A good year for imprecise probability

Gregory Wheeler

June 18, 2008

The theory of imprecise probability (IP) has been slow to catch on in philosophy, although there are signs that this is beginning to change. Gert de Cooman's link between IP and logic, presented in a terrific talk at progic 2007, is sure to draw interest from philosophical logicians; a general result in de Cooman and Miranda (2007) showing how Finetti's exchangeability theorem drops out as a special case should receive attention from philosophers of science; and a proposal by Carl Wagner (2007) to get around weak book arguments by replacing the classical theory of *linear* previsions with IP's lower/upper previsions suggests some interest in Bayesian epistemology.

Even so, the theory of imprecise probability is loaded with surprises, which may explain the spare use of the theory so far. But 2007 was a watershed year for imprecise probability, owing to a paper by Teddy Seidenfeld, Mark Schervish, and Jay Kadane (2007). I want to call attention to this result, which builds on their work on imprecise probability over the last two decades.

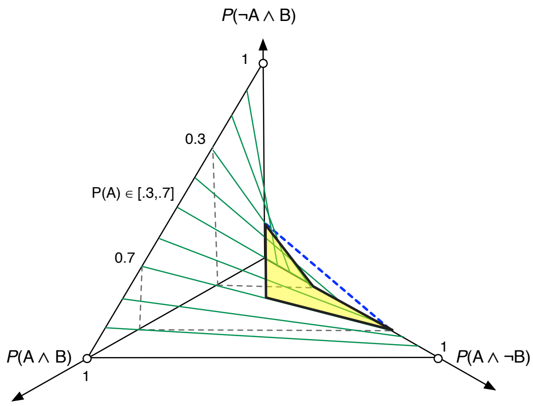
For those accustomed to thinking about probabilistic *independence* and *coherence* in terms of a single probability distribution, it can be a shock to discover how different these notions behave in an imprecise probability setting.

Consider two binary variables, A and B, but rather than assign sharp probabilities to each we assign intervals, $P(A) \in [0.3, 0.7]$ and $P(B) \in [0, .5]$. We may specify a full joint distribution over four atomic states, taking $x_1 = P(A \wedge B), x_2 = P(A \wedge \neg B), x_3 = P(\neg A \wedge B)$, and $x_4 = P(\neg A \wedge \neg B)$, and represent the constraints on A and B within a 3-dimensional unit simplex.

In the figure below we picture a cross-section of this tetrahedron. When $x_1 = 1$, the probability that A and B are both 'true' is 1, which is the extremal point on the lower left of the tetrahedron; when $x_4 = 1$, the probability that both A and B are 'false' is 1, which is the vertex whose label is omitted from the figure. And so on.

We see that the marginal probability of $P(A)$ is represented along the axis representing values for $P(A \wedge B)$ and $P(\neg A \wedge B)$, which is labeled '$P(A) \in [.3, .7]$' on the left, as well as along the axis representing values for $P(A \wedge \neg B)$ and $P(\neg A \wedge \neg B)$, whose label is omitted. $P(B) \in [0, .5]$ may likewise be represented in a similar fashion by the remaining pair of edges.

Within this polytope the points where A and B are *not* dependent are those on the curved 'surface of independence', and the shaded quadrilateral on the surface of independence identifies points where A and B are independent of one another *and* satisfy the constraints in our example, namely that $P(B)$ takes values between 0 and 1/2, and $P(A)$ takes values between 0.3 and 0.7.



Notice that there is a thick, dashed line connecting the extremal point denoting $(P(B) = 0.5, P(A) = 0.3)$ to the extremal point denoting $(P(B) = 0, P(A) = 0.7)$. The endpoints of this line lie on the surface of independence, but all points between denote values in which A and B are dependent. There is quite a lot of mischief that follows from this observation.

For instance, if you think that degrees of belief are closed under linear transformation, then you'll be surprised to find that while the extremal values of A and B are independent, a convex mixture of A and B may *fail* to be independent. Our figure above is a counterexample.

Another surprise concerns the limits of pairwise comparisons of options in an (imprecise) decision problem. Peter Walley's theory of imprecise probability (Walley 1991) inherits an idea from Savage, and also from Von Neumann and Morgenstern, whose axiomatic treatment of rational choice relies upon binary comparisons of pairs of options (gambles). These axiom systems yield a coherent probability assessments in the sense that such assessments avoid sure loss.

However, we might wonder whether binary comparison is too strong a condition. There are well-known objections to Von Neumann and Savage's axioms that describe an otherwise rational agent who nevertheless is unwilling to choose among 'equivalent' options but is not indifferent to their outcomes. And although this has appeared to some to be an elicitation problem, Isaac Levi (1980) observed that pairwise comparison may fail to distinguish between relevantly different options is a more direct challenge to this condition. Let's pursue Levi's point further.

