# DOMINATION FOR FINITE PROPER SCORING RULES

ALEXANDER R. PRUSS

Abstract. Scoring rules measure the deviation between a forecast, which assigns degrees of confidence to various events, and reality. Strictly proper scoring rules have the property that for any forecast, the mathematical expectation of the score of a forecast $p$ by the lights of $p$ is strictly better than the mathematical expectation of any other forecast $q$ by the lights of $p$. It has recently been shown that any strictly proper scoring rule that is continuous on the probabilities has the property that the score for any forecast that does not satisfy the axioms of probability is strictly dominated by the score for some probabilistically consistent forecast. I shall show that in the case of a finite score, continuity on the regular probabilities—those that assign assign a non-zero value to every point—suffices for the result, and more generally it suffices to have continuity at the regular probabilities and at any infinite-scoring irregular ones.

## 1. The main results

Scoring rules measure the deviation between a forecast, which assigns degrees of confidence or credence to various events, and reality. Strictly proper scoring rules have the property that for any forecast, the mathematical expectation of the score of a forecast $p$ by the lights of $p$ is strictly better than the mathematical expectation of any other forecast $q$ by the lights of $p$. Forecasts need not satisfy the axioms of probability, but under some continuity conditions, the score of a forecast that does not satisfy the axioms of probability is strictly dominated by the score of a forecast by that does satisfy these axioms. This result has been interpreted by epistemologists as supporting the idea that reasonable forecasts will always be probabilistically consistent (e.g., [4], [2], [6]).

To be precise, let $\Omega$ be a finite sample space, encoding the situations being forecast. Let $\mathcal{C}$ be the set of all forecasts or credence functions, i.e., functions from the power set of $\Omega$ to the reals. Let $\mathcal{P}$ be the set of those credences that satisfy the axioms of probability. An *accuracy scoring rule* is a function $s$ from a set $\mathcal{F} \supseteq \mathcal{P}$ of credence function to $[-\infty, M]^{\Omega}$ for some finite $M$, where $A^B$ is the set of functions from $B$ to $A$, and where $s(c)(\omega)$ represents the epistemic utility of having credence $c$ when in fact we are at $\omega \in \Omega$. Higher accuracy scores are better in this terminology and measure closeness to truth.

Given a probability $p \in \mathcal{P}$ and an extended real function $f$ on $\Omega$, let $E_p f$ be the expected value with respect to $p$ defined technically as follows to avoid computing $0 \cdot (-\infty)$.

$$E_p f = \sum_{\omega \in \Omega, p(\{\omega\}) \neq 0} p(\{\omega\}) f(\omega).$$

A scoring rule $s$ is *proper* on $\mathcal{F} \supseteq \mathcal{P}$ provided that for every $p \in \mathcal{P}$ and every $c \in \mathcal{F}$, we have $E_p s(p) \geq E_p s(c)$, *strictly proper* on $\mathcal{F}$ provided the inequality is always strict when $p \neq c$, and *quasi-strictly proper* there provided that it is proper and the inequality is strict when $p \in \mathcal{P}$ and $c \in \mathcal{F} \backslash \mathcal{P}$.

Given propriety, if an agent adopts a probability function $p$ as their forecast, then by the agent's lights there can be no improvement in the expected score from switching to a different forecast. Strict propriety captures the idea that an agent who has adopted a probability function $p$ as their forecast will think other forecasts to be inferior. Proper and strictly proper scoring rules have been widely studied: for a few examples, see [1], [3], [6], [8], [11]. Some of the literature concerns inaccuracy scores, which measure deviation from truth or epistemic disutility, but one can easily translate: $-s$ is a proper, strictly proper or quasi-strictly proper inaccuracy scoring rule on $\mathcal{F}$ if and only if $s$ is respectively a proper, strictly proper or quasi-strictly proper scoring rule on $\mathcal{F}$.

