Notes on Inference Devices

Santa Fe Institute

Edward G. Huang

Summer 2018

1 Abstract

2 Introduction

2.1 Notation and Definitions

This manuscript utilizes standard notation taken from set theory and vector algebra. We clarify notation specific to Turing machine theory and inference devices.

Turing Machine Notation

$\mathbb{B}^*$	The space of all finite bit strings.
Λ	Symbol alphabet of a Turing Machine.
σ	A symbol on a Turing Machine tape.
Q	Set of finite states of a Turing Machine.
Δ	Transition function of a Turing Machine.
k	Number of tapes of a Turing Machine. The first tape is assumed to be read-only.
η	Non-halting state of a Turing Machine.

Inference Device Notation

U	Set of possible histories of the universe.
u	A history of the universe in U.
X	Setup function of an ID that maps $U \to X(U)$. A binary question concerning $\Gamma(u)$.
x	A binary question and a member of image $X(U)$.
Y	Single-valued conclusion function of an ID that maps $U \to \{-1, 1\}$. A binary answer of an ID for $X(u) = x$.
y	A single-valued answer, and member of image $Y(U) = \{0, 1\}$.
Γ	A function of the actual values of a physical variable over U, equivalent to $\Gamma(u) = S(t_i)(u)$.
γ	Possible value of a physical variable, a member of the image $\Gamma(U)$.
δ	Probe of any variable V parameterized by $v \in V$ such that : $\delta_v(v') = \begin{cases} 1 & \text{if } v = v' \\ -1 & \text{otherwise} \end{cases}$
$\wp$	Set of probes over $\Gamma(U)$.
$\mathcal{D}$	An inference device, consisting of functions (X, Y).
ξ	A function $\Gamma(U) \to \overline{X}$.
Γ^{-1}	Inverse. Given a function Γ over U, $\Gamma^{-1} = \bar{\Gamma} \equiv \{\{u : \Gamma(u) = \gamma\} : \gamma \in \Gamma(U)\}$.
$\overline{\Gamma}$	Given a function Γ over U, the partition of U given by Γ^{-1}.
$>$	Weak inference: a device $\mathcal{D}$ weakly infers Γ *iff* $\forall \gamma \in \Gamma(U), \exists x \in X(U)$ s.t. $\forall u \in U$, $X(u) = x \implies Y(u) = \delta_\gamma(\Gamma(u))$.
$\gg$	Strong inference: a device (X, Y) strongly infers a function (S, T) over U *iff* $\forall \delta \in \wp(T)$ and all $s \in S(U)$, $\exists x$ *such that* $X(u) = x \implies S(u) = s, Y(u) = \delta(T(u))$.
$\mathcal{C}_\mu(\Gamma; \mathcal{D})$	Inference complexity.
μ	Measure defined for $u \in U$.

2.2 Turing Machines

Deterministic Turing Machines

Arora and Barak denote a Turing Machine (TM) as $T = (\Lambda, Q, \Delta)$ containing:

1. An *alphabet* Λ of a finite set of symbols that T's tapes can contain. We assume that Λ contains a special blank symbol B, start symbol S, and the symbols 0 and 1.

2. A finite set Q of possible states that T's register can be in. We assume that Q contains a special start state q_s and a special halt state q_h.

3. A transition function $\Delta : Q \times \Lambda^k \to Q \times \Lambda^{k-1} \times \{L, \mathcal{S}, R\}^k$, where $k \geq 2$, describing the rules T use in performing each step. The set $\{L, \mathcal{S}, R\}$ denote the actions *Left, Stay,* and *Right*, respectively.

Suppose T is in state $q \in Q$ and $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ are the symbols on the k tapes. Then $\Delta(q, (\sigma_1, \ldots, \sigma_k)) = (q', (\sigma_2^{'}, \ldots, \sigma_k^{'}), z)$ where $z \in \{L, \mathcal{S}, R\}^k$ and at the next step the σ symbols in the last $k-1$ tapes will be replaced by the σ' symbols, the machine will be in state q, and the k heads will move *Left, Right* or *Stay.* This is illustrated in Figure 2.1.

