MOVING BEYOND SETS OF PROBABILITIES

GREGORY WHEELER
FRANKFURT SCHOOL OF FINANCE & MANAGEMENT
g.wheeler@fs.de

forthcoming in STATISTICAL SCIENCE

Abstract

The theory of lower previsions is designed around the principles of coherence and sure-loss avoidance, thus steers clear of all the updating anomalies highlighted in Gong and Meng's "Judicious Judgment Meets Unsettling Updating: Dilation, Sure Loss, and Simpson's Paradox" except dilation. In fact, the traditional problem with the theory of imprecise probability is that coherent inference is too complicated rather than unsettling. Progress has been made simplifying coherent inference by demoting sets of probabilities from fundamental building blocks to secondary representations that are derived or discarded as needed.

Keywords: Desirable gambles, lower previsions, imprecise probabilities, dilation

Professors Gong and Meng's (2021) lucid and thought-provoking article views imprecise probability through the lens of three updating rules, highlighting discrepancies in inference between the *generalized Bayes rule*, on the one hand, and *Dempster's rule* and its dual, the *geometric rule*, on the other. In doing so, Gong and Meng vividly illustrate two important points, namely (*i*) inferential anomalies involving imprecise probabilities ought to be viewed as a helpful warning sign that some structural uncertainty looms in one's model, and (*ii*) such uncertainty is different in kind to sampling variability and therefore not resolved by updating with additional data.

Even so, the route Gong and Meng take to arrive at these two conclusions risks leaving the impression that the theory of imprecise probability is wobblier than it is. Specifically, in writing that,

> "in the world of imprecise probabilities, not only must we live with imperfections, but also accept intrinsic contradictions",

Gong and Meng suggest little has changed from the days of C.A.B. Smith's outline for inference with lower and upper personal "pignic odds" (Smith 1961), a proposal that Savage and de Finetti deemed "not fit for characterizing a new, weaker kind of coherent behaviour" (de Finetti and Savage 1962).

In my remarks, I would like to offer a corrective to the notion that inference with imprecise probabilities is plagued by inherent contradictions. On the contrary, for the contemporary theory of *lower previsions* (Walley 1991; Troffaes and de Cooman 2014), which includes lower probabilities as a special case, coherence preservation under inference is inviolable. Yet, once sure-loss avoidance is promoted to a fundamental principle, both Dempster's rule and the geometric rule fall by the wayside—except in specific, benign circumstances where their application is guaranteed to avoid sure loss.

1 Sure Loss Avoidance & Coherence

Whether to accept sure-loss avoidance as fundamental will depend on what you get from the theory of lower previsions in return. Gong and Meng rightly observe that if lower and upper previsions are interpreted as acceptable one-sided betting odds, with lower previsions denoting the maximum buying price you would pay for a gamble and upper previsions denoting the minimum selling price you would accept for that gamble, then it is natural to accept sure-loss avoidance as a principle of rationality. They nevertheless contrast this *direct* interpretation of a lower prevision, as a representation of your disposition to bet on a collection of gambles, with an *indirect* interpretation that regards a lower prevision as a summary of the set of probabilities that are compatible with an incompletely specified model. This indirect interpretation is central to Bayesian sensitivity analysis, but it has also played an important role in the historical development of imprecise probabilities more generally.

For instance, Smith showed that every coherent lower prevision may be understood as the lower envelope of some set of linear previsions, a result that Walley later strengthened to a characterization (1991, §3.3): specifically, a lower prevision $\underline{\mathrm{P}}$ *avoids sure loss* if and only if there is a linear prevision P such that $P(X) \geq \underline{\mathrm{P}}(X)$, for all gambles X on a fixed domain, and $\underline{\mathrm{P}}$ is a *coherent lower prevision* if and only if there is a set of linear previsions $\mathbb{P}$ such that $\underline{\mathrm{P}}$ is the lower envelope of $\mathbb{P}$, that is $\underline{\mathrm{P}}(X) = \inf\{P(X) : P \in \mathbb{P}\}$, for all X on a similarly shared domain. When the range of X is restricted to $\{0, 1\}$, X works as an indicator function and $\underline{\mathrm{P}}(X)$ as a lower probability. Such sure-loss avoidance and coherence conditions extend to conditional lower previsions, too.

