Asymmetry and the Geometry of Reason

for blind review

Abstract

The geometry of reason is the view that the underlying topology for credence functions is a metric space, on the basis of which axioms and theorems of epistemic utility for partial beliefs are formulated. It implies that Jeffrey conditioning must cede to an alternative form of conditioning. The latter fails a long list of plausible expectations. One solution to this problem is to reject the geometry of reason and accept information theory in its stead. Information theory comes fully equipped with an axiomatic approach which covers probabilism, standard conditioning, and Jeffrey conditioning. It is not based on an underlying topology of a metric space, but uses a non-commutative divergence instead of a symmetric distance measure. I show that information theory, despite initial promise, also fails to accommodate basic epistemic intuitions.

1 Introduction

There are various ways in which epistemic norms for partial beliefs are justified. Three important standards of justification in the literature are pragmatics, accuracy, and evidence.

A partial belief, as opposed to a full belief, expresses uncertainty whether or not a proposition is true. In formal epistemology, this uncertainty is captured in mathematical models. Examples for epistemic norms are probabilism (partial beliefs are numerically most effectively represented as probabilities), Bayesian conditionalization (a rational agent updates beliefs using conditional probabilities if the updating scenario permits it), the principle of indifference (if there is no further information, mutually disjoint and jointly exhaustive events are equiprobable), and additional updating methods beyond standard conditioning (Jeffrey conditioning, affine constraints). Whether the justification for (or rejection of) these norms cites pragmatic, alethic, or evidential reasons, one important ingredient of the formal model is scoring rules.

This paper investigates scoring rules for partial beliefs that exclusively reward and penalize on epistemic grounds. This aligns my paper roughly with the school that privileges alethic justification for epistemic norms, but it should be noted that the debate over scoring rules also has important implications for pragmatists and evidentialists, who hold that epistemic norms are rooted in decision theory and the appropriate relationship between beliefs and the evidence on which they are based, respectively.

> **Example 1: Trichotomy.** A game between the home team and the away team ends in a win (for the home team), a loss, or a tie.

I am restricting myself to finite algebras of propositions, as in example 1, where exactly one of three random outcomes takes place. Let these outcomes (or possible worlds) be ξ_1, ξ_2, ξ_3. Let the agent's report over these three outcomes be $c = (c_1, c_2, c_3)^\intercal$ (the transpose $^\intercal$ symbol merely turns the list of numbers into a vector). Another restriction for this paper shall be that the c_i are non-negative real numbers so that the vector c is located in the non-negative orthant $\mathcal{D}_0$ of an $n = 3$ dimensional vector space. The agent is penalized for reporting c according to a loss function which she wants to minimize (she has no other concerns). Her penalty is

$$S(\xi_i, c) \tag{1}$$

once it is established that ξ_i is the realized outcome. To discourage any dishonesty on her part, a common requirement is that the scoring rule be proper. The propriety of a scoring rule ensures that an agent reports the same distribution according to which she thinks the random process selects the outcome. A scoring rule S is strictly proper if and only if

$$\sum_{i=1}^{n} c_i S(\xi_i, c) < \sum_{i=1}^{n} c_i S(\xi_i, \hat{c}) \text{ for all } \hat{c} \in \mathcal{D}_0 \setminus \{c\} \tag{2}$$

A scoring rule S is proper if and only if (2) is true with a $\leq$ symbol rather than $<$. In the following, I will consistently say 'proper' and 'propriety' in abbreviation for 'strictly proper' and 'strict propriety.' Propriety guarantees that an agent is motivated to report the distribution that they deem to be the one according to which the random outcomes are generated. Propriety significantly narrows down the set of acceptable scoring rules. John McCarthy showed in a seminal paper that propriety requires the existence of a convex entropy function, for which the scoring rule is a type of derivative (see McCarthy, 1956).

The problem I am addressing in this paper is whether there are further restrictions on rationally acceptable scoring rules. McCarthy's theorem leaves open the possibility for symmetric and non-symmetric scoring rules. A symmetric scoring rule assigns as much loss to a reported credence c when the true distribution is $\bar{c}$ as it does when the report is $\bar{c}$ and the true distribution is c. Richard Pettigrew has recently defended symmetric scoring rules (for example in Pettigrew, 2016, 80).

