Accuracy representation results and estimates

Catrin Campbell-Moore*

April 4, 2022

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.

In this working document, full references etc are not yet included.

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,

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

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

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 $p \in [0,1]$,

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

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 $\mathfrak{s}_v$, 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 $\mathfrak{s}_v(v) = 0$.

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. 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 redescribe this as:

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

(if $x < v$, the switching limits and absolute value signs cancel out)

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

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

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

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

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 $x \in [0,1]$,*

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v v - t \ \ \lambda(\mathrm{d}t). \qquad \textit{eq. } (9)$$

2. *For all $x, y \in [0,1]$,*

$$\mathfrak{a}_v(y) - \mathfrak{a}_v(x) = \int_x^y v - t \ \ \lambda(\mathrm{d}t). \tag{11}$$

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

$$\mathrm{Exp}_p\mathfrak{a}(p) - \mathrm{Exp}_p\mathfrak{a}(x) = \int_x^p (p-t) \ \lambda(\mathrm{d}t) \tag{12}$$

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*

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v (v-t)m(t)\,\mathrm{d}t \tag{13}$$

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

$$\mathfrak{a}_v(v) - \mathfrak{a}_v(x) = \int_x^v \mathfrak{a}_v'(t)\,\mathrm{d}t \tag{14}$$

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

$$t\mathfrak{a}_1'(t) + (1-t)\mathfrak{a}_0'(t) = 0 \tag{15}$$

By manipulating eq. (15)

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

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. (14), we obtain eq. (13).

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:

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v (v-t)\ \mathrm{d}\lambda(t) \tag{17}$$

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:

$$\varphi_{\mathfrak{a}}(p) := \mathrm{Exp}_p\mathfrak{a}(p) = p\mathfrak{a}_1(p) + (1-p)\mathfrak{a}_0(p) \tag{18}$$

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

$$\operatorname{Exp}_p \mathfrak{a}(p) - \operatorname{Exp}_p \mathfrak{a}(x) = \varphi_{\mathfrak{a}}(p) - \varphi_{\mathfrak{a}}(x) - (p - x)\varphi'_{\mathfrak{a}}(x) \tag{19}$$

If it is not differentiable, then we have the same form, but with φ' as some sub-gradient.

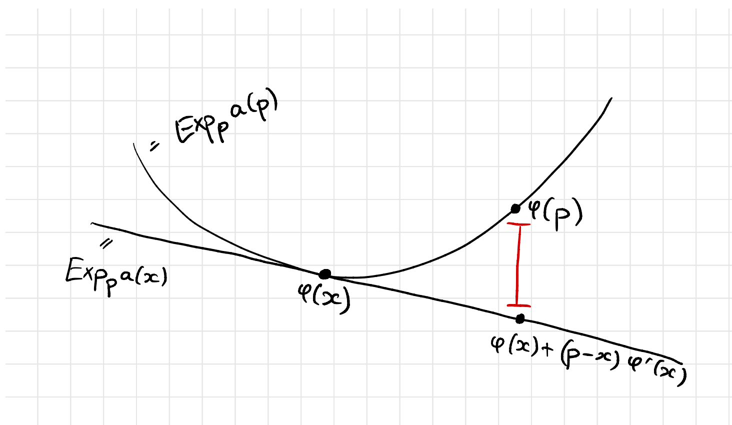

Figure 1: Divergence diagram

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

$$\operatorname{Exp}_p \mathfrak{a}(x) = p\mathfrak{a}_1(x) + (1 - p)\mathfrak{a}_0(x) \tag{20}$$

is a linear function of p (we could name it, e.g., $f_x(p) = \operatorname{Exp}_p \mathfrak{a}(x)$). 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 $f_x(p) = \operatorname{Exp}_p \mathfrak{a}(x)$ a subtangent of it at x.