Figure 2.2.1. The transition function Δ for a k-tape Turing Machine

$(q, (\sigma_1, \ldots, \sigma_k))$				$(q', (\sigma_2^{'}, \ldots, \sigma_k^{'}), z)$				
Input symbol	Work/output symbol read	$\cdots$	Current state	New work/output tape symbol	$\cdots$	Move work/output tape	$\cdots$	New state
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
σ_1	σ_i	$\ddots$	q	$\sigma_i^{'}$	$\ddots$	z_i	$\ddots$	$q^{'}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$

Remark: Λ can be reduced to $\mathbb{B} = \{0, 1\}$ and k can be reduced to 1 without loss of computational power. Then, any Turing Machine can be expressed as a partial recursive function mapping $\mathbb{B}^* \to \mathbb{B}^* \cup \eta$, where η is the undefined non-halting output. Since $|\mathbb{B}^* \times \mathbb{B}^* \cup \eta| = |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$, the set of all Turing Machines is countably infinite.

Non-deterministic Turing Machines

Non-deterministic Turing Machines (NDTM) differ from deterministic Turing Machines by having two transition functions Δ_0, Δ_1 and a special state q_{accept}. From Arora and Barak:

> When a NDTM M computes a function, we envision that at each computational step M makes an arbitrary choice as to which of its two transition functions to apply. For every input x, we say that $M(x) = 1$ if there exists some sequence of these choices (which

we call nondeterministic choices of M) that would make M reach q_{accept} on input x. Otherwise - if every sequence of choices makes M halt without reaching q_{accept} - then we say that $M(x) = 0$.

If $M(x) = 1$, we say that M accepts the input x. There are two ways to interpret the choice of update function to use in a NDTM. We can either assume that the NDTM chooses updates that will lead to an accepting state, or we can assume that the machine branches out into its choices such that it has a "computation tree" and if any of the branches reaches the accepting state then the machine accepts the input. From this second interpretation, the computational power of DTMs to NDTMs is analogous to the computational complexity of P to NP.

3 Weak Inference

Definition Two functions Γ_1 and Γ_2 with the same domain U are **(functionally) equivalent** iff the inverse functions Γ_1^{-1} and Γ_2^{-1} induce the same partitions of U, i.e., iff $\overline{\Gamma_1} = \overline{\Gamma_2}$.

Lemma *A function Γ can be weakly inferred by a device $\mathcal{D}$ if $|\Gamma^{-1}(\gamma)| \geq 2$ for any $\gamma \in \Gamma(U)$.*

Proof Let $U := \mathbb{N}$. Enumerate $\gamma \in \Gamma$ as $1, 2, \ldots, i, \ldots, n$ and define $V^i = \{u : \Gamma^{-1}(i) = u\}$. Then enumerate each element in V^i as $i_1, i_2, \ldots, i_j, \ldots, i_m$. Continue to define X and Y as

$$X(i_j) = \begin{cases} a_i & \text{if } j = 2 \\ b_i & \text{otherwise} \end{cases} \qquad Y(i_j) = \begin{cases} -1 & \text{if } j = 2 \\ 1 & \text{otherwise.} \end{cases}$$

Then for each $i \in \Gamma(U)$ choose $x = b_i$ to force $Y(X^{-1}(b_i)) = \delta_i(\Gamma(X^{-1}(b_i)) = i) = 1$. □

Corollary *A function Γ can be weakly inferred by a device $\mathcal{D}$ if $|\Gamma(U)| \geq 3$.*

Proof Let $U := \mathbb{N}$. Enumerate $\gamma \in \Gamma(U)$ as $1, 2, \ldots, i, \ldots, n$ and define $V^i = \{u : \Gamma^{-1}(i) = u\}$. Then enumerate each element in V^i as $i_1, i_2, \ldots, i_j, \ldots, i_m$. Continue to define X and Y as

$$X(i_j) = \begin{cases} a & \text{if } i = 1 \text{ and } j = 1 \\ b & \text{if } i = 2 \text{ and } j = 1 \\ c & \text{if } i = 3 \text{ and } j = 1 \\ d & \text{otherwise.} \end{cases} \qquad Y(i_j) = \begin{cases} 1 & \text{if } i = 1 \text{ and } j = 1 \\ -1 & \text{otherwise.} \end{cases}$$