The claim at the heart of my paper is that a defence of symmetry reveals a more fundamental misapprehension about partial beliefs and their relationships to each other. The misapprehension is that there is a geometry of partial beliefs which can be visualized. It is tempting to view a credence $c = (c_1, c_2, c_3)^\intercal$, for example, as a vector in 3-dimensional space and then evaluate its distance to other credences in terms of its metric distance to them. I will call this view, following Hannes Leitgeb and Richard Pettigrew (see Leitgeb and Pettigrew, 2010, 210), the geometry of reason.

Thomas Mormann explicitly warns against the assumption that the metrics for a geometry of logic is Euclidean by default: “All too often, we rely on geometric intuitions that are determined by Euclidean prejudices. The geometry of logic, however, does not fit the standard Euclidean metrical framework” (see Mormann, 2005, 433; also Miller, 1984). Mormann concludes in his article “Geometry of Logic and Truth Approximation,”

> Logical structures come along with ready-made geometric structures that can be used for matters of truth approximation. Admittedly, these geometric structures differ from those we are accustomed with, namely, Euclidean ones. Hence, the geometry of logic is not Euclidean geometry. This result should not come as a big surprise. There is no reason to assume that the conceptual spaces we use for representing our theories and their relations have a Euclidean structure. On the contrary, this would appear to be an improbable coincidence. (Mormann, 2005, 453.)

For Pettigrew, the geometry of reason stands out as an appealing account because its associated scoring rule, the Brier score, is (up to linear transformations) unique once symmetry is required. Pettigrew even has an argument why this uniqueness in its own right has appealing features. I will provide a more detailed summary of the various requirements that one might have with respect to scoring rules. An alternative to the geometry of reason emerges from this analysis, which I will call information theory.

Information theory, just like the geometry of reason, has an associated scoring rule: the Log score. The two different scoring rules, Brier score and Log score, license different updating methods in dynamic partial belief theory. They agree on Bayesian updating in the standard case (using conditional probabilities), but they disagree on updating in a Jeffrey-type updating scenario. Both scoring rules agree on recommending probabilism. I will argue that information theory is the better Bayesian, because Bayesian standard conditioning is smoothly generalized in information theory to more general updating situations (the technical term is affine constraints, of which Jeffrey-type updating scenarios are a special case). The generalization for the geometry of reason is bumpy at best, implausible at worst. I will present this view in full detail in the main body of the paper.

The Log score is asymmetric, but unlike the Brier score not unique among its asymmetric peers. According to Pettigrew's independent argument why it is a good idea to have a unique scoring rule this would count against the Log score and against information theory. The Log score, however, is unique in fulfilling a locality requirement that arguably commands as much plausibility as symmetry. Yet the tenor of my paper is that Pettigrew's independent argument for uniqueness is suspect (in defence of Pettigrew, many of the claims in his book *Accuracy and the Laws of Credence* do not depend on it) and that neither the Brier score's symmetry nor the Log score's locality is sufficient to make them uniquely superior to other scoring rules.

I believe that there are serious problems with the geometry of reason, to the point where I would reject it as a plausible formal account of partial beliefs. As I will show, however, there are serious problems with information theory and how it accommodates epistemic intuitions as well. These are not insurmountable. My hope is that a further detachment from geometry can give us a better understanding of why information theory has the odd features that I will highlight in the paper.

2 Features of Scoring Rules

2.1 List of Features and Preliminaries

Consider the following list of features for a scoring rule SR.

propriety The SR encourages an agent to report the distribution which is her best guess for what generates the random event.

geometry The divergence function associated with the SR is a metric. Consequently, credence functions can be 'visualized' with a distance defined between them.

information The entropy function associated with the SR fulfills Shannon's axioms for an entropy function.

symmetry The expected penalty for a reported distribution with respect to the true distribution is the same when the roles of the reported and true distribution are reversed.

locality How a distribution scores when an event takes place depends only on the credence assigned by the distribution to this event.

horizon The divergence function associated with the SR has a tendency to measure distributions near the centre as being closer together than distributions near the extremes, all else being equal.

conditioning The SR licenses standard conditioning.

reductio-resistance The SR is not led ad absurdum by licensing implausible updating methods.

univocal dominance The SR is uniquely superior to all other SRs in order to address the Bronfman objection.

