

Evaluating Imprecise Forecasts

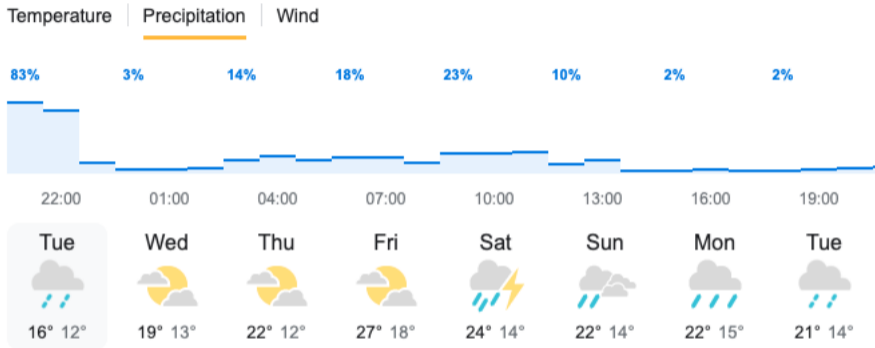
Jason Konek

11 July 2023

Department of Philosophy, Bristol
ISIPTA 2023, Oviedo

Motivation




Probabilistic Forecasting



Penalise “inaccurate forecasts” using **loss functions**

- **Ex Ante:** incentivise “careful” and “honest” forecasts
- **Ex Post:** sift accurate/inaccurate forecasters; how to improve

Imprecise Forecasting

			
X	3	2	-1
Y	-50	-10	10

Is X desirable? Is Y desirable?

Ex Ante **Ex Post**

Ex Ante: Prediction Markets



Will Russia and Ukraine sign or announce an agreement to end the current conflict in Ukraine?

A Before 1 June 2022	1%	↓ 1
B Between 1 June 2022 and 31 July 2022	4%	↓ 1
C Between 1 August 2022 and 30 September 2022	7%	—
D Between 1 October 2022 and 30 November 2022	9%	—
E Not before 1 December 2022	79%	↑ 2

Probabilities as of 11 May 2022

When will Russia and Ukraine sign or announce an agreement to end the current conflict in Ukraine?

Opened: 14 Oct 2022, Suspends 30 Sep 2024

	Current Forecast	1-Week Change
A Before 1 April 2023	1%	↓ 1
B Between 1 April 2023 and 30 September 2023	7%	—
C Between 1 October 2023 and 31 March 2024	13%	↓ 1
D Between 1 April 2024 and 30 September 2024	18%	—
E Not before 1 October 2024	61%	↑ 2

Probabilities as of 6 January 2023

Implied Median: 18 Dec 2024

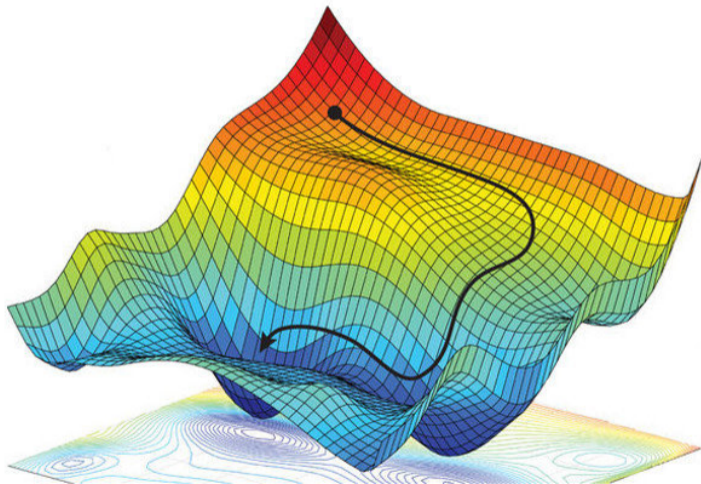
Source: Good Judgment Inc

- Traders can change market IP forecasts; pay score of status quo forecast in exchange for score of updated forecast
- Incentivise reporting of “best” IP model from class of admissible models

Ex Post: Machine Learning



Label	Description
0	T-shirt/top
1	Trouser
2	Pullover
3	Dress
4	Coat
5	Sandal
6	Shirt
7	Sneaker
8	Bag
9	Ankle boot



IP Loss Functions

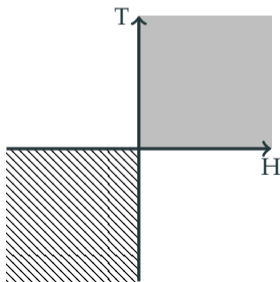
Let Ω be a finite possibility space.