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

$$\operatorname{Exp}_p \mathfrak{a}(x) = \varphi_{\mathfrak{a}}(x) + (p - x)\varphi'_{\mathfrak{a}}(x) \tag{21}$$

and eq. (19). If $\varphi_{\mathfrak{a}}$ is not differentiable, then one can take the slope of $\operatorname{Exp}_p \mathfrak{a}(x)$ and observe it is a subgradient of $\varphi_{\mathfrak{a}}$ by propriety; that will play the role of $\varphi'_{\mathfrak{a}}$. □

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

$$\mathfrak{d}(p, x) = \varphi(p) - \varphi(x) - (p - x)\varphi'(x) \tag{22}$$

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

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

$$\mathfrak{a}_v(x) = \varphi(x) + (v - x)\varphi'(x) \tag{23}$$

Proof. $\mathfrak{a}_v(x) = \mathrm{Exp}_v \mathfrak{a}(x)$. And from eq. (19), using the fact that $\varphi(v) = \mathrm{Exp}_v \mathfrak{a}(v)$

$$\mathrm{Exp}_v \mathfrak{a}(x) = \varphi(x) + (v - x)\varphi'(x) \tag{24}$$

□

Remark. There is an alternative proof that goes directly via rearrangments of eq. (15) 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:*

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

That is, the error-score is:

$$\mathfrak{s}_v(x) = \varphi(v) - \varphi(x) - (v - x)\varphi'(x) \tag{26}$$

4 Relationships between Bregman divergences and the Schervish form

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

$$\int_x^p (p - t)\varphi''(t)dt = \varphi(p) - \varphi(x) - (p - x)\varphi'(x) \tag{27}$$

Proof. Integration by Parts. □

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.

$$\begin{aligned} \varphi'(x) &= \mathfrak{a}_1(x) - \mathfrak{a}_0(x) + x\mathfrak{a}_1'(x) + (1 - x)\mathfrak{a}_0'(x) && \text{product rule} && (28) \\ &= \mathfrak{a}_1(x) - \mathfrak{a}_0(x) && \text{eq. (15)} && (29) \end{aligned}$$

And from eq. (15),

$$\mathfrak{a}_1'(x) - \mathfrak{a}_0'(x) = \frac{\mathfrak{a}_0'(x)}{-x} = m(x). \tag{30}$$

□

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

Part II
Estimates

Literature references:

- The estimates-based framework is the original Savage etc. It's actually Schervish (1989) who pushes it to focus on 0/1 values.
- The "Schervish form for estimates" from is already in Schervish et al. (2014), see Lemma 1. However:
  - They only show one direction: that scoring rules with Schervish form are proper.
  - They show that the converse fails in some cases when infinite values are allowed. See

5 Accuracy of Estimates

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 (previsions) for random variables.

Setup 5.1. Consider a fixed random variable $V : \Omega \to \mathrm{Re}$ which takes some finitely many possible values in $\mathsf{Values} \subseteq \mathrm{Re}$.

Let $v_{\min}$ and $v_{\max}$ be the minimum and maximum values.

An accuracy measure for V gives, for each $k \in \mathsf{Values}$, a measure of accuracy, $\mathfrak{a}_k$.

$$\mathfrak{a}_k : [v_{\min}, v_{\max}] \to \mathrm{Re}\,.$$

Definition 5.2 (Propriety)**.** $\mathfrak{a}$ is (strictly) proper iff for any p probabilistic over values of V, $\mathrm{Exp}_p\mathfrak{a}(x)$ is (uniquely) maximised at $x = \mathrm{Exp}_p[V]$.

$$\mathrm{Exp}_p\mathfrak{a}(x) = \sum_k p[V = k]\mathfrak{a}_k(x) \tag{31}$$

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

- V takes values in a compact $[a, b]$.
  - This can probably quite easily be removed *if* we want to have a strictly proper $\mathfrak{a}$ which is independent of a choice of V. So, we have $\mathfrak{a} : \mathrm{Re} \to \mathrm{Re}$ such that for every V it is proper. That should mean that our representation result applies everywhere. But it's a stronger "extensionality" assumption.
- V only takes finitely many values.

- This should probably be removable. Need to do integrals rather than sums. But does this get us into substantial subtlety territory (Schervish et al., 2014)?

