IP Scoring Rules: Foundations and Applications

Jason Konek
Department of Philosophy, University of Bristol
Jason.Konek@bristol.ac.uk

University of BRISTOL

Overview

- Drawing on the work of de Finetti [1] and Savage [2], contemporary Bayesians like Joyce [3, 4], Schervish *et al.* [5] and Pettigrew [6] use scoring rules, together with resources from decision theory, to show that traditional Bayesian methods provide rational strategies for securing accurate estimates.
- To extend these justifications to the IP framework, we need *IP Scoring Rules*.
- My research has three principal objectives:
  1. Provide an axiomatic characterization of reasonable IP scoring rules, which aim to measure the alethic (accuracy-related) virtues of IP distributions.
  2. Use IP scoring rules to derive accuracy-centred justifications for existing IP methods.
  3. Engineer new IP priors, updating rules, deference principles and aggregation procedures with desirable alethic properties.
- The aim here is twofold:
  - Respond to impossibility theorems for IP scoring rules by Seidenfeld *et al.* [7], Mayo-Wilson and Wheeler [8], and Schoenfield [9].
  - Highlight an application of IP scoring rules to IP aggregation.

IP Scoring Rules

- An IP scoring rule is a loss function $\mathcal{I}$ which maps an IP distribution $\mathcal{C}$ (lower probabilities, lower previsions, sets of probabilities, etc.) and a world w to a non-negative real number, $\mathcal{I}(\mathcal{C}, w)$.
- In the accuracy-centred tradition, we take $\mathcal{I}(\mathcal{C}, w)$ to measure the extent to which $\mathcal{C}$ fails to avoid various types of alethic (accuracy-related) error.
- Inspired by Levi [10, 11], in [12] I proposed that IP scoring rules for sets of probabilities should take the following form:

$$\mathcal{I}_\alpha(\mathcal{C}, w) = \alpha \cdot \min_{c \in \mathcal{C}} I(c, w) + (1 - \alpha) \cdot \max_{c \in \mathcal{C}} I(c, w)$$

- IP distributions are subject to two principal types of alethic error: *generalised type I and type II error.*
- A set of probabilities $\mathcal{C}$ *avoids generalised type I error* to the extent that it *leaves open* accurate probability functions.
  - [12] measures generalised type I error by $\min_{c \in \mathcal{C}} I(c, w)$, where I is any (precise) strictly proper scoring rule.
- A set of probabilities $\mathcal{C}$ *avoids generalised type II error* to the extent that it *rules out* inaccurate probability functions.
  - [12] measures generalised type II error by $\max_{c \in \mathcal{C}} I(c, w)$.
- α measures the degree to which you prioritise avoiding type I over type II error, or vice versa.

Impossibility Theorems

- Seidenfeld *et al.* [7], Mayo-Wilson and Wheeler [8], and Schoenfield [9] show that any continuous IP scoring rule $\mathcal{I}$ renders some IP distribution dominated.
- Some authors conclude that IP methods cannot be motivated by accuracy-centred considerations ([9, p. 14], [8, p. 15]).
- Others are doubtful about the usefulness of continuous real-valued IP scoring rules for elicitation ([7, p. 1257]).
- *These reactions are premature.*
- Choose an interval forecast $[a, b]$ of event E. Then $\mathcal{I}_\alpha$ with

$$\alpha = \frac{-b + ab - \sqrt{ab - a^2 b - ab^2 + a^2 b^2}}{a - b}$$

  is the *unique* IP scoring rule of the form proposed in [12] that renders $[a, b]$ non-dominated.
- Each $\mathcal{I}_\alpha$ determines a curve of non-dominated interval forecasts.

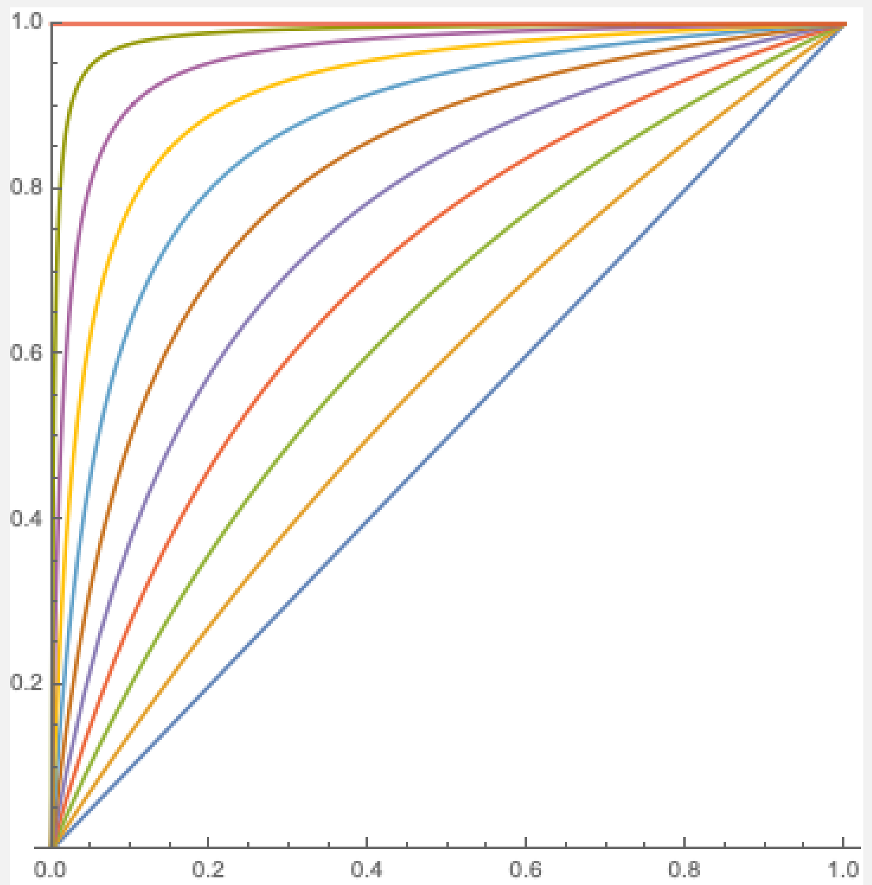

Fig 1: Non-dominated interval forecasts wrt $\mathcal{I}_\alpha$ for a range of α values between 0.5 and 1.