The probabilities $\mathcal{P}$ can be identified with the set of points of $\mathbb{R}^n$ with non-negative coordinates summing to 1, where $n$ is the number of points in $\Omega$. Specifically, given $p \in \mathcal{P}$, we can think of $p$ as a vector $\hat{p}$ with the coordinates $\hat{p}_i = p(\{i\})$. This embedding provides $\mathcal{P}$ with a topology. Also, the score $s(c)$ of $c$ is a function from $\Omega$ to $[-\infty, M]$, and the set of all such functions can be identified with the product space $[-\infty, M]^n$ and thereby gets a topology. We then say that $s$ is probability-continuous at $p$ provided that $p$ is a probability, and $s$ restricted to the probabilities is continuous at $p$ in this topology. This is equivalent to requiring that whenever $(p_n)$ is a sequence of probabilities converging to a probability $p$, then $\lim_n s(p_n)(\omega) = s(p)(\omega)$ for every $\omega \in \Omega$.

Predd, et al. [8] showed that if $s$ is a probability-continuous strictly proper scoring rule that can be expressed as a sum of single-proposition scores, then for any non-probability $c$, there is a probability $p$ such that $s(c)$ is strictly dominated by $s(p)$, i.e., $s(c)(\omega) < s(p)(\omega)$ for all $\omega$ in our accuracy setting. In other words, any forecaster whose forecast fails to be a probability can find a forecast that is a probability and that is strictly better no matter what. The more recent Pettigrew [7], Nielsen [5] and Pruss [9] theorem shows that this is true without the restriction to scores that are sums of single-proposition scores. Nielsen's proof also extended the result to the quasi-proper case. A philosophical upshot of these results is that we can get an argument in favor of probabilistic consistency in one's credence assignments under much weaker conditions than the additivity assumed by Predd, et al.

Say that a proper scoring rule $s$ defined on the probabilities has *the domination property* provided that any extension of $s$ to a quasi-proper scoring rule on all credences makes every non-probability be $s$-dominated by some probability, i.e., for any $c \notin \mathcal{P}$, there is a $p \in \mathcal{P}$ such that $s(c)(\omega) < s(p)(\omega)$

for all $\omega$. Pruss [10] gave a set of necessary and sufficient conditions for a proper scoring rule to have the domination property. However, one of these conditions is geometric and difficult to apply.

The purpose of this paper is to give a sufficient condition that is of philosophical interest. As a motivating philosophical point, note that one might think that a scoring rule that models reasonable intuitions about epistemic value could have a discontinuity when the probability in some proposition hits zero or one. For one might have the Cartesian intuition that there is something particularly valuable about being *certain* of a truth, and that this value is not just the limit of the value of high probability in the truth. And, perhaps even more intuitively, we might have the converse intuition that there is something particular disvaluable about being certain of a falsehood, again in a discontinuous way. Thus a reasonable philosophical condition on a scoring rule is that the scoring rule be continuous at the regular probabilities—those that assign non-zero values to all events (and hence that do not assign one to any event other than the trivial event of the whole space).

It turns out that this condition is insufficient on its own for the domination property, but becomes sufficient when we require the scores to be finite. More precisely, we have the following, where we say that the score of $c$ is finite just in case $s(c)(\omega)$ is finite for all $\omega$.

**Theorem 1.** *Suppose $s$ is probability-continuous on the regular probabilities and for any sequence $(p_n)$ in $\mathcal{P}$ that converges to a non-regular probability $p$ such that $s(p_n)$ is finite for all $n$ while $s(p)$ is not finite, we have $\lim_n E_{p_n} s(p_n) = E_p s(p)$. Then $s$ has the domination property: for any extension of $s$ to a quasi-strictly proper scoring rule on all credences and any non-probability $c$, there is a probability $p$ such that $s(c)(\omega) < s(c)(\omega)$ for all $\omega$.*

By [10, Lemma 3], if $p_n$ converges to $p$ and $s(p_n)$ converges to $r$, we have $\lim_n E_{p_n} s(p_n) = E_p r$, so the limit condition in Theorem 1 is met if $s$ is continuous at $p$. Thus:

**Corollary 1.** *If $s$ is probability-continuous on the regular probabilities and at all the irregular probabilities where it's infinite, then it has the domination property.*

In particular, if $s$ is probability-continuous on the regular probabilities and finite everywhere, then it has the domination property.

Note that as long as $s$ has any quasi-strictly proper extension to non-probabilities, the score of any regular probability must be finite, since if $p$ is regular and $s(p)(\omega) = -\infty$, then $E_p s(p) = -\infty$ and so we cannot have $E_p s(c) < E_p s(p)$. Further, note that by [10, Lemma 3], $\lim_n E_{p_n} s(p_n)$ always exists for a proper scoring rule $s$ if $(p_n)$ is a convergent sequence of probabilities.