Here is a brief summary with evaluative annotation (plausibility, for example, means that in conclusion to my arguments in this paper I find the requirement plausible). If my evaluation is endorsed, only the Log score qualifies as a rationally acceptable scoring rule and information theory (as opposed to the geometry of reason) is vindicated; however, the violation of HORIZON must be explained, which is beyond the purview of this paper.

property	**Brier score**	**Log score**	**other scores**	**annotation**
propriety	yes	yes	yes	commonly accepted
geometry	yes	no	no	implausible
information	no	yes	no	plausible
symmetry	yes	no	no	implausible
locality	no	yes	no	weakly plausible
horizon	no	long story	perhaps	plausible
conditioning	yes	yes	perhaps	plausible
reductio-resistance	no	yes	perhaps	plausible
univocal dominance	yes	yes	perhaps	implausible

In the following, let $\mathcal{A}$ be the comprehensive algebra over a finite set of events Ω. This means that all possible combinations of events in terms of negation, union, and intersection are in $\mathcal{A}$. If $\Omega = \{A, B\}$, for example, then $\mathcal{A}$ contains all possible combinations of

$$\bigcup_{i=1}^{4} \omega_i^{j(i)} \tag{3}$$

where

$$\omega_1 = \neg A \cap \neg B \qquad \omega_2 = \neg A \cap B \qquad \omega_3 = A \cap \neg B \qquad \omega_4 = A \cap B \tag{4}$$

and $j(i)$ either negates ω_i or not. If the cardinality of Ω is m, then the cardinality of $\mathcal{A}$ is $n = 2^{2^m}$. For our example $\Omega = \{A, B\}$, $\mathcal{A}$ contains 16 elements; any credence function can therefore be viewed as a vector in the positive orthant of $\mathbb{R}^{16}$. An orthant is a generalization of a quadrant in $\mathbb{R}^2$. I will write $\mathcal{D}_0$ if the orthant includes vectors which have elements that equal zero; $\mathcal{D}$ if all elements of the vector are greater than zero. The zero vector itself is not an element of $\mathcal{D}_0$. Requiring that credence functions are finite, or non-negative, or positive in all elements, or regular in other ways is artificial in order to aid discussion. Examinations of what happens when these regularity conditions are weakened are always welcome.

Probability functions are a strict subset of credence functions. In the vector space $\mathbb{R}^n$ they form a $(2^m - 1)$-dimensional simplex. For two events, it is a three-dimensional tetrahedron in $\mathbb{R}^{16}$;[TBD] for three events, it is a 7-simplex in $\mathbb{R}^{256}$. The reason why probability functions are of much lesser dimension than general credence functions is because they are constrained by logical entailment relationships and Kolmogorov's axioms. In the following, I will ignore the logical entailment relationships. This means that I will assume that credence functions obey these entailment relationships as probability functions do. This means that I can concentrate on a case such as the trichotomoy in example 1 and restrict my attention to credence functions in n-dimensional space. All arguments defending probabilism in this paper do not only justify probabilistic credence functions over non-probabilistic ones that obey the logical entailment relationships, but ad fortiorem also over non-probabilistic ones that disobey the logical entailment relationships.

Possible worlds ξ_i corresponding to ω_i are not credence functions, but they can be embedded by defining the vector elements $\xi_k \in \{0, 1\}$ depending on whether ω_i is (1) or is not (0) negated in (3), $k = 1, \ldots, n$. This is somewhat artificial, especially the choice of the number 1, and any account of epistemic norms must prove itself to be robust if this number is changed to something else that makes sense (see Howson, 2008, 20; for a response Pettigrew, 2016, part I, chapter 6).

2.2 Scoring Rules, Entropy, Divergence

Bruno de Finetti showed that the probability functions form the convex hull of possible worlds so embedded (see de Finetti, 2017, subsection 3.4). It is relatively straightforward to see that for any vector c in the vector space of credence functions, there is a vector p in the convex hull of probability functions which is closer to all possible worlds than c. If c is not an element of the convex hull, then the vector p is strictly closer to all possible worlds.

There are two important points to make about de Finetti's theorem. (1) It shows in what sense probability functions are privileged over other credence functions. This can be cashed out in terms of pragmatics (see Savage, 1971) or accuracy (see Joyce, 1998) (perhaps also in terms of evidence; I am not familiar with the literature). (2) There is a need to define what it means for one credence function to be close to another credence function. One could simply use metric distance. I am hoping to make the case in what follows that this is implausible.

