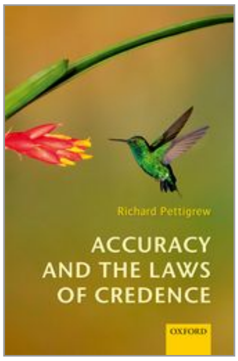
University Press Scholarship Online
Oxford Scholarship Online



Accuracy and the Laws of Credence
Richard Pettigrew

Print publication date: 2016
Print ISBN-13: 9780198732716
Published to Oxford Scholarship Online: May 2016
DOI: 10.1093/acprof:oso/9780198732716.001.0001

The accuracy argument for Probabilism

Richard Pettigrew

DOI:10.1093/acprof:oso/9780198732716.003.0008

Abstract and Keywords

This chapter summarizes the results of the investigation so far and gives the final version of the accuracy argument for Probabilism.

Keywords: Accuracy, Probabilism, Dominance principle

We are now in a position to state my final version of the accuracy argument for Probabilism. It runs as follows:

> (I^*_P) **Veritism** The sole fundamental source of epistemic value is accuracy.
> (II^*_P) **Brier Alethic Accuracy** The inaccuracy of a credence function at a world is measured by its Brier score at that world. This is a consequence of the following axioms:
> (i) **Perfectionism** The accuracy of a credence function at a world is its proximity to the vindicated credence function at that world.
> (ii) **Squared Euclidean Distance** Distance between credence functions is measured by squared Euclidean distance.
> This is a consequence of the following axioms:

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

- **Perfectionism** (cf. Chapter 4)
- **Divergence Additivity** (cf. Section 4.1)
- **Divergence Continuity** (cf. Section 4.2)
- **Decomposition** (cf. Section 4.3)
- **Symmetry** (cf. Section 4.4)
- **Theorem 4.4.1**

(iii) **Alethic Vindication** The vindicated credence function at a world is the omniscient credence function at that world.

(III^*_P) **Immodest Dominance** (cf. Section 2.2)

(IV^*_P) **Theorems I.D.5** and **I.D.7** and **I.A.2** (which together entail **Theorem 4.3.4**) and **Proposition I.B.2**.

Therefore,

(V^*_P) **Probabilism**

This argument avoids the Bronfman objection because the second premise says (p.81) that the inaccuracy of a credence function at a world is given by its Brier score at that world. But it is worth noting that, for *any* additive and continuous strictly proper inaccuracy measure $\mathfrak{I}$—that is, any inaccuracy measure that is generated by an additive Bregman divergence $\mathfrak{D}$ in line with Alethic Vindication, so that $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$—if we replace the second premise in the argument just given with the claim that the inaccuracy of a credence function c at a world w is given by $\mathfrak{I}(c, w)$, then Theorem 4.3.4 is sufficiently general that the argument will go through. The point is that the Bronfman objection shows that the second premise of our accuracy argument must restrict the number of legitimate inaccuracy measures to one; and Theorem 4.3.4 will only let our argument go through if that single legitimate inaccuracy measure is additive, continuous, and strictly proper; but it is not necessary for the success of the argument that it is the Brier score in particular that is taken to be the single legitimate measure of inaccuracy. I take the Brier score to be that single legitimate measure because the claim that it is follows from Perfectionism, Divergence Additivity, Divergence Continuity, Decomposition, Symmetry, and Alethic Vindication, each of which I endorse. But others might endorse a different characterization of a different single legitimate measure of inaccuracy. And so long as that measure is additive, continuous, and strictly proper, that is sufficient to establish Probabilism. This will be true also throughout the remainder of the book. I will assume that the Brier score is the unique legitimate measure of inaccuracy in a number of places. But all of the arguments

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

for the various laws of credence that we will give will go through just as well if we instead take another additive and continuous strictly proper inaccuracy measure to be the unique legitimate such measure.

(p.82) Appendix I: The mathematical results

In this appendix, we outline the mathematical results that underpin the arguments that we have been considering in this part of the book. These arguments contain two mathematical components: the first is a characterization of the legitimate measures of accuracy; the second is a proof that, relative to any one of these measures, the credence functions deemed irrational by Immodest Dominance are precisely those that violate Probabilism. In this chapter we will proceed towards these two components as follows:

- We begin with a geometric characterization of the probabilistic credence functions on a set $\mathscr{F}$ as the members of the (closed) convex hull of the omniscient credence functions on $\mathscr{F}$ at the worlds relative to $\mathscr{F}$.
- We characterize the inaccuracy measures that satisfy the conditions imposed in Chapter 4 (except Symmetry).
    - We show that they are those generated by a sort of divergence called an additive Bregman divergence (Theorem 4.3.3).
    - We establish that they are the additive and continuous strictly proper inaccuracy measures (Theorem 4.3.5).
- We show that, if we add Symmetry, then we characterize (positive linear transformations of) the Brier score (Theorem 4.4.1).
- We establish two general results about additive Bregman divergences that will be crucial throughout the book (Theorems I.D.5 and I.D.7).
- We use both of these crucial results, together with our geometrical characterization of the probabilistic credence functions, to establish Theorem 4.3.4.
- We then note in Proposition I.B.2 that squared Euclidean distance is an additive Bregman divergence, and thus the Brier score is an additive and continuous strictly proper inaccuracy measure (Theorem 2.2.2). Combined with Theorem 4.3.4, this gives us Theorem 1.0.2, which is what is required for our final version of the argument for Probabilism—that is, the version stated in Chapter 7.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

Throughout this appendix, if $\mathscr{F} = \{X_1, \ldots, X_n\}$, we will treat a credence function $c : \mathscr{F} \to [0,1]$ as the following vector in $\mathbb{R}^n$:

$$(c(X_1), \ldots, c(X_n))$$

(p.83) We will abuse notation and denote this vector c. Thus,

- $\mathscr{B}_{\mathscr{F}} = [0,1]^n$ is the set of all credence functions on $\mathscr{F}$.
- $\mathcal{P}_{\mathscr{F}} \subseteq \mathscr{B}_{\mathscr{F}} = [0,1]^n$ is the set of all probability functions on $\mathscr{F}$.
- $\mathscr{V}_{\mathscr{F}} \subseteq \mathcal{P}_{\mathscr{F}} \subseteq \mathscr{B}_{\mathscr{F}} = [0,1]^n$ is the set of all omniscient credence functions on $\mathscr{F}$.

I.A Characterizing the probabilistic credence functions

As far as I know, the following geometric characterization of $\mathcal{P}_{\mathscr{F}}$ is originally due to de Finetti (1974). To state it, we need a definition:

Definition I.A.1 (**Convex hull**) *Suppose $\mathscr{X}$ is a finite subset of* $\mathbb{R}^n$. *Then let $\mathscr{X}^+$ be the* convex hull of $\mathscr{X}$.