By the main theorem of [10], the limit condition in the Theorem is necessary for the domination property except in the trivializing case where $E_p s(p) = -\infty$ for some probability $p$. In that trivializing case, there is no quasi-strictly proper extension to non-probabilities, since we cannot have $E_p s(c) < E_p s(p)$.

There do exist unbounded scoring rules that are strictly proper on the probabilities, continuous on the regular probabilities, but do not satisfy the limit condition and do not have the trivializing condition and hence do not have the domination property. For instance, let $\Omega = \{1, 2\}$ and let $\delta_1$ be the probability such that $\delta_1(\{1\})$. Let $s(p)(i) = \log p(\{i\})$ for $i = 1, 2$, whenever $p \neq \delta_1$, and let $s(\delta_1)(1) = 1$ and $s(\delta_1)(2) = -\infty$. It's easy to see that $E_p s(q)$ is the same as for the logarithmic rule $\lambda$ (where $\lambda(p)(i) = \log p(\{i\})$ in all cases) except when $p = q = \delta_1$ in which case $E_p s(q) = 1 > 0 = E_p \lambda(q)$. Thus $s$ is strictly proper because $\lambda$ is. The rule $s$ does not have the trivializing condition and it is continuous except at $\delta_1$. But it does not satisfy the limiting condition since $\lim E_{p_n} s(p_n) = E_p \lambda s(p)$ if $p_n \to p = \delta_1$ and the $p_n$ are regular, but $E_{\delta_1} \lambda(\delta_1) \neq E_{\delta_1} s(\delta_1)$.

Our result implies that if a philosopher can argue that the correct scoring rule should be strictly proper, finite and continuous on the regular probabilities, then they have an argument for probabilism—the thesis that an agent's credences should be probabilities—because it is very plausible that one is not rationally permitted to have credences that are dominated by another set of credences. And finiteness has some plausibility to it. If the score of, say, certainty in a falsehood is infinitely bad, then we cannot say that it's worse to be sure of two falsehoods than of one.

## 2. Proofs

Without loss of generality suppose $\Omega = \{1, ..., n\}$. Identify a score $s(p)$ with the extended-real "vector" $(s(p)(1), \ldots, s(p)(n))$ in $[-\infty, M]^n$. Let $F$ be the set of finite scores of probabilities, i.e., $F = \{s(p) : p \in \mathcal{P}$ and $s(p)$ is finite$\}$. On our identification, $F$ is a subset of Euclidean space $\mathbb{R}^n$. Write $\langle v, z \rangle$ for the inner product on $\mathbb{R}^n$. Note that $E_p s(p) = \langle \hat{p}, s(p) \rangle$ if $s(p)$ is finite, given our identification.

**Lemma 1.** *Let $C \subseteq \mathbb{R}^n$ for $n \geq 2$ and let $f$ be a function defined from an open subset $U$ of the sphere $S^{n-1} = \{v \in \mathbb{R}^n : |v| = 1\}$ to $C$ such that $\langle v, z \rangle \leq \langle v, f(v) \rangle$ for all $v \in U$ and $z \in C$. Suppose that $v_0 \in U$ and there is a point $z_1 \neq f(v_0)$ in the closed convex hull of $C$ such that $\langle v_0, z_1 \rangle = \langle v_0, f(v_0) \rangle$. Then $f$ is not continuous at $v_0$.*

*Proof.* Without loss of generality, assume that $C$ is closed and convex, since the inequality $\langle v, z \rangle \leq \langle v, f(v) \rangle$ extends to all $z$ in the closed convex hull of $C$. We now have $z_1 \in C$.

Let $V = \{v \in \mathbb{R}^n \backslash \{0\} : v/|v| \in U\}$. This is also open. Letting $f(v) = f(v/|v|)$ for $v \in V \backslash U$ extends $f$ to a function on $V$, which is continuous at

$v$ if and only if the original function was as well, and continuous to satisfy the inequality $\langle v, z\rangle \le \langle v, f(v)\rangle$ for $z \in C$.