Then for each $i \in \Gamma(U)$ to force $Y(X^{-1}(x)) = \delta_i(\Gamma(X^{-1}(x)) = i)$ choose $x = a$ if $i = 1$, $x = c$ if $i = 2$, or else choose $x = b$. □

Countable Inference Theorem *A set of inequivalent countable functions A^* can be weakly inferred by a device if each function $A_i \in A^*$ is independently inferrable.*

Proof Let $U := \mathbb{N}$ and fix any A^*. Let a_i, a_j, a_k represent any distinct three elements in $A_i(U)$. Write $V = \{1, 2, 3\} \subset U$. The following table represents all possible combinations of a_i, a_j, and a_k over V.

u	$X(u)$	$Y(u)$	$A_1(u)$	$A_2(u)$	$A_3(u)$	$A_4(u)$	$A_5(u)$	$\dots$
1	1	-1	a_i	a_i	a_i	a_i	a_i	$\dots$
2	2	-1	a_i	a_j	a_j	a_j	a_i	$\dots$
3	3	1	a_i	a_k	a_j	a_i	a_j	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Note that setting $X(u) = u$, $Y(1) = -1$, $Y(2) = -1$, and $Y(3) = 1$ immediately satisfies weak inference for any function A_i that is functionally equivalent to A_1, A_2, A_3 and A_4 over V. This holds regardless of the values of $Y(u)$ and $A_{i\in\{1,2,3,4\}}(u)$ that follow for $u > 3$. This is shown by selecting x for the following cases:

Case 1: $\overline{A_i(V)} = \overline{A_1(V)}$.
Choose $x = 3$ for $a = a_i$ or otherwise choose $x = 2$.

Cases 2, 3, 4: $\overline{A_i(V)} = \overline{A_2(V)}$ or $\overline{A_3(V)}$ or $\overline{A_4(V)}$.
Choose $x = 2$ for $a = a_i$ or otherwise choose $x = 1$.

Now we need to guarantee weak inference for A_i that are functionally equivalent to A_5 over V. Enumerate each $A_i \in A^* : \overline{A_i(V)} = \overline{A_5(V)}$ as $B_4, B_5, \dots, B_i, \dots, B_n$. To satisfy weak inference for any A_i and B_i, define X, Y more explicitly as

$$X(u) = u, \qquad Y(u) = \begin{cases} -1 & \text{if } u = 1, 2 \text{ or if } B_i(u) = a_j, a_k : X(u) = i \\ 1 & \text{if } u = 3 \text{ or if } B_i(u) = a_i : X(u) = i \\ -1 & \text{otherwise.} \end{cases}$$

For each $a \in B_i(U)$ to force $Y(X^{-1}(x)) = \delta_a(B(X^{-1}(x)))$ choose $x = i$ if $a = a_i$ or otherwise choose $x = 1$. $\square$

4 Strong Inference

In the next three examples we examine strong inference of integer-valued functions.

Example Let $T(U) = \{0, 1\}$ and $S(U) = \{0, 1, 2\}$. We construct (X, Y) in the table at the left such that it strongly infers (S, T). The right table indicates x for each s, δ such that the definition of strong inference is satisfied:

u	$X(u)$	$Y(u)$	$S(u)$	$T(u)$
1	1	1	0	0
2	2	-1	0	0
3	3	1	1	0
4	4	-1	1	0
5	5	1	2	1
6	6	-1	2	1