Thus,

$$X^+ := \left\{ \sum_{x \in X} \lambda_x x : \lambda_x \geq 0 \text{ and } \sum_{x \in X} \lambda_x = 1 \right\}$$

So $\mathscr{X}^+$ is the set of all mixtures or weighted sums or convex combinations of the elements of $\mathscr{X}$.

Equivalently, $\mathscr{X}^+$ is the smallest convex set containing $\mathscr{X}$.[1]

We can now state de Finetti's characterization result:

Theorem I.A.2 (**De Finetti's Characterization Theorem**)
$P_F = V_F^+$

That is, the probability functions are precisely the mixtures of the omniscient credence functions. We see this in Figure 0.1, where we illustrate the accuracy argument for No Drop. In that case, the omniscient credence functions are v_{w_1}, v_{w_2}, and v_{w_3}. And the set of probability functions is the set of credence functions that satisfies No Drop; and that, in turn, is the set of credence functions represented by vectors in the triangle whose vertices are v_{w_1}, v_{w_2}, and v_{w_3}; and this is, of course, the convex hull of $V_F = \{v_{w_1}, v_{w_2}, v_{w_3}\}$. Similarly we can see it in Figures 1.1 and 1.2, where the line from v_{w_1} to v_{w_2} is the convex hull of $\{v_{w_1}, v_{w_2}\}$ in the former, and the triangle with vertices $v_{w_1}, v_{w_2}, v_{w_3}$ is the convex hull of $\{v_{w_1}, v_{w_2}, v_{w_3}\}$ in the latter.

Proof of Theorem I.A.2.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

- First, we show that $V_F^+ \subseteq P_F$. Due to the definition of the convex hull, this follows from the following two claims, which are easily verified:
  (i) $\mathcal{V}_{\mathscr{F}} \subseteq \mathcal{P}_{\mathscr{F}}$. That is, every omniscient credence function is a probability function.
  (p.84) (ii) $\mathcal{P}_{\mathscr{F}}$ is convex. That is, if $p, p' \in \mathcal{P}_{\mathscr{F}}$ are probabilityfunctions, so is any mixture $\lambda p + (1 - \lambda)p'$, where $0 \leq \lambda \leq 1$.

- Second, we show that $P_F \subseteq V_F^+$. Suppose p is in $\mathcal{P}_{\mathscr{F}}$. And suppose that p^* is a probabilistic extension to $\mathscr{F}^*$. Now, for each $w \in \mathscr{W}_{\mathscr{F}}$, let w^* be the unique extension of w to $\mathscr{F}^*$. Then $w^* \in \mathscr{W}_{\mathscr{F}}^*$. And $w \mapsto w^*$ is a bijection. Now, for all X in $\mathscr{F}$
$$p(X) = p^*(X) = \sum_{\substack{w \in W_F \\ p^*(w^*) > 0}} p^*(w^*) p^*(X|w^*)$$
where the conditional probability is defined using the ratio form, as usual—so
$$p^*(X|w^*) := \frac{p^*(X \,\&\, w^*)}{p^*(w^*)}$$
Furthermore, if $p^*(w^*) > 0$, then
  – $p^*(X|w^*) = 1 \Leftrightarrow X$ is true at $w^* \Leftrightarrow X$ is true at w
  – $p^*(X|w^*) = 0 \Leftrightarrow X$ is false at $w^* \Leftrightarrow X$ is false at w
Thus, $p^*(X|w^*) = v_w(X)$, if $p^*(w^*) > 0$. So, for all X in $\mathscr{F}$,
$$p(X) = \sum_{w \in W_F} p^*(w^*) v_w(X)$$
Thus, p is in V_F^+, as required.

This completes our proof. □

I.B Characterizing legitimate inaccuracy measures (without Symmetry)

We now characterize the inaccuracy measures that satisfy the conditions (other than Symmetry) laid down in Chapter 4. We first show that they are precisely those generated by a class of divergences called the additive Bregman divergences (Theorem 4.3.3); then we show that they are precisely the additive and continuous strictly proper inaccuracy measures (Theorem 4.3.5).

First, we must define the class of additive Bregman divergences. Throughout, we let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ be vectors in $[0,1]^n$.

Definition I.B.1 (**Additive Bregman divergence**) *Suppose* $\mathfrak{D}$: $[0,1]^n \times [0,1]^n \to [0, \infty]$. *Then*

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

- $\mathfrak{D}$ is a divergence *if* $\mathfrak{D}(\mathbf{x}, \mathbf{y}) \geq 0$ *for all* $\mathbf{x}, \mathbf{y} \in [0,1]^n$ *with equality iff* $\mathbf{x} = \mathbf{y}$.
- $\mathfrak{D}$ is additive *if there is* $\mathfrak{d} : [0,1] \to [0, \infty]$ *such that* (p.85)

$$\mathrm{D}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathrm{d}(x_i, y_i)$$

We say that $\mathfrak{d}$ *is a one-dimensional divergence and* $\mathfrak{D}$ *is generated by* $\mathfrak{d}$.

- $\mathfrak{D}$ is an additive Bregman divergence *if* $\mathfrak{D}$ *is an additive divergence and, if* $\mathfrak{D}$ *is generated by* $\mathfrak{d}$, *there is* $\varphi: [0, 1] \to \mathbb{R}$ *such that*

  (i) φ *is continuous, bounded, and strictly convex on* $[0, 1]$;

  (ii) φ *is continuously differentiable on* $(0, 1)$;

  (iii) *For all* $x, y \in [0,1]$,

$$\mathrm{d}(x, y) = \varphi(x) - \varphi(y) - \varphi'(y)(x - y)$$

  where we define $\varphi' = \lim_{x \to i} \varphi'(x)$ *for* $i = 0,1$.

Thus, if we let $\Phi(\mathrm{x}) = \sum_{i=1}^{n} \varphi(x)$, *we have*:[2]

$$\mathrm{D}(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) - \Phi(\mathbf{y}) - \nabla\Phi(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})$$

In this case, we say that $\mathfrak{d}$ *is the one-dimensional Bregman divergence generated by* φ *and* $\mathfrak{D}$ *is the additive Bregman divergence generated by* φ.

Figure 7.1 gives a useful illustration of the role of φ in an additive Bregman divergence.