Let us start more modestly with a scoring rule and restrict ourselves to probability functions $\mathcal{P} \subset \mathcal{D}_0$. Probabilism, after all, is not where the geometry of reason and information theory disagree. I have defined scoring rules in equation (1) and the associated propriety in equation (2). McCarthy characterizes proper scoring rules in a theorem whose proof he omits (see McCarthy, 1956, 654). Thankfully, Arlo Hendrickson and Robert Buehler provide a proof (see Hendrickson and Buehler, 1971, 1918).

Definition 2.1 *Let $\bar{D}$ be a convex subset of $\mathcal{D}$, H be a function $H : \bar{D} \to \mathbb{R}$, and $q, q^* \in \bar{D}$*

such that

$$H(p) \leq \langle p - q, q^* \rangle + H(q) \text{ for all } p \in \bar{D}. \tag{5}$$

Then q^* is a supergradient of H at q relative to $\bar{D}$.

$\langle ., . \rangle$ is the inner product of two vectors (or the matrix product if dual spaces are used). The supergradient is the gradient wherever the function is differentiable.

Definition 2.2 *A function $f : V \subset \mathbb{R}^k \to \mathbb{R}$ is homogeneous of degree d if and only if*

$$f(\alpha x) = \alpha^d f(x) \text{ for all } \alpha > 0. \tag{6}$$

Use of Euler's Homogeneous Function Theorem (see Zill, 2011) allows for the proof of McCarthy's Theorem. Let ∇H be the gradient of the function H if it exists, i.e.

$$\nabla H(x) = \left(\frac{\partial H}{\partial x_1}(x), \ldots, \frac{\partial H}{\partial x_n}(x) \right)^{\intercal} \tag{7}$$

and

$$S(x) = \left(S(\xi_1, x), \ldots, S(\xi_n, x)\right)^{\intercal}, \Xi = \{\xi_1, \ldots, \xi_n\} \tag{8}$$

Theorem 2.1 (McCarthy's Theorem) *A scoring rule $S : \Xi \times \mathcal{P} \to \mathbb{R} \cup \{-\infty, \infty\}$ is proper if and only if there exists a function $H : \mathcal{D}_0 \to \mathbb{R}$ which is (a) homogeneous of the first degree, (b) convex,*[TBD] *and (c) such that S is a supergradient of H relative to $\mathcal{D}_0$ at p for all $p \in \mathcal{P}$.*

If H is differentiable, then $\nabla H(p) = S(p)$. It is important to take the partial derivative of H as a function defined on $\mathcal{D}_0$, not just as a function defined on $\mathcal{P}$. As Hendrickson and Buehler point out, this is the error in Marschak, 1959, 97, where the Log score appears to be a counterexample to McCarthy's theorem. None of this is new, but it gives us leverage for what is to follow. Not only is McCarthy's Theorem a powerful characterization theorem for proper scoring rules, it also associates an entropy function H and a divergence D with each scoring rule. With McCarthy's result in hand, scoring rules now come as triplets of scoring rules, entropy functions, and divergences.

Definition 2.3 *The entropy H associated with S is defined as per McCarthy's Theorem (see theorem 2.1); the divergence associated with S is defined to be*

$$D_S(c\|\hat{c}) = H(c) - H(\hat{c}) - \langle c - \hat{c}, S(\hat{c})\rangle \tag{9}$$

Here are two example, the Log score and the Brier score. All summation indices go from 1 to n. If $x \in \mathcal{P}$, then $\sum_k x_k = 1$. The loss function or scoring rule for the Log score is

$$\text{(LS)}\ S(\xi_i, x) = \left(\ln \sum_k x_k\right) - \ln x_i \tag{10}$$

For the Brier score, it is

$$\text{(BS)}\ S(\xi_i, x) = \frac{2x_i}{\sum_k x_k} - 1 - \sum_j \left(\frac{x_j}{\sum_k x_k}\right)^2 \tag{11}$$

The corresponding entropy functions are

$$\text{(LS)}\;\; H(x) = -\sum_i x_i \ln \frac{x_i}{\sum_k x_k} \tag{12}$$

$$\text{(BS)}\;\; H(x) = \sum_i x_i \left(\frac{2x_i}{\sum_k x_k} - 1 - \sum_j \left(\frac{x_j}{\sum_k x_k} \right)^2 \right) \tag{13}$$