The question then is whether the inferential capabilities that one would need when approximating a true but unknown probability distribution can be subsumed under the machinery developed for lower previsions based on a direct, behavioral interpretation. Walley argued that it does (1991, §2.10) and I agree, with one qualification.

That qualification, a benefit of hindsight, is to concede that managing coherence conditions for conditional lower previsions is complicated when those conditions are tied to a set of linear previsions in the (customary) manner sketched above. One reason why is that the familiar equivalence between additive probability and linear previsions does not carry over to lower probability and lower previsions. A linear prevision is simply the expectation calculated by taking the integral with respect to a given probability, and this equivalence licenses Bayesians to treat "degrees of belief" expressed over a language of events as fundamental. However, an analogous one-to-one correspondence between lower probability and lower previsions does not hold. Specifically, unlike linear previsions, two lower previsions can agree in values for all *events*, and therefore express the same lower probabilities, but still express different values over *gambles*. This one-to-many relationship means that commonplace probabilistic intuitions can go haywire in the context of imprecise probabilities, resulting in some forms of reasoning that are valid for precise probabilities being invalid for imprecise probabilities.

Since Walley's chef-d'œuvre, simpler and more unified inference methods for conditional lower previsions have been developed (Troffaes and de Cooman 2014), but they have come about by abandoning the notion that sets of probabilities are elemental. Whereas the Old Testament approach to imprecise probabilities closely links lower previsions to sets of probabilities, thereby setting a difficult path for coherent inference to follow, the New Testament puts coherence and inference first but demotes (closed convex) sets of probabilities to derivable or dispensable representations, as need be. Instead, *desirable* or *acceptable gambles* are treated as fundamental, where a gamble X on a set of possibilities is a real-valued map from those possibilities, interpreted as the gain or loss that you associate with each possible state. Then, $\underline{P}(X)$ represents the supremum price you are willing to pay in exchange for the gamble X, and a conditional lower prevision of the gamble X given the $\{0,1\}$-gamble G, $\underline{P}(X|G)$ is your lower prevision for X contingent on the event G occurring ($G = 1$), which is "called-off" otherwise ($G = 0$).

Briefly, and to just give a flavor, there are four simple yet constructive axioms for a coherent set $\mathbb{D}$ of desirable gambles. The first two, which are rationality axioms, mandate that you ought to (*i*) never accept a gamble you cannot win (i.e., do not include in $\mathbb{D}$ an X whose vector of values is everywhere negative), and (*ii*) always accept a gamble you cannot lose. The 0-gamble denotes *status quo ante*, and there are variants of these axioms which include, rather than exclude, 0-gambles among a coherent set of gambles—a difference reflected, even if only loosely observed, in the terminology used to refer to the strict desirability of gambles or merely to their acceptability. The second pair of axioms are closure conditions, encoding the properties of a linear scale for evaluating gambles, namely (*iii*) positive scale invariance and (*iv*) a combination rule whereby if X and Y are each acceptable gambles, then $X+Y$ ought to be acceptable to you, too.

The generalized Bayes rule in this scheme is simply

$$\underline{P}\left(G\left[X - \underline{P}(X|G)\right]\right) = 0 \tag{1}$$

where it is assumed that both $\underline{P}(G) > 0$ and the contingent gamble $G\left[X - \underline{P}(X|G)\right]$ are in $\mathbb{D}$. Methods for conditioning and updating on zero-probabilities have been simplified, too (De Bock and de Cooman 2015).

2 What price for generality?

The New Testament's full embrace of modeling uncertainty in terms of the rationality of beliefs and behavioral dispositions might appear to go too far, even among those who otherwise favor the Bayesian approach. Yet, the contemporary theory of lower previsions is a general framework attuned to foundational issues of the kind that Gong and Meng raise, and as such includes traditional linear previsions as a special case, much like first-order logic includes propositional logic as a special case. Lower previsions offer an alternative way of conceiving and working with probability models, not an alternative to probability altogether. Viewed in this light, it is perhaps less surprising to find that sets of probabilities are derivable from, rather than foundational to, lower previsions.

