Self-reference and Self-reproduction of Evidence

Michael Evans and Constantine Frangakis

October 25, 2024

1. Introduction.

Defining and measuring evidence in statistical problems has been discussed at length in the literature. The most well-known framework assumes the principles of sufficiency and conditionality, and suggest that evidence is the likelihood function (Birnbaum).

From the more general study of functions on statements (events), we know that there are conditions that determine if a function is *definable* or *computable* in arithmetic and logic. Below we argue why an "evidence" function should reasonably satisfy certain conditions, which make it one of the not fully definable and not fully computable functions in arithmetic and logic.

2. Definability and self referenced evidence functions.

Suppose that a random variable takes values $x = 1, 2, ...$, and the data are generated from one of two possible distributions, $p_1(\cdot)$ and $p_2(\cdot)$, denoted as $\{p\}$. We explore the consequence of assuming existence of a function $Ev(\{p\}, x)$, with input $\{p\}$ and an observation x, and an output that has basic properties of what can be meant by "evidence for p_1 vs. p_2 based on x".

One important property appears to be that evidence should be applicable, not just to x, but also to events that can be described based on x. This, however brings up certain consequences as implied from the limitations of logic.

To describe this, consider any event $S(\cdot)$ (with $\cdot$ a dummy variable taking values as $x = 1, ...$) that can be spelled out with the language of arithmetic and logic. Assign to every such $S(\cdot)$ its Gödel number (a unique identifier, Gödel 1933). For example, suppose the statement "$\cdot \geq 10$", has Gödel number, say 43; then we define $S(43, \cdot)$ to be "$\cdot \geq 10$". We can arrange all this in the Tarski-Gödel countable list $S(m, \cdot)$ (Table 1, left), with the implied events $S(m, x)$ for specific $x = 1, ...$. For example, $S(43, x)$ for $x = 8$ is "$8 \geq 10$".

	1	...	x	...
$\vdots$				...
$S(m, \cdot)$			$S(m, x)$	...
$\vdots$				...

	1	...	x	...
$\vdots$				...
$Ev\{S(m, \cdot)\}$			$Ev\{S(m, x)\}$	...
$\vdots$				...

Table 1: Events (left) and evidence based on events (right).

1. Suppose there are at least two distinct events, $S(m_1, x_1)$ (call it $S1$) and $S(m_2, x_2)$ (call it $S2$) for which the evidence is different, $Ev(S1) \neq Ev(S2)$ (we suppress $\{p\}$ as it is always a factor).

2. Define a event that depends on x as follows

$$\text{New}(x) = \begin{cases} S1 \text{ if } Ev\{S(x,x)\} = Ev(S2) \\ S2 \text{ otherwise} \end{cases}$$

3. The event function New(x) has an interesting property. Suppose we want the evidence function $Ev(S)$ to depend not in any arbitrary way on the *string* expressing event S, but to depend on the actual meaning of S. To do this, we assume that if two event functions $S(a, \cdot)$ and $S(b, \cdot)$ have the same set $\{x\}$ for which they are true (or false) then $Ev(S(a, x)) = Ev(S(b, x))$ for every x (equivariance invariant).

But under this assumption, we can see that the sequence $Ev(\text{New}(x)), x = 1, ...$ is different from the sequence $Ev(S(m, x)), x = 1, ...$ of every $S(m, x), x = 1, ...$ This means that such evidence function cannot be stated in the language of arithmetic and logic.

One resolution of this can be that the statement New is a measurable event by both $\{p\}$, but is not any of $S(m,)$, and is rather a limiting event in the uncountable sigma algebra of natural numbers. The problem there is that the form of a candidate evidence function may be determined mostly by the implied restrictions of the uncountable sets that are outside 1st-order logic and difficult to interpret. A possible approach would be to require that an evidence function has desired properties in the 1st order sets and not necessarily in the dominant uncountable sets outside.

3. Computability of self referenced evidence functions.

4. Self-reproduction of evidence function.

Denote “evidence” for a distribution $a(\cdot)$ vs $b(\cdot)$ based on x, as $Ev\{a(.), b(.), x\}$; if evidence means everything we can say about that comparison, then we can capture that meaning by saying that <u>if Ev introspects on its own product</u> (self reference), <u>it should give the same result</u>:

$$Ev\Big[a(.), b(.), \boxed{Ev\{a(.), b(.), x\}}\Big] \text{ should be equal to } \boxed{Ev\{a(.), b(.), x\}} \tag{1}$$

Mathematically this property is called idempotency, but its meaning here is ”self-awareness” so it can reproduce itself.

To fix ideas, say the numerical (non distributional) part of the output of $Ev\{a(.), b(.), x_1\}$ is any comparison $f(a(x_1), b(x_1))$. The application of $Ev\{a(.), b(.), f(a(x_1), b(x_1))\}$ means to (i) find the set of observations that give the same f as $f(a(x_1), b(x_1))$, say $\{x_1, ..., x_k\}$; (ii) find the two probabilities:

$$a\{f(a(X), b(X)) = f(a(x_1), b(x_1))\} = a(x_1) + ... + a(x_k)$$
$$b\{f(a(X), b(X)) = f(a(x_1), b(x_1))\} = b(x_1) + ... + b(x_k),$$
$$\text{and calculate } f\{a(x_1) + ... + a(x_k), b(x_1) + ... + b(x_k)\}$$

Self reproduction (idemptotency) here then means :

$$f\{a(x_1) + ... + a(x_k),\ b(x_1) + ... + b(x_k)\} = f\{a(x_1), b(x_1)\}, \tag{2}$$

which will also equal to f applied to any of $x_2, ..., x_k$, by construction.

From here on, for simplicity say there are just two observations, x_1, x_2, that give the same f, that is, $f(a(x_1), b(x_1)) = f(a(x_2), b(x_2))$, or more compactly, $f(a1, b1) = f(a2, b2)$, i.e., $(a1, b1)$ and $(a2, b1)$ are pairs of probabilities that are in the same equivalence class of Ev. Then $f(a1 + a2, b1 + b2)$ (the Ev() in the LHS of (2) applied on the first evaluation), is also equal to $f\{a1, b1\}$, the RHS of (2). Fix that value to $f0$, and let $B(a)$ indicate the value of b that preserves $f(a, b) = f0$ for a given a (we assume uniqueness).

So, $f0 = f(a, B(a)) = f(a, B(a))$ (identity), which by idempotency, is $f(a + a, B(a) + B(a))$. But by definition of B, at the same contour value $f0$, the latter must equal also to $f(a + a, B(a + a))$. So, $B(2a) = 2B(a)$.

Add $B(2a) = 2B(a)$ to $B(2d) = 2B(d)$ for small d, to get $B(2a + 2d) = 2B(a) + 2B(d)$, subtract $B(2a)$: $B(2a + 2d) - B(2a) = 2B(d)$. Dividing over $2d$, and assuming the limit $B(2d)/(2d)$ exists, it will be constant over a. So, from the left had side limit, $B'(2a)$ is a constant when $f = f0$. So, $B(a)$ is linear in a, say $B(a) = \lambda a + \gamma$. Then, for any number N, $f(Na, B(Na)) = f(Na, (Na)\lambda + \gamma)$, so $f(a, B(a)) = f(a, a\lambda + \gamma/N)$, so $\gamma = 0$. Therefore, f is fixed at f_0 when $B(a) = \lambda_{f0} a$, or $\boxed{b/a = \lambda_{f0}}$.

5. Stable and unstable ancillaries: a review

Consider a *nominal* model $p(x \mid \theta)$ for data x, and denote the minimal sufficient statistic (MSS) likelihood ratio $\ell r(x)$ (θ is suppressed). Suppose the model has B maximal ancillary statistics: $\text{Max}_1(x), ..., \text{Max}_B(x)$, where maximal means not a function of any other ancillary (Basu). Following Durbin's suggestion, we focus only on ancillaries that are functions of the data through the MSS $\ell r(x)$.

Review on stability with respect to each other, and conditionality, and transitivity.

6. Stable likelihood and evidence

We want to consider the problem where some of the above (nominal) maximal ancillaries break and are no longer ancillary. This can represent physical or sampling settings where an ancillary statistic can be part of the measurement device or process. To represent this, we consider a more *practical* model in which we only know the following:

(1) only some but not all of the above (candidate) maximals remain true ancillary statistics, and we do not know which;

(2) for every nominal maximal, we know the nominal likelihood function given that maximal; however, by (1), this function may not be the correct likelihood function;

(3) for the nominal maximals that remain ancillaries, the nominal conditional likelihood from (2), namely, $p\{x \mid \text{Max}(x), \theta\}$, is still the correct conditional likelihood, however we do not know which are these maximals.

With this problem, we still want to find and condition on any uninformative variation, so we task ourselves to find functions that are still definitely ancillary.

One way to achieve this is to construct a statistic A^{stable} that is both, a deterministic function of Max_1 alone, and,..., and a deterministic function of Max_B alone.

Proposition. Any statistic A^{stable} as above is an ancillary statistic in the practical model, and the likelihood function $p(x \mid A^{stable}(x), \theta\}$ is the same in the nominal as in the practical models.

Proof. Construction of A^{stable} means that for each observation x, the event identified by $A^{stable}(x)$, is, for every $\text{Max}_k(\cdot)$, an exact union of some of its disjoint events. Therefore, $A^{stable}(x)$ will be also an exact union of some disjoint events of the maximal ancillary statistic(s) that remain(s) ancillary. Since for such statistic, a union of disjoint events is still ancillary, $A^{stable}(x)$ is an ancillary event as well. The result follows.