Computationally-based Agnostic Induction/Inference

Terrence L. Fine
School of Electrical & Computer Engrng
Cornell University

UCB on 29 November 2007

Background to Inductive Inference

- A long-studied but less well understood probabilistic notion relates to *indeterminacy*---the given facts or *evidence* (expressed in a formal language) only partially *deductively* support the *conclusion* or *hypothesis*.

Standard Statistical Model

- ``Evidence" maps to ``Data"
- ``Hypothesis" maps to ``Model" and/or unobserved data.
- Not precise correspondence---our approach is more general.
- We do not start with a likelihood function or prior over models.

- There is no logical deduction from the evidence *e* to the hypothesis *h*.
- Motivated by early work of Keynes, we let *h*|*e* represent the *degree of inductive support* lent to *h* by *e*.

- Numerical assessments of partial degrees of support were first proposed by Leibniz.
- This was meant for the area of legal trials (Leibniz' concern) but more generally that of *inductive reasoning or inference,* including statistical inference.

- The ``standard approach'' is to assume that there exists a conditional probability P and $h|e >= h'|e'$ if and only if $P(h|e) >= P(h'|e')$.

- Where does the conditional probability come from in a general setting?
- Perhaps it comes from an argument made by Solomonoff in the early 1960s when he derived a universal semicomputable prior.
- However, assigning any numerical value, of whatever origin, to the support lent by evidence to hypothesis will automatically make accessible a complete ordering according to this support.

- Inductive relations have an inherent indefiniteness or imprecision that may not be expressible by standard numerical probability.
- We hold that standard probability and upper and lower probability are too expressive to model inductive support.
- Nor need all pairs $h|e$ and $h'|e'$ be comparable.

- We assume only the existence of a *partial ordering* between pairs *h*|*e* and *h'*|*e'* that is understood as ``the inductive support lent by evidence *e* to hypothesis *h* is as at least as great as that lent by *e'* to *h'*.

- Our proposal places the thinking of Keynes in the context of computational or algorithmic considerations that relate to Kolmogorov complexity.
- We depart from a real-valued assessment of inductive support and consider instead an ordering of levels of inductive support.

Syntax

- We take for our simple formal language one with the *syntax* of the set $\mathcal{B} = \{0, 1\}^*$ of all finite-length binary-valued strings.
- This set of strings is closed under *concatenation*, an operation allowing us to form a new string from two given strings.

Semantics

- The formal language does not preserve explicitly the "true-false" semantics of the original language.
- We expect that statements in the natural language will be *expanded* before being encoded into the formal language so as to appear with their various meanings.

- For example, if the word "swan" appears in the natural language statement then there will be a full presentation about swans including other text usages, images, and video information on swans will all be encoded in the formal language.

- We can formally carry out this process through an *oracle machine* consisting of a *Turing machine* T and an *oracle* G.
- When a request is made by T to G regarding a particular string corresponding to a natural language statement, then G expands this string by providing a coded collection of other natural language statements that explain the given string.

- These latter represent the range of meanings of the given string.
- *Meaning* is *emergent* from the large collection of interrelated materials (e.g., dictionary definitions of "swan", images and videos containing "swans") identified by the oracle.

Understanding and Reasoning

- Reasoning is modeled by the selection of an effective computational procedure (a model of brain functioning) to transform given information in the form of a string into a *conclusion* string.
- Such conclusions may only be provisional steps in a reasoning process.

- **The reasoning process is carried out by a partial recursive function that we think of as a *Turing machine*.**
- **Such a model of reasoning need be no more unique than are human reasoners.**
- **Arbitrarily choose a Turing machine T.**
- **We will subsequently examine the effect of other choices.**

- **Inductive support will be based upon a notion of *explanation*.**
- **Explanation, in turn, will be based upon how we compute one statement from another.**
- **The concept of computation is that of *universal Turing machines* (UTMs) denoted T, T_1, T_2, etc.**

- **Modeling reasoning plausibly requires a, not necessarily unique, choice of TM.**
- **In what follows we assume that reasoning relies upon a *universal Turing machine (UTM)*.**
- **A UTM (they are all equivalent in a sense to be clarified later) is the most powerful machine in that it is an ideal general purpose computer.**

Explanations

- **String p is an *explanation* for a given string h if and only if**

  $p \in \mathcal{B}$ **and** $T(p) = h \in \mathcal{B},$ **is defined.**
- **Equivalently, we can reason from an explanation p to a conclusion h.**

- String p is a ***supplementary explanation*** for a given string h, in the presence of evidence e, only if

  $p \in \mathcal{B}$ and $T(pe) = h \in \mathcal{B}$, is defined,

  where pe is the concatenation of the two strings.
- We can reason from e augmented by p to h.