A gamble $X : \Omega \rightarrow \mathbb{R}$ is an uncertain reward. We will treat them as elements $X = \langle x_1, \dots, x_n \rangle$ of \mathbb{R}^n .

A set $\mathcal{D} \subseteq \mathbb{R}^n$ is a **coherent set of almost desirable gambles** if and only if it satisfies the following five axioms:

AD1. If $X < 0$ then $X \notin \mathcal{D}$ (where $X < 0 \Leftrightarrow x_i < 0$ for all $i \leq n$)

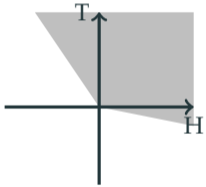
AD2. If $X \geq 0$ then $X \in \mathcal{D}$ (where $X \geq 0 \Leftrightarrow x_i \geq 0$ for all $i \leq n$)



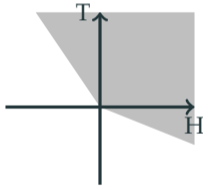
AD3. If $X \in \mathcal{D}$ and $\lambda > 0$ then $\lambda X \in \mathcal{D}$

AD4. If $X, Y \in \mathcal{D}$ then $X + Y \in \mathcal{D}$

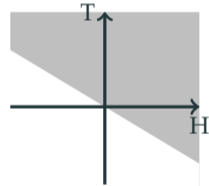
AD5. If $X + \epsilon \in \mathcal{D}$ for all $\epsilon > 0$ then $X \in \mathcal{D}$



convex cone including $\mathbb{R}_{\geq 0}^2$
excluding $\mathbb{R}_{< 0}^2$



larger set:
more committal model



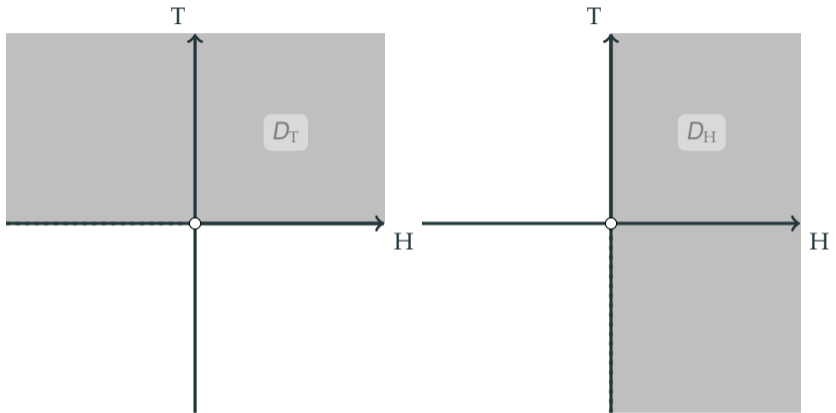
open half-space:
precise prob model

Ideal Sets of Gambles

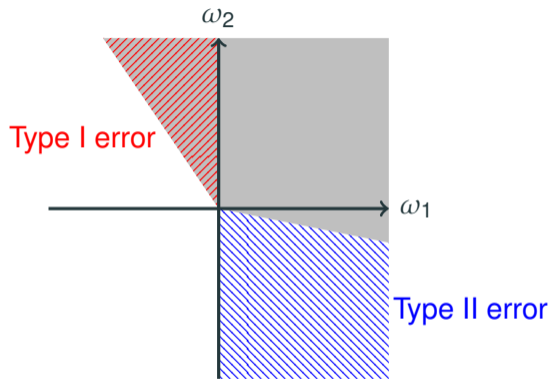
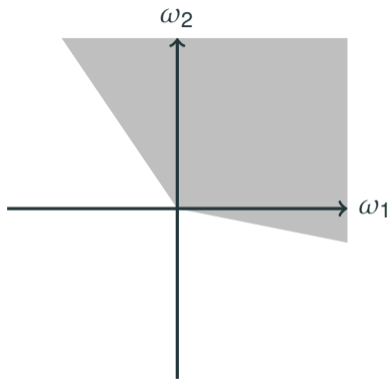
The ideal set of almost desirable gambles if ω_i is true is given by

$$\mathcal{D}_i = \{X \mid x_i \geq 0\}$$

\mathcal{D}_i contains all and only the gambles that are *in fact* almost desirable at ω_i .