To help illustrate, we follow Seidenfeld *et. al.* in defining a *choice function* C over a closed set of feasible options O, $C : O \mapsto \mathcal{P}(O)$, where $C(O)$ identifies that subset of O that are admissible options to the decision problem given by O.

$C(O)$ is a *coherent choice function* iff there is a non-empty set $\mathbf{S} = \{\langle p_1, u_1\rangle, \ldots, \langle p_n, u_n\rangle\}$ of probability-utility pairs such that the admissible options under C are exactly those that are Bayes with respect to some $\langle p, u\rangle \in \mathbf{S}$. In other words, for each admissible option o under $C(O)$, there is a pair $\langle p, u\rangle$ in $\mathbf{S}$ such that o maximizes the p-expected u-utility over O. These are called the *Bayes admissible-* or *acceptable options* in O with respect to S. *Unacceptable options* (with respect to S) are defined as $O - C(O)$.

Returning to Levi's observation, consider now the following example from (Schervish et al. 2003).

▶ **Example 1 (Schervish et al. 2003)** Suppose a binary decision, ω_0 or ω_1, concerning three options, $O = \{f, g, h\}$, where utility is determinate: $\mathbf{u}(f(\omega_0)) = \mathbf{u}(g(\omega_1)) = 0, \mathbf{u}(f(\omega_1)) = \mathbf{u}(g(\omega_0)) = 1.0, \mathbf{u}(h(\omega_0)) = \mathbf{u}(h(\omega_1)) = 0.4$. (In words: obtaining f and relinquishing g yields expected utility 0, obtaining f and relinquishing g yields expected utility 1.0, and option h yields a sure 0.4.) Let uncertainty over states be indeterminate, where $P(\omega_1) \in [0.25, 0.75]$. (In words: the probability that state ω_1 obtains is between 0.25 and 0.75, inclusive.) There are three decision rules for this problem.

1. *Maximin*: Maximize minimum expected utility over options O. Maximin induces a preference ordering over options and only $\{h\}$ is admissible from the set O. However, the von-Neumann-Morgenstern independence postulate does not hold.

2. *Maximality*: All three options $\{f, g, h\}$ are admissible under maximality, since no option dominates the others in pairwise comparisons. There is no preference ordering over O, but instead admissibility is given by pairwise comparisons of all options in O.

3. *Coherent Choice*: Only options $\{f, g\}$ are admissible from the set O under coherent choice, since both f and g maximize expected utility for some probability in the convex set $\mathbb{P} = \{P : 0.25 \leq P(\omega_1) \leq 0.75\}$. This rule neither induces an ordering over options nor reduces to pairwise comparisons.

Since h is uniformly dominated by some (convex) mixture of f and g, h is never a 'Bayes'-admissible option with respect to $\mathbb{P}$, i.e., there is always some mixture of f and g that maximize expected utility. More precisely, the mixed alternatives $q_x := xf \oplus (1-x)g$, for $0.4 < x < 0.6$, uniformly dominates option h. But each mixed alternative q_x maximizes expected utility with respect to the convex, closed hull of the options set O for exactly one probability, namely $P(\omega_0) = 1/2$. ◀

The distinction operating behind Example 1 is Levi's (1980) distinction between 'second best' and 'second worse'. Option h is second worse. If instead h were replaced by h' with constant 0.6 (i..e, $\mathbf{u}(h'(\omega_1)) = 0.6$), then h' would be second best in the set $\{f, g, h'\}$. Levi observed that in pairwise comparisons, h versus f or h versus g appears equivalent to h' versus f, or h' versus g. Hence, pairwise comparison cannot distinguish second best from second worse options.

The neat trick in Seidenfeld et al. (2007) is that they use this finding to establish that each set of probabilities has a unique coherent choice function. Moreover, they provide 4 axioms for the theory of choice functions that are necessary for coherence, and are jointly sufficient for coherence when using a set of probability / almost-state-independent utility pairs.

What's more, their axioms do not rely upon convexity! This is a significant departure for characterizing imprecise probability and marks a fundamental breakthrough. Given the problems that stem from convexity, many of which uncovered by Seidenfeld *et. al.*, it is remarkable that we now have an axiomatic characterization of imprecise probability that dispenses with this condition.

We reproduce the axiom system of Seidenfeld et al. (2007) here:

Axiom 1: (*a*) You cannot promote an unacceptable option into an acceptable option by adding to the choice set of options, and (*b*) you cannot promote an unacceptable option into an acceptable option by deleting unacceptable options from the choice set of options.