Let $z_0 = f(v_0)$. Fix $\varepsilon > 0$. Let $w_\varepsilon = v_0 + \varepsilon(z_1 - z_0)$. Suppose $\varepsilon$ is sufficiently small that the vector $w_\varepsilon$ is in $V$. I claim that $|f(w_\varepsilon) - z_0| \ge |z_1 - z_0|$. If that's correct, that implies $f$ is discontinuous at $v_0$, since $w_\varepsilon \to v_0$ as $\varepsilon \to 0+$.

To prove the claim, without loss of generality by rotating $C$ if necessary about the origin, we can assume that $v_0$ lies on the first coordinate axis, i.e., $v_0 = (1, 0, \ldots, 0)$. Since $\langle v_0, z_0\rangle = \langle v_0, z_1\rangle$, the first coordinate of $z_0$ is the same as that of $z_1$. Thus by a rotation that leaves the first coordinate axis fixed, we can assume that $z_i = (x, y_i, 0, \ldots, 0)$ for $i = 1, 2$ and for some fixed $x$, $y_0$ and $y_1$, with $y_1 > y_0$. Let $t = y_1 - y_0 = |z_1 - z_0|$. Note that $w_\varepsilon = (1, \varepsilon t, 0, \ldots, 0)$.

Write $f(w_\varepsilon) = (\alpha_1, \ldots, \alpha_n)$. Note that $\alpha_1 = \langle v_0, f(w_\varepsilon)\rangle \le \langle v_0, f(v_0)\rangle = x$.

Since $\langle w_\varepsilon, z_1\rangle \le \langle w_\varepsilon, f(w_\varepsilon)\rangle$, we have:

$$x + \varepsilon t y_1 \le \alpha_1 + \varepsilon t \alpha_2 \le x + \varepsilon t \alpha_2.$$

Hence $\alpha_2 \ge y_1$. But $|f(w_\varepsilon) - z_0| \ge |\alpha_2 - y_0|$, and $|\alpha_2 - y_0| \ge y_1 - y_0 = |z_1 - z_0|$ since $\alpha_2 \ge y_1 > y_0$. Thus $|f(w_\varepsilon) - z_0| \ge |z_1 - z_0|$, as desired. □

*Proof of Theorem 1.* Assume the non-trivializing condition that $E_p s(p)$ is finite for all probabilities $p$, and hence that $s(p)$ is finite for any regular $p$.

By the main theorem of [10], to get the domination property under the other conditions assumed in our Theorem 1, we only need to show that $F$ is dense in $\partial^+ \operatorname{Conv} F$, where $\operatorname{Conv} F$ is the convex hull of $F$ and $\partial^+ U$ is the "positive-facing boundary" of $U$, i.e., the set of topological boundary points $z$ of $U$ such that there is a vector $v \in (0, \infty)^n$ (i.e., a positive-facing vector) with the property that $\langle v, w\rangle \le \langle v, z\rangle$ for all $w \in F$, where $\langle \cdot, \cdot\rangle$ is the dot product on $\mathbb{R}^n$.

We now show that this is true in the special case where $s$ is strictly proper. Suppose $z_1 \in \partial^+ \operatorname{Conv} F$. Let $v$ be a positive-facing vector such that $\langle v, w\rangle \le \langle v, z_1\rangle$ for all $w \in F$. Rescaling if necessary, we may suppose $v = \hat{p}$ for a regular probability $p$. If $s(p) = z_1$, then we are done, so suppose $s(p) \ne z_1$.

I now claim that $\langle \hat{p}, z_1\rangle = \langle \hat{p}, s(p)\rangle$. For we have $\langle v, s(p)\rangle \le \langle v, z_1\rangle$ and $v = \hat{p}$, and conversely $z_1$ is a limit of convex combinations of $s(q_i)$ for $q_i \in \mathcal{P}$, and $\langle \hat{p}, s(q_i)\rangle \le \langle \hat{p}, s(p)\rangle$.