Type 1 and 2 Error



Type I and type II error for world ω_1

$\mathcal{E}_i = T1 \cup T2$ is \mathcal{D} 's **error set** at ω_i

- The set of gambles that \mathcal{D} mischaracterizes at ω_i .

Inaccuracy is a measure of error.

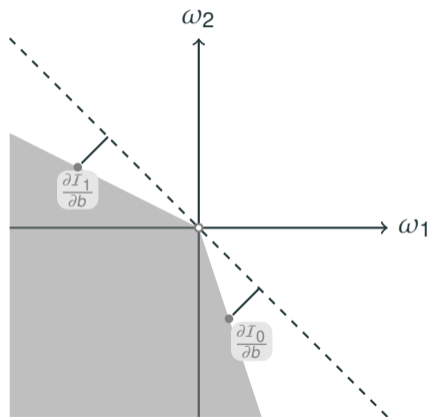
The inaccuracy of \mathcal{D} at ω_j , $\mathcal{I}(\mathcal{D}, \omega_j)$, is the measure of \mathcal{E}_j according to an appropriate measure ν_j :

$$\mathcal{I}(\mathcal{D}, \omega_j) = \mathcal{I}_j(\mathcal{D}) = \nu_j(\mathcal{E}_j)$$

Assume that ν_j is finite and absolutely continuous with respect to the product Lebesgue measure μ . In that case

$$\mathcal{I}_j(\mathcal{D}) = \int_{\mathcal{E}_j} |\phi_j| \, d\mu$$

Main Result



If v_i is finite and absolutely continuous with respect to μ , for all $i \leq n$, then the following two conditions are equivalent:

1. There is some $h : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ s.t. for all $i \leq n$

$$\delta I_i(b, h) = \int_{\mathbb{R}^{n-1}} \frac{\partial I_i}{\partial b} h \, d\lambda < 0$$

First variation—calculus of variations analogue of directional derivative

2. $0 \notin \text{posi}(\{\phi_i(\cdot, b(\cdot)) \mid i \leq n\})$

Loss functions for (infinitely many) precise previsions

Suppose that for all $i \leq n$

$$\mathcal{I}_i(\mathcal{D}) = \int_{\mathcal{E}_i} |g_i| d\mu$$

Then for any probability mass function $p : \Omega \rightarrow \mathbb{R}$ and any $\mathcal{D} \neq \mathcal{D}_p$

$$\sum_{i \leq n} p_i \mathcal{I}_i(\mathcal{D}_p) < \sum_{i \leq n} p_i \mathcal{I}_i(\mathcal{D})$$

unless both $\mathcal{D} \setminus \mathcal{D}_p$ and $\mathcal{D}_p \setminus \mathcal{D}$ are sets of measure zero.

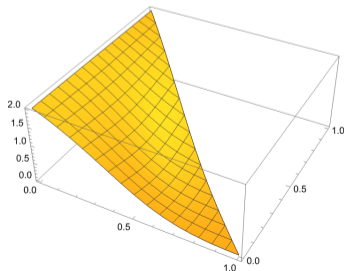
\mathcal{I} is a **strictly proper scoring rule**.

Loss functions for (infinitely many) precise previsions

Example. Let ρ be the normal distribution on the Borel σ -algebra $\mathfrak{B}(\mathbb{R})$ with mean 0 and standard deviation 5. Let μ be the product measure $\rho \times \rho \times \rho$ on $\mathfrak{B}(\mathbb{R}^3)$. In that case

$$\mathcal{I}_i(\mathcal{D}_\rho) = \frac{5 \left(1 - \frac{p_i}{\sqrt{p_1^2 + p_2^2 + p_3^2}} \right)}{2\pi}$$

This is a non-additive analogue of the Spherical score.



Loss functions for non-maximal sets of almost desirable gambles

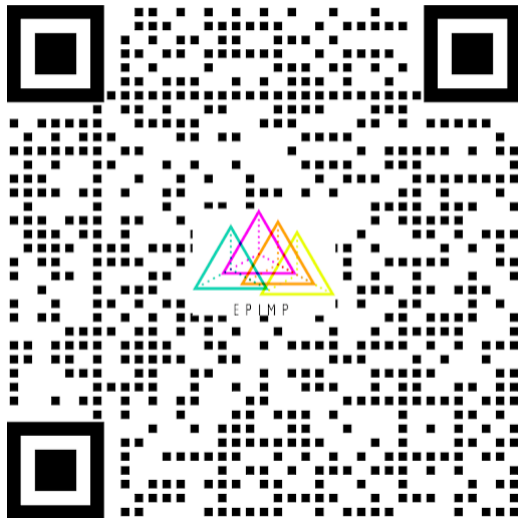
Suppose that there are $\lambda \geq \gamma > 0$ such that

$$I_i(\mathcal{D}) = \int_{\mathcal{E}_i} \begin{cases} -\lambda x_i & \text{if } x_i < 0 \\ \gamma x_i & \text{if } x_i \geq 0 \end{cases} d\mu$$

for all $i \leq n$. Then $\mathcal{D}_b = \{\langle g_1, g_2, g_3 \rangle \mid g_3 \geq b(g_1, g_2)\} \subseteq \mathbb{R}^3$ is admissible, *i.e.*, not dominated by some $\mathcal{D} \neq \mathcal{D}_b$, iff there are $\alpha, \beta \geq 0$ s.t.

$$b(x_1, x_2) = \begin{cases} \frac{-\gamma(\alpha x_1 + \beta x_2)}{\lambda} & \text{if } x_1 \geq 0, x_2 \geq 0 \\ \frac{-\lambda(\alpha x_1 + \beta x_2)}{\gamma} & \text{if } x_1 < 0, x_2 < 0 \\ \frac{-\gamma(\alpha \lambda x_1 + \beta \gamma x_2)}{\gamma} & \text{if } x_1 < 0, x_2 \geq 0, \alpha \lambda x_1 + \beta \gamma x_2 < 0 \\ \frac{-\lambda(\alpha \lambda x_1 + \beta \gamma x_2)}{\lambda} & \text{if } x_1 < 0, x_2 \geq 0, \alpha \lambda x_1 + \beta \gamma x_2 \geq 0 \\ \frac{-\gamma(\alpha \gamma x_1 + \beta \lambda x_2)}{\gamma} & \text{if } x_1 \geq 0, x_2 < 0, \alpha \gamma x_1 + \beta \lambda x_2 < 0 \\ \frac{-\lambda(\alpha \gamma x_1 + \beta \lambda x_2)}{\lambda} & \text{otherwise} \end{cases}$$

Loss functions for non-maximal sets of almost desirable gambles



Come to the poster!

Evaluating Imprecise Forecasts

Jason Konec

University of Bristol

Keywords: Scoring rules, loss functions, forecasting, lower previsions, sets of almost desirable gambles



Background

$\mathcal{D} = \{d_1, \dots, d_n\}$ is a finite possibility space
 \mathcal{F} is the power set of \mathcal{D} . Elements of \mathcal{F} are events.
 Upper- and lower-probability forecasts for events

Event	Upper	Lower
\mathcal{D}	1	0
$\{d_1\}$	0.5	0.2
$\{d_2\}$	0.5	0.2
$\{d_1, d_2\}$	0.8	0.4
$\{d_1, d_3\}$	0.8	0.4
$\{d_2, d_3\}$	0.8	0.4
$\{d_1, d_2, d_3\}$	1	0

Forecasts are assessed relative to an "ignorance" or "ignorance" loss function L .
 L is a function from \mathcal{F} to \mathbb{R} .
 L is an assignment of precise forecasts