$s \setminus \delta$	δ_0	δ_1
0	1	2
1	3	4
2	6	5

Example Let $T(U) = \{1, 2, 3\}$ and $S(U) = \{1, 2, 3, 4, 5\}$. Again, we construct (X, Y) in the table at the left such that it strongly infers (S, T). The right table indicates x for each s, δ such that the definition of strong inference is satisfied:

u	$X(u)$	$Y(u)$	$S(u)$	$T(u)$
1	1	1	1	1
2	2	-1	1	1
3	3	1	2	1
4	4	-1	2	1
5	5	1	3	2
6	6	-1	3	2
7	7	1	4	2
8	8	-1	4	2
9	9	1	5	3
10	10	-1	5	3

$s \setminus \delta$	δ_1	δ_2	δ_3
1	1	2	2
2	3	4	4
3	6	5	6
4	8	7	8
5	10	10	9

Example Let $T(U) = \{1, 2, 3\}$ and $S(U) = \{1, 2\}$. In this example, the inferred function $f : S \to T, f(s) = T(S^{-1}(s))$ is not single-valued.

u	$X(u)$	$Y(u)$	$S(u)$	$T(u)$
1	1	-1	1	1
2	2	-1	1	2
3	3	1	1	3
4	4	-1	2	1
5	5	-1	2	2

$s \setminus \delta$	δ_1	δ_2	δ_3
1	2	1	1
2	5	4	4

5 Inference of Turing Machines

Theorem *A deterministic Turing machine* (Λ, Q, Δ) *can be strongly inferred by a device iff*

$$\forall s \in S(U),\ |S^{-1}(s)| \geq 2.$$

This holds for both the representation of a Turing machine as a partial recursive function and the representation as an update function.

Proof First examine the partial function case. Let f be the partial recursive function that describes the given Turing machine tuple. Let $U := \mathbb{N}$. Choose any convenient single valued surjective function $S : U \to \mathbb{B}^*$ and define $T : U \to \mathbb{B}^* \cup \eta$ by $T(u) = f(S(u))$ as the single-valued function mapping U to the halting and non-halting outputs of f. Then f can be written as the single-valued mapping $S \to T$ by $f(s) = T(S^{-1}(s))$.

Enumerate the elements of $S(U)$ as $1, 2, \ldots, s, \ldots$. Let $V^s = \{u : S^{-1}(s) = u\}$ for $s \in S(U)$. Similarly enumerate the elements of V^s as $s_1, s_2, \ldots, s_{|V^s|}$. Then define X and Y as follows:

$$X(s_i) = \begin{cases} a_s & \text{if } i = 1 \\ b_s & \text{otherwise} \end{cases} \qquad Y(s_i) = \begin{cases} 1 & \text{if } i = 1 \\ -1 & \text{otherwise} \end{cases}$$

Note that the condition $|S^{-1}(s)| \geq 2$ is required to guarantee $Y(V^s) = \{1, -1\}$. For each pair $(s, \delta_{t \in T(U)})$, to force $S(u) = s$ and $Y(u) = \delta_t(T(u))$, choose $x = a_s$ if $t = T(s_1)$ or otherwise choose $x = b_s$. Since the choices of s and t were arbitrary, this holds for all (s, δ_t) pairs.

Now consider the update function that describes the given Turing machine. Recall that the update function is written as $\Delta : Q \times \Lambda^k \to Q \times \Lambda^{k-1} \times \{L, \mathcal{S}, R\}^k$, $k \geq 2$. Consider a convenient single-valued surjective function $S : U \to Q \times \Lambda^k$ representing the possible inputs for a Turing Machine and a corresponding single-valued $T : U \to Q \times \Lambda^{k-1} \times \{L, \mathcal{S}, R\}^k$ as $T(u) = \Delta(S(u))$. Observe that Δ can be written as the single-valued function $\Delta(s) = T(S^{-1}(s))$. Then define V^s, X, Y, and choose x for each (s, δ_t) as described in the preceding portion of the proof. Hence, the claim holds for the update function of a Turing machine.

To show that the condition is necessary for either representation, suppose that $|V^s| < 2$ for some s. If $V^s = \varnothing$ then there exists no x that can force $S = s$. If $|V| = 1$, then we can assign $Y(s_i) = y \in \{-1, 1\}$. However, whichever value is assigned, there exists a t such that $\delta_t(T(s_i)) \neq Y(s_i)$ since $|T(U)| \geq 2$. $\square$

Remark: Conventionally all functions over U must have a range of at least two elements. This implies that Turing machines that never halt cannot be strongly inferred.

