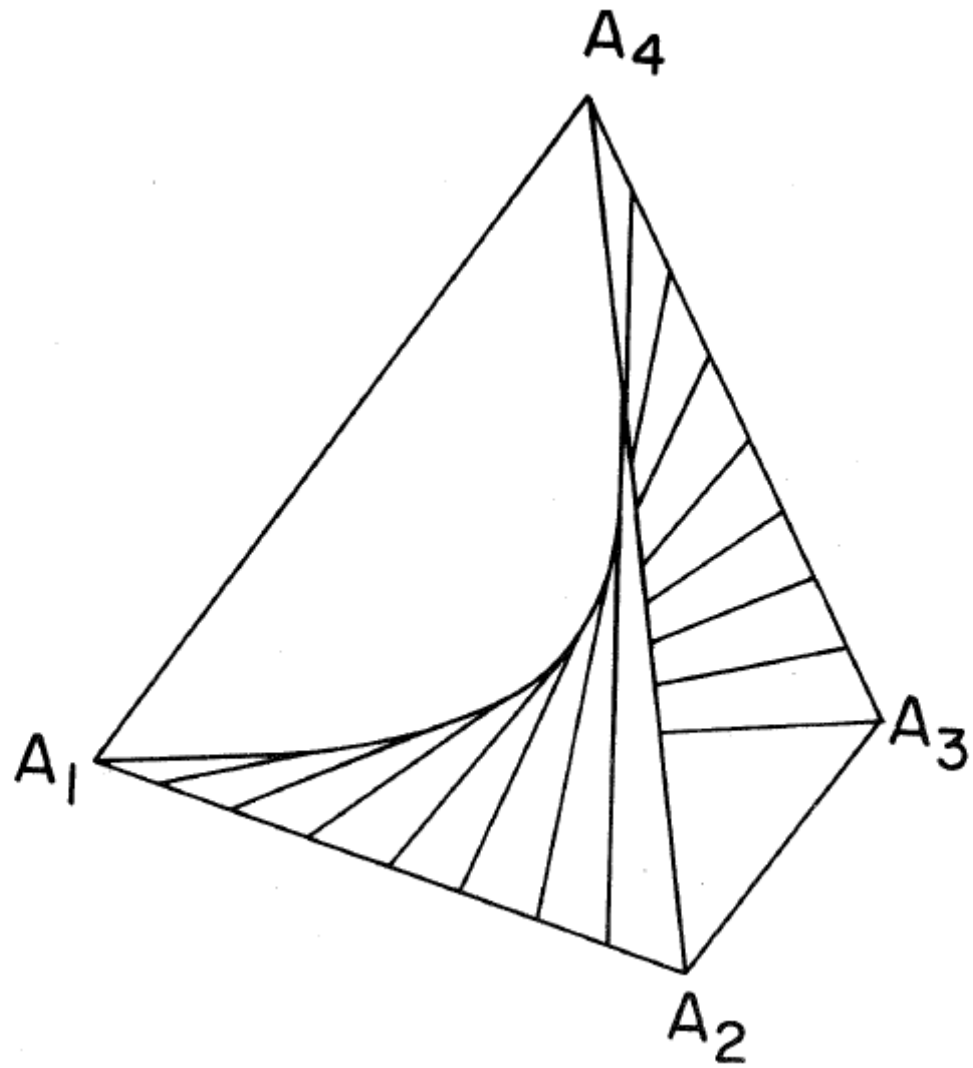
This creates a choice function for the team whose admissible options are precisely those that are admissible for at least one expert.

**Note: With binary choice problems, this is equivalent to a Pareto rule –
If everyone strictly prefers o_1 over o_2 , then so does the team.**

The proposal results in the team's choice function equal to one given by a set of probabilities, \wp_T , formed by taking the union of the experts' individual opinions, P_i ($i = 1, \dots, n$): $\wp_T = \cup_i P_i$ and using E-admissibility with respect to \wp_T .

