Accuracy representation results and estimates

Catrin Campbell-Moore*

December 2, 2024

This document goes through some standard results about strictly proper measures of accuracy and representation theorems (Schervish and Savage/Bregman). It also presents the slightly less well studied case of measuring the accuracy of estimates of random variables.

The spirit of the document is to include proofs, but to make things simple in order to make the central ideas of the proofs across, often at the cost of generality. A number of restrictive assumptions are made throughout.

Full analysis of what restrictions can be dropped in the estimates case requires further work.

Contents

*

Part I
Accuracy of a credence

How accurate is a credence value in a proposition, say 0.6, when the proposition is true? We give a measure.

I am not here considering the accuracy of an entire credence function at a world, but just of a single proposition.

1 Definitions

Setup 1.1. We give an accuracy measure to describe how accurate a credence is in a proposition when it is true/false. Formally, we have two accuracy measures,

$$\mathfrak{a}_1 : [0,1] \to \text{Re} \tag{1}$$
$$\mathfrak{a}_0 : [0,1] \to \text{Re} \tag{2}$$

△

Remark (Infinite accuracy)**.** One can often allow infinite values at end-points. In particular, one can allow infinite inaccuracy at the maximally far-away points (this assumes that credences can only take values in $[0,1]$, if credences can take values in Re, then we cannot have infinite values and keep truth-directendess — we can always get worse.) See discussion about infinity, and various other assumptions and their relationships in Schervish et al. (2009). △

Definition 1.2. $\mathfrak{a}$ is *(strictly) proper* iff for any $p \in [0,1]$,

$$\text{Exp}_p\mathfrak{a}(x) := p\mathfrak{a}_1(x) + (1-p)\mathfrak{a}_0(x) \tag{3}$$

obtains a (unique) maximum at $x = p$. △

Definition 1.3. $\mathfrak{a}$ is *(strictly) truth-directed* iff If $v < x < y$ or $y < x < v$ then $\mathfrak{a}_v(x) > \mathfrak{a}_v(y)$ △

Proposition 1.4. *(Strict) propriety entails (strict) truth-directedness.*

This is Schervish (1989, Lemma A1). I include a proof in appendix A. I leave this outside the main body of the paper because truth directedness is incredibly plausible, and certainly more plausible than propriety as a constraint on measurements of accuracy. (Note that this is different when one is interested in elicitation directly rather than, as philosophers usually are, measurements of the epistemic value of credences.)

Remark. Sometimes it would be nicer to think directly about a loss function, $\mathfrak{s}_v$, with

$$\mathfrak{s}_v(x) = \mathfrak{a}_v(v) - \mathfrak{a}_v(x) \tag{4}$$

$\mathfrak{s}_v(x)$ measures the difference between the accuracy of perfection and the accuracy of the given credence. Note that this picture requires $\mathfrak{s}_v(v) = 0$.

This is sometimes called a "scoring rule", although that terminology is also simply used for inaccuracy, i.e., negative accuracy.

By giving a strictly proper measure $\mathfrak{s}_v$, one can arbitrarily choose values for self-accuracy $\mathfrak{a}_v(v)$ to obtain a strictly proper accuracy measure by

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \mathfrak{s}_v(x) \tag{5}$$

The representations are actually really directly characterising $\mathfrak{s}$. We can talk about strict propriety etc directly of $\mathfrak{s}$. This is actually more commonly done in the literature.

The literature such as Pettigrew (2016) works with *inaccuracy*, but I work with accuracy because it more closely ties with the philosophical presentation of trying to maximise the good of having accurate credences. Say something about inaccuracy vs scoring rules vs loss functions... △

We present two representation results for accuracy measures.

2 Schervish

2.1 Schervish form

The central result Schervish (1989, Theorem 4.2)

Theorem 2.1. $\mathfrak{a}$ *is (strictly) proper iff there is some measure* λ *(and values* $\mathfrak{a}_v(v)$*) such that for every* $x \in [0, 1]$,

$$\mathfrak{a}_0(x) = \mathfrak{a}_0(0) - \int_0^x t \;\; \lambda(\mathrm{d}t) \tag{6}$$

$$\mathfrak{a}_1(x) = \mathfrak{a}_1(1) - \int_x^1 1 - x \; \lambda(\mathrm{d}x) \tag{7}$$

(for strictness, it should assign positive value to each interval)

Setup 2.2. When $a > b$ define the integral

$$\int_a^b f(x)\,\mathrm{d}x = -\int_b^a f(x)\,\mathrm{d}x$$

(i.e., if it's "wrong-way-around" integration limits, just take negative).

Note then we can re-describe the Schervish form as:

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v v - t \;\; \lambda(\mathrm{d}t). \tag{8}$$

i.e.,

$$\mathfrak{s}_v(x) = \int_x^v v - t \;\; \lambda(\mathrm{d}t). \tag{9}$$

where $\mathfrak{s}_v(x) := \mathfrak{a}_v(v) - \mathfrak{a}_v(x)$. (if $x < v$, the switching limits and absolute value signs cancel out) △

Lemma 2.3. *A useful fact, then, is*

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v v - t \;\; \lambda(\mathrm{d}t). \tag{10}$$

Remark. If working with inaccuracy, or the scoring rule, the signs are cleanest writing it as

$$\mathfrak{s}_v(x) = \int_v^x t - v \;\; \lambda(\mathrm{d}t). \tag{11}$$

△

2.2 Any such $\mathfrak{a}$ is proper

Lemma 2.4. *The following are equivalent:*

1. *Schervish form: for $v \in \{0, 1\}$ and any $x \in [0, 1]$,*

$$\mathfrak{a}_v(v) - \mathfrak{a}_v(x) = \int_x^v v - t \;\; \lambda(\mathrm{d}t). \qquad \textit{eq. } (10)$$

2. *For all $x, y \in [0, 1]$, (i.e., replacing $v \in \{0, 1\}$ by a general $y \in [0, 1]$)*

$$\mathfrak{a}_v(y) - \mathfrak{a}_v(x) = \int_x^y v - t \;\; \lambda(\mathrm{d}t). \tag{12}$$

3. *For all $x, p \in [0, 1]$,*

$$\mathrm{Exp}_p \mathfrak{a}(p) - \mathrm{Exp}_p \mathfrak{a}(x) = \int_x^p (p - t)\, \lambda(\mathrm{d}t) \tag{13}$$

Proof Sketch. These follow from quite simple manipulations. To obtain item 1, or item 2 from item 3, note that $\mathrm{Exp}_v \mathfrak{a}(x) = \mathfrak{a}_v(x)$. A full proof is included in appendix B. □

Proposition 2.5. *If $\mathfrak{a}$ has Schervish form it is (strictly) proper.*

Proof. Suppose $x < p$. Then for any $t \in [x, p]$, $p - t > 0$, so $\int_x^p (p-t)\lambda(\mathrm{d}t) > 0$.

Suppose $x > p$. Then for any $t \in [x, p]$, $p - t < 0$, so $\int_p^x (p-t)\lambda(\mathrm{d}t) < 0$. But, $\mathrm{Exp}_p \mathfrak{a}(p) - \mathrm{Exp}_p \mathfrak{a}(x)$ switches the integral bounds, i.e., involves $\int_x^p$, which is then positive by specification of wrong-way-around integrals. □

2.3 Schervish's representation result

We prove it simply for the absolutely continuous case in order to keep the proof easy to follow. The general result holds (Schervish, 1989, Theorem 4.2)

Proposition 2.6. *If $\mathfrak{a}$ is strictly proper and absolutely continuous, then there is a positive function m with*

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v (v-t)m(t)\,\mathrm{d}t \tag{14}$$

Proof. For absolutely continuous $\mathfrak{a}$, by the fundamental theorem of calculus

$$\mathfrak{a}_v(v) - \mathfrak{a}_v(x) = \int_x^v \mathfrak{a}_v'(t)\,\mathrm{d}t \tag{15}$$

By propriety, $\mathrm{Exp}_t \mathfrak{a}(s)$ has a maximum at $s = t$, so the derivative at this point is 0.

$$t\mathfrak{a}_1'(t) + (1-t)\mathfrak{a}_0'(t) = 0 \tag{16}$$

By manipulating eq. (16)

$$\frac{\mathfrak{a}_0'(t)}{-t} = \frac{\mathfrak{a}_1'(t)}{1-t} \tag{17}$$

Define the function m by $m(t) = {}^{\mathfrak{a}_0'(t)}/_{-t}$. So $\mathfrak{a}_0'(t) = -tm(t)$ and $\mathfrak{a}_1'(t) = (1-t)m(t)$. So, by replacing these in eq. (15), we obtain eq. (14).

By proposition 1.4, $\mathfrak{a}$ is strictly truth-directed, so $\mathfrak{a}_0'(t) < 0$ and $\mathfrak{a}_1'(t) > 0$. Thus, m is positive. □

Remark. When it is not absolutely continuous we can obtain a representation of the form:

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v (v-t)\ \mathrm{d}\lambda(t) \tag{18}$$

we just can't push the measure λ into a mass function. The proof of this is (Schervish, 1989, Theorem 4.2) and instead takes the Radon Nikodym derivatives of $\mathfrak{a}_0$ and $\mathfrak{a}_1$ relative to $\mathfrak{a}_1 - \mathfrak{a}_0$.

Schervish also shows that the finiteness assumptions can be relaxed, △

3 Bregman divergences

3.1 Entropy and Bregman Divergence

Definition 3.1. Define the *entropy* of $\mathfrak{a}$ as:

$$\varphi_{\mathfrak{a}}(p) := \operatorname{Exp}_p \mathfrak{a}(p) = p\mathfrak{a}_1(p) + (1-p)\mathfrak{a}_0(p) \tag{19}$$

△

Proposition 3.2. *If $\mathfrak{a}$ is proper, then $\varphi_{\mathfrak{a}}$ is convex and if it is differentiable, then:*

$$\operatorname{Exp}_p \mathfrak{a}(p) - \operatorname{Exp}_p \mathfrak{a}(x) = \varphi_{\mathfrak{a}}(p) - \varphi_{\mathfrak{a}}(x) - (p-x)\varphi'_{\mathfrak{a}}(x) \tag{20}$$

If it is not differentiable, then we have the same form, but with $\varphi'_{\mathfrak{a}}$ as some sub-gradient.

Also for $v \in \{0,1\}$,

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - (\varphi_{\mathfrak{a}}(v) - \varphi_{\mathfrak{a}}(x) - (v-x)\varphi'_{\mathfrak{a}}(x)) \tag{21}$$

And in fact

$$\mathfrak{a}_v(x) = \varphi_{\mathfrak{a}}(x) + (v-x)\varphi'_{\mathfrak{a}}(x) \tag{22}$$

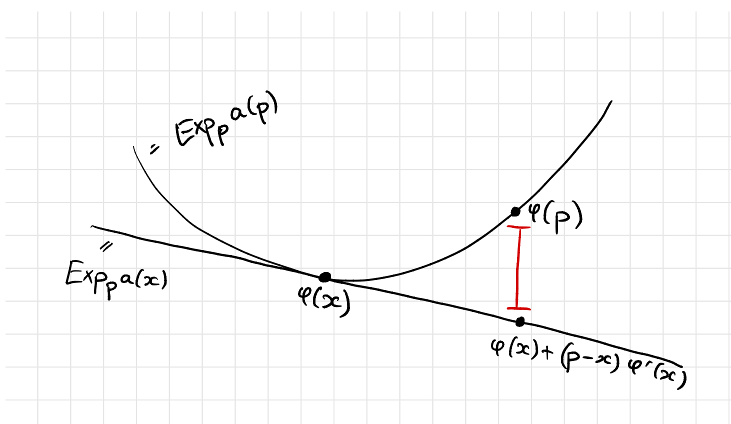

Figure 1: Divergence diagram

Proof. By strict propriety, $\operatorname{Exp}_p \mathfrak{a}(x) < \operatorname{Exp}_p \mathfrak{a}(p) = \varphi_{\mathfrak{a}}(p)$. And

$$\operatorname{Exp}_p \mathfrak{a}(x) = p\mathfrak{a}_1(x) + (1-p)\mathfrak{a}_0(x) \tag{23}$$

is a linear function of p (we could name it, e.g., $f_x(p) = \mathrm{Exp}_p \mathfrak{a}(x)$). So we have a linear function entirely lying below $\varphi_{\mathfrak{a}}$ and touching it just at p. Therefore, $\varphi_{\mathfrak{a}}$ is convex, with $f_x(p) = \mathrm{Exp}_p \mathfrak{a}(x)$ a subtangent of it at x.

If $\varphi_{\mathfrak{a}}$ is differentiable at x, then the subtangent at x, which is equal to $\mathrm{Exp}_x \mathfrak{a}(p)$, is given by:

$$\mathrm{Exp}_p \mathfrak{a}(x) = \varphi_{\mathfrak{a}}(x) + (p - x)\varphi'_{\mathfrak{a}}(x) \tag{24}$$

and eq. (20). If $\varphi_{\mathfrak{a}}$ is not differentiable, then one can take the slope of $\mathrm{Exp}_p \mathfrak{a}(x)$ and observe it is a sub-gradient of $\varphi_{\mathfrak{a}}$ by propriety; that will play the role of $\varphi'_{\mathfrak{a}}$.

Equation (25) follows immediately, putting $p \in \{0, 1\}$ and observing that $\mathrm{Exp}_v \mathfrak{a}(v) = \mathfrak{a}_v(v)$ to get

$$\mathfrak{a}_v(v) - \mathfrak{a}_v(x) = (\varphi_{\mathfrak{a}}(v) - \varphi_{\mathfrak{a}}(x) - (v - x)\varphi'_{\mathfrak{a}}(x)) \tag{25}$$

and then rearranging

□

Definition 3.3. A *Bregman divergence* associated with a convex function φ is:

$$\mathfrak{d}(p, x) := \varphi(p) - \varphi(x) - (p - x)\varphi'(x) \tag{26}$$

△

So this tells us that $\mathrm{Exp}_p \mathfrak{a}(p) - \mathrm{Exp}_p \mathfrak{a}(x)$ is a Bregman divergence.

Corollary 3.4. *If $\mathfrak{a}$ is strictly proper, then*

$$\mathfrak{a}_v(x) = \varphi_{\mathfrak{a}}(x) + (v - x)\varphi'_{\mathfrak{a}}(x) \tag{27}$$

where $\varphi_{\mathfrak{a}}$ the entropy for $\mathfrak{a}$, i.e., as in eq. (19)*.*

Proof. $\mathfrak{a}_v(x) = \mathrm{Exp}_v \mathfrak{a}(x)$. And from eq. (20), using the fact that $\varphi(v) = \mathrm{Exp}_v \mathfrak{a}(v)$

$$\mathrm{Exp}_v \mathfrak{a}(x) = \varphi(x) + (v - x)\varphi'(x) \tag{28}$$

□

Remark. There is an alternative proof that goes directly via rearrangments of eq. (16) using the definition of entropy, but that proof doesn't directly show that it is convex. △

We also have the converse,

Proposition 3.5. *$\mathfrak{a}$ is strictly proper iff there is a convex function φ (with values $\mathfrak{a}_v(v)$) where:*

$$\mathfrak{a}_v(x) := \mathfrak{a}_v(v) - (\varphi(v) - \varphi(x) - (v - x)\varphi'(x)) \tag{29}$$

That is, the error-score is:

$$\mathfrak{s}_v(x) = \varphi(v) - \varphi(x) - (v - x)\varphi'(x) \tag{30}$$

4 Relationships between Bregman divergences and the Schervish form

Lemma 4.1. *For any twice-differentiable* φ*,*

$$\int_x^p (p-t)\varphi''(t)dt = \varphi(p) - \varphi(x) - (p-x)\varphi'(x) \tag{31}$$

Proof. Integration by parts tells us that $\int_x^p u(t)v'(t)\,\mathrm{d}t = [u(p)v(p) - u(x)v(x)] - \int_x^p v(t)u'(t)\,\mathrm{d}t$. We apply this with $u(t) := (p-t)$, and $v(t) := \varphi'(t)$, observing that $u' = -1$. So:

$$\int_x^p (p-t)\varphi''(t)\,\mathrm{d}t \tag{32}$$

$$= [(p-p)\varphi'(p) - (p-x)\varphi'(x)] - \int_x^p \varphi'(t) \times (-1)\,\mathrm{d}t \quad \text{Integration by parts} \tag{33}$$

$$= \int_x^p \varphi'(t)\,\mathrm{d}t - (p-x)\varphi'(x) \tag{34}$$

$$= \varphi(p) - \varphi(x) - (p-x)\varphi'(x) \tag{35}$$

□

We can also do this with a measure rather than the mass function when λ is a measure associated with the distribution function φ'.

Lemma 4.2. *For an accuracy measure, the* m *from Schervish and* φ *the entropy, we have:* $m(t) = \varphi''(t)$*.*

Proof.

$$\begin{aligned} \varphi'(x) &= \mathfrak{a}_1(x) - \mathfrak{a}_0(x) + x\mathfrak{a}_1'(x) + (1-x)\mathfrak{a}_0'(x) && \text{product rule} \quad (36) \\ &= \mathfrak{a}_1(x) - \mathfrak{a}_0(x) && \text{eq. (16)} \quad (37) \end{aligned}$$

And from eq. (16),

$$\mathfrak{a}_1'(x) - \mathfrak{a}_0'(x) = \frac{\mathfrak{a}_0'(x)}{-x} = m(x). \tag{38}$$

□

So $\varphi''(x) = m(x)$.

Part II
Estimates

5 Accuracy of Estimates

add more discussion

We want to consider not only credences, which are truth-value estimates, or evaluated as good or bad with their "closeness to the truth-value of 0/1", but also the accuracy of one's general estimates, or forecasts, for random variables more generally.

The random variable might, for example, be representing the utility of taking some action; or it might be the number of millimetres of rain next week. The agent will provide an estimated value. Lets say that she estimates the value to be 10, and its true value turns out to be 30. How accurate was her estimate? This is specified by an accuracy measure. $\mathfrak{a}_v : \mathrm{Re} \to \mathrm{Re}$ with $\mathfrak{a}_v(x)$ describing the accuracy of providing an estimated value of x for V when the true value is v. Observe that, like in the earlier setting, we are assuming that accuracy measures are finite.

Setup 5.1. We assume that Ω is finite.

A *random variable* is a function $V : \Omega \to \mathrm{Re}$.

A *probability function* is $p : \Omega \to [0, 1]$ with $\sum_{w \in \Omega} p(w) = 1$. (That is, we're presenting it as a probability mass function; fine since Ω is finite). △

Definition 5.2 (Expected accuracy)**.** For p probabilistic,

$$\mathrm{Exp}_p[\mathfrak{a}_{V(\cdot)}(x)] := \sum_{w \in \Omega} p(\omega)\mathfrak{a}_{V(\omega)}(x) \tag{39}$$

$$\mathrm{Exp}_p[V] := \sum_{w \in \Omega} p(\omega)V(\omega) \tag{40}$$

△

Definition 5.3 (Propriety)**.** $\mathfrak{a}$ is *(strictly) proper for* V iff for any probability p, $\mathrm{Exp}_p[\mathfrak{a}_{V(\cdot)}(x)]$ is (uniquely) maximised at $x = \mathrm{Exp}_p[V]$. △

6 Schervish form for estimates

Schervish's representation very naturally extends to consider accuracy of a value as an estimate of any random variable. This is provided in Schervish et al. (2014, eq 1), inspired by Savage (1971, eq 4.3). It just applies the same integral form as in eq. (10) but allows the limits to be the true values of the variable, which may not be 0 or 1.[1]

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k - t) \;\; \lambda(\mathrm{d}t) \tag{41}$$

Schervish et al. (2014, Lemma 1) show that all $\mathfrak{a}$ that have this form are strictly proper. They require that λ is finite on every bounded interval and is mutually absolutely continuous wrt the Lebesgue measure.[2] They do not assume

[1]It was Jason Konek who suggested this form to me and asked whether the Schervish representation result extends to this setting.

[2]their proof of proposition 6.1 only requires the one direction, that the Lebesgue measure is absolutely continuous wrt λ.

that Ω is finite; instead they define (strict) propriety by restricting to p where $\mathrm{Exp}_p[V]$ is finite and $\mathrm{Exp}_p[\mathfrak{a}_V(x)]$ is finite for some x.

Proposition 6.1 (Schervish et al. (2014, Lemma 1))**.** *Suppose $\mathfrak{a}$ has the form:*

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k-t) \;\; \lambda(\mathrm{d}t) \tag{42}$$

where λ is finite on every bounded interval; then $\mathfrak{a}$ is proper. If λ gives positive measure to every non-generate interval, then $\mathfrak{a}$ is strictly proper.

We give the proof in the more restrictive setting where Ω is finite, i.e., there are only finitely many possible values that V can take, and λ is finite on every bounded interval. check what I actually use in the writing of it!

Proof. Suppose $\mathfrak{a}$ is of form eq. (42). First observe that for any k, x and y,

$$\mathfrak{a}_k(y) - \mathfrak{a}_k(x) = \int_x^y k - t \;\; \lambda(\mathrm{d}t) \tag{43}$$

as in the proof of $1 \implies 2$ in appendix B.

Let p be probabilistic. Let $e = \mathrm{Exp}_p[V]$.

$$\mathrm{Exp}_p[\mathfrak{a}(e)] - \mathrm{Exp}_p[\mathfrak{a}(x)] \tag{44}$$

$$= \sum_w p(w) \times (\mathfrak{a}_{V(w)}(e) - \mathfrak{a}_{V(w)}(x)) \tag{45}$$

$$= \sum_w p(w) \times \left(\int_x^e V(w) - t \;\; \lambda(\mathrm{d}t) \right) \tag{46}$$

$$= \int_x^e \left(\sum_w p(w) \times (V(w) - t) \right) \;\; \lambda(\mathrm{d}t) \tag{47}$$

$$= \int_x^e (e-t) \;\; \lambda(\mathrm{d}t) \tag{48}$$

If $x < e$, then $e - t > 0$ for all $t \in [x, e]$, and thus this integral is positive.

If $x > e$, then $e - t < 0$ for all $t \in [e, x]$, so

$$\int_e^x (e-t) \;\; \lambda(\mathrm{d}t) < 0;$$

and thus eq. (48)> 0 because the integral limits are switched, as in setup 2.2

Thus, for any $x \neq e$, $\mathrm{Exp}_p[\mathfrak{a}(e)] - \mathrm{Exp}_p[\mathfrak{a}(x)] = \int_x^e (e-t) \;\; \lambda(\mathrm{d}t) > 0$. So we know that $\mathrm{Exp}_p[\mathfrak{a}(x))]$ is maximised at $x = e$, as required. ☐

We will show the converse: that any strictly proper $\mathfrak{a}$ has this form, at least given some restrictive assumptions.

We first will make use of a lemma

Definition 6.2. $\mathfrak{a}$ is *(strictly) value-directed* iff If $k < x < y$ or $y < x < k$ then $\mathfrak{a}_k(x) > \mathfrak{a}_k(y)$ △

Proposition 6.3. *(Strict) propriety entails (strict) value-directedness.*

Again we relegate the proof to the appendix because we find its fiddlyness outweighs its philosophical interest, as, for accuracy measures, value directedness can be directly motivated.

Setup 6.4. Let $v_{\min}$ and $v_{\max}$ be $\min(\mathsf{range}(V))$ and $\max(\mathsf{range}(V))$, observing that V reaches its minimum and maximum by the assumption that Ω is finite, so it takes only finitely many possible values. △

Theorem 6.5. *Assume $\mathfrak{a}_k$ is absolutely continuous for each k.*

If $\mathfrak{a}$ is (strictly) proper for r.v. V, then there is a (strictly) positive function m such that for every $k \in \mathsf{range}(V)$ and $x \in [v_{\min}, v_{\max}]$

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k-t)m(t)\,\mathrm{d}t \tag{49}$$

Moreover,

$$m(t) = \frac{\mathfrak{a}_k'(t)}{k-t}$$

for all $k \in \mathsf{range}(V)$.

The proof is a mild extension of the proof of proposition 2.6.

Proof. For $k \in \mathsf{range}(V)$, define $m_k : [v_{\min}, v_{\max}] \to \mathrm{Re}$ by:

$$m_k(t) := \frac{\mathfrak{a}_k'(t)}{k-t}. \tag{50}$$

Observe that m_k is (strictly) positive by (strict) value-directedness.

For absolutely continuous $\mathfrak{a}$, by the fundamental theorem of calculus

$$\mathfrak{a}_k(k) - \mathfrak{a}_k(x) = \int_x^k \mathfrak{a}_k'(t)\,\mathrm{d}t \tag{51}$$

Thus we immediately get (recalling the definition of m_k):

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k-t)m_k(t)\,\mathrm{d}t \tag{52}$$

The important part is that m_k does not in fact depend on the choice of k. That is, we need to show that $m_k = m_r$ for $k, r \in \mathsf{range}(V)$. For this, we use the strict propriety.

We first show it for certain t:

Sublemma 6.5.1. *For $t \in \mathsf{ConvHull}(\{k, r\})$, we have $m_k(t) = m_r(t)$*

Proof. Wlog suppose $k < r$. Let $\omega_k \in \Omega$ s.t. $V(\omega_k) = k$ and $\omega_r \in \Omega$ s.t. $V(\omega_r) = r$. Since we have assumed $t \in \mathsf{ConvHull}(\{k, r\})$, we con consider p^* with

$$p^*(\omega) = \begin{cases} \frac{t-k}{r-k} & \omega = \omega_k \\ \frac{r-t}{r-k} & \omega = \omega_r \\ 0 & \text{otherwise} \end{cases} \tag{53}$$

Observe that p^* is probabilistic, i.e., these are $\geqslant 0$ and sum to 1.

Now, observe that $\mathrm{Exp}_{p^*}[V] = t$. So by propriety, the function:

$$\mathrm{Exp}_{p^*}\mathfrak{a}(x) = \frac{t-k}{r-k}\mathfrak{a}_r(x) + \frac{r-t}{r-k}\mathfrak{a}_k(x) \tag{54}$$

is maximised at $x = t$, so its derivative is 0 at t. I.e.:

$$\frac{t-k}{r-k}\mathfrak{a}'_r(t) + \frac{r-t}{r-k}\mathfrak{a}'_k(t) = 0 \tag{55}$$

By manipulating eq. (55) we can see that:[3]

$$\frac{\mathfrak{a}'_r(t)}{r-t} = \frac{\mathfrak{a}'_k(t)}{k-t}. \tag{60}$$

I.e., we have shown that $m_k(t) = m_r(t)$ when $t \in \mathsf{ConvHull}(\{r, k\})$. □

We still need to show $m_{k'}(t) = m_{r'}(t)$ for any $t \in [v_{\min}, v_{\max}]$ and $k', r' \in \mathsf{range}(V)$. (We use k' and t' as we need to use the lemma for other values too). We assume wlog that $k' \leqslant r'$.

If $k' \leqslant t \leqslant r'$, then we can immediately apply the lemma to obtain that $m_{k'}(t) = m_{r'}(t)$.

If $t \leqslant k' \leqslant r'$, then consider also $v_{\min}$, and we know that $v_{\min} \leqslant t \leqslant k' \leqslant r'$. So we can apply the lemma now with $v_{\min}$:

$$m_{r'}(t) = m_{v_{\min}}(t) \qquad \text{Lemma with } k = v_{\min}, r = r' \tag{61}$$

$$m_{k'}(t) = m_{v_{\min}}(t) \qquad \text{Lemma with } k = v_{\min}, r = k' \tag{62}$$

$$\text{Thus, } m_{k'}(t) = m_{r'}(t) \tag{63}$$

Similarly, if $k' \leqslant r' \leqslant t \leqslant v_{\max}$, we can similarly use the lemma to show that $m_{r'}(t) = m_{\max}(t)$ and $m_{k'}(t) = m_{\max}(t)$, so $m_{k'}(t) = m_{r'}(t)$.

We can thus simply put $m(t) := m_k(t)$, independent of choice of k. □

[3] From eq. (55) we get:

$$\frac{(t-k)\mathfrak{a}'_r(t) + (r-t)\mathfrak{a}'_k(t)}{r-k} = 0 \tag{56}$$

$$(t-k)\mathfrak{a}'_r(t) + (r-t)\mathfrak{a}'_k(t) = 0 \tag{57}$$

$$(r-t)\mathfrak{a}'_k(t) = (k-t)\mathfrak{a}'_r(t) \tag{58}$$

$$\frac{\mathfrak{a}'_r(t)}{r-t} = \frac{\mathfrak{a}'_k(t)}{k-t} \tag{59}$$

Corollary 6.6. *For p probabilistic with* $\mathrm{Exp}_p[V] = e$*,*

$$\mathrm{Exp}_p\mathfrak{a}(e) - \mathrm{Exp}_p\mathfrak{a}(x) = \int_x^e (e-t)m(t)\,\mathrm{d}t \tag{64}$$

(Schervish, 1989) does not need to assume absolute continuity. We conjecture that this holds in the estimates setting too. See appendix C.

Corollary 6.7. *Suppose each $\mathfrak{a}_k$ is* absolutely *continuous. Then $\mathfrak{a}$ has the form eq.* (42) *iff $\mathfrak{a}$ is strictly proper.*

Proof. From theorem 6.5 and proposition 6.1 □

Remark (Assumptions)**.** We are making some quite strong assumptions here. We list various of the assumptions and their statuses: CHECK!

- Restriction to estimates evaluated being in $[v_{\min}, v_{\max}]$
  - For theorem 6.5, this restriction is essential: Strict propriety doesn't give us any control over what $\mathfrak{a}_k(x)$ looks like for $x \notin [v_{\min}, v_{\max}]$, so something like the Schervish representation isn't going to be applicable. However, if we impose value-directedness in general we can still, for example, obtain accuracy-dominance results without any further control over how $\mathfrak{a}$ looks outside of $[v_{\min}, v_{\max}]$
- Ω is finite, so that V only takes finitely many values.
  - This should probably be removable. Need to do integrals rather than sums.
  - In the writing of the proof of theorem 6.5, we have used the fact that V obtains its maximum and minimum values; but this is inessential. Replace the range of estimates that we have control over the form of (for theorem 6.5) to be $\mathsf{ConvHull}(\mathsf{range}(V))$. Rewrite the end part of that proof to just pick some $v_{\min} \leqslant t \leqslant k, r$ or $v_{\max} \geqslant t \geqslant k, r$.
  - It does get more delicate though because we now need to talk about whether p must be merely finitely additive, or countably additive, etc.
- Accuracy values are finite
  - Schervish (1989) also applies allowing some infinities, so we conjecture that the same would apply here too. We need to avoid having both $+\infty$ and $-\infty$ allowed, as this causes expectations to be undefined (Schervish et al., 2009). Probably it's fine to allow $-\infty$ so long as it only appears at endpoints, i.e., $\mathfrak{a}_k(x) = -\infty \implies x \in \{v_{\min}, v_{\max}\}$; which we'd get by value-directedness anyway if we were to directly assume it, which I think as accuracy measures go we'd be happy to do so. Recall the status of this stuff... doesn't it *follow* from strict propriety that it's only infty at endpoingts?

- Interestingly, though, if we want to use the same accuracy measure for all possible random variables, we will consider $\mathfrak{a}_k(x)$ for all $x \in \mathrm{Re}$, so this must be finite, as no such real-valued x is an end-point for all possible variables. So the restriction can be motivated.
- Consider Schervish et al. (2014, example 1) showing a case where it's important that we assume the measure to be finite on any bounded interval, equivalent [check] to the accuracy measure only being infinity at endpoints.

- One dimensionality! We're just looking at scoring a single real-valued variable at a time. It's all one-dimensional!
  - We can push it up to finitely-many multiple variables simultaneously by just using additivity. But really it would now be natural to do this with infinitely many variables. So we're in accuracy-for-infinitely-many-propositions territory! This is exactly the sort of thing that Schervish et al. (2014) are considering. See also Kelley and Walsh for accuracy measures for infinitely many propositions.
- For theorem 6.5 we assume that $\mathfrak{a}$ is absolutely continuous.
  - I conjecture that the *absolute* continuity is not required for theorem 6.5, just as it is not in fact required for proposition 2.6 in Schervish (1989, Theorem 4.2); of course we won't be able to push it to Schervish form with a mass function, but will need to stay in the measure setting and use Radon-Nikodyn derivatives. See appendix C for an attempted proof.
  - The simple continuity part is also probably inessential, as it is for Schervish (1989, Theorem 4.2) because everything is anyway one-sided continuous by value-directedness. However, as Schervish et al. (2009) point out, continuity is essential for dominance results!

For proposition 6.1, Schervish et al. (2014) use an assumption that λ is absolutely continuous wrt the Lebesgue measure.

△

7 Bregman results

There is a difficulty facing the Bregman strategy which is that there is now no unique definition of entropy.[4]

However, we can still get the representation by going via the Schervish representation result.

[4]For a variable V which takes values $0, 0.5, 1$, consider $p_1[V=1]=0.5$, $p_1[V=.5]=0$, $p_1[V=0]=0.5$, or $p_2[V=1]=0$, $p_2[V=.5]=0.5$, $p_2[V=0]=0$. $\mathrm{Exp}_{p_1}[V]=\mathrm{Exp}_{p_2}[V]=0.5$. But it may be that $\mathrm{Exp}_{p_1}\mathfrak{a}(0.5)\neq\mathrm{Exp}_{p_2}\mathfrak{a}(0.5)$.

Proposition 7.1. *Assume:*

- $\mathfrak{a}$ *is continuously differentiable. What weakenings would work? Need its derivative to be integrable.*
- $\mathfrak{a}$ *is absolutely continuous. NB this follows from cts diff, but perhaps not from the relevant weakenings.*

$\mathfrak{a}$ *is strictly proper iff there is a strictly convex function* φ *(and values* $\mathfrak{a}_v(v)$*) where:*

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - [\varphi(v) - \varphi(x) - (v-x)\varphi'(x)] \tag{65}$$

That is, the error-score is:

$$\mathfrak{s}_v(x) = \varphi(v) - \varphi(x) - (v-x)\varphi'(x) \tag{66}$$

Proof. We show the $\implies$ direction: Assume $\mathfrak{a}$ is strictly proper.

Since $\mathfrak{a}$ is abs cts, by theorem 6.5, we have some positive m with

$$\mathfrak{a}_v(v) - \mathfrak{a}_v(x) = \int_x^v (v-t)m(t)\,\mathrm{d}t \tag{67}$$

By the assumption that $\mathfrak{a}$ is continuously differentiable, we know that m is continuous, so we can find φ where $\varphi'' = m$. [Weakenings?]

φ is (strictly) convex since its second derivative, m, is (strictly) positive.

Then, using lemma 4.1, i.e., just by integration by parts, we know that

$$\int_x^v (v-t)m(t)dt = \varphi(v) - \varphi(x) - (v-x)\varphi'(x) \tag{68}$$

Equation (69) follows immediately from eq. (68) and eq. (67). □

The expectation form is also equivalent:

Corollary 7.2. *Also iff*

$$\mathrm{Exp}_p\mathfrak{a}(e) - \mathrm{Exp}_p\mathfrak{a}(x) = \varphi(p) - \varphi(x) - (p-x)\varphi'(x) \tag{69}$$

where $e := \mathrm{Exp}_p[V]$.

Proof.

$$\mathrm{Exp}_p\mathfrak{a}(e) - \mathrm{Exp}_p\mathfrak{a}(x) = \sum_v p[V=v]\ (\mathfrak{a}_v(e) - \mathfrak{a}_v(x)) \tag{70}$$

$$= \sum_v p[V=v]\varphi(v) - \varphi(x) - (v-x)\varphi'(x) \tag{71}$$

□

Remark. For any individual pair k and r, we can choose some φ where $\varphi(k) = \varphi(r) = 0$ or where $\varphi(k) = \mathfrak{a}_k(k)$ and $\varphi(r) = \mathfrak{a}_r(r)$; but we cannot generally choose a single φ with $\varphi(v) = 0$ for all φ. (Since φ must be convex, there can only be two values where $\varphi(v) = 0$!). △

Question: how close is it to Savage (1971)?

References

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

Leonard J Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

Mark J Schervish. A general method for comparing probability assessors. *The Annals of Statistics*, 17.4:1856–1879, 1989.

Mark J Schervish, Teddy Seidenfeld, and Joseph B Kadane. Proper scoring rules, dominated forecasts, and coherence. *Decision Analysis*, 6(4):202–221, 2009.

Mark J Schervish, Teddy Seidenfeld, and Joseph B Kadane. Dominating countably many forecasts. *The Annals of Statistics*, 42(2):728–756, 2014.

Teddy Seidenfeld, Mark J Schervish, and Joseph B. Kadane. Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning*, 53 (8):1248–1261, 2012.

Part III
Appendix

A Propriety entails truth/value directedness

A.1 Truth directedness

Proof of proposition 1.4. Take $0 \leqslant z < y \leqslant 1$. We will show that $\mathfrak{a}_1(y) > \mathfrak{a}_1(z)$ and $\mathfrak{a}_1(y) < \mathfrak{a}_1(z)$.

By strict propriety,

$$\mathrm{Exp}_y \mathfrak{a}(y) > \mathrm{Exp}_y \mathfrak{a}(z) \tag{72}$$

$$\text{So, } \mathrm{Exp}_y[\mathfrak{a}(y) - \mathfrak{a}(z)] > 0 \tag{73}$$

$$\text{So, } y \times (\mathfrak{a}_1(y) - \mathfrak{a}_1(z)) + (1-y) \times (\mathfrak{a}_0(z) - \mathfrak{a}_0(z)) > 0 \tag{74}$$

Let

$$c = \mathfrak{a}_1(y) - \mathfrak{a}_1(z) \tag{75}$$

$$d = \mathfrak{a}_0(y) - \mathfrak{a}_0(z) \tag{76}$$

So from eq. (74)

$$yc + (1-y)d > 0 \tag{77}$$

Similarly, by strict propriety,

$$\mathrm{Exp}_z \mathfrak{a}(z) > \mathrm{Exp}_z \mathfrak{a}(y) \tag{78}$$

$$\text{So, } \mathrm{Exp}_z[\mathfrak{a}(y) - \mathfrak{a}(z)] < 0 \tag{79}$$

$$\text{So, } z \times (\mathfrak{a}_1(y) - \mathfrak{a}_1(z)) + (1-z)(\mathfrak{a}_0(z) - \mathfrak{a}_0(z)) \tag{80}$$

$$zc + (1-z)d < 0 \qquad \text{definition of } c,\, d \tag{81}$$

From eqs. (77) and (81)

$$yc + (1-y)d > zc + (1-z)d \tag{82}$$
$$\text{So, } (y-z)c > (y-z)d \tag{83}$$
$$\text{Thus, } c > d \qquad \text{since } y > z \tag{84}$$

Thus

$$c = yc + (1-y)c > yc + (1-y)d > 0 \tag{85}$$

using $c > d$ for the first inequality and eq. (77) for the second.

Thus $c > 0$. I.e., $\mathfrak{a}_1(y) - \mathfrak{a}_1(z) > 0$, so $\mathfrak{a}_1(y) > \mathfrak{a}_1(z)$.

Similarly, Thus

$$d = yd + (1-y)d < yc + (1-y)d < 0 \tag{86}$$

using $c > d$ for the first inequality and eq. (81) for the second.

Thus $d < 0$. I.e., $\mathfrak{a}_0(y) - \mathfrak{a}_0(z) > 0$, so $\mathfrak{a}_0(y) < \mathfrak{a}_0(z)$. □

A.2 Value directedness

Proof of proposition 6.3. Suppose r and k are in the range of possible values of V (with $r \neq k$). Consider a, b in between r and k, so in $[r, k]$ or $[k, r]$, and $e \in \{r, k\}$.

For x between r and k define $p_x = \frac{x-r}{k-r}$. Observe that $\mathrm{Exp}_{p_x} V = x$.

By strict propriety, $\mathrm{Exp}_{p_a}\mathfrak{i}(b) > \mathrm{Exp}_{p_a}\mathfrak{i}(a)$ and $\mathrm{Exp}_{p_b}\mathfrak{i}(b) < \mathrm{Exp}_{b}\mathfrak{i}(a)$. So $\mathrm{Exp}_{p_a}(\mathfrak{i}(b) - \mathfrak{i}(a)) > \mathrm{Exp}_{p_b}(\mathfrak{i}(b) - \mathfrak{i}(a))$. I.e.:

$$\frac{a-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-a}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{87}$$
$$> \frac{b-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-b}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{88}$$

So

$$\frac{a-b}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) > \frac{a-b}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{89}$$

Suppose $a > b > e$. Then:

$$\frac{1}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) > \frac{1}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{90}$$

Thus

$$\mathfrak{i}(b, e) - \mathfrak{i}(a, e) \tag{91}$$
$$= \mathrm{Exp}_{p_e}\mathfrak{i}(b) - \mathrm{Exp}_{p_e}\mathfrak{i}(a) \tag{92}$$
$$= \frac{e-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-e}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{93}$$
$$< \frac{b-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-b}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{94}$$
$$< 0 \tag{95}$$

With eq. (93) to eq. (94) being because $e < b$ and there is less weight on something positive and more on something negative.

Similarly, if $a < b < e$. Then

$$\frac{1}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) > \frac{1}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{96}$$

So the step from eq. (93) to eq. (94) nonetheless holds with signs reversed. This shows value directedness whenever a, b, e are between r and k

By choosing appropriate r, we thus show that whenever b moves directly towards k, accuracy improves. □

B Schervish equivalences

Proof of lemma 2.4.

- $1 \implies 2$:

$$\mathfrak{a}_v(y) - \mathfrak{a}_v(x) = \left(\mathfrak{a}_v(v) - \int_y^v v - t \ \ \lambda(\mathrm{d}t)\right) - \left(\mathfrak{a}_v(v) - \int_x^v v - t \ \ \lambda(\mathrm{d}t)\right) \tag{97}$$

$$= \left(\int_x^v v - t \ \ \lambda(\mathrm{d}t)\right) - \left(\int_y^v v - t \ \ \lambda(\mathrm{d}t)\right) \tag{98}$$

$$= \int_x^y v - t \ \ \lambda(\mathrm{d}t) \tag{99}$$

- $2 \implies 3$:

$$\mathrm{Exp}_p\mathfrak{a}(p) - \mathrm{Exp}_p\mathfrak{a}(x) \tag{100}$$

$$= p \times (\mathfrak{a}_1(p) - \mathfrak{a}_1(x)) + (1-p) \times (\mathfrak{a}_0(p) - \mathfrak{a}_0(x)) \tag{101}$$

$$= p \times \left(\int_x^p 1 - t \ \ \lambda(\mathrm{d}t)\right) + (1-p) \times \left(\int_x^p 0 - t \ \ \lambda(\mathrm{d}t)\right) \quad \text{by item 2} \tag{102}$$

$$= \int_x^p (p \times (1-t) + (1-p) \times (0-t)) \ \ \lambda(\mathrm{d}t) \tag{103}$$

$$= \int_x^p (p - t) \ \ \lambda(\mathrm{d}t) \tag{104}$$

- $3 \implies 1$: put p as either 0 or 1, i.e., v, and simply observe that:

$$\mathrm{Exp}_v\mathfrak{a}(v) = \mathfrak{a}_v(v) \text{ and } \mathrm{Exp}_v\mathfrak{a}(x) = \mathfrak{a}_v(x) \tag{105}$$

It then follows immediately from rearranging.

□