The analogy to logic goes a bit further. Consider some differences between propositional logic, which dates back two millennia, and first-order logic, which is just over a century old. Both the syntax and semantics of first-order logic work very differently than the syntax and semantics of propositional logic. First-order logic admits syntactically well-formed "open" sentences which are nevertheless uninterpretable, semantically, until "closed" under quantification. There is no such thing as a syntactically well-formed formula of propositional logic that is semantically uninterpretable, however. Every formula of propositional logic is interpreted by evaluating all logically exhaustive combinations of its interpretations, such as may be displayed in a truth table, which is impossible to do for first-order logic. As for inference, propositional logic is decidable whereas first-order logic is not. Yet, if one were to maintain that semantic interpretability and syntactic well-formedness were inseparable properties of logical formulas, truth tables fundamental to model theory, or decidability essential to logic itself, the world of first-order logic would be regarded as imperfect and contradictory, too. We generally don't take that view, however, and similar slack should be afforded to lower previsions—or so I would argue. Space prohibits more than a gesture here, but a paper-sized treatment appears elsewhere (Wheeler 2021).

The main point is this. Trouble for imprecise probabilities rarely comes in the form of inherent contradictions, but instead is more apt to arise from seeking to preserve consistency at all costs. Disjunction, for instance, is missing from the vocabulary of desirable gambles, and is tricky to deal with. Recent work using desirable gamble *sets* to construct choice functions (De Bock and de Cooman 2019) offers a promising avenue to address this deficiency, however. This extension offers the capability to say of a set of gambles that at least one is desirable without necessarily identifying which it is. Accommodating set-based choice also suggests a means, in a coherence preserving setting, to address problems of the kind that motivate the use of belief functions.

3 Dilation and Association

Which brings us to dilation. Dilation occurs when the interval estimate of an event E is properly included in the interval estimate of E conditional on every element of some measurable partition $\mathcal{B}$. As Gong and Meng point out, in such cases, updating by the generalized Bayes rule on *any* value of $\mathcal{B}$ would render your initial estimate of E less precise. Should you update or instead refuse information that would resolve your uncertainty about $\mathcal{B}$? Would you be willing to pay some amount, however small, to remain ignorant? With dilation, one could be forgiven for thinking, *so much for consistency*.

Yet, the notion that you can be better off with less information is not unheard of in the theory of games. Akerlof's study of market failures in the used car market, circa 1970, is a prime example. A customer will not know, but a used-car salesman will know, which cars on the lot are lemons. Wary of being fleeced, a customer will refuse to pay more than the going rate for a bad car, if not refuse to trade altogether. For the salesman then there is a disadvantage knowing more about the quality of the cars on the lot than his customers do, as no car, good or bad, can command a good-car price.

Akerlof's demonstration of *adverse selection* is an example of a strategic interaction in which information asymmetry backfires on the player with more information. Some textbook treatments of adverse selection maintain that negative-valued information cannot occur in single-person decision problems, however, as act-state independence would rule out the type of act-state dependence that the customer on the car lot fears will be used against him. But this is only true for single-person decision problems with additive probabilities. Dilation illustrates a form of state uncertainty, which lower previsions capture, that is sufficient to break the independence condition that ordinary decision problems take for granted. Put more carefully, dilation examples do not explicitly rule out that the pair of events in question are dependent. And a cleverly designed dilation example will prey on intuitions that are misleading in an imprecise probabilities setting, particularly those to do with structural properties of independence and association.

At root, dilation is not so much an updating paradox as a result of reasoning as if stochastic independence holds when it does not. Although Gong and Meng remark that "generalized notions of association and independence...are yet to be defined for sets of probabilities", there are several logically distinct notions of independence for imprecise probabilities (Couso, Moral, and Walley 1999). Here reference to an explicit set of probabilities helps. For instance, for an ordinary additive probability $p \in \mathbb{P}$ and events A, B, you know that if B is irrelevant to A with respect to this p, that is, if $p(A|B) = p(A)$, then A is irrelevant to B, and the joint distribution of A and B is the product of the pair of marginal distributions. But each step in this sequence of valid inferences is invalid for imprecise probabilities. Irrelevance for lower previsions is not symmetric, and even when both A is irrelevant to B and B irrelevant to A, the set of joint distributions might not factorize. The converse of each is valid, however, pointing to a range of strong to weak independence concepts.

Wily dilation examples are often constructed to satisfy weaker notions of irrelevance without satisfying full, factorized stochastic independence, and will in fact include a distribution in $\mathbb{P}$ for which the pair of events are positively associated and another distribution for which they are negatively associated (Pedersen and Wheeler 2014), an observation that is easily adapted to include asymmetric cases in which one event dilates another but not vice versa (Pedersen and Wheeler 2019).

But if this explains what dilation is, what should be done about it? I agree with Gong and Meng that dilation alone is not a problem, anymore than an open formula of first-order logic is itself a problem. But instead of opting for an alternative updating rule, and braving the hazards they bring, I prefer to stick to the generalized Bayes rule and simply select an appropriate decision rule. In fact, returning to the questions above that suggest you might be rationally compelled by dilating probabilities to either ignore information or even pay someone to avoid it, such injunctions depend crucially on your choice of decision rule. In fact, some decision rules for imprecise probabilities preserve the principle that no decision maker should be made worse off, in expectation, from receiving free information (Pedersen and Wheeler 2015).

To be fair, a remnant of the updating anomalies that Gong and Meng discuss carries over to decision making with imprecise probabilities. There is no single decision rule that is unequivocally best, and the current state of the art is far less tidy. A complaint might then be lodged that this only kicks the inference can down the pick-the-right-decision-rule road, and there is a kernel of truth to this. But, that is a discussion to save for another day.

In closing, I commend Gong and Meng for their valuable contribution and wish to stress once more how much I agree with them in the main. Lower previsions afford much greater expressive capacity and, as a consequence, pull apart some notions that are unitary concepts in standard, additive probability models. Thus, it is a natural response to view updating anomalies like dilation as a helpful pointer to some of the novel implications that follow from uncertainty in such settings.

References

Couso, I., S. Moral, and P. Walley (1999). Examples of independence for imprecise probabilities. In G. de Cooman (Ed.), *Proceedings of the First Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Ghent, Belgium.

De Bock, J. and G. de Cooman (2015). Conditioning, updating and lower probability zero. *International Journal of Approximate Reasoning 67*, 1–36.

De Bock, J. and G. de Cooman (2019). Interpreting, axiomatising and representing coherent choice functions in terms of desirability. In J. De Bock, C. P. de Campos, G. de Cooman, E. Quaeghebeur, and G. Wheeler (Eds.), *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*, Volume 103 of *Proceedings of Machine Learning Research*, Thagaste, Ghent, Belgium, pp. 125–134.

de Finetti, B. and L. J. Savage (1962). Sul modo di scegliere le probabilità iniziali. *Biblioteca del Metron, Serie C 1*, 81–154.

Gong, R. and X.-L. Meng (2021). Judicious judgment meets unsettling updating: Dilation, sure loss, and simpson's paradox. *Statistical Science*.

Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis 79*(6), 1305–1342.

Pedersen, A. P. and G. Wheeler (2015). Dilation, disintegrations, and delayed decisions. In *Proceedings of the 9th Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Pescara, Italy, pp. 227–236.

Pedersen, A. P. and G. Wheeler (2019). Dilation and asymmetric relevance. In J. De Bock, C. P. Campos, G. de Cooman, E. Quaeghebeur, and G. Wheeler (Eds.), *Proceedings of Machine Learning Research*, Volume 103 of *Proceedings of the 11th Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, pp. 324–326.

Smith, C. A. B. (1961). Consistency in statistical inference (with discussion). *Journal of the Royal Statistical Society 23*, 1–37.

Troffaes, M. C. M. and G. de Cooman (2014). *Lower Previsions*. Chichester, West Sussex: Wiley and Sons.

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

Wheeler, G. (2021). A gentle approach to imprecise probability. In T. Augustin, F. Cozman, and G. Wheeler (Eds.), *Reflections on the Foundations of Statistics: Essays in Honor of Teddy Seidenfeld*, Theory and Decision Library A. Springer.