The reader can verify that

$$\nabla H(x) = S(x) \tag{14}$$

for both the Log score and the Brier score. This is where we need the entropy to be defined on $\mathcal{D} \supset \mathcal{P}$ in order to avoid Marschak's error from above. For the Log score, the divergence is the Kullback-Leibler divergence,

$$D_{\text{LS}}(p\|q) = \sum_i p_i \ln \frac{p_i}{\sum_i p_k} - \sum_i p_i \ln \frac{q_i}{\sum_k q_k} \tag{15}$$

For the Brier score, the divergence is

$$D_{\text{BS}}(p\|q) = \sum_i p_i \left[\sum_j \left(\frac{q_j}{\sum_k q_k} - \delta_{ij} \right)^2 - \sum_j \left(\frac{p_j}{\sum_k p_k} - \delta_{ij} \right)^2 \right] \tag{16}$$

where δ_{ij} is the Kronecker delta.

The Brier score divergence is symmetric for probability functions and thus can operate as a distance function on them. For both Log score and Brier score, the divergence is the difference between the expected loss of reported q when p is the true distribution and the entropy of p, which is the expected loss of reported p when p is the true distribution.

In section$^{\text{TBD}}$ I will show using convex conjugates that only the Brier score (and its close relatives) fulfill SYMMETRY. In section$^{\text{TBD}}$ I will show that only the Log score fulfills LOCALITY. These are not new results, although the proof using convex conjugates is original. Once these results are established, I have the tools to address Pettigrew's argument for UNIVOCAL DOMINANCE. The core pieces of the paper are sections$^{\text{TBD}}$, where I show that the Brier score violates REDUCTIO-RESISTANCE and both Brier score and Log score violate HORIZON.

3 Symmetry

The Brier score and its close relatives (linear transformations, which only differ from the Brier score in the sense that they bill/pay in a different currency and provide a different initial penalty/reward) are the only proper scoring rules fulfilling SYMMETRY.

4 Locality

4.1 Rewarding Uncertainty About Non-Realized Outcomes

The Brier score, the Spherical score (another proper scoring rule), and many other scoring rules depend on all components of the vector p representing a probabilistic credence function. A scoring rule fulfilling the LOCALITY requirement only depends on the probability assigned to the event that is the realized outcome (for a characterization of local scoring rules that are local in a less restrictive sense see Dawid et al., 2012). In subsection 4.3 I show, using Leonard J. Savage's proof (see Savage, 1971, 794), that the only scoring rule fulfilling LOCALITY are the Log score and its not relevantly different close relatives.

> **Example 2: Tokens.** Casey draws from a bag with n kinds of tokens in it: colour 1, colour 2, ..., colour n. Tatum reports the forecast $(p_1, \ldots, p_n)$. Tatum's forecast agrees with the axioms of probability.

Let p_1 be fixed and colour 1 be the realized outcome. If the Brier score is used, Tatum's penalty T depends on $p_2, \ldots, p_{n-1}$ and is

$$T(p_2, \ldots, p_{n-1}) = 1 - 2p_1 + \sum_{i=1}^{n-1} p_i^2 + \left(1 - \sum_{i=1}^{n-1} p_i\right)^2 \tag{17}$$

T reaches its minimum where $p_i = \frac{1-p_1}{n-1}$ for $i = 2, \ldots, n-1$. The higher the entropy of Tatum's non-realized probabilities, the less stinging Tatum's penalty. The Brier score thus penalizes Tatum (1) for not correctly identifying colour 1 as the realized outcome, but also (2) for reporting variation in the non-realized probabilities. Even though this is the Brier score, it has a ring of information theory to it. The Log score depends only on the realized probability. I.J. Good appears to have favoured such a scoring rule (see Good, 1952, 112).

Because I feel the intuitive appeal of information theory, I consider LOCALITY to be only weakly plausible. There is a sense in which you may want to reward a forecaster not only for assigning a high probability to the realized outcome, but also for uncertainty about the outcomes that were not realized. Of course, doing so sometimes results in a greater loss for Tatum than for Casey even if Tatum assigned a higher probability to the realized outcome. As an example, let Tatum's forecast be $(0.12, 0.86, 0.02)$ and Casey's be $(0.10, 0.54, 0.36)$.