6 Inference Complexity

Definition Let $\mathcal{D}$ be an inference device and Γ be a function over U where $X(U)$ and $\Gamma(U)$ are countable and $\mathcal{D} > \Gamma$. Let the **size** of $\gamma \in \Gamma(U)$ be written as $\mathcal{M}_{\mu:\Gamma(\gamma)} = -\ln[\int_{\Gamma^{-1}(\gamma)} d\mu(u)1]$ such that $d\mu$ denotes a measure over U. Then the **inference complexity** of Γ with respect to $\mathcal{D}$ and measure μ is defined as:

$$\mathcal{C}_\mu(\Gamma;\mathcal{D}) \triangleq \sum_{\delta \in \wp(\Gamma)} \min_{x:X=x \implies Y=\delta(\Gamma)} [\mathcal{M}_{\mu,X}(x)]$$

Incompressibility Theorem *Let c be a positive integer. For each fixed y, every finite set A of cardinality m has at least $m(1-m^{-c})+1$ elements x with $\mathcal{C}(x|y) \geq \log m - c$.*

Proof The number of programs of length less than $\log m - c$ is

$$\sum_{i=0}^{\log m - c - 1} 2^i = 2^{\log m - c} - 1.$$

Hence, there are at least $m - m2^{-c} + 1$ elements in A which have no program of length less than $\log m - c$. □

Inference Incompressibility Theorem *Let c be a positive integer. Let A^* be a countable set of inferable functions and $\mathcal{D}$ be a device that infers all functions $a \in A^*$. There exists A^* and $\mathcal{D}$ such that every finite subset A of A^* with cardinality m has at least $m(1-m^{-c})+1$ elements $a \in A$ such that $\mathcal{C}_\mu(a;\mathcal{D}) \geq \log m - c$.*

Proof Fix any convenient A^* and choose any subset $A \subseteq A^*$ with cardinality m. Let $U := \mathbb{B}^*$. Let $X : U \to \mathbb{B}^*$ and $Y : U \to \{-1, 1\}$ be the X and Y defined in the Countable Inference Theorem when given A. Note that X in this case maps $\mathbb{N} \to \mathbb{B}^*$ instead of $\mathbb{N} \to \mathbb{N}$ but the mappings are equivalent. Then $\mathcal{D} = (X,Y) > A$ by construction.

Take $d\mu(u) = \ell(u)$ where $\ell(b)$ is the length of a bit string b. Recall that the inference complexity of $a \in A$ with respect to $\mathcal{D}$ is

$$\mathcal{C}_\mu(a;\mathcal{D}) = \sum_{\delta \in \wp(a)} \min_{x:X=x \implies Y=\delta(a)} [\mathcal{M}_{\mu,X}(x)] : \mathcal{M}_{\mu,X}(x) = -\ln \sum_{X^{-1}(x)} \ell(u)$$

Now suppose that all $A_i \in A$ is functionally equivalent to A_5 over V. Then the maximum inference complexity of any A_i with respect to $\mathcal{D}$ is

$$\mathcal{C}_\mu(A_i;\mathcal{D}) = \sum_{\delta \in \wp(A_i)} \min_{x:X=x \implies Y=\delta(a)} -\ln \sum_{X^{-1}(x)} \ell(A_i^{-1}(a))$$

$$= \sum_{\delta \in \wp(A_i)} \min_{x:X=x \implies Y=\delta(a)} -\ln A_i^{-1}(a) = -\ln 1 - \ln(A_i^{-1}(a_i)) = -\ln(u).$$

Since a unique u is used by $\mathcal{D}$ to infer every $A_i \in A$, we can apply the pigeonhole principle. The number of bit strings of length less than $m-c$ is

$$\sum_{i=0}^{m-c-1} 2^i = 2^{m-c-1}.$$

Hence, there are at least $m - m2^{-c} + 1$ functions a in A which have inference complexity greater than $\ln \log m - c$. $\square$