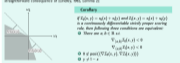
is events in \mathcal{F} . \mathcal{L} is the space of all such assignments.
 Inclusion assignments of precise forecasts by a scoring rule is then function $S: \mathcal{L} \times \mathcal{F} \rightarrow \mathbb{R}$.
 A scoring rule $S: \mathcal{L} \times \mathcal{F} \rightarrow \mathbb{R}$ is strictly proper iff

$$\sum_{d \in \mathcal{F}} S(d, A) < \sum_{d \in \mathcal{F}} S(d, A) - S(d, A)$$

for any probability function p on \mathcal{D} and any $A \in \mathcal{F}$.
 • Brier Score: $S(d, A) = \sum_{d \in \mathcal{F}} (d(A) - p(d))^2$
 • Log Score: $S(d, A) = \sum_{d \in \mathcal{F}} -\log(p(d))$
 • Spherical Score: $S(d, A) = \sum_{d \in \mathcal{F}} \left(\frac{1 - |d(A) - p(d)|}{\sqrt{d(A)^2 + p(d)^2}} \right)$

Strictly proper scoring rules: admissibility

An assignment of forecasts $\mu: \mathcal{F} \rightarrow \mathbb{R}$ is admissible relative to a scoring rule S if and only if it is not weakly dominated (in terms of S) by any other assignment ν .
 • For μ to be admissible, it must be that for all ν ,
 $\sum_{d \in \mathcal{F}} \nu(d) S(d, A) < \sum_{d \in \mathcal{F}} \mu(d) S(d, A)$



Conclusion: A pair of forecasts μ and ν for \mathcal{D} is admissible if and only if μ is not weakly dominated by ν .

Sets of almost desirable gambles

A gamble $g: \mathcal{D} \rightarrow \mathbb{R}$ is an uncertain reward which pays out in those states. We will treat them as elements of $\mathbb{R}^{\mathcal{D}}$.
 A set $\mathcal{D} \subseteq \mathbb{R}^{\mathcal{D}}$ is a subset set of almost desirable gambles if it satisfies

- $g \geq 0$ for all $g \in \mathcal{D}$ (non-negativity)
- $g \geq 0$ for all $g \in \mathcal{D}$ (non-negativity)
- $g \geq 0$ for all $g \in \mathcal{D}$ (non-negativity)
- $g \geq 0$ for all $g \in \mathcal{D}$ (non-negativity)
- $g \geq 0$ for all $g \in \mathcal{D}$ (non-negativity)

Challenges

- Suppose that for all $\mu \in \mathcal{D}$, we have $\mu \geq 0$ and $\mu \geq 0$ for all $\mu \in \mathcal{D}$.
 • It is not clear how to assess the admissibility of μ .
- It is not clear how to assess the admissibility of μ .
- It is not clear how to assess the admissibility of μ .
- It is not clear how to assess the admissibility of μ .

References:
 • Dawid, G. (1984). Bayesian Epistemology. In: *Bayesian Epistemology*.
 • Seidenfeld, R., Kadane, J. B., & Schervish, P. (1997). *Decision Making with Incomplete Information*.
 • Seidenfeld, R., Kadane, J. B., & Schervish, P. (1997). *Decision Making with Incomplete Information*.

Scoring imprecise forecasts

The ideal set of almost desirable gambles \mathcal{D} is the set of all gambles g such that $g \geq 0$ and $g \geq 0$ for all $g \in \mathcal{D}$.
 • This is a set of almost desirable gambles \mathcal{D} .
 • The set of almost desirable gambles \mathcal{D} is the set of all gambles g such that $g \geq 0$ and $g \geq 0$ for all $g \in \mathcal{D}$.



Linear previsions and non-additivity

Prevision following constraints:
 • $\mu(A) \geq 0$ for all $A \in \mathcal{F}$.
 • $\mu(\mathcal{D}) = 1$.
 • $\mu(A) \leq \mu(B)$ if $A \subseteq B$.
 • $\mu(A) + \mu(B) \leq \mu(A \cup B)$.
 • $\mu(A) + \mu(B) \geq \mu(A \cap B)$.

It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

IP scoring rules: admissibility

It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

Example 8

Suppose that $\mathcal{D} = \{d_1, d_2, d_3\}$ and that for some $\mu \geq 0$ and $\mu \geq 0$ for all $\mu \in \mathcal{D}$.
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

Linear previsions and non-additivity

It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .



IP scoring rules: admissibility

It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

Example 8

Suppose that $\mathcal{D} = \{d_1, d_2, d_3\}$ and that for some $\mu \geq 0$ and $\mu \geq 0$ for all $\mu \in \mathcal{D}$.
 • It is not clear how to assess the admissibility of μ .
 • It is not clear how to assess the admissibility of μ .

References