- Accuracy values are finite, including at endpoints.
  - By strict value directedness, it's only the finite at endpoint that's the real restrictive assumptions. This excludes, e.g., the log score. It can be justified if

- 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 one would want to do this with infinitely many variables. So we're in accuracy-for-infinitely-many-propositions territory! Look at (Schervish et al., 2014), also Kelley and Walsh.

- For theorem 6.3 we will also assume that $\mathfrak{a}$ is absolutely continuous.
  - This should surely be inessential (at least the strong *absolute* continuity part). Although 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 somehow. See Schervish (1989, Theorem 4.2)
  - 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!

Remark (Literature)**.** The scoring rules for estimates setting is actually the historically more basic one, e.g., Savage (1971) and OTHER REFS. Schervish (1989), though, is restricted to 0/1-valued.

Remark (Terminology)**.** I'm talking about accuracy for "estimates". In the imprecise probability etc literature, usually talk about "previsions". Should I change it?

6 Schervish for estimates

Schervish's representation very naturally extends to consider accuracy of a value as an estimate of any random variable.

NB: this "Schervish form" in eq. (32) is already used in Schervish et al. (2014, eq1).[1]

[1]With a measure rather than a mass function; we can pull it into a mass function because of the assumption of absolute continuity.

Remark (Literature: Seidenfeld et al)**.** The Schervish et al. (2014, Lemma 1) gives one direction of this result: that such $\mathfrak{a}$ are strictly proper. They don't have the converse. They say some delicate things regarding infinities. They discuss the example of the log score saying it doesn't have this form. Does this constitute a counterexample to the converse in the setting with infinities? I'm not sure. They certainly show that nice probability things go badly to infinity with the log score. (Schervish et al., 2014, Example 1)

We first will make use of a lemma

Definition 6.1. $\mathfrak{a}$ is *(strictly) value-directed* iff If $k < x < y$ or $y < x < k$ then $\mathfrak{a}_k(x) > \mathfrak{a}_k(y)$

Proposition 6.2. *(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.

Main result: the Schervish representation for strictly proper measures of accuracy of estimates.

Theorem 6.3. *If $\mathfrak{a}$ is strictly proper and absolutely continuous, then there is positive m such that for any k some possible value of the variable and x lying in the convex hull of the possible values of V,*

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k-t)m(t)\,\mathrm{d}t \tag{32}$$

Moreover,

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

for any k value of V.

Proof. Take k, r possible values of V. And x between them.

Consider a probability function assigning probability $\frac{k-t}{k-r}$ to $[V = r]$ and $\frac{t-r}{k-r}$ to $[V = k]$. Note that $\mathrm{Exp}_p[V] = t$. So by propriety,

$$\mathrm{Exp}_p\mathfrak{a}(x) = \frac{t-r}{k-r}\mathfrak{a}_k(x) + \frac{k-t}{k-r}\mathfrak{a}_r(x) \tag{33}$$

is maximised at $x = t$, so its derivative is 0 at x,

$$\frac{x-r}{k-r}\mathfrak{a}'_k(x) + \frac{k-x}{k-r}\mathfrak{a}'_r(x) = 0 \tag{34}$$

By manipulating eq. (34)

$$\frac{\mathfrak{a}'_k(t)}{k-t} = \frac{\mathfrak{a}'_r(t)}{r-t} \tag{35}$$

Define

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

Using eq. (35), this doesn't depend on the choice of k. So that $\mathfrak{a}'_k(t) = (k-t)m(t)$ for all k. And observe that m is positive by value-directedness.

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

$$\mathfrak{a}_k(k) - \mathfrak{a}_k(x) = \int_x^k \mathfrak{a}'_k(t)\,\mathrm{d}t \tag{36}$$

Thus

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k-t)m(t)\,\mathrm{d}t \tag{37}$$

□

Corollary 6.4. *For p probabilistic with $\mathrm{Exp}_p[V] = e$,*

$$\mathrm{Exp}_p\mathfrak{a}(e) - \mathrm{Exp}_p\mathfrak{a}(x) = \int_x^e (e-t)m(t)\,\mathrm{d}t \tag{38}$$

Proposition 6.5. *Any such $\mathfrak{a}$ is strictly proper.*

Schervish et al. (2014, Lemma 1) give a more general version of this proposition.

Proof. Let p be probabilistic.

$$\begin{aligned}
&\mathrm{Exp}_p[\mathfrak{a}(\mathrm{Exp}_p[V])] - \mathrm{Exp}_p[\mathfrak{a}(x)] && (39)\\
&= \sum_w p(w) \times (\mathfrak{a}_{V(w)}(\mathrm{Exp}_p[V]) - \mathfrak{a}_{V(w)}(x)) && (40)\\
&= \sum_w p(w) \times \left(\int_x^{\mathrm{Exp}_p[V]} V(w) - t \;\; \lambda(\mathrm{d}t)\right) && (41)\\
&= \int_x^{\mathrm{Exp}_p[V]} \left(\sum_w p(w) \times (V(w) - t)\right) \;\; \lambda(\mathrm{d}t) && (42)\\
&= \int_x^{\mathrm{Exp}_p[V]} \big(\mathrm{Exp}_p[V] - t\big) \;\; \lambda(\mathrm{d}t) && (43)
\end{aligned}$$

If $x < \mathrm{Exp}_p[V]$, then $\mathrm{Exp}_p[V] - t > 0$ for all $t \in [x, \mathrm{Exp}_p[V]]$, and thus this integral is positive.

If $x > \mathrm{Exp}_p[V]$, then $\mathrm{Exp}_p[V] - t < 0$ for all $t \in [\mathrm{Exp}_p[V], x]$, so $\int_{\mathrm{Exp}_p[V]}^x \big(\mathrm{Exp}_p[V] - t\big) \;\; \lambda(\mathrm{d}t) < 0$; and thus eq. (43)$> 0$ because the integral limits are switched. □

Please check for updates here or by emailing me

Draft April 4, 2022

7 Bregman results

There is a difficulty facing the Bregman results which is that there is now no unique definition of entropy.[2]

However, we can still get the representation by using the Schervish representation and eq. (27) to take any φ with $\varphi'' = m$ and then see that for $e = \mathrm{Exp}_p[V]$,

$$\mathrm{Exp}_p\mathfrak{a}(e) - \mathrm{Exp}_p\mathfrak{a}(x) = \varphi(e) - \varphi(x) - (e-x)\varphi'(x) \tag{44}$$

or

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - (\varphi(k) - \varphi(x) - (k-x)\varphi'(x)) \tag{45}$$

i.e.,

$$\mathfrak{s}_k(x) = \varphi(k) - \varphi(x) - (k-x)\varphi'(x) \tag{46}$$

Remark. It needn't be that $\varphi(k) = \mathfrak{a}_k(k)$. We can ensure that $\varphi(k) = 0$ at two chosen values of k, but not everywhere simultaneously (it must be convex).

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.

Part III
Appendix

A Propriety entails truth/value directedness

[2]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 $\mathrm{Exp}_{p_1}\mathfrak{a}(0.5) \neq \mathrm{Exp}_{p_2}\mathfrak{a}(0.5)$.

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,

$$\begin{aligned} &\mathrm{Exp}_y \mathfrak{a}(y) > \mathrm{Exp}_y \mathfrak{a}(z) && (47)\\ \text{So, } &\mathrm{Exp}_y[\mathfrak{a}(y) - \mathfrak{a}(z)] > 0 && (48)\\ \text{So, } &y \times (\mathfrak{a}_1(y) - \mathfrak{a}_1(z)) + (1-y) \times (\mathfrak{a}_0(z) - \mathfrak{a}_0(z)) > 0 && (49) \end{aligned}$$

Let

$$c = \mathfrak{a}_1(y) - \mathfrak{a}_1(z) \tag{50}$$

$$d = \mathfrak{a}_0(y) - \mathfrak{a}_0(z) \tag{51}$$

So from eq. (49)

$$yc + (1-y)d > 0 \tag{52}$$

Similarly, by strict propriety,

$$\begin{aligned} &\mathrm{Exp}_z \mathfrak{a}(z) > \mathrm{Exp}_z \mathfrak{a}(y) && (53)\\ \text{So, } &\mathrm{Exp}_z[\mathfrak{a}(y) - \mathfrak{a}(z)] < 0 && (54)\\ \text{So, } &z \times (\mathfrak{a}_1(y) - \mathfrak{a}_1(z)) + (1-z)(\mathfrak{a}_0(z) - \mathfrak{a}_0(z)) && (55)\\ &zc + (1-z)d < 0 \qquad \text{definition of } c,\, d && (56) \end{aligned}$$

From eqs. (52) and (56)

$$\begin{aligned} &yc + (1-y)d > zc + (1-z)d && (57)\\ \text{So, } &(y-z)c > (y-z)d && (58)\\ \text{Thus, } &c > d \qquad \text{since } y > z && (59) \end{aligned}$$

Thus

$$c = yc + (1-y)c > yc + (1-y)d > 0 \tag{60}$$

using $c > d$ for the first inequality and eq. (52) 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

$$d = yd + (1-y)d < yc + (1-y)d < 0 \tag{61}$$

using $c > d$ for the first inequality and eq. (56) 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 6.2. 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.:

$$\begin{aligned} &\frac{a-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-a}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) && (62) \\ &> \frac{b-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-b}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) && (63) \end{aligned}$$

So

$$\frac{a-b}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) > \frac{a-b}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{64}$$

Suppose $a > b > e$. Then:

$$\frac{1}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) > \frac{1}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{65}$$

Thus

$$\begin{aligned} &\mathfrak{i}(b,e) - \mathfrak{i}(a,e) && (66) \\ &= \mathrm{Exp}_{p_e} \mathfrak{i}(b) - \mathrm{Exp}_{p_e} \mathfrak{i}(a) && (67) \\ &= \frac{e-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-e}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) && (68) \\ &< \frac{b-r}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) + \frac{k-b}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) && (69) \\ &< 0 && (70) \end{aligned}$$

With eq. (68) to eq. (69) being because $e < b$ and there is less weight on something positive and more on something negative.

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

$$\frac{1}{k-r}(\mathfrak{i}_k(b) - \mathfrak{i}_k(a)) > \frac{1}{k-r}(\mathfrak{i}_r(b) - \mathfrak{i}_r(a)) \tag{71}$$

So the step from eq. (68) to eq. (69) 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. • $1 \implies 2$:

$$\mathfrak{a}_v(y) - \mathfrak{a}_v(x) = \left(\mathfrak{a}_v(v) - \int_y^v v - t \ \lambda(\mathrm{d}t)\right) - \left(\mathfrak{a}_v(v) - \int_x^v v - t \ \lambda(\mathrm{d}t)\right) \tag{72}$$

$$= \left(\int_x^v v - t \ \lambda(\mathrm{d}t)\right) - \left(\int_y^v v - t \ \lambda(\mathrm{d}t)\right) \tag{73}$$

$$= \int_x^y v - t \ \lambda(\mathrm{d}t) \tag{74}$$

• $2 \implies 3$:

$$\mathrm{Exp}_p \mathfrak{a}(p) - \mathrm{Exp}_p \mathfrak{a}(x) \tag{75}$$

$$= p \times (\mathfrak{a}_1(p) - \mathfrak{a}_1(x)) + (1-p) \times (\mathfrak{a}_0(p) - \mathfrak{a}_0(x)) \tag{76}$$

$$= p \times \left(\int_x^p 1 - t \ \lambda(\mathrm{d}t)\right) + (1-p) \times \left(\int_x^p 0 - t \ \lambda(\mathrm{d}t)\right) \quad \text{by item 2} \tag{77}$$

$$= \int_x^p (p \times (1-t) + (1-p) \times (0-t)) \ \lambda(\mathrm{d}t) \tag{78}$$

$$= \int_x^p (p-t) \ \lambda(\mathrm{d}t) \tag{79}$$

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

$$\mathrm{Exp}_v \mathfrak{a}(v) = \mathfrak{a}_v(v) \text{ and } \mathrm{Exp}_v \mathfrak{a}(x) = \mathfrak{a}_v(x) \tag{80}$$

It then follows immediately from rearranging.

□

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.