Corollary *There exist A^* and $\mathcal{D}$ such that for all functions of any subset $A \subseteq A^*$, $\mathcal{C}_\mu(A_i;\mathcal{D}) = 0$*

Proof Recall the functions and spaces defined in the Countable Inference Theorem and take $\mu(u) = \ell(u)$. Then choose $A^* : \overline{A_i(V)} \in A^* = \overline{A_{j\in\{2,3,4\}}(V)}$. Then the inference complexity for any function $A_i \in A$ is

$$\mathcal{C}_\mu(A_i;\mathcal{D}) = \sum_{\delta \in \wp(A_i)} \min_{x:X=x \implies Y=\delta(A_i)} -2\ln \sum_{X^{-1}(x)} \ell(u) = \sum_{\delta \in \wp(f)} \min_{x:X=x \implies Y=\delta(A_i)} -\ln 1 = -2\ln 1 = 0.$$

$\square$

7 Physical Knowledge

Definition Consider an inference device (X,Y) defined over U, a function Γ defined over U, a $\gamma \in \Gamma(U)$, and a subset $W \subseteq U$. We say that (X,Y) **(physically) knows** $\Gamma = \gamma$ over W iff $\exists \xi : \Gamma(U) \to \overline{X}$ such that

i) $\forall \gamma' \in \Gamma(U), u \in \xi(\gamma') \implies \delta_{\gamma'}(\Gamma(u)) = Y(u)$
ii) $\varnothing \neq \xi(\gamma) \cap W \subseteq Y^{-1}(1)$
iii) For all $\gamma' \neq \gamma$, $\varnothing \neq \xi(\gamma') \cap W \subseteq Y^{-1}(-1)$.

Then the **knowledge complexity** of Γ with respect to $\mathcal{D}$ and measure μ is defined as:

$$\mathcal{K}_\mu(\Gamma; \mathcal{D}) \triangleq \sum_{\gamma \in \Gamma(U)} \min_{|\xi(\gamma)| : \mathcal{D} \text{ knows } \Gamma = \gamma} [\mathcal{M}_{\mu;\xi}(\gamma)]$$

$$\mathcal{M}_{\mu;\xi}(\gamma) = -\ln \int_{\xi(\gamma)} d\mu(u)$$

Example Let $U := \{1,2,3\}$ and take $W = U$. The following table and formulas demonstrate physical knowledge for $\gamma \in \Gamma : |\Gamma(U)| = 2$.

u	$X(u)$	$Y(u)$	$\Gamma(u)$
1	1	1	1
2	2	-1	1
3	3	1	2
4	4	-1	2

$$\xi_1(\gamma) = u$$

$$\xi_2(\gamma) = u$$

Theorem *For every function Γ there exists a device $\mathcal{D}$ such that $\mathcal{D}$ physically knows all $\gamma \in \Gamma(U)$.*

Theorem *Let c be a positive integer. Let A^* be a countable set of inferable functions and $\mathcal{D}$ be a device that infers all functions $a \in A^*$. Every finite subset A of A^* with cardinality m has at least $m(1 - m^{-c}) + 1$ elements a such that $\mathcal{K}_\mu(a; \mathcal{D}) \geq \log m - c$.*

Proof Fix any convenient A^* and choose any subset $A \subseteq A^*$ with cardinality m. Let $U := \mathbb{B}^*$. Let $X : U \to \mathbb{B}^*$ and $Y : U \to \{-1, 1\}$ be the X and Y defined in the Countable Inference Theorem when given A. X in this case maps $\mathbb{N} \to \mathbb{B}^*$ instead of $\mathbb{N} \to \mathbb{N}$ but the mappings are equivalent. Then $\mathcal{D} = (X,Y) > A$ by construction. Take $d\mu(u) = \ell(u)$ where $\ell(b)$ is the length of a bit string b.

8 Future Work

9 Acknowledgements

Wolpert DