Accuracy representation results and estimates

Catrin Campbell-Moore*

January 7, 2021

1 Definitions

Setup 1.1. For $v \in \{0, 1\}$,

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

measuring the accuracy of a credence x in a proposition, when it is true/false.

Remark (Infinite accuracy)**.** We can allow infinite values at end-points. We should not allow both sided $\pm\infty$. But to keep things simple I rule it out by fiat for now.

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

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

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

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

Proposition 1.4. *Propriety entails truth-directedness.*

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

$$\mathfrak{s}_v(x) = \mathfrak{a}_v(v) - \mathfrak{a}_v(x) \tag{2}$$

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \mathfrak{s}_v(x) \tag{3}$$

Note that this picture requires $\mathfrak{s}_v(v) = 0$. Choice of $\mathfrak{a}_v(v)$ is free compatibly with propriety.

The representations are really directly characterising $\mathfrak{s}$. We can talk about strict propriety etc directly of $\mathfrak{s}$.

---

*

2 Schervish

2.1 Schervish form

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

$$\mathfrak{a}_0(x) = \mathfrak{a}_0(0) - \int_0^x t \;\; \lambda(\mathrm{d}t) \tag{4}$$

$$\mathfrak{a}_1(x) = \mathfrak{a}_1(1) - \int_x^1 1 - x \; \lambda(\mathrm{d}x) \tag{5}$$

Remark. We might also phrase it concisely as:

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_{\min\{v,x\}}^{\max\{v,x\}} |v - t| \;\; \lambda(\mathrm{d}t). \tag{6}$$

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:

$$\mathfrak{a}_v(x) = \mathfrak{a}_v(v) - \int_x^v v - t \;\; \lambda(\mathrm{d}t). \tag{7}$$

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

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

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

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

Lemma 2.3. *If* $\mathfrak{a}$ *has Schervish form (eq.* (7)*) iff*

$$\mathrm{Exp}_p\mathfrak{a}(p) - \mathrm{Exp}_p\mathfrak{a}(x) = \int_x^p (p - t)\; \lambda(\mathrm{d}t) \tag{9}$$

Proof. $\Longleftarrow$ follows from $\mathrm{Exp}_v\mathfrak{a}(x) = \mathfrak{a}_v(x)$. For $\Longrightarrow$ observe

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

Proposition 2.4. *If* $\mathfrak{a}$ *has Schervish form it is (strictly) proper.*

Proof. When $x < p$, $p - t > 0$ for any $t \in [x, p]$, so relevant $\int_p^x$ is negative, and therefore $\int_x^p$ is positive. □

2.3 The representation result

We prove it simply for the absolutely continuous case.

Proposition 2.5. *If $\mathfrak{a}$ is strictly proper and absolutely continuous, then it has Schervish form.*

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

$$\mathfrak{a}_v(v) - \mathfrak{a}_v(x) = \int_x^v \mathfrak{a}_v'(t)\,\mathrm{d}t \tag{10}$$

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

$$t\mathfrak{a}_1'(t) + (1-t)\mathfrak{a}_0'(t) = 0 \tag{11}$$

By manipulating eq. (11)

$$\frac{\mathfrak{a}_0'(t)}{-t} = \frac{\mathfrak{a}_1'(t)}{1-t} \tag{12}$$

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 we have:

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

Also observe that m is positive by truth-directedness. □

Remark. Without absolute continuity it still holds but with a measure rather than mass function.

3 Bregman divergences

3.1 Entropy and Bregman Divergence

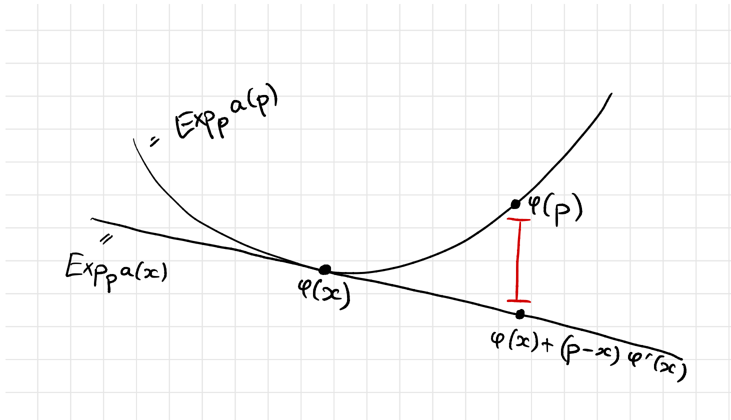
Definition 3.1. Define the *entropy* of $\mathfrak{a}$ as:

$$\varphi(p) := \mathrm{Exp}_p\mathfrak{a}(p) = p\mathfrak{a}_1(p) + (1-p)\mathfrak{a}_0(p) \tag{14}$$

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

$$\mathrm{Exp}_p\mathfrak{a}(p) - \mathrm{Exp}_p\mathfrak{a}(x) = \varphi(p) - \varphi(x) - (p-x)\varphi'(x) \tag{15}$$

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



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

$$\mathrm{Exp}_p \mathfrak{a}(x) = p\mathfrak{a}_1(x) + (1-p)\mathfrak{a}_0(x) \tag{16}$$

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 φ and touching it just at p. Therefore, φ is convex, with $f_x(p) = \mathrm{Exp}_p \mathfrak{a}(x)$ a subtangent of it at x.

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

$$\mathrm{Exp}_p \mathfrak{a}(x) = \varphi(x) + (p-x)\varphi'(x) \tag{17}$$

and eq. (15). If φ is not differentiable, then one can take the slope of $\mathrm{Exp}_p \mathfrak{a}(x)$ and observe it is a subgradient of φ by propriety; that will play the role of φ'. □

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

$$\mathfrak{d}(p, x) = \varphi(p) - \varphi(x) - (p-x)\varphi'(x) \tag{18}$$

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*

$$\mathfrak{a}_v(x) = \varphi(x) + (v-x)\varphi'(x) \tag{19}$$

Proof. $\mathfrak{a}_v(x) = \mathrm{Exp}_v \mathfrak{a}(x)$. And from eq. (15), using the fact that $\varphi(v) = \mathrm{Exp}_v \mathfrak{a}(v)$

$$\mathrm{Exp}_v \mathfrak{a}(x) = \varphi(x) + (v-x)\varphi'(x) \tag{20}$$

□

Remark. There is an alternative proof that goes directly via rearrangments of eq. (11) 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:*

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

That is, the error-score is:

$$\mathfrak{s}_v(x) = \varphi(v) - \varphi(x) - (v-x)\varphi'(x) \tag{22}$$

4 Bregman divergences and the Schervish form

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

$$\int_x^p (p-t)\varphi''(t)dt = \varphi(p) - \varphi(x) - (p-x)\varphi'(x) \tag{23}$$

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. $m(t) = \varphi''(t)$.

Proof.

$$\begin{aligned} \varphi'(x) &= \mathfrak{a}_1(x) - \mathfrak{a}_0(x) + x\mathfrak{a}_1'(x) + (1-x)\mathfrak{a}_0'(x) && \text{product rule} \quad (24) \\ &= \mathfrak{a}_1(x) - \mathfrak{a}_0(x) && \text{eq. (11)} \quad (25) \end{aligned}$$

And from eq. (11),

$$\mathfrak{a}_1'(t) - \mathfrak{a}_0'(t) = \frac{\mathfrak{a}_0'(t)}{-t}. \tag{26}$$

□

Estimates

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.

Setup 5.1. Consider a fixed variable V which takes some possible values in $\mathsf{Values} \subseteq \mathrm{Re}$.

Accuracy measures give, for each $k \in \mathsf{Values}$, a measure of accuracy, $\mathfrak{a}_k$.

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

$$\mathrm{Exp}_p\mathfrak{a}(x) = \sum_k p[V = k]\mathfrak{a}_k(x) \tag{27}$$

Remark (Extensionality)**.** I quite like a different setup which assumes some kind of strong extensionality. The accuracy score should not be variable dependent. But that for any variable, V, the same accuracy score is used. Since we can choose variables across Re, that then means we have $\mathfrak{a}_k(x)$ defined for any $k \in \mathrm{Re}$ and $x \in \mathrm{Re}$. But since this is stronger, I won't assume it and relativise everything to a choice of V.

Remark (Just one-dimensional)**.** Note that I am only considering accuracy for a single real-valued variable at a time. We're still doing 1D stuff, it's just replacing the truth values of 1 and 0 (or indicator variables) by arbitrary real-valued variables.

6 Schervish for estimates

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

Theorem 6.1. *$\mathfrak{a}$ is proper (for V) iff there is positive m such that for any k some value of the variable and x lying in the convex hull of the possible values of V,*

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - \int_x^k (k - t)m(t)\,\mathrm{d}t \tag{28}$$

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 $\frac{k-x}{k-r}$ to $[V=r]$ and $\frac{x-r}{k-r}$ to $[V=k]$. Note that $\mathrm{Exp}_p[V]=x$. So by propriety,

$$\mathrm{Exp}_p\mathfrak{a}(t)=\frac{x-r}{k-r}\mathfrak{a}_k(t)+\frac{k-x}{k-r}\mathfrak{a}_r(t) \tag{29}$$

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

$$\frac{x-r}{k-r}\mathfrak{a}'_k(x)+\frac{k-x}{k-r}\mathfrak{a}'_r(x)=0 \tag{30}$$

By manipulating eq. (30)

$$\frac{\mathfrak{a}'_k(t)}{k-t}=\frac{\mathfrak{a}'_r(t)}{r-t} \tag{31}$$

Define

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

Using eq. (31), 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

$$\mathfrak{a}_k(k)-\mathfrak{a}_k(x)=\int_x^k \mathfrak{a}'_k(t)\,\mathrm{d}t \tag{32}$$

Thus

$$\mathfrak{a}_k(x)=\mathfrak{a}_k(k)-\int_x^k (k-t)m(t)\,\mathrm{d}t \tag{33}$$

□

Corollary 6.2. *For p probabilistic with* $\mathrm{Exp}_p[V]=e$*,*

$$\mathrm{Exp}_p\mathfrak{a}(e)-\mathrm{Exp}_p\mathfrak{a}(x)=\int_x^e (e-t)m(t)\,\mathrm{d}t \tag{34}$$

Proposition 6.3. *Any such $\mathfrak{a}$ is strictly proper.*

7 Bregman results

There is a challenge facing the Bregman results which is that there is now no unique definition of entropy.[1]

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

However, we can use the Schervish representation and eq. (23) to take any φ with $\varphi'' = m$ and then see that for $e = \mathrm{Exp}_p[V]$,

$$\mathrm{Exp}_p\mathfrak{a}(e) - \mathrm{Exp}_p\mathfrak{a}(x) = \varphi(e) - \varphi(x) - (e-x)\varphi'(x) \tag{35}$$

or

$$\mathfrak{a}_k(x) = \mathfrak{a}_k(k) - (\varphi(k) - \varphi(x) - (k-x)\varphi'(x)) \tag{36}$$

i.e.,

$$\mathfrak{s}_k(x) = \varphi(k) - \varphi(x) - (k-x)\varphi'(x) \tag{37}$$

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).