- Since T is a UTM, for any pair e, h of evidence and hypothesis strings, there are infinitely many supplementary explanation strings p.

Support Sets

- Let the ***range set***

  $$\mathcal{R}(p) = \{h' : (\exists e)\ T(pe) = h'\}.$$
- The ***explanatory support set*** $\mathcal{P}(h|e)$ is the set of all supplements p to e such that

  $$\mathcal{P}(h|e) = \{p : p \in \mathcal{B},\ T(pe) = h,\ \mathcal{R}(p) \text{ is countably infinite.}\}$$

- $\mathcal{P}(h|e)$ can also be truncated to a finite set, as follows. First observe that

  $$(\forall h)(\exists h^*)(\forall e)\ T(h^*e) = h \text{ and } |h^*| \leq c + |h| + 2\log|h|.$$
- Why then consider supplementary explanations longer than h^*?
- However, we do not take this step.

Properties of $\mathcal{P}(h|e)$

Assume a UTM T.

- **For all h, e, $\mathcal{P}(h|e)$ is a countably infinite set.**
- **If $h \neq h'$, then for any e, $\mathcal{P}(h|e) \perp \mathcal{P}(h'|e)$ (disjoint sets).**

- **For a given UTM T, the mapping from ordered pairs $h|e$ to support sets $\mathcal{P}(h|e)$ is many-to-one.**
- **There can exist $e \neq e', h \neq h'$ such that $\mathcal{P}(h|e) = \mathcal{P}(h'|e')$.**

- **There is no Turing machine that given as input any triple h, e and integer k produces as output the k-th element of the set $\mathcal{P}(h|e)$ in a given effective enumeration of strings.**
- **Restated, $\mathcal{P}(h|e)$ is not an effectively computable function of h, e.**

Comparative Inductive Support

- **Our object of interest is $\{h|e\}$, a set of ordered pairs of hypothesis h and evidence e, with $h|e$ read as "the degree of inductive support lent by e to h".**
- **We neither assume nor expect that $h|e$ will have a numerical representation or evaluation.**

- We assume only the existence of a ***partial order*** $h|e \succsim h'|e'$ that is read "the inductive support lent by e to h is at least as great as that lent by e' to h'".

Understanding $\succsim$ through Representation

- The "standard approach" is to assume that there exists a conditional probability P and

$$h|e \succsim h'|e' \iff P(h|e) \geq P(h'|e').$$

- P in Carnap's approach comes from choosing a prior on statements in first-order predicate logic that has invariance properties such as under interchange of individuals.

Representation through a Partial Order

- We ***postulate*** that there exists a partial order $\succsim_{\mathcal{P}}$ and a homomorphism from $\{h|e\}$ to $\{\mathcal{P}(h|e)\}$ that maps the order $\succsim$ to $\succsim_{\mathcal{P}}$ through

$$h|e \succsim h'|e' \iff \mathcal{P}(h|e) \succsim_{\mathcal{P}} \mathcal{P}(h'|e').$$

- This postulate is needed because we have established that there is a many-to-one mapping from $\{h|e\}$ to $\{\mathcal{P}(h|e)\}$
- This approach is directly motivated by work in mathematical psychology associated with Suppes and Luce [10, 13].

- We postulate that it is only the ***lengths of strings*** in $\mathcal{P}(h|e)$ that are relevant to inductive support.
- Other possible distinctions between p, q are: sequence composition; whether p is a permutation of q, etc.
- These have not been considered owing to my lack of intuition as to their inductive consequences.

- A shorter supplementary explanation provides more inductive support than does a longer one—Occam's razor.
- We allow for ***multiple explanations*** and do not insist upon a single 'best' one—Epicurus.
- A greater number of multiple explanations provides increased robustness to an inference.

- For p to be a supplementary explanation in an analysis of the degree to which e supports h, $T(pe)$ must depend upon e.
- We can revisit this and consider other syntactical distinctions between strings if we can identify an intuition regarding their inductive significance.

Axiom of Sequence Representation

- The preceding motivates us to reduce $\mathcal{P}(h|e)$ to a multiset

$$\mathcal{L}_{h|e} = \{|p| : p \in \mathcal{P}(h|e)\},$$

of the repeated lengths of strings,

- and in turn to a mathematically equivalent sequence or function

$$L_{h|e} : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$$

with $L_{h|e}(i) = \|\{|p| : |p| = i,\ p \in \mathcal{P}(h|e)\}\|.$

- The desired homomorphism is now to a partial order $\succsim_L$ on the set $\{L_{h|e}\}$.
- **Axiom I:** $h|e \succsim h'|e' \iff L_{h|e} \succsim_L L_{h'|e'}$.
- This representation provides a link to the work on case-based reasoning by Gilboa and Schmeidler [4].

Axiom of Dominance

- **Axiom 2:**

  $\Big((\forall i)\ L_{h|e}(i) \geq L_{h'|e'}(i)\Big)$

  $\Rightarrow L_{h|e} \succsim_L L_{h'|e'} \Rightarrow h|e \succsim h'|e'$.

  If, in addition, there is an index j such that $L_{h|e}(j) > L_{h'|e'}(j)$, then $L_{h|e} \succ_L L_{h'|e'}$ and $h|e \succ h'|e'$.
- This dominance axiom is motivated by the inductive desirability of multiple explanations.

- Replacing one-sided implication by a two-sided implication would provide too stringent a notion of inductive support that would only rarely be applicable.

Axiom of Redistribution

- Our philosophical position is that a large number of smaller supplementary explanations is more convincing than a similar number of longer supplementary explanations. We capture this through an Axiom 3 of *redistribution* defined as follows.

- **Given any $L = \{L(i)\}$, choose indices $m < n$ and define L' as**

  $(\forall i \neq m, n)\ L'(i) = L(i)$
  and $L'(m) = L(n),\ L'(n) = L(m).$

  (Hence the two sequences agree except at indices m, n where they exchange values.)

- **We postulate that**

$$L(m) > L(n) \Rightarrow L \succ_L L';$$
$$L(m) < L(n) \Rightarrow L' \succ_L L.$$

Axiom of Additive Combination

- **We introduce an Axiom 4 of *additive combination* that is drawn from Gilboa and Schmeidler [4], p.67.**

- $L_1 \succsim_L L_2\ (L_1 \succ_L L_2)$ **and** $L_3 \succsim_L L_4$
  $\Rightarrow L_1{+}L_3 \succsim_L L_2{+}L_4\ (L_1{+}L_3 \succ_L L_2{+}L_4).$

- **Adding such sequences together will be shown to be meaningful for sets of hypotheses.**

- **Equating $L_2 = L_3$ implies that $\succsim_L$ is *transitive*.**

- **It does not imply that $\succsim_L$ is complete, as assumed in Gilboa and Schmeidler [4].**

- **Nor will we require completeness.**

- **Theorem 1 establishes the Consistency of Axioms 1,2,3,4, in that they are consistent with $\succsim_L$ being a complete order.**

- **This complete order can be generated by a probability mass function on strings that assigns all strings of the same length the same probability and strings of greater length a lower probability.**

Axiom of Stochastic Dominance

- **Stochastic dominance holds between random variables, X stochastically dominates Y, if the corresponding cumulative distribution functions satisfy**

$$(\forall z \in \mathbb{R})\ F_X(z) \leq F_Y(z).$$

- **In our context we define an *unnormalized cdf***

$$(\forall i)\ F_{h|e}(i) = F_{L_{h|e}}(i) = \sum_{j \leq i} L_{h|e}(j).$$

- **If $i < 0$ then $F_{h|e}(i) = 0$.**

- **Axiom 5 is given by**

$$(\forall i)\ F_{L_{h|e}}(i) \geq F_{L_{h'|e'}}(i) \Rightarrow L_{h|e} \succsim_L L_{h'|e'}$$
$$\iff h|e \succsim h'|e'.$$

- **If, in addition,**

$$(\exists j\ F_{L_{h|e}}(j) > F_{L_{h'|e'}}(j) \Rightarrow L_{h|e} \succ_L L_{h'|e'}$$
$$\iff h|e \succ h'|e'.$$

- **For hypothesis h given evidence e to be more strongly inductively supported than is $h'|e'$ requires that there be a larger number of shorter supplementary explanations for the former than for the latter.**

- **Theorem 2 asserts that when Axiom 5 holds for $\succsim$ then the partial order also satisfies Axioms 1, 2, 3, 4, and is a reflexive partial order.**

Degrees of Comparative Inductive Support

- **We use our representation to introduce a series of successively stronger notions of comparative inductive support by defining a family $\succ_\gamma$ as follows.**

- **For $\gamma \geq 1$,**

$$(\forall i)\ F_{h|e}(i) \geq F_{h'|e'}(i+\gamma) \Rightarrow h|e \succ_\gamma h'|e'.$$

- **Lemma 1 asserts:**
- **For $\gamma \geq 1$, the partial ordering is strict.**
- **Furthermore,**

$$(\gamma > \gamma') \Rightarrow \left(h|e \succ_\gamma h'|e' \Rightarrow h|e \succ_{\gamma'} h'|e'\right),$$

but not conversely.
- **We will use this notion of degrees of support to examine the question of the robustness of agnostic inductive support to changes in the UTM by which it is defined.**

Robustness with Respect to Choice of UTM

- **Theorem 3 asserts that given any two UTMs T_1, T_2, there are constants c_{12}, c_{21} such that**

$$(\forall p_1)(\exists p_2)\ |p_2| \leq c_{12} + |p_1|$$
$$\textbf{and } T_1(p_1) = q \implies T_2(p_2) = q;$$
$$(\forall p_2)(\exists p_1)\ |p_1| \leq c_{21} + |p_2|$$
$$\textbf{and } T_2(p_2) = q \implies T_1(p_1) = q.$$

- **A sketch of the proof of this theorem is based upon the existence of a self-limiting string t_{12} (Gödel number), which when prepended to a given program p_1, results in a concatenation p_2 that informs UTM T_2 to emulate T_1 and provides it with input p_1.**

Implication for Robustness

- **We consider the effect on comparative inductive support of a change from UTM T_1 to UTM T_2.**
- **Let L_i denote the sequence of ordered string length occupancies corresponding to use of T_i for $\mathcal{P}(h|e)$; L_i' denotes the corresponding sequence for $\mathcal{P}(h'|e')$.**

- **It follows then that for any two UTMs T_1, T_2, there exists a constant c such that for all $h|e, h'|e'$ and $\gamma > 2c$,**

$$L_1 \succ_\gamma L_1^{'} \Rightarrow L_2 \succ_{\gamma-2c} L_2^{'}.$$

- **If $h|e$ is sufficiently more strongly supported than $h'|e'$, then a change from UTM T_1 to a "neighboring" T_2 will preserve the weaker conclusion that $h|e$ is still at least as strongly supported as $h'|e'$.**

Extending the Partial Order

- **Any extension of $\succsim, \succ, \succ_\gamma$ will introduce a comparison that contradicts the Axiom of Stochastic Dominance.**
- **Pending a satisfactory additional axiom, we expect that comparative inductive support is only a partial ordering.**
- **We expect there to exist incomparable pairs $h|e$ and $h'|e'$.**

- **There is room for an extension to *subsets of hypotheses.***
- **Assume that we are given specific evidence e. Then we know from Lemma 1 that $\mathcal{P}(h|e)$ and $\mathcal{P}(h'|e)$ are disjoint sets for $h \neq h'$ and common evidence e.**

- **This suggests that to a family of hypotheses H we associate**

$$\mathcal{P}(H|e) = \bigcup_{h\in H} \mathcal{P}(h|e)$$

and define $L_{H|e} = \sum_{h\in H} L_{h|e}.$

- **By the disjointness of the support sets for common e, no programs are double-counted in our summation.**

- Of course, the new object $L_{H|e}$ is again a sequence of nonnegative integers and is ordered by our original $\succsim_L$ without any need for an extended definition.
- Given a set H of hypotheses, and common evidence e, we extend $\succsim$ to sets of hypotheses through

$$H|e \succsim H'|e' \iff L_{H|e} \succsim_L L_{H'|e'}.$$

- For common evidence e we find that

$$H \supseteq H' \Rightarrow (\forall i \geq 0)\ L_{H|e}(i) \geq L_{H'|e} \Rightarrow H|e \succsim H'|e.$$

- This is a plausible conclusion. A set of hypotheses is more strongly supported than any of its proper subsets.

Of What Use is Inductive Support?

- We can make inferences and reach conclusions based upon degrees of comparative inductive support.
- We can also use inductive support to make rational decisions dependent upon consequences, although our current ability to do this is rather anemic.

Inferences, Plausibility, Conclusions

- Perhaps we have reason to focus on e and on a subset $\mathcal{H}$ of hypotheses.
- Agnostic comparative inductive support provides insight or understanding into the plausibility of statement h when we assume or know e.
- If, say, $h|e \succ_{100} h'|e$, then we might discard the relatively implausible h' and reduce $\mathcal{H}$.

- We can say that $h|e$ is ***undominated in*** $\mathcal{H}$ if there does not exist $h'|e' \in \mathcal{H}$ such that $h'|e' \succ h|e$.
- Our notion of comparative inductive support enables us to reduce $\mathcal{H}$ to its subset of undominated pairs $h|e$.
- This is akin to the use of ***admissibility*** in statistical analysis to restrict consideration of possible decision rules.

- While this reduction might be helpful it does not take into account the ***consequences*** of choosing an hypothesis h given evidence e.

Inductively Supported Decision-Making

- If we think of this as ***personal or individual*** decision-making then we need to assume that the individual accepts the assessments of agnostic comparative inductive support as their personal beliefs.
- Such acceptance might be even more plausible for ***group-based*** decision-making.

Components of the Decision-Making Problem

States of the World: Assume a set $\mathcal{H} = \{h\}$ of hypotheses that describe the "states of the world" or the alternatives that concern us.

Evidence or Data: Evidence $e \in \mathcal{E}$ is known to the decision-maker holds throughout the decision process.

Set of Consequences: There is a set $\mathcal{C} = \{c\}$ of the consequences of our actions or decisions. This set comes equipped with a complete order $\succsim_C$ that is anti-symmetric. We will assume $\mathcal{C} = \{c_1, \ldots, c_m\}$ to be a finite set with, without loss of generality,

$$i > j \iff c_i \succ_C c_j.$$

Decision Rule/Function: A decision is to be made about which element of $\mathcal{H}$ to select ("is true", although we lack the semantics to make this a natural statement). This decision is a function d of given evidence from a set $\mathcal{E} = \{e\}$,

$$d : \mathcal{E} \to \mathcal{H}, \; d \in \mathcal{D}, \; d(e) = \eta.$$

Consequences of Decisions: The consequence of a decision is provided by a ***gain*** function

$$g : \mathcal{H} \times \mathcal{H} \to \mathcal{C}.$$

When we decide $d(e) = \eta$, based on provided evidence $e \in \mathcal{E}$ and the "correct" hypothesis is h then we are awarded consequence $g(\eta, h) = c \in \mathcal{C}$.

Resulting Gamble: Given evidence e and a decision η we are awarded a ***gamble***

$$G(\eta) = \bigcup_{c \in \mathcal{C}} (\{h : g(\eta, h) = c\}, c)$$

that is a union of ordered pairs (H_c, c) consisting of the subset H_c of hypotheses paying off with consequence c when we make decision $d(e)$ and the consequence c itself.

Restated, there is a partition $\{H_c : c \in \mathcal{C}\}$ of $\mathcal{H}$ and a gamble in our setting is a set

$$G = \bigcup_{c \in \mathcal{C}} (H_c, c),$$

and a gamble is a subset of $\mathcal{H} \times \mathcal{C}$.

Objective: To partially order by $\succsim_{\mathcal{G}}$ the set $\mathcal{G} = \{G\}$ of gambles in accordance with the decision maker's preferences.

Preferences for Gambles

- **The preference order $\succsim_{\mathcal{G}}$ is in effect a joint order constructed from a combination of *independent type* out of the marginal order $\succsim$ on subsets $H|e$ of the form $\{h : g(d(e), h) = c\}$ and the marginal order on individual consequences given by $\succsim_{\mathcal{C}}$.**
- **This brings us to *comparative probability*, albeit in the context of partial orders.**

- **If $\mathcal{A}$ and $\mathcal{B}$ are families of subsets of sets Ω_A, Ω_B with non-null-equivalent elements $A_i \in \mathcal{A}, B_i \in \mathcal{B}$, then a comparative probability order $\succsim_{\mathcal{A}\times\mathcal{B}}$ of *independent type* satisfies**

$$(\forall A \in \mathcal{A})(\forall B_1, B_2 \in \mathcal{B})$$
$$(A, B_1) \succsim_{\mathcal{A}\times\mathcal{B}} (A, B_2) \iff$$
$$(\mathcal{A}, B_1) \succsim_{\mathcal{A}\times\mathcal{B}} (\mathcal{A}, B_2);$$

and similarly if we interchange As and Bs.

- ***Marginals* are preserved in the joint order if**

$$A \succsim_{\mathcal{A}} A' \iff (A, \Omega_B) \succsim (A', \Omega_B),$$

and similarly for the subsets in $\mathcal{B}$.

- **The existence of such joint orders depends upon characteristics of the marginal orders and has been studied by many, perspicuously by Kaplan [5] and Kaplan and Fine [6].**

- **We can narrow the class of acceptable partial preference order $\succsim_{\mathcal{G}}$ through dominance axioms of the kind discussed above.**
- **What is missing, however, from our approach is the key notion of a mixture distribution.**
- **While a mixture axiom seems innocent to many, it is indeed substantive, as evidenced by the work it does.**

- **In von Neumann-Morgenstern approaches to utility and subjective probability pioneered by de Finetti and Jimmie Savage, we must extend our known preferences to preferences between mixtures of those gambles we care about.**
- **We leave this matter unfinished.**