Definition [strict partial order $\langle$]: For option sets O_1 and O_2 and unacceptable options $R(\cdot)$, $O_1 \langle\, O_2$ iff $O_1 \subseteq R[O_1 \cup O_2]$.

Given axiom 1, $\langle$ is a strict partial order over pairs of *sets* of options.

The next axiom governs mixtures. Notation: where o is an option in the set O_1 and α is a real number, $\alpha O_1 \oplus (1-\alpha)o$ is the set of pointwise mixtures, $\alpha o_1 \oplus (1-\alpha)o$ for $o_1 \in O$.

Axiom 2: (a) *Independence* is defined for the binary relation $\langle$ over sets of options. For $0 < \alpha \leq 1$ and an option o, $O_1 \langle O_2$ iff $\alpha O_1 \oplus (1-\alpha)o \langle \alpha O_2 \oplus (1-\alpha)o$. (In words, if and only if the pointwise mixtures of O_1 are a subset of the unacceptable pointwise mixtures of O_1 or O_2, then the option set O_1 is a subset of the unacceptable option set $O_1 \cup O_2$.) (b) *Mixtures* is defined in terms of the closed convex set of options, $H(O)$. If $o \in O$ and $o \in R(H(O))$, then $o \in R(O)$. (In words, an unacceptable option from a mixed set of options is also unacceptable before mixing.)

The next axiom provides an Archimedian condition. Let A_n and B_n (for $n = 1, \ldots$) be sets of options converging pointwise to options sets A and B. Let N be an option set.

Axiom 3: (a) If for each n, $B_n \langle A_n$ and $A \langle N$, then $B \langle N$; (b) If for each n, $B_n \langle A_n$ and $N \langle B$, then $N \langle A$.

The final axiom governs state-neutrality and dominance, which is expressed in terms of *horse lotteries* (Anscombe and Aummann 1963). Consider a set $\mathbb{P} = \{P_1, \ldots, P_n\}$ of probability distribution on states ω_0, ω_1. (This of course may be generalized to a finite space of ω_m states.) When $\underline{P}$ is the smallest, nonzero coordinate of $\mathbb{P}$, define a *constant horse lottery act* $\mathbf{a} = \underline{P}\mathbf{1} \oplus (1-\underline{P})\mathbf{0}$. Let $\{\mathbf{0}, \mathbf{1}\}$ be a set of prizes, where the constant act $\mathbf{1}$ is better than, and the constant act $\mathbf{0}$ is worse than, all other constant acts. Consider horse lotteries $h_1 = \beta_{1_j}\mathbf{1} \oplus (1-\beta_{1_j})\mathbf{0}$ and $h_2 = \beta_{2_j}\mathbf{1} \oplus (1-\beta_{2_j})\mathbf{0}$, for $j = 1, 2$. Seidenfeld et al. (2007) then define weak dominance in terms of horse lotteries h_1, h_2, and axiom 4 in terms of weak dominance:

Definition [Weak dominance]: h_2 weakly dominates h_1 if $\beta_{2_j} \geq \beta_{1_j}$, for $j = 1, \ldots, n$.

Axiom 4: Suppose o_2 weakly dominates o_1, and a is an option different from these two. (a) If $o_2 \in O$ and $a \in R(\{o_1\} \cup O)$, then $a \in R(O)$; (b) If $o_1 \in O$ and $a \in R(O)$, then $a \in R(\{o_2\} \cup O \setminus \{o_1\})$. (In words, ($a$) says that when a weakly dominated option is removed from the set of options O, other inadmissible options remain inadmissible; (b) says that when an option o_1 is replaced by another o_2 that weakly dominates it, other inadmissible options remain inadmissible.) Given axiom 1, (4a) says that when an option is replaced in the option set by one that it weakly dominates, other admissible options remain admissible, and (4b) says that adding a weakly dominated option cannot promote an inadmissible option to an admissible one.

Seidenfeld et al. (2007) show that axioms 1-4 are necessary for a coherent choice function, and the core of their paper—which I hope that I've motivated you to study—is a result showing that these 4 axioms are jointly sufficient.

References

Anscombe, F. J. and Aummann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205.

de Cooman, G. and Miranda, E. (2007). Symmetry of models versus models of symmetry. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* King's College Publications, Lon.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.

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 Intenational Symposium on Imprecise Probabilities: Theories and Applications*, pages 496–510. Carleton Scientific.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2007). Coherent choice functions under uncertainty. In *Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA)*, Prague, Czech Republic.

Wagner, C. G. (2007). The Smith-Walley interpretation of subjective probability: An appreciation. *Studia Logica*, 86:343–350.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.