Even though Tatum assigned 12% to colour 1 while Casey assigned 10%, and Casey drew a token of colour 1, Tatum is penalized more severely at 1.5144 compared to Casey at 1.2312 using the Brier score.

4.2 Bronfman Objection

Here is how LOCALITY may still work in favour of information theory against the geometry of reason. To set the scene, I should mention that there are at least two publication anomalies in the study of scoring rules. I already mentioned the earlier one: McCarthy omitted the proof to one of its most important theorems. The proof, using Euler's Theorem, is not trivial and was published almost twenty years later by Hendrickson and Buehler. The later publication anomaly is that Aaron Bronfman wrote an excellent article about a problem with using supervaluationist semantics to justify probabilism (see Bronfman, 2009). Then he decided not to publish it. The manuscript has circulated and is available online.

It is called "A Gap in Joyce's Argument for Probabilism." Jim Joyce provides a non-pragmatic (i.e. alethic) vindication of probabilism by demonstrating that given a particular proper scoring rule, any non-probabilistic credence c is dominated by a probabilistic credence function p (depending on c) in the sense that p is strictly closer to all possible worlds (and therefore more accurate) than c. No probabilistic credence function is dominated in this way (see Joyce, 1998). The proof is a version of de Finetti's theorem referred to in subsection 2.2.

I owe the following characterization of Bronfman's objection to Pettigrew (see chapter 5 in Pettigrew, 2016; for the Pater Peperium case see Paul, 2016).

> **Example 3: Pater Peperium.** I must choose between three sandwich options: Marmite, cheese, and Pater Peperium (or Gentleman's Relish).

I have eaten cheese sandwiches before and feel indifferent about them. I have never had a Marmite or Pater Peperium sandwich, but know that people either love Marmite and hate Pater Peperium or vice versa. There appears to be nothing irrational about choosing the cheese sandwich even though either way (whether I am of the love-marmite-hate-pater-peperium or hate-marmite-love-pater-peperium type) there is a better sandwich to choose.

Joyce has shown that for any proper scoring rule, a non-probabilistic credence function is accuracy dominated by a probabilistic credence function. The Bronfman objection is that you can show that there is always another proper scoring rule (Bronfman shows that only having two candidate quadratic loss scoring rules suffices to make this point) by which moving from the accuracy dominated credence function to the probabilistic credence function results in a loss at some possible world. Unless we settle on a unique scoring rule to do the accounting, Joyce's non-pragmatic vindication of probabilism is undermined.

Pettigrew uses Bronfman's objection to propose UNIVOCAL DOMINANCE. It is an appealing feature of a scoring rule to have some claim to uniqueness in order to address Bronfman's objection. The Brier score has this claim: it is (up to linear transformation, which do not make a relevant difference) the only proper scoring rule which fulfills SYMMETRY. Unfortunately for Pettigrew, the Brier score is not the only score with this kind of claim. The Log score is the only proper scoring rule which fulfills LOCALITY. We could now haggle over which uniqueness claim is stronger. In some ways, this paper is meant to undermine the intuitive appeal of SYMMETRY altogether. I will not, however, push UNIVOCAL DOMINANCE and LOCALITY as joint justification for the Log score, as Pettigrew pushes UNIVOCAL DOMINANCE and SYMMETRY as joint justification for the Brier score.

Pettigrew's argument is suspect (he by no means is unaware of its tenuous appeal and reiterates that many of his results stand even if UNIVOCAL DOMINANCE is implausible). Let a uniqueness claim have dependent and independent reasons. The dependent reasons justify the uniqueness on account of the features that the object of the uniqueness claim exhibits. The independent reasons make no reference to these features, but provide a reason to have a successful candidate for winning the uniqueness contest. I do not see how these independent reasons add to the epistemic justification for the uniqueness claim.