Proposition I.B.2 *Squared Euclidean distance is an additive Bregman divergence.*

Proof. Let $\varphi(x) = x^2$. Thus, $\Phi(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 = \| x \|_2^2$. Then

$$\mathrm{d}(x, y) = \varphi(x) - \varphi(y) - \varphi'(y)(x - y) = x^2 - y^2 - 2y(x - y) = (x - y)^2$$

and

$$\mathrm{D}(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) - \Phi(\mathbf{y}) - \nabla\Phi(\mathbf{y})(\mathbf{x} - \mathbf{y}) = \sum_{i=1}^{n} (x_i - y_i)^2 = \| \mathbf{x} - \mathbf{y} \|_2^2$$

as required. □

Next, we reiterate the definition of additive and continuous strictly proper inaccuracy measures: (p.86)

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

Definition I.B.3 (**Additive and continuous strictly proper inaccuracy measures**)

• *A scoring rule is a function* $\mathfrak{s} : \{0,1\} \times [0,1] \to [0, \infty]$ *with* $\mathfrak{s}(0,0) = \mathfrak{s}(1, 1) = 0$.

• *A scoring rule* $\mathfrak{s}$ *is* strictly proper *if, for each* $0 \leq p \leq 1$,

$$p\mathfrak{s}(1, x) + (1 - p)\mathfrak{s}(0, x)$$

is minimized uniquely as a function of x *at* $x = p$.

• *A scoring rule* $\mathfrak{s}$ *is* continuous *if* $\mathfrak{s}(i, x)$ *is a continuous function of* x *for* $i = 0,1$.

• *An inaccuracy measure* $\mathfrak{I}$ *is* additive, continuous, and strictly proper *if there is a continuous strictly proper scoring rule* $\mathfrak{s}$ *such that, if* c *is defined on* $\mathcal{F}$,

$$I(c, w) = \sum_{X \in F} \mathfrak{s}(v_w(X), c(X))$$

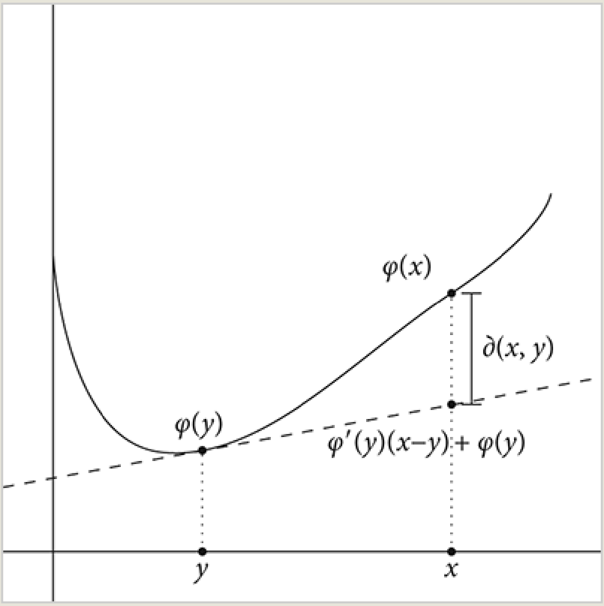


Figure 7.1 As this diagram shows, if $\mathfrak{d}$ is generated by φ, then $\mathfrak{d}(x,y)$ is the difference between the value of two different functions at y. The first function is φ, which takes value $\varphi(x)$ at x. The second is the tangent to φ taken at y, which takes value $\varphi'(y)(x - y) + \varphi(y)$ at x. Thus, $\mathfrak{d}(x, y) = \varphi(x) - \varphi(y) - p'(y)(x - y)$.

Next we prove the crucial connection between additive Bregman divergences and additive and continuous strictly proper inaccuracy measures. Much of this result and its proof depends on (Predd et al. 2009).

Theorem I.B.4

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

(1) *Suppose* $\mathfrak{s}$ *is a continuous strictly proper scoring rule. Then there is a onedimensional Bregman divergence* $\mathfrak{d}$ *such that,*

(a) *For* $i = 0,1$ *and* $x \in [0,1]$,

$$\mathfrak{d}(i, x) = \mathfrak{s}(i, x)$$

(b) *For* $x, y \in [0,1]$ (p.87)

$$\mathfrak{d}(x, y) = \mathrm{Exp}_{\mathfrak{s}}(y|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)$$

where $\mathrm{Exp}_{\mathfrak{s}}(y|x) := x\mathfrak{s}(1,y) + (1-x)\,\mathfrak{s}\,(0,y)$.

(2) *Suppose* $\mathfrak{d}$ *is a one-dimensional Bregman divergence. Then define the following scoring rule*: $\mathfrak{s}(i, x) := \mathfrak{d}(i, x)$. *Then* $\mathfrak{s}$ *is a continuous strictly proper scoring rule.*

(3) *Suppose* $\mathfrak{d}$ *is a one-dimensional divergence that is continuous in its second argument. Then define the following scoring rule*: $\mathfrak{s}(i, x) := \mathfrak{d}(i, x)$. *Then, if, for all* $x,y \in [0,1]$,

$$\mathfrak{d}(x, y) = \mathrm{Exp}_{\mathfrak{s}}(y|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)$$

then $\mathfrak{s}$ *is a continuous strictly proper scoring rule and* $\mathfrak{d}$ *is a one-dimensional Bregman divergence.*

Proof of Theorem I.B.4(1). Suppose $\mathfrak{s}$ is a strictly proper scoring rule. Then define $\varphi : [0,1] \to \mathbb{R}$ as follows:

$$\varphi(x) := -\mathrm{Exp}_{\mathfrak{s}}(x|x) = -x\mathfrak{s}(1, x) - (1-x)\mathfrak{s}(0, x)$$

Then, as proven in (Predd et al. 2009),

(i) φ is continuous, bounded, and strictly convex on [0,1];
(ii) φ is continuously differentiable on (0,1).

As above, we extend φ' to the boundaries as follows: $\varphi'(i) = \lim_{x\to i} \varphi'(x)$ for $i = 0,1$. With this extension in hand, a little calculus shows that, for all $x \in [0,1]$,

$$\varphi'(x) = \mathfrak{s}(0, x) - \mathfrak{s}(1, x)$$

And we can use that to show that, for $i = 0,1$ and $x \in [0,1]$,

$$\mathfrak{s}(i, x) = -\varphi(x) - \varphi'(x)(i - x)$$

We also have $\varphi(0) = \varphi(1) = 0$, since $\mathfrak{s}(0,0) = \mathfrak{s}(1,1) = 0$. Thus, if $\mathfrak{d}$ is the onedimensional Bregman divergence generated by φ, it follows that, for $i = 0,1$ and $x \in [0,1]$,

$$\mathfrak{d}(i, x) = \varphi(i) - \varphi(x) - \varphi'(x)(i - x) = -\varphi(x) - \varphi'(x)(i - x) = \mathfrak{s}(i, x)$$

as required.

Moreover:

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

$$\begin{aligned}&\mathrm{Exp}_{\mathfrak{s}}(y|x)\\&= x\mathfrak{s}(1, y) + (1-x)\mathfrak{s}(0, y)\\&= x\mathfrak{d}(1, y) + (1-x)\mathfrak{d}(0, y)\\&= x[\varphi(1) - \varphi(y) - \varphi'(y)(1-y)] + (1-x)[\varphi(0) - \varphi(y) - \varphi'(y)(0-y)]\\&= -\varphi(y) - \varphi'(y)(x-y)\end{aligned}$$

(p.88) So, in particular, $\mathrm{Exp}_{\mathfrak{s}}(x|x) = -\varphi(x)$. Thus,

$$\mathfrak{d}(x, y) = \varphi(x) - \varphi(y) - \varphi'(y)(x-y) = \mathrm{Exp}_{\mathfrak{s}}(y|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)$$

as required.

Proof of Theorem I.B.4(2). Suppose $\mathfrak{d} : [0,1] \times [0,1] \to [0, \infty]$ is a one-dimensional Bregman divergence. Then let $\mathfrak{s}(i,x) := \mathfrak{d}(i,x)$. Then, by the calculation above, if $x,y \in [0,1]$

$$\mathfrak{d}(x, y) = \mathrm{Exp}_{\mathfrak{s}}(y|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)$$

Considered as a function of y, $\mathrm{Exp}_{\mathfrak{s}}(y|x)$ is minimized when $\mathfrak{d}(x,y)$ is minimized; and, considered as a function of y, $\mathfrak{d}(x,y)$ is minimized uniquely at $x = y$, since $\mathfrak{d}$ is a divergence. So $\mathfrak{s}$ is a strictly proper scoring rule.

Proof of Theorem I.B.4(3). Suppose $\mathfrak{d} : [0,1] \times [0,1] \to [0, \infty]$ is a one-dimensional divergence. Then let $\mathfrak{s}(i, x) := \mathfrak{d}(i, x)$. Now suppose that, for all $x,y \in [0,1]$,

$$\mathfrak{d}(x, y) = \mathrm{Exp}_{\mathfrak{s}}(y|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)$$

Then, again, considered as a function of y, $\mathrm{Exp}_{\mathfrak{s}}(y|x)$ is minimized when $\mathfrak{d}(x,y)$ is minimized; and, considered as a function of y, $\mathfrak{d}(x, y)$ is minimized uniquely at $x = y$, since $\mathfrak{d}$ is a divergence. So ?? is a strictly proper scoring rule.

This completes the proof. □

At this point, we are ready to prove Theorem 4.3.3.[3]

Theorem 4.3.3 *Suppose Alethic Vindication, Perfectionism, Divergence Additivity, Divergence Continuity, and Decomposition. Then, if $\mathfrak{I}$ is a legitimate inaccuracy measure, there is an additive Bregman divergence $\mathfrak{D}$ such that $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$.*

Proof of Theorem 4.3.3. Suppose $\mathfrak{I}$ is an inaccuracy measure. By Perfectionism and Vindication, there is a divergence $\mathfrak{D}$ such that, if c is defined on $\mathscr{F}$ and $|\mathscr{F}| = n$,

$$I(c, w) = \mathfrak{D}(v_w, c)$$

By Additivity, there is a one-dimensional divergence $\mathfrak{d} : [0,1] \times [0,1] \to [0, \infty]$ such that, for all $\mathbf{x}, \mathbf{y} \in [0,1]^n$,

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

$$\mathrm{D}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathrm{d}(x_i, y_i)$$

So,

$$\mathrm{I}(c, w) = \sum_{X \in \mathrm{F}} \mathrm{d}(v_w(X), c(X))$$

By Continuity, $\mathfrak{d}$ is continuous in its first and second argument.

(p.89) Thus, to prove our theorem, it suffices to show that $\mathfrak{d}$ is a one-dimensional Bregman divergence. And, by Theorem I.B.4(3), it therefore suffices to show that, if we let $\mathfrak{s}(i,x) := \mathfrak{d}(i,x)$, then, for all $x,y \in [0,1]$

$$\mathrm{d}(x, y) = \mathrm{Exp}_{\mathrm{s}}(y|x) - \mathrm{Exp}_{\mathrm{s}}(x|x)$$

Thus, suppose $x,y \in [0,1]$. Then choose a rational number $q = \frac{m}{n}$ that is arbitrarily close to x. Now consider a set $\mathscr{F}$ containing n propositions and a world w in $\mathscr{W}_{\mathscr{F}}$ at which m of those propositions are true. And let $c(X) = y$, for all X in $\mathscr{F}$. So $c^w(X) = \frac{m}{n} = q$, for all $X \in \mathscr{F}$. Then, by Decomposition,

$$I(c, w) = \alpha \mathrm{D}(c^w, c) + \beta \mathrm{J}(c^w, w)$$

That is,

$$\mathrm{D}(v_w, c) = \alpha \mathrm{D}(v_w, c^w) + \beta \mathrm{D}(c^w, c)$$

Now:

$$\begin{aligned} \mathrm{D}(v_w, c) &= \sum_{X \in \mathrm{F}} \mathrm{d}(v_w(X), c(X)) \\ &= m\mathrm{d}(1, y) + (n-m)\mathrm{d}(0, y) \\ &= n[q\mathrm{d}(1, y) + (1-q)\mathrm{d}(0, y)] \\ &= n\mathrm{Exp}_{\mathrm{s}}(y|q) \end{aligned}$$

And

$$\begin{aligned} \mathrm{D}(c^w, c) &= \sum_{X \in \mathrm{F}} \mathrm{d}(c^w(X), c(X)) \\ &= n\mathrm{d}(q, y) \end{aligned}$$

And

$$\begin{aligned} \mathrm{D}(c^w, c) &= \sum_{X \in \mathrm{F}} \mathrm{d}(v_w(X), c^w(X)) \\ &= m\mathrm{d}(1, q) + (n-m)\mathrm{d}(0, q) \\ &= n[q\mathrm{d}(1, q) + (1-q)\mathrm{d}(0, q)] \\ &= n\mathrm{Exp}_{\mathrm{s}}(q|q) \end{aligned}$$

Thus,

$$\beta \mathrm{d}(q, y) = \mathrm{Exp}_{\mathrm{s}}(y|q) - \alpha \mathrm{Exp}_{\mathrm{s}}(q|q)$$

Now, suppose $x = q = 1$ and $y \neq 1$. Then

$$\beta \mathrm{d}(1, y) = \mathrm{Exp}_{\mathrm{s}}(y|1) - \alpha \mathrm{Exp}_{\mathrm{s}}(1|1)$$

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

(p.90) then, since $\mathrm{Exp}_{\mathfrak{s}}(y|1) = \mathfrak{d}(1,y)$ and $\mathrm{Exps}(1|1) = \mathfrak{d}(1, 1) = 0$ (since $\mathfrak{d}$ is a divergence), we have $\mathfrak{d}(1,y) = \beta\mathfrak{d}(1,y)$. Thus, since $y \neq 1$, $\beta = 1$. Next, let $x = q = y = \frac{1}{2}$. Then

$$\mathrm{d}\left(\frac{1}{2}, \frac{1}{2}\right) = \mathrm{Exp}_{\mathrm{s}}\left(\frac{1}{2}\middle|\frac{1}{2}\right) - \alpha\mathrm{Exp}_{\mathrm{s}}\left(\frac{1}{2}\middle|\frac{1}{2}\right)$$

Then, since $\mathfrak{d}(1/2, 1/2) = 0$, we have $\alpha = 1$. Finally, since $\mathfrak{d}$ is continuous in its first argument,

$$\mathrm{d}(x, y) = \mathrm{Exp}_{\mathrm{s}}(y|x) - \mathrm{Exp}_{\mathrm{s}}(x|x)$$

as required. □

I.C Characterizing legitimate inaccuracy measures (with Symmetry)

In this section, we show that we can characterize the positive linear transformations of the Brier score if we add Symmetry to the other conditions considered in Chapter 4.

Theorem I.C.1 *Suppose $\mathfrak{D}$ is an additive Bregman divergence. And suppose that, for all $\mathbf{x}, \mathbf{y} \in [0, 1]^n$*

$$\mathrm{D}(\mathbf{x}, \mathbf{y}) = \mathrm{D}(\mathbf{y}, \mathbf{x})$$

Then $\mathrm{D}(\mathbf{x}, \mathbf{y}) = \alpha\sum_{i=1}^{n}(x_i - y_i)^2 = \|\mathbf{x} - \mathbf{y}\|_2^2$ *for some* $\alpha > 0$.

That is, the only symmetric additive Bregman divergence is squared Euclidean distance (up to positive linear transformation).

Proof. This proof adapts a proof of a similar fact in (Selten, !998). Suppose $\mathfrak{D}(\mathbf{x}, \mathbf{y}) = \mathfrak{D}(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in [0, 1]^n$. Then $\mathfrak{d}(x, y) = \mathfrak{d}(y, x)$ for all $x, y \in [0, 1]$. By Theorem I.B.4, if $\mathfrak{s}(i, x) := \mathfrak{s}(i, x)$, then

$$\mathrm{d}(x, y) = \mathrm{Exp}_{\mathrm{s}}(y|x) - \mathrm{Exp}_{\mathrm{s}}(x|x)$$

Thus, for all $x, y \in [0, 1]$,

$$\mathrm{Exp}_{\mathrm{s}}(y|x) - \mathrm{Exp}_{\mathrm{s}}(x|x) = \mathrm{Exp}_{\mathrm{s}}(x|y) - \mathrm{Exp}_{\mathrm{s}}(y|y)$$

In particular,

(i) $\mathrm{Exp}_{\mathfrak{s}}(1|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x) = \mathrm{Exp}_{\mathfrak{s}}(x|1) - \mathrm{Exp}_{\mathfrak{s}}(1|1)$
(ii) $\mathrm{Exp}_{\mathfrak{s}}(0|x) - \mathrm{Exp}_{\mathfrak{s}}(x|x) = \mathrm{Exp}_{\mathfrak{s}}(x|0) - \mathrm{Exp}_{\mathfrak{s}}(0|0)$

Now, since $\mathfrak{d}(x, y) = \mathfrak{d}(y, x)$, we have $\mathfrak{d}(1, 0) = \mathfrak{d}(0, 1)$. Let $\alpha = \mathfrak{d}(1, 0) = \mathfrak{d}(0, 1)$. So (i) and (ii) give us:

(i′) $\alpha(1 - x) - \mathrm{Exp}_{\mathfrak{s}}(x|x) = \mathfrak{s}(\mathrm{l}, x)$
(ii′) $\alpha x - \mathrm{Exp}_{\mathfrak{s}}(x|x) = \mathfrak{s}(0,x)$

(p.91) Thus,

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

$$\begin{aligned}\mathrm{Exp}_{\mathfrak{s}}(x|x) &= x\mathfrak{s}(1,x) + (1-x)\mathfrak{s}(0,x)\\ &= x[\alpha(1-x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)] + (1-x)[\alpha x - \mathrm{Exp}_{\mathfrak{s}}(x|x)]\\ &= \alpha x(1-x) + \alpha(1-x)x - \mathrm{Exp}_{\mathfrak{s}}(x|x)\\ &= 2\alpha x(1-x) - \mathrm{Exp}_{\mathfrak{s}}(x|x)\end{aligned}$$

So

$$\mathrm{Exp}_{\mathfrak{s}}(x|x) = \alpha x(1-x)$$

Thus, substituting this into (i′) and (ii′) gives:

(i″) $\mathfrak{s}(1,x) = \alpha(1-x) - \alpha x(1-x) = \alpha(1-x)^2$
(ii″) $\mathfrak{s}(0,x) = \alpha x - \alpha x(1-x) = \alpha x^2$

Thus, $\mathfrak{d}(x,y) = (x-y)^2$ and $\mathrm{D}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}(x_i - y_i)^2 = \|\mathrm{x}-\mathrm{y}\|_2^2$. $\square$

As a corollary to this, we get Theorem 4.4.1.

I.D Two theorems concerning additive Bregman divergences

In this section, we prove two theorems concerning additive Bregman divergences. These will prove crucial throughout the remainder of the book. In the context of this first part of the book, they are the key lemmas in the proof of Theorem 4.3.4.

In order to state them in the generality we will need later in the book, we must introduce some standard terminology from topology.

Definition I.D.1 (**Open ball**) *Given* $\mathbf{x} \in \mathbb{R}^n$ *and* $\varepsilon > 0$, *we let*

$$B_\varepsilon(\mathbf{x}) = \{\mathbf{y} \in R^n : \|\mathbf{x}-\mathbf{y}\|_2 < \varepsilon\}$$

$B_\varepsilon(\mathbf{x})$ *is called the* open ball with centre $\mathbf{x}$ and radius ε.

Definition I.D.2 (**Open and closed sets**) *Suppose* $\mathscr{X} \subseteq \mathbb{R}^n$. *Then*:

- $\mathscr{X}$ is an open set *if, for all* $\mathbf{x} \in \mathscr{X}$, *there is* $\varepsilon > 0$ *such that* $B_\varepsilon(\mathbf{x}) \subseteq \mathscr{X}$. (*For example, any open ball is open.*)
- $\mathbf{x}$ is a limit point of $\mathscr{X}$ *if, for all* $\varepsilon > 0$, $B_\varepsilon(\mathbf{x}) \cap \mathscr{X} \neq \varnothing$. *Equivalently,* $\mathbf{x}$ *is a limit point of* $\mathscr{X}$ *if there is a sequence* $\langle \mathbf{x}_i \rangle_{i=1}^{\infty}$ *of elements of* $\mathscr{X}$ *and* $\lim_{i\to\infty} \mathbf{x}_i = \mathbf{x}$.
- $\mathscr{X}$ *is a closed set if* $\mathscr{X}$ *contains all limits points of* $\mathscr{X}$.
- The closure of $\mathscr{X}$ (*denoted* $\mathrm{cl}(\mathscr{X})$) *is the smallest closed set that contains* $\mathscr{X}$. *If* $\mathscr{X}$ *is closed, then* $\mathrm{cl}(\mathscr{X}) = \mathscr{X}$.

Proposition I.D.3 *If* $\mathscr{X}$ *is finite,* $\mathscr{X}^+$ *is closed.*

(p.92) **Proposition I.D.4** *If* $\mathscr{X}$ *is convex*, $\mathrm{cl}(X)$ *is convex.*

Thus, if $\mathscr{X}$ is finite, $\mathrm{cl}(\mathscr{X}^+) = \mathscr{X}^+$.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

Theorem I.D.5 *Suppose $\mathfrak{D}$ is an additive Bregman divergence. And suppose $\mathscr{X} \subseteq [0,1]^n$. Then, if $\mathbf{z} \notin \mathrm{cl}(\mathscr{X}^+)$, then there is $\mathbf{z}^* \in \mathscr{X}^+$ such that, for all $\mathbf{x} \in \mathscr{X}$, $\mathfrak{D}(\mathbf{x}, \mathbf{z}^*) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$.*

Proof of Theorem I.D.5. We begin by proving the following lemma, which is a slight weakening of Theorem I.D.5. Then we use continuity considerations to strengthen it to give the full strength of Theorem I.D.5.

Lemma I.D.6 *Suppose $\mathfrak{D}$ is an additive Bregman divergence. And suppose $\mathscr{X} \subseteq [0,1]^n$. Then, if $\mathbf{z} \notin \mathrm{cl}(\mathscr{X}^+)$, then there is $\pi_{\mathbf{z}} \in \mathrm{cl}(\mathscr{X}^+)$ such that, for all $\mathbf{x} \in \mathscr{X}$, $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$.*

Proof of Lemma I.D.6. Suppose $\mathbf{z} \notin \mathrm{cl}(\mathscr{X}^+)$. We consider two cases separately: (A) $\nabla\Phi(\mathbf{z})$ is finite; and (B) $\nabla\Phi(\mathbf{z})$ is infinite.[4] Our proof relies heavily on (Predd et al. 2009).

CASE (A): $\nabla\,\Phi(\mathbf{z})$ is finite. Consider $\mathfrak{D}(\mathbf{x}, \mathbf{z})$ as a function of $\mathbf{x}$ on cl($\mathscr{X}^+$): it is bounded, continuous, and strictly convex. Thus, since cl($\mathscr{X}^+$) is convex and closed, there is a unique $\pi_{\mathbf{z}} \in \mathscr{X}^+$ such that, for all $\mathbf{x} \in \mathrm{cl}(\mathscr{X}^+)$,

$$\mathrm{D}(\pi_{\mathbf{z}}, \mathbf{z}) \leq \mathrm{D}(\mathbf{x}, \mathbf{z})$$

(7.1)

Now we show that, for all $\mathbf{x} \in \mathrm{cl}(\mathscr{X}^+)$,

$$\mathrm{D}(\mathbf{x}, \pi_{\mathbf{z}}) \leq \mathrm{D}(\mathbf{x}, \mathbf{z})$$

Suppose $\mathbf{x} \in \mathrm{cl}(X^+)$.Then let $0 \leq \varepsilon \leq 1$. Then, since $\mathrm{cl}(\mathscr{X}^+)$ is convex, $(1-\varepsilon)\pi_{\mathbf{z}}+\varepsilon\mathbf{x} \in \mathrm{cl}(\mathscr{X}^+)$. Then, by (71), we have

$$\mathrm{D}(\pi_{\mathbf{z}}, \mathbf{z}) \leq \mathrm{D}((1-\varepsilon)\pi_{\mathbf{z}} + \varepsilon\mathbf{x}, \mathbf{z})$$

But $(1-\varepsilon)\pi_{\mathbf{z}} + \varepsilon\mathbf{x} = \pi_{\mathbf{z}} + \varepsilon(\mathbf{x} - \pi_{\mathbf{z}})$, so

$$0 \leq \mathrm{D}(\pi_{\mathbf{z}} + \varepsilon(\mathbf{x} - \pi_{\mathbf{z}}), \mathbf{z}) - \mathrm{D}(\pi_{\mathbf{z}}, \mathbf{z})$$

(7.2)

But

$$\begin{aligned} &\mathrm{D}(\pi_{\mathbf{z}} + \varepsilon(\mathbf{x} - \pi_{\mathbf{z}}), \mathbf{z}) - \mathrm{D}(\pi_{\mathbf{z}}, \mathbf{z}) \\ &= \Phi(\pi_{\mathbf{z}} + \varepsilon(\mathbf{x} - \pi_{\mathbf{z}})) - \Phi(\pi_{\mathbf{z}}) - \nabla\,\Phi(\mathbf{z})\varepsilon(\mathbf{x} - \pi_{\mathbf{z}}) \end{aligned}$$

(p.93) So

$$\lim_{\varepsilon \to 0} \frac{\mathrm{D}(\pi_{\mathbf{z}} + \varepsilon(\mathbf{x} - \pi_{\mathbf{z}}), \mathbf{z}) - \mathrm{D}(\pi_{\mathbf{z}}, \mathbf{z})}{\varepsilon} = (\nabla\,\Phi(\pi_{\mathbf{z}}) - \nabla\,\Phi(\mathbf{z})) \cdot (\mathbf{x} - \pi_{\mathbf{z}})$$

Thus, by (7.2),

$$(\nabla\,\Phi(\pi_{\mathbf{z}}) - \nabla\,\Phi(\mathbf{z})) \cdot (\mathbf{x} - \pi_{\mathbf{z}}) \geq 0$$

But we can also show that

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

$$\mathfrak{D}(\mathbf{x}, \mathbf{z}) - \mathfrak{D}(\pi_{\mathbf{z}}, \mathbf{z}) - \mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) = (\nabla\,\Phi\,(\pi_{\mathbf{z}}) - \nabla\,\Phi\,(\mathbf{z}))\cdot(\mathbf{x} - \pi_{\mathbf{z}})$$

So

$$\mathfrak{D}(\mathbf{x}, \mathbf{z}) \geq \mathfrak{D}(\pi_{\mathbf{z}}, \mathbf{z}) + \mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}})$$

But, since $\mathbf{z} \notin \mathrm{cl}(\mathscr{X}^+)$ and $\pi_{\mathbf{z}} \in \mathrm{cl}(\mathscr{X}^+)$, $\mathbf{z} \neq \pi_{\mathbf{z}}$. Thus, since $\mathfrak{D}$ is a divergence, $\mathfrak{D}(\pi_{\mathbf{z}}, \mathbf{z}) > 0$. So

$$\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) > \mathfrak{D}(\mathbf{x}, \mathbf{z})$$

as required.

CASE (B): $\nabla\Phi(\mathbf{z})$ is infinite. It follows that $\mathbf{z}$ lies on the boundary of $[0, 1]^n$. Thus, it lies on an $(n-1)$-dimensional face of $[0, 1]^n$. Let $\mathscr{T}$ be the lowest dimensional face of $[0, 1]^n$ on which $\mathbf{z}$ lies. There are three cases:

- If $\mathscr{T} \cap \mathscr{X} = \emptyset$, then let $\pi_{\mathbf{z}}$ be any element in $\mathrm{int}(\mathrm{cl}(\mathscr{X}^+))$. Then, since $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) < \infty$ and $\mathfrak{D}(\mathbf{x}, \mathbf{z}) = \infty$ for all $\mathbf{x} \in \mathscr{X}$, we have $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$ for all $\mathbf{x} \in \mathscr{X}$.
- If $\mathscr{T} \cap \mathscr{X} \neq \emptyset$ and $\dim(\mathscr{T}) > 0$, then $\mathbf{z} \in \mathrm{int}(\mathscr{T})$: if $\mathbf{z} \notin \mathrm{int}(\mathscr{T})$, then there is a lower-dimensional face $\mathscr{T}'$ of $[0, 1]^n$ such that $\mathbf{z} \in \mathscr{T}'$. Now, for all $\mathbf{x} \in \mathscr{T} \cap \mathscr{X}$, $\mathfrak{D}(\mathbf{x}, \notin\mathbf{z}) < \infty$. Thus, by the first part of the proof, there is $\pi_{\mathbf{z}} \in \mathscr{T} \cap \mathrm{cl}(\mathscr{X}^+)$ such that $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) < \mathfrak{D}(\mathbf{x}, \mathbf{z}) < \infty$ for all $\mathbf{x} \in \mathscr{T} \cap \mathscr{X}$. However, we still have $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}) = \infty$ for $\mathbf{x} \in \mathscr{T} - \mathscr{X}$. So let
$$\pi_{\mathbf{z}}^{\varepsilon} := (1-\varepsilon)\mathbf{z}' + \frac{\varepsilon}{|\mathrm{X} - \mathrm{T}|}\sum_{\mathbf{x}\in \mathrm{X}-\mathrm{T}} \mathbf{x}$$
Then $\pi_{\mathbf{z}}^{\varepsilon} \in \mathrm{int}([0, 1]^n)$, so $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}^{\varepsilon}) < \infty = \mathfrak{D}(\mathbf{x}, \mathbf{z})$ for $\mathbf{x} \in \mathscr{X} - \mathscr{T}$. Moreover, we can choose ε small enough so that $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}^{\varepsilon}) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$ for all $\mathbf{x} \in \mathscr{X} \cap \mathscr{T}$, as required.
- If $\mathscr{T} \cap \mathscr{X} \neq \emptyset$ and $\dim(\mathscr{T}) = 0$, then $\mathbf{z} \in \mathscr{X}$.

This completes the proof of Lemma I.D.6.
□

We now wish to strengthen this result by showing that there is $\mathbf{z}^*$ in $\mathscr{X}^+$ such that $\mathfrak{D}(\mathbf{x}, \mathbf{z}^*) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$ for all $\mathbf{x} \in \mathscr{X}$.

CASES (A): Since there is a positive lower bound on the difference between $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}})$ and $\mathfrak{D}(\mathbf{x}, \mathbf{z})$, and since $\mathfrak{D}$ is continuous in its second argument, there is $\mathbf{z}^* \in \mathrm{int}(\mathrm{cl}(\mathscr{X}^+))$ such that $\mathfrak{D}(\mathbf{x}, \mathbf{z}^*) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$. It follows that $\mathbf{z}^* \in \mathscr{X}^+$.

(p.94) CASE (B): Since there is a positive lower bound on the difference between $\mathfrak{D}(\mathbf{x}, \pi_{\mathbf{z}}^{\varepsilon})$ and $\mathfrak{D}(\mathbf{x}, \mathbf{z})$, and since $\mathfrak{D}$ is continuous in

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

its second argument, there is $\mathbf{z}^* \in \text{int}(\text{cl}(\mathscr{X}^+))$ such that $\mathfrak{D}(\mathbf{x}, \mathbf{z}^*) < \mathfrak{D}(\mathbf{x}, \mathbf{z})$. It follows that $\mathbf{z}^* \in \mathscr{X}^+$.

This completes the proof of Theorem I.D.
5. □

Theorem I.D.7 *Suppose $\mathfrak{D}$ is an additive Bregman divergence. And suppose $\mathscr{X} \subseteq [0,1]^n$. Then, if $\mathbf{z} \in \mathscr{X}^+$ then $\mathbf{z}$ is a convex combination of finitely many elements of $\mathscr{X}$. Thus, there are $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathscr{X}$ and $\lambda_1, \ldots, \lambda_n \geq 0$ with $\sum_{i=1}^n \lambda_i = 1$ such that $\mathbf{z} = \sum_{i=1}^n \lambda_i \mathbf{x}_i$. Then, if $\mathbf{y}, \mathbf{y}' \in [0,1]^n$,*

$$\sum_{i=1}^n \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y}) < \sum_{i=1}^n \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y}') \Longleftrightarrow \mathrm{D}(\mathbf{z}, \mathbf{y}) < \mathrm{D}(\mathbf{z}, \mathbf{y}')$$

Proof of Theorem I.D.7 Suppose $\mathbf{z} \in \mathscr{X}^+$. Note that the set of convex combinations of finitely many elements of $\mathscr{X}$ is convex and contains all elements of $\mathscr{X}$. Thus, $\mathscr{X}^+$ is the set of convex combinations of finitely many elements of $\mathscr{X}$. Thus, since $\mathbf{z} \in \mathscr{X}^+$, there are $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathscr{X}$ and $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\sum_{i=1}^n \lambda_i = 1$ and $\mathbf{z} = \sum_{i=1}^n \lambda_i \mathbf{x}_i$. Then suppose $\mathbf{y}, \mathbf{y}' \in [0,1]^n$. Then