- How $\mathcal{I}_\alpha$ prioritises avoiding type I over type II error (or vice versa) fixes the stock of available interval forecasts.
- *No problem for providing accuracy-centered justifications for existing IP methods.*
  - Certain ways of prioritising type I over type II error-avoidance justify certain IP methods (ones that recommend more imprecise interval forecasts), while others justify others (ones that recommend more precise interval forecasts).
- Continuous real-valued IP scoring rules still potentially useful for elicitation.
  - Elicit $[a, b]$ independently; extract α; $\mathcal{I}_\alpha$ is elicitation compatible for all

$$\left[x, \frac{\alpha^2 x}{1 - 2\alpha + \alpha^2 - x + 2\alpha x}\right]$$

Aggregating Interval Forecasts

- Nau [13], Kriegler *et al.* [14], Stewart and Quintana [15], and others specify IP aggregation principles.
  - **Good**: satisfy some *prima facie* desirable axioms
  - **Bad**: deliver *dominated* aggregates
- An alternative engineered using IP scoring rules:
  EU Aggregation (unique α)*: If n interval forecasts $[x_1, y_1], \ldots, [x_n, y_n]$ for E are all uniquely non-dominated relative to $\mathcal{I}_\alpha$, then any reasonable aggregate must take the following form:*

$$\left[z, \frac{\alpha^2 z}{1 - 2\alpha + \alpha^2 - z + 2\alpha z}\right]$$

  where $\min_i x_i \leq z \leq \max_i x_i$.
- Application to interval forecasts from Kriegler *et al.* [14] for the melting of the Greenland ice sheet by 2200 (MGIS).

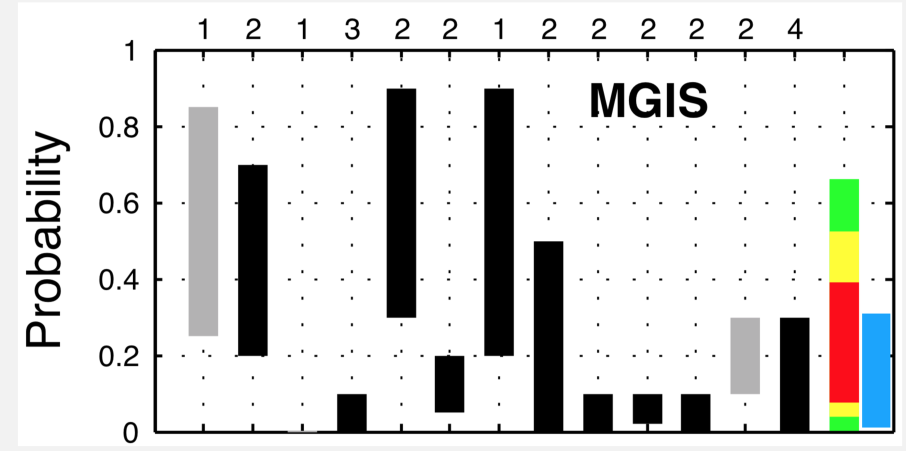

Fig 2: EU aggregate w/ uniform weights: $[0.0153442, 0.318378]$ *(blue).*

- **Advantages**: non-dominated aggregates; robust against outliers.

References

[1] Bruno de Finetti. *Theory of Probability. A Critical Introductory Treatment.* John Wiley & Sons, 1974.
[2] L.J. Savage. *The Foundations of Statistics.* Dover, New York, 1972.
[3] J Joyce. A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603, 1998.
[4] James M Joyce. Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In Franz Huber and Christoph Schmidt-Petri, editors, *Degrees of Belief*, volume 342. Springer, Dordrecht, 2009.
[5] Mark Schervish, Teddy Seidenfeld, and Jay Kadane. Proper scoring rules, dominated forecasts, and coherence. *Decision Analysis*, 6(4):202–221, 2009.
[6] Richard Pettigrew. *Accuracy and the Laws of Credence.* Oxford University Press, Oxford, 2016.
[7] T Seidenfeld, M J Schervish, and J B Kadane. Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning*, 53:1248–1261, 2012.
[8] Conor Mayo-Wilson and Gregory Wheeler. Accuracy and imprecision: A mildly immodest proposal. *Philosophy and Phenomenological Research*, 2015.
[9] Miriam Schoenfield. The accuracy and rationality of imprecise credences. *Nous*, 2015.
[10] Isaac Levi. *Gambling with Truth.* Knopf, New York, 1967.
[11] Isaac Levi. *Decisions and Revisions.* Cambridge University Press, Cambridge, 1984.
[12] Jason Konek. Epistemic conservativity and imprecise credence. *Philosophy and Phenomenological Research*, 2019.
[13] Robert Nau. The aggregation of imprecise probabilities. *Journal of Statistical Planning and Inference*, 105(1):265–282, 2002.
[14] Elmar Kriegler. Imprecise probability assessment of tipping points in the climate system. *Proceedings of the National Academy of Sciences*, 106(13):5041–5046, 2009.
[15] R.T. Stewart and I.O. Quintana. Probabilistic opinion pooling with imprecise probabilities. *J Philos Logic*, 47:17–45., 2018.