In particular, this proposal allows the team to respect shared agreements among the experts regarding both unconditional probabilities and probabilistic independence, as those affect the set of probabilities used for decision making.

Recall that the surface of probabilistic independence between two events is saddle shaped, and without convex regions except where one or the other event is given a constant probability.



Example 3: Let the normal form decision problem be a choice among three options – three treatment plans $\{T_1, T_2, T_3\}$ defined over 4 states $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ with determinate utility outcomes given in the following table. That is, the numbers in the table are the utility outcomes for the options (rows) in the respective states (columns)

	ω_1	ω_2	ω_3	ω_4
T_1	0.00	0.00	1.00	1.00
T_2	1.00	1.00	0.00	0.00
T_3	0.99	-0.01	-0.01	0.99

The team is comprised by two medical experts, each with determinate probability distributions P_1 and P_2 , defined by the following table.

Distribution P_3 is the .50-.50 (convex) mixture of P_1 and P_2 .

	ω_1	ω_2	ω_3	ω_4
P_1	0.08	0.32	0.12	0.48
P_2	0.48	0.12	0.32	0.08
P_3	0.28	0.22	0.22	0.28

Note that (for $i = 1, 2, 3$), under probability P_i , only option T_i maximizes expected utility from the option set of $\{T_1, T_2, T_3\}$.

Without convexity – that is, using only the set comprised by the two (extreme) distributions $\wp_T = \{P_1, P_2\}$ – option T_3 is the sole unacceptable option from among the three options $\{T_1, T_2, T_3\}$. But with *Strong Independence*, where the convex set is the convex hull of \wp_T , all three options are admissible.

Now, interpret these states as the cross product of a binary medical partition with a binary meteorological partition.

Label the medical partition as A (= allergic) and NA (not-allergic);

Label the meteorological partition as S (sunny) and C (cloudy).

Specifically: $\omega_1 = A \& S$ $\omega_2 = A \& C$ $\omega_3 = NA \& S$ $\omega_4 = NA \& C$.

T₁	0.00	0.00	1.00	1.00
T₂	1.00	1.00	0.00	0.00
T₃	0.99	-0.01	-0.01	0.99

For expert P₁, the two events are independent: P₁(A) = .4 and P₁(S) = .2.

Likewise, for P₂, the events are independent: P₂(A) = .6 and P₂(S) = .8.

And for distribution P₃ the events A and S are positively correlated:

$$.56 = P_3(A | S) > P_3(A) = .5,$$

as with each distribution P that is a (non-trivial) mixture of P₁ and P₂.

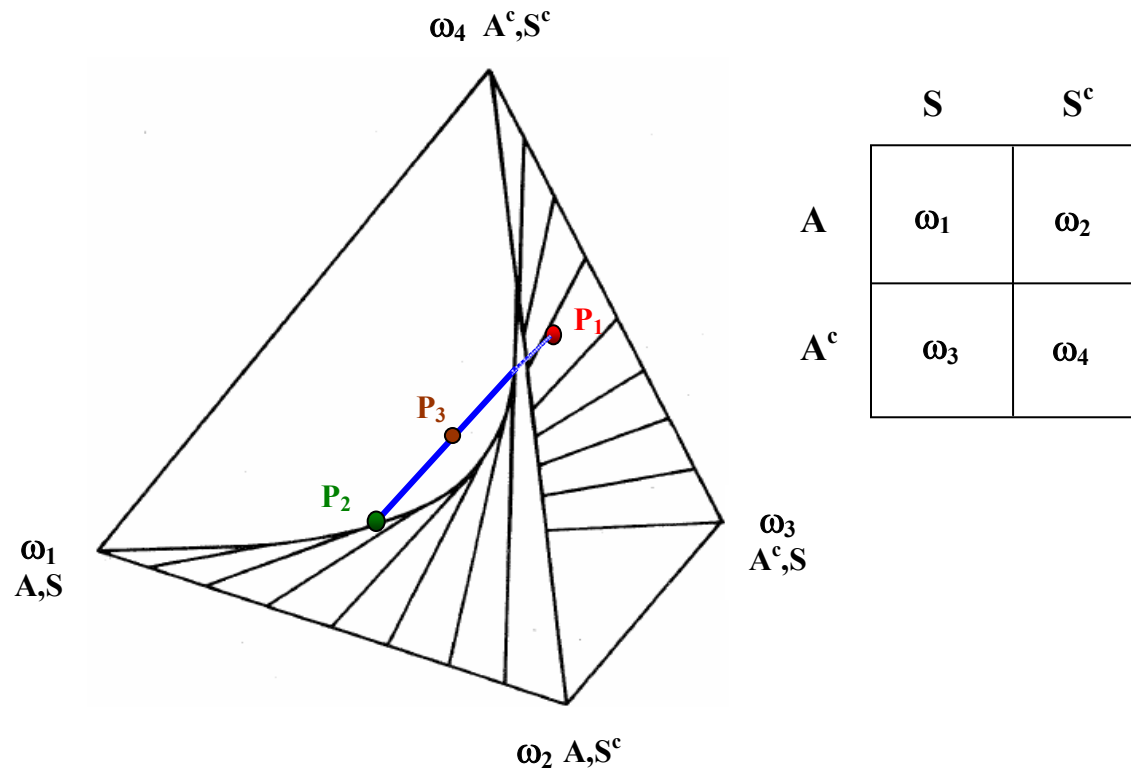


Diagram of the *Strongly Independent* (but not *Completely Independent*) IP set that arises by convexifying the two expert opinions, $\{P_1, P_2\}$.

Continuing with the story, the three options have the following interpretations:

T_1 and T_2 are ordinary medical options, with outcomes that depend solely upon the patient's allergic state.

T_3 is an option that makes the allocation of medical treatment a function of the meteorological state, with a “fee” of 0.01 utile assessed for that input. That is, T_3 is the option

“ T_1 if cloudy and T_2 if sunny, with a fee of 0.01 for that input.”

Suppose that P_1 represents the opinion of medical Expert₁, and P_2 represents the opinion of medical Expert₂. Without convexity of the probabilities for the team, T_3 is inadmissible.

Inadmissibility of T_3 captures the shared agreement between the two medical experts that T_3 is unacceptable from the choice of three $\{T_1, T_2, T_3\}$. And it captures the pre-systematic understanding that under T_3 you pay to use medically irrelevant inputs in order to determine the medical treatment.

With convexity of the set generated by P_1 and P_2 , as required for *Strong Independence*, then T_3 is admissible.

Each local Bayes model for T_3 in this decision requires a *positive association* between the medical and meteorological events.

What do you think of a medical “expert” who holds such beliefs?

- Thus we have an operational distinction between *Strong* and *Complete independence* within IP theory.

***E* and *F* are completely independent *if and only if* for no (normal form) decisions is it admissible to “buy” information about the one in order to determine a decision whose outcome depends solely on the other.**

Aside: From a *bargaining* point-of-view, it makes sense for each expert to accept T_3 . That option, T_3 , allows each to think that, with probability .8, his/her medical view will decide the treatment allocation for the patient. But that is not a consensus or indeterminate point-of-view.

Summary of Section 1:

Weak Independence (the symmetrized *Weak Irrelevance* relation) fails key graphoid properties necessary for the use of directed and undirected graphical representations of independence.

Strong Independence (where the extreme points of the convex set \wp factorize) compels admissibility of acts that spend resources on using irrelevant inputs for determining a terminal decision.

Complete Independence has neither of these failings! But in order to use Complete Independence with a matching IP decision theory – for example, the Bayes- or *E*-admissibility rule, decisions do not reduce to pairwise comparisons among the feasible options!

2. *Conditional Independence with Null Events.*

Here we examine similar concerns with independence given null events.

Aside: From a decision-theoretic perspective, neither Savage's nor deFinetti's theory affords any guidance about conditional probability given a null event.

That is, if $P(F) = 0$, $P(\bullet | F)$ is not compelled to be a coherent probability function. What is called for is a non-Archimedean theory, as sketched below.

Following Krauss (1968) and Dubins (1975) define a full conditional (finitely additive) probability on algebra \mathcal{A} ,

$$P(\bullet|\bullet): B \times (B \setminus \phi) \rightarrow [0,1]$$

by these four axioms.

For each $\phi \neq C \in \mathcal{F}$,

Axiom 1: $P(C | C) = 1$ (Note: This condition fails for improper rcd's.)

Axiom 2: $0 \leq P(\bullet|\bullet) \leq 1$

Axiom 3: $P(A \cup B | C) = P(A|C) + P(B|C)$ whenever $A \cap B = \phi$

Axiom 4: $P(A \cap B | C) = P(A | B \cap C)P(B|C)$ whenever $B \cap C \neq \phi$

And we can consider the analog to weak irrelevance as follows:

Weak Conditional Irrelevance of F to E, given C:

$$P(E | C) = P(E | F, C) \text{ whenever } F \cap C \neq \phi.$$

Alas, this is not a symmetric notion of irrelevance!

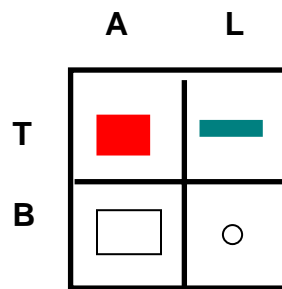
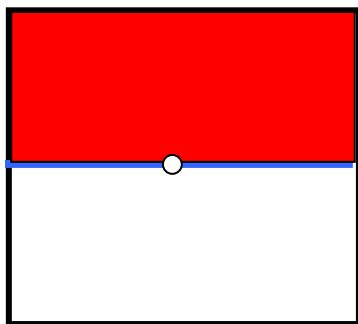
Example 4: Consider uniform (Lebesgue) measure on the unit square, with the following 2 x 2 partition.

TOP: $\{(x,y): y \geq 1/2 \text{ and excluding } (1/2, 1/2)\}$

BOTTOM: $\{(x,y): y < 1/2 \text{ and including } (1/2, 1/2)\}$

LINE: $\{(x,y): y = 1/2 \}$

AREA: Unit Square excluding Line



**$P(\text{Line} \mid \text{Top}) = P(\text{Line} \mid \text{Bottom}) = 0$. But $P(\text{Top} \mid \text{Line}) = 1 \neq P(\text{Bottom} \mid \text{Line}) = 0$.
So Top/Bottom is irrelevant to Area/Line, but not conversely!**

Not only does this account of irrelevance fail symmetry, also it fails the graphoid property of Weak Union.

- **Weak Union: If $(E \perp [F, G])$ then $(E \perp F)$.**

As a result, the symmetrized independence relation (= irrelevance in both directions) also fails Weak Union.

Hence, independence in Krauss/Dubins' full-conditional probabilities fails to satisfy a basic graphical representation, e.g., d -separation also fails.

Let E and F live in the same tier of probability if

$$0 < P(E \mid E \cup F) < 1.$$

One diagnosis of the problem we see in Example 4 is that we have insufficient information about tiers preserved in the real-valued full conditional probability function.

The conditional probability measure $P(\bullet \mid E, F)$ may collapse some of the tiers of null events relative to the conditional probability $P(\bullet \mid E)$.

One approach to a satisfactory account of irrelevance with null events, then, is to replace Dubins' real-valued, full conditional probabilities with a non-Archimedean conditional probability function.

- See M. Troffaes (in preparation) for helpful discussion of some *non-standard* full conditional probabilities.
- A review of other approaches is given in Cozman & Seidenfeld (2007).
- Another, less dramatic solution is to use *lexicographic*, full conditional probabilities. *Conjecture*: An adequate theory of non-Archimedean full conditional probability is given by a lexicographic approach.

Summary

1. We examined different concepts of irrelevance and independence with IP theory. In order to defend the concept of *complete independence* – all distributions in the IP set \wp factor – we investigated several decision rules for use in IP theory.

One of these decision rules,

***E-admissibility* – admissible options are Bayes for some probability in set \wp supports *complete irrelevance*. That notion satisfies the usual graphoid properties that we know from the familiar case where probability is determinate, i.e. where \wp is a single distribution.**

2. There are parallel problems with irrelevance and independence arising within the framework of (Krauss/Dubins) full conditional real-valued probabilities. I suggest the use of non-Archimedean probability theory to address these problems.

Selected References

- Cano, J.E., M. Delgado and S. Moral. (1993) “An Axiomatic framework for the propagation of Uncertainty in Directed Acyclic Networks” *International Journal of Approximate Reasoning* 8, 253-280.
- Cozman (2000) “Credal Networks” *Artificial Intelligence*, 120: 199-233.
- Cozman, F.G. and Walley, P. (2005) “Graphoid properties of epistemic irrelevance and independence,” *Annals Math. and A.I.* 45: 173-195.
- Cozman, F.G. and Seidenfeld, T. (2007) “Independence for Full Conditional Measures, Graphoids and Networks,” U. Sao Paulo, Computer Science Technical Report.
<http://www.poli.usp.br/p/fabio.cozman/Publications/Report/ind10.pdf>
- Dubins, L.E. (1975) “Finitely additive conditional probabilities, conglomerability, and disintegrations,” *Ann Probability* 3: 89-99.
- de Finetti, B. (1974) *Theory of Probability* (2 volumes) New York: Wiley
- Gilboa, I. and Schmeidler, D. (1989) “Maxmin expected utility with non-unique prior,” *J. Math.Econ.* 18: 141-153.
- Kadane, J.B., Schervish, M.J., and Seidenfeld, T. (2004) “A Rubinesque theory of decision,” *IMS Lecture Notes Monograph* 45: 1-11.
- Kindler, J. (1983) “A General Solution Concept for Two Person, Zero Sum Games,” *J. Optimization Theory and Applications* 40: 105-119.
- Levi, I. (1974) “On indeterminate probabilities” *J.Phil.* 71: 391-418.
- I. Levi (1980) *Enterprise of Knowledge* Cambridge: MIT Press.

- Pearce, D. (1984) “Rationalizable Strategic Behavior and the Problem of Perfection”**
Econometrica 52: 1029-1050.
- San Martin, E., M. Mouchart, J.-M. Rolin, (2005) “Ignorable Common Information, Null Sets and Basu’s First Theorem”** *Sankhya:The Indian Journal of Statistics*, 67, 674–698.
- Savage, L.J. (1954) *The Foundations of Statistics*. New York: Wiley.**
- Schervish, M.J., Seidenfeld, T, Kadane, J.B., and Levi, I. (2003) “Extensions of expected utility theory and some limitations of pairwise comparisons”** In *Proceedings of the Third ISIPTA* (J-M Bernard, T.Seidenfeld, and M.Zaffalon, eds.) 496-510. Carleton Scientific.
- Seidenfeld, T., Schervish, M.J., and Kadane, J.B. (1995) “A representation of partially ordered preferences,”** *Ann Stat.* 23: 2168-2217.
- Seidenfeld, T., M.J.Schervish, and J.B.Kadane (2007) “Coherent Choice Functions.”** In De Cooman, Vejnarova, and Zaffalon (eds.) ISIPTA-07, Proceedings. SIPTA: Prague: pp. 385-394. Available on-line at www.sipta.org.
- Sen, A. (1977) “Social choice theory: a re-examination,”** *Econometrica* 45: 53-89.
- Troffaes, M. (in preparation) “Conditional Independence via Non-Archimedean Probabilities,”** Dept. of Statistics, Durham University
- Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.**