Let $v_0 = \hat{p}/|\hat{p}|$. Let $U$ be a neighborhood of $v_0$ such that all the coordinates of all the vectors in $U$ are strictly positive. For $x = (x_1, \ldots, x_n) \in U$, let $\check{x}$ be the regular probability such that $\check{x}(\{i\}) = x_i / \sum_{k=1}^n x_k$. Let $f(x) = s(\check{x})$. Then $f$ satisfies the conditions of the Lemma with $C = F$. Hence $f$ is discontinuous at $v_0$, which implies $s$ is discontinuous at $p$, a contradiction.

It remains to prove the result in the case where $s$ is only quasi-strictly proper. Fix a non-probability $c_0$. Let $F$ be the set of $s(p)$ for $p \in \mathcal{P}$, and for $p \in \mathcal{P}$ let

$$f(p) = E_p s(p) - E_p s(c_0).$$

Note that $E_p s(p) = \sup_{z \in F} E_p z$ by propriety, and hence $p \mapsto E_p s(p)$ is lower semicontinuous, since it is the supremum of continuous functions (this uses the finiteness of all the members of $F$). Thus $f$ is lower semicontinuous. Moreover, by quasi-strict propriety, $f$ is strictly positive on $\mathcal{P}$, and hence by lower semicontinuity and the compactness of $\mathcal{P}$ we have $\varepsilon = \min f > 0$. Let $t$ be any strictly proper probability-continuous scoring rule such that $|t(c)(i)| < \varepsilon/2$ for all $c$ and $1 \leq i \leq n$. Let $s'(p) = s(p) + t(p)$ for $p \in \mathcal{P}$ and $s'(c) = s(c_0) + \varepsilon/2$ for $c \notin \mathcal{P}$. Then

$$E_p s'(p) > E_p s(p) - \frac{\varepsilon}{2} \geq E_p s(c_0) + \varepsilon - \frac{\varepsilon}{2} = E_p s'(c_0),$$

where the first inequality follows from the bounds on $t$ and the second follows by the fact that $\varepsilon = \min f$. Moreover, $s'$ is strictly proper on the probabilities as there it is the sum of a proper scoring rule and a strictly proper one. Hence $s'$ is strictly proper, and it is finite everywhere and continuous on the regular probabilities. Thus by what we have proved before, $s'(c_0)$ is strictly dominated by $s'(p)$ for some $p$. Then:

$$s(c_0)(i) = s'(c)(i) - \frac{\varepsilon}{2} < s'(p)(i) - \frac{\varepsilon}{2} < s(p)(i) + \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = s(p)(i),$$

and so $s(c_0)$ is strictly dominated by $s(p)$ as desired. □

## References


[1] A. P. Dawid, and M. Musio. 2014. "Theory and applications of proper scoring rules", *Metron* 72:169–183.

[2] H. Greaves, and D. Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility", *Mind* 115:607–632.

[3] T. Gneiting, and A. E. Raftery. 2007. "Strictly proper scoring rules, prediction, and estimation", *Journal of the American Statistical Association* 102:359—378.

[4] J. M. Joyce. 1998. "A Nonpragmatic Vindication of Probabilism", *Philosophy of Science* 65:575–603.

[5] Michael Nielsen. 2022. "On the Best Accuracy Arguments for Probabilism", *Philosophy of Science* **89**:621–630.

[6] R. Pettigrew. 2016. *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

[7] R. Pettigrew. Forthcoming. "Accuracy-First Epistemology Without the Additivity Axiom", *Philosophy of Science*.

[8] Joel B. Predd, Robert Seiringer, Elliott H. Lieb, Daniel N. Osherson, H. Vincent Poor, and Sanjeev R. Kulkarni. 2009. "Probabilistic Coherence and Proper Scoring Rules", *IEEE Transactions on Information Theory* 55:4786–4792.

[9] Alexander R. Pruss. 2021. "Proper Scoring Rules and Domination". `https://arxiv.org/abs/2102.02260`

[10] Alexander R. Pruss. 2023. "Necessary and Sufficient Conditions for Domination Results for Proper Scoring Rules", *Review of Symbolic Logic*, forthcoming. `https://arxiv.org/abs/2103.00085`

[11] R. L. Winkler, J. Munoz, J. L. Cervera, J. M. Bernardo, G. Blattenberger, J. B. Kadane, D. V. Lindley, A. H. Murphy, R. M. Oliver, and D. Ríos-Insua. 1996. "Scoring rules and the evaluation of probabilities", *Test* 5:1—60.