Probability Aggregation and Optimal Scoring

Christian J. Feldbacher-Escamilla

Summer 2017

Project Information

Publication(s):

- Feldbacher-Escamilla, Christian J. and Schurz, Gerhard (minor revisions). "Meta-Inductive Probability Aggregation". In: *manuscript*.
- Feldbacher-Escamilla, Christian J. and Schurz, Gerhard (2020b). "Optimal Probability Aggregation Based on Generalized Brier Scoring". In: *Annals of Mathematics and Artificial Intelligence* 88.7, pp. 717–734. DOI: 10.1007/s10472-019-09648-4.

Talk(s):

- Feldbacher-Escamilla, Christian J. and Schurz, Gerhard (2020a-09-25/2020-09-21). *Meta-Induction, Probability Aggregation, and Optimal Scoring*. Conference. Presentation (contributed). KI2020. 3rd German Conference on Artificial Intelligence. University of Bamberg: Fachbereich Künstliche Intelligenz.
- Feldbacher-Escamilla, Christian J. (2017a-08-21/2017-08-26). *Probability Aggregation and Optimal Scoring*. Conference. Presentation (contributed). ECAP 9. European Congress of Analytic Philosophy. LMU Munich: MCMP.
- Feldbacher-Escamilla, Christian J. (2017b-07-14/2017-07-16). *Probability Aggregation and Optimal Scoring*. Conference. Presentation (contributed). Joint Session of the Aristotelian Society and the Mind Association: The Open Session. University of Edinburgh: Aristotelian Society, Mind Association.

Project(s):

- DFG funded research unit *New Frameworks of Rationality* (SPP1516); subproject: *The Role of Meta-Induction in Human Reasoning*.

Heroes

Intro

The problem of induction, simply put: How to justify at least some inductive methods.

E.g., how to justify: $Pc_1, Pc_2, \ldots, Pc_n \mid\!\sim Pc_{n+1}$

Problem: Not deductively valid; inductive justification (e.g. by reference to past success) is circular.

Meta-inductive approach: No proof of validity *per se*, but validity *per compartionem*.

Result: Inductive methods are justified as best available alternatives.

This result can be also used for probability aggregation.

Contents

Meta-Induction

Some More Details

Let's consider a series of events $e_1, e_2, \ldots$ with outcomes in $[0, 1]$.

Now, consider prediction methods for the event outcomes: $pred_1, \ldots, pred_n$ of the form $pred_i(e_t) \in [0, 1]$

A simple prediction method for binary events would be, e.g., a binarized likelihood method: $pred(e_t) = 1$ if $\frac{E_1 + \cdots + E_{t-1}}{t-1} \geq 0.5$ otherwise $pred(e_t) = 0$

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	$\ldots$
E_i	0	0	1	1	1	1	0	
$pred_1$	1	0	0	0	1	1	1	

Now, assume that past predictions and event outcomes (E's) are available.

Then we can evaluate prediction methods according to their success.

Hume's Challenge: Induction as Mere Custom?

"There is no object, which implies the existence of any other if we consider these objects in themselves, and never look beyond the ideas which we form of them.

[. . .]

We have no other notion of cause and effect, but that of certain objects, which have been always conjoined together. . . . We cannot penetrate into the reason of the conjunction.

[. . .]

All our reasonings concerning causes and effects are derived from nothing but custom; and that belief is more properly an act of the sensitive, than of the cogitative part of our natures."

(Hume, Enquiry, 1748)

Reichenbach's Approach: Induction as Best Alternative

1. "If we cannot realize the sufficient conditions of success, we shall at least realize the necessary conditions." (p.348)
2. "Let us introduce the term "predictable" for a world which is sufficiently ordered to enable us to construct a series with a limit." (p.350)
3. "The principle of induction [i.e. the straight rule which transfers the observed frequency to the limit] has the quality of leading to the limit, if [the world is predictable]." (p.353)
4. "Other methods [might also] indicate to us the value of the limit." (p.353)
5. "The inductive principle will do the same;" (p.355)
6. [Hence, asymptotical convergence with the inductive principle is a necessary condition.]

(Reichenbach 1938)

Problem: Assumption that the frequency of E_i is limited.

An Expansion: Meta-Induction

1. Nothing in Reichenbach's argument excludes that God-guided clairvoyants may be predictively much more successful than the object-inductivist.
2. He was well aware of this problem, and he remarked that if successful future-teller existed, then the inductivist would recognize this by applying induction to the success of prediction methods.
3. But he did neither show nor even attempt to show that by this meta-inductivistic observation the inductivist could have equally high predictive success as the future-teller.
4. Skilful application of results from machine learning serve this aim.

(cf. Schurz 2008, p.281)

The Meta-Inductive Recipe

How to cook up $pred_{MI}$:

- We measure the past success of a method by inverting the loss.

E_i	0	0	0
$pred_1$	1	0	1
$pred_2$	0	0	1

$\Rightarrow$

success
0.33
0.66

- We measure the attractivity of a method for the *MI*-method ($pred_{MI}$) by cutting off worse than *MI*-performing methods.

$pred_{MI}$	0.66
$pred_1$	0.33
$pred_2$	0.66

$\Rightarrow$

attractivity
0.0
0.66

- We calculate weights out of the attractivities.

	attractivity
$pred_1$	0.0
$pred_2$	0.66

$\Rightarrow$

weight
0.0
1.0

- We define $pred_{MI}$ by attractivity-based weighting of predictions $pred_i$.

Formal Details

$$success(pred_i, t) = \frac{\sum_{k=1}^{t} 1 - loss(pred_i(e_k), E_k)}{t}$$

$$attractivity(pred_i, t+1) = \begin{cases} success(pred_i, t), & \text{if } success(pred_i, t) \geq \\ & success(pred_{MI}, t) \\ 0, & \text{otherwise} \end{cases}$$

$$weight(pred_i, t+1) = \frac{attractivity(pred_i, t+1)}{\sum_{k=1}^{n} attractivity(pred_k, t+1)}$$

$$pred_{MI}(e_{t+1}) = \sum_{k=1}^{n} weight(pred_k, t+1) \cdot pred_k(e_{t+1})$$

Application to the Problem of Induction

Main result of the meta-inductive research programme: long-run optimality; In the long run $pred_{MI}$'s performs at least as good as any other method, if loss is convex:

$$lim_{t \longrightarrow \infty} success(pred_{MI}, t) - success(pred_i, t) \geq 0, \quad \text{for all } 1 \leq i \leq n$$

By this success-based induction is justified (*per comparationem*).

Hence, given the past success of inductive methods as, e.g., the so-called *straight rule*, a success-based choice of these methods is also justified.

Provisos: garbage in $\Rightarrow$ garbage out, $pred_{MI}$ is "parasitical", optimality of $pred_{MI}$ holds only for the long run and only for real-valued predictions, the number of object-methods has to be finite, etc.

Meta-Inductive Probability Aggregation

Intro

"Consider a group of people [. . .] supposed to have the same utility function, at least for the consequences to be considered in the present context, but their personal probabilities are not necessarily the same. The group of people is placed in a situation in which it must, acting in concert, choose an act f from a finite set of available acts F[.] The situation just described will be called a group decision problem."

(cf. Savage 1972, p.172)

Intro

Savage's investigation of statistical opinion pooling rules or his "model of group decision" triggered a vast amount of literature (Savage 1972, chpt.10.2).

A lot of it is collected in (Genest and Zidek 1986).

It was also expanded to the Bayesian framework (cf., e.g., Mongin 2001).

The main underlying problem is the question of how to aggregate probability functions:

$$Pr_1, \ldots, Pr_n \implies Pr_{aggr}$$

An example for the relevance of such an aggregation is the reference class problem: How to deal with different statistics based on overlapping reference classes? $Pr_{sample_1}, Pr_{sample_2} \implies Pr_{aggr}$

Probability Aggregation Constraints

Similarly to the famous impossibility results in social choice theory (cf., e.g., Arrow 1963), impossibility results hold also for probability aggregation:

E.g., the impossibility of combining linear aggregation with Bayesian update or independence constraints.

Therefore, e.g., Bayesian orthodoxy cannot be fully met by linear weighting.

However, important aggregation properties are jointly satisfiable:

- **(U)** Universality: *aggr* allows as input any *Pr* (probability function).
- **(A)** Anonymity: *aggr* cannot identify any specific input: $aggr(Pr_1, \ldots, Pr_n) = aggr(Pr_1, \ldots, Pr_n, Pr_{n-1}) = \ldots$
- **(S)** Systematicity: $Pr_{\{1,\ldots,n\}}(\varphi) = aggr^*(Pr_1(\varphi), \ldots, Pr_n(\varphi))$
  where $aggr^*(Pr_1(\varphi), \ldots, Pr_n(\varphi)) = aggr(Pr_1, \ldots, Pr_n)(\varphi)$; similarly for *Pr*;

Aggregation Properties

Relevance of these properties:

(U), **(A)**, and **(S)** jointly characterize linear weighting aggregation methods:

$$Pr_{aggr}(X_j = x_k) = \sum_{i=1}^{n} w_i \cdot Pr_i(X_j = x_k)$$

However, there is a problem of underdetermination of linear probability aggregation:

By these formal constraints alone the question of how to choose the weights w_i is not settled.

As it turns out, meta-induction can be applied also here:

- It allows for the success-based calculation of weights.
- It can provide a justification for using these weights.

Meta-Inductive Probability Aggregation

The meta-inductive framework is based on prediction games, i.e. series of events $(e_1, e_2, \dots)$.

However, one can also try to base it on probabilistic versions of these.

Instead of the series $e_1, e_2, \dots$ we now have a series of probabilities or random variables: $X_1, X_2, \dots$

We assume that their value space is finite $x_1, \dots, x_q$.

The true outcome (binary states): $val(X_t = x_{1 \leq k \leq q}) = 0/1$

$Pr_1, \dots, Pr_n$ are the probabilities of object-forecasters.

Example

- $X_1, X_2, \ldots$: Random variable about the weather
- x_1: ⛆ (rain, exclusively)
- x_2: ☼ (sun, exclusively)
- x_3: ☁ (wind, exclusively)
- $Pr_1, \ldots, Pr_n$: the predictions of n different weather forecasters

	t_1	t_2	t_3	...
X_t	⛆	☁	☼	...
I.e.:	1.0⛆ 0.0☁ 0.0☼	0.0⛆ 1.0☁ 0.0☼	0.0⛆ 0.0☁ 1.0☼	...
Pr_1	0.5⛆ 0.25☁ 0.25☼	0.25⛆ 0.5☁ 0.25☼	0.15⛆ 0.75☁ 0.1☼	...
Pr_2	0.1⛆ 0.9☁ 0.0☼	0.85⛆ 0.1☁ 0.05☼	0.5⛆ 0.1☁ 0.4☼	...

Meta-Inductive Probability Aggregation

Meta-induction can be employed in order to construct success-based weights.

Note that this cannot be done simply by calculating weights for each value of the value space separately (this would be realised, e.g., by running parallel prediction games).

Reason: One easily ends up with an inconsistency (Pr_{aggr} would no longer be guaranteed to be a probability function).

However, there is a nice way to implement meta-induction in probability aggregation: Namely by considering a game about the prediction of the true value of each round of the past.

Meta-Inductive Probability Aggregation

Here is how it works:

$$success(Pr_i, t) = \frac{\sum_{k=1}^{t} 1 - loss(Pr_i(X_k = x_{k^*}), val(X_k = x_{k^*}))}{t}$$

... where k^* points to that value of $x_1, \dots, x_q$ which turned out to be the true value at k (i.e. that k^*, such that $val(X_k = x_{k^*}) = 1$)

Given this success-measure, we can, again, define a measure for attractivity which in turn allows for defining success-based weights.

Results of Meta-Inductive Probability Aggregation

These weights serve for meta-inductive linear weighted probability aggregation.

Such aggregation is provably long-term optimal.

Provisos: Same as for meta-induction in general plus: No Bayesian update.

However: The long-term optimality holds not only for distance measures as, e.g., that one proposed by Brier, but for all convex distance measures (e.g. also for normalized relative entropy etc.).

So, meta-induction provides an optimality argument for a wide range of probabilistic distance/scoring measures.

Summary

- Meta-induction allows for a wide range of interesting applications.
- One of them is probability aggregation.
- Here, e.g., formal constraints characterize linear weighting as adequate probability aggregation.
- However, the choice of the weights remains underdetermined.
- Meta-inductive optimality results suggest to apply success/attractivity-based weights.

References I

Feldbacher-Escamilla, Christian J. and Schurz, Gerhard (2020b). "Optimal Probability Aggregation Based on Generalized Brier Scoring". In: *Annals of Mathematics and Artificial Intelligence* 88.7, pp. 717–734. DOI: 10.1007/s10472-019-09648-4.

— (minor revisions). "Meta-Inductive Probability Aggregation". In: *manuscript*.

Genest, Christian and Zidek, James V. (1986-02). "Combining Probability Distributions: A Critique and an Annotated Bibliography". In: *Statistical Sciences* 1.1, pp. 114–135.

Hume, David, ed. (1748/2007). *An Enquiry Concerning Human Understanding*. Oxford World's Classics. (ed. by Millican, Peter). Oxford University Press.

Mongin, Philippe (2001). "The Paradox of the Bayesian Experts". In: *Foundations of Bayesianism*. Ed. by Corfield, David and Williamson, Jon. Dordrecht: Springer Science+Business Media, pp. 309–338.

Reichenbach, Hans (1938). *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.

Savage, Leonard J. (1972). *The Foundations of Statistics*. Second Revised Edition. New York: Dover Publications.

Schurz, Gerhard (2008). "The Meta-Inductivist's Winning Strategy in the Prediction Game: A New Approach to Hume's Problem". In: *Philosophy of Science* 75.3, pp. 278–305. DOI: 10.1086/592550.