$$\begin{aligned}
&\sum_{i=1}^n \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y}) - \sum_{i=1}^n \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y}') \\
&= \sum_{i=1}^n \lambda_i [\mathrm{D}(\mathbf{x}_i, \mathbf{y}) - \mathrm{D}(\mathbf{x}_i, \mathbf{y}')] \\
&= \sum_{i=1}^n \lambda_i \left[(\Phi(\mathbf{x}_i) - \Phi(\mathbf{y}) - (\mathbf{x}_i - \mathbf{y}) \cdot \nabla \Phi(\mathbf{y})) - (\Phi(\mathbf{x}_i) - \Phi(\mathbf{y}') - (\mathbf{x}_i - \mathbf{y}') \cdot \nabla \Phi(\mathbf{y}'))\right] \\
&= \sum_{i=1}^n \lambda_i \left[\Phi(\mathbf{y}') - \Phi(\mathbf{y}) - (\mathbf{x}_i - \mathbf{y}) \cdot \nabla \Phi(\mathbf{y}) + (\mathbf{x}_i - \mathbf{y}') \cdot \nabla \Phi(\mathbf{y}')\right] \\
&= \Phi(\mathbf{y}') - \Phi(\mathbf{y}) - \sum_{i=1}^n \lambda_i (\mathbf{x}_i - \mathbf{y}) \cdot \nabla \Phi(\mathbf{y}) + \sum_{i=1}^n \lambda_i (\mathbf{x}_i - \mathbf{y}') \cdot \nabla \Phi(\mathbf{y}') \\
&= \Phi(\mathbf{y}') - \Phi(\mathbf{y}) - \left(\sum_{i=1}^n \lambda_i \mathbf{x}_i - \mathbf{y}\right) \cdot \nabla \Phi(\mathbf{y}) + \left(\sum_{i=1}^n \lambda_i \mathbf{x}_i - \mathbf{y}'\right) \cdot \nabla \Phi(\mathbf{y}') \\
&= \Phi(\mathbf{y}') - \Phi(\mathbf{y}) - (\mathbf{z} - \mathbf{y}) \cdot \nabla \Phi(\mathbf{y}) + (\mathbf{z} - \mathbf{y}') \cdot \nabla \Phi(\mathbf{z}) \\
&= [\Phi(\mathbf{z}) - \Phi(\mathbf{y}) - (\mathbf{z} - \mathbf{y}) \cdot \nabla \Phi(\mathbf{y})] - [\Phi(\mathbf{z}) - \Phi(\mathbf{y}') - (\mathbf{z} - \mathbf{y}') \cdot \nabla \Phi(\mathbf{y}')] \\
&= \mathrm{D}(\mathbf{z}, \mathbf{y}) - \mathrm{D}(\mathbf{z}, \mathbf{y}')
\end{aligned}$$

(p.95) Thus,

$$\sum_{i=1}^n \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y}) < \sum_{i=1}^n \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y}') \Longleftrightarrow \mathrm{D}(\mathbf{z}, \mathbf{y}) < \mathrm{D}(\mathbf{z}, \mathbf{y}')$$

as required.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

This completes the proof of Theorem I.D.
7. □

As a corollary of Theorem I.D.7, we have:

Theorem I.D.8 *Suppose* $\mathfrak{D}$ *is an additive Bregman divergence. And suppose* $\mathcal{X} \subseteq [0, 1]^n$. *Then, if* $\mathbf{z} \in \mathcal{X}^+$, *then* $\mathbf{z}$ *is a convex combination of finitely many elements of* $\mathcal{X}$. *Thus, there are* $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$ *and* $\lambda_1, \ldots, \lambda_n \geq 0$ with $\sum_{i=1}^n \lambda_i = 1$ *such that* $\mathbf{z} = \sum_{i=1}^n \lambda_i \mathbf{x}_i$. *Then, if* $\mathbf{y} \in [0, 1]^n$ *and* $\mathbf{y} \neq \mathbf{z}$,

$$\sum_{i=1}^{n} \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{z}) < \sum_{i=1}^{n} \lambda_i \mathrm{D}(\mathbf{x}_i, \mathbf{y})$$

Proof. If $\mathbf{y} \neq \mathbf{z}$, then $\mathfrak{D}(\mathbf{z}, \mathbf{z}) = 0 < \mathfrak{D}(\mathbf{z}, \mathbf{y})$. □

For our purposes in this book, the set $\mathcal{X}$ is the set of ideal credence functions and $\mathfrak{D}$ is one of the divergences that gives rise to a legitimate inaccuracy measure. Thus interpreted, together, Theorems I.D.5 and I.D.7 show that rationality requires an agent to have a credence function in the closure of the convex hull of the set of ideal credence functions.

- Theorem I.D.5 says that any credence function outside this set is accuracy dominated by a credence function in the convex hull.
- Theorem I.D.8 then shows that, given any credence function inside the convex hull, it will be the case that any probabilistic extension of that credence function to $\mathcal{F}^*$ expects the credence function to be most accurate amongst credence functions on $\mathcal{F}$.

Theorem 4.3.4 is now a corollary of these two theorems where we let $\mathcal{X}$ be the set of omniscient credence functions $\mathcal{V}_{\mathcal{F}} = \{v_w : w \in \mathcal{W}_{\mathcal{F}}\}$.

Note that we do not require the set $\mathcal{X}$ to be finite. This will be important in Part II of this book, where we consider an argument in which the set of ideal credence functions is not necessarily finite. (p.96)

Notes:

([1]) A subset $\mathcal{Z} \subseteq \mathbb{R}^n$ is *convex* iff, whenever $z, z' \in \mathcal{Z}$, $\lambda z + (1 - \lambda)z' \in \mathcal{Z}$ for $0 \leq \lambda \leq 1$. Thus, if $\mathcal{Z}$ is convex and $\mathcal{X} \subseteq \mathcal{Z}$, then $\mathcal{X}^+ \subseteq \mathcal{Z}$.

([2]) As usual, for a function $\psi : \mathbf{R}^n \to \mathbf{R}$ and a vector $\mathbf{z} = (z_1, \ldots, z_n)$ in $\mathbf{R}^n$, we have

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017

$$\nabla \Psi(\mathrm{z}) = \left(\frac{\partial}{\partial z_1}\Psi(\mathrm{z}), \ldots, \frac{\partial}{\partial z_n}\Psi(\mathrm{z})\right)$$

providing each $\partial/\partial z_i \Psi(\mathbf{z})$ is defined. Also, for two vectors x, y in $\mathbf{R}^n$, we define their dot product x • y as follows:

$$\mathrm{x} \cdot \mathrm{y} = \sum_{i=1}^{n} x_i y_i$$

([3]) It is worth noting that this theorem is a sort of converse to Theorem 4 from (DeGroot & Fienberg 1983).

([4]) As noted above, $\nabla \Phi(\mathbf{z}) = \left(\frac{\partial}{\partial z_1}\Phi(\mathbf{z}), \ldots, \frac{\partial}{\partial z_n}\Phi(\mathbf{z})\right)$ is a vector. Thus, when we say that it is finite, we mean that each coordinate is finite; and when we say that it is infinite, we mean that some coordinate is infinite.

Access brought to you by:

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: The University of British Columbia Library; date: 24 December 2017