4.3 Proof of Locality Uniqueness for the Log Score

Let there be a function $f : (a, b) \to \mathbb{R}$ ($a, b \in \mathbb{R}$ with $a < b$) for which the following is true: for every $x \in (a, b)$ there exists a linear function $L_x : \mathbb{R} \to \mathbb{R}$ such that $L_x(x) = x$ and

$$L_x(y) < f(y) \text{ for all } y \in (a, b) \tag{18}$$

The conditions basically say that f has a subgradient at every point in its domain. Then the function is strictly convex, i.e.

$$f(\lambda y + (1 - \lambda)\hat{y}) < \lambda f(y) + (1 - \lambda)f(\hat{y}) \tag{19}$$

for all $0 < \lambda < 1$ and for all $y, \hat{y}$ in the domain of f. A theorem of convex analysis tells us that such a function is almost everywhere differentiable (i.e. the set of points where it is not differentiable is countable; see theorem 25.5 in Rockafellar, 1997).

Define $f_i(p_i) = S(\xi_i, p_i)$ for a scoring rule fulfilling LOCALITY and PROPRIETY. f_i are functions from $[0,1] \to \mathbb{R} \cup \{-\infty, \infty\}, i = 1, \ldots, n$. Propriety requires

$$\sum_{i=1}^{n} p_i f_i(p_i) > \sum_{i=1}^{n} q_i f_i(p_i) \text{ for all } p, q \in \mathcal{P} \tag{20}$$

where $p = (p_1, \ldots, p_n)^\intercal$ and $q = (q_1, \ldots, q_n)^\intercal$. Let $0 < k < 1$ be arbitrary, but fixed, and define

$$g_k(x) = x f_1(x \cdot k) + (1 - x) f_2((1 - x) \cdot k) \tag{21}$$

Let $\alpha, \beta \in (0,1)$ with $\alpha \neq \beta$. (20) for the two distributions $p = (\alpha \cdot k, (1-\alpha) \cdot k, \frac{1-k}{n-2}, \ldots, \frac{1-k}{n-2})$ and $q = (\beta \cdot k, (1 - \beta) \cdot k, \frac{1-k}{n-2}, \ldots, \frac{1-k}{n-2})$ implies

$$g_k(\alpha) > (\alpha \cdot k) f_1(\beta \cdot k) + ((1 - \alpha) \cdot k) f_2((1 - \beta) \cdot k) \tag{22}$$

References

Bronfman, Aaron. "A Gap In Joyce's Argument for Probabilism.", 2009. University of Michigan: unpublished manuscript.

Dawid, A Philip, Steffen Lauritzen, and Matthew Parry. "Proper Local Scoring Rules on Discrete Sample Spaces." *Annals of Statistics* 40, 1: (2012) 593–608.

De Finetti, Bruno. *Theory of Probability.* Chichester, UK: Wiley, 2017.

Good, I.J. "Rational Decisions." *Journal of the Royal Statistical Society. Series B (Methodological)* 14, 1: (1952) 107–114.

Hendrickson, Arlo, and Robert Buehler. "Proper Scores for Probability Forecasters." *Annals of Mathematical Statistics* 42, 6: (1971) 1916–1921.

Howson, Colin. "De Finetti, Countable Additivity, Consistency and Coherence." *British Journal for the Philosophy of Science* 59, 1: (2008) 1–23.

Joyce, James. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65, 4: (1998) 575–603.

Leitgeb, Hannes, and Richard Pettigrew. "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science* 77, 2: (2010) 201–235.

Marschak, Jacob. "Remarks on the Economics of Information." Technical report, Cowles Foundation for Research in Economics, Yale University, 1959.

McCarthy, John. "Measures of the Value of Information." *Proceedings of the National Academy of Sciences* 42, 9: (1956) 654–655.

Miller, David. "A Geometry of Logic." In *Aspects of Vagueness*, edited by Heinz Skala, Settimo Termini, and Enric Trillas, Dordrecht, Holland: Reidel, 1984, 91–104.

Mormann, Thomas. "Geometry of Logic and Truth Approximation." *Poznan Studies in the Philosophy of the Sciences and the Humanities* 83, 1: (2005) 431–454.

Paul, L.A. *Transformative Experience.* Oxford, UK: Oxford University, 2016.

Pettigrew, Richard. *Accuracy and the Laws of Credence*. Oxford, UK: Oxford University, 2016.

Rockafellar, Ralph. *Convex Analysis*. Princeton, N.J: Princeton University, 1997.

Savage, Leonard. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66, 336: (1971) 783–801.

Zill, Dennis. *Multivariable Calculus*. Sudbury, MA: Jones and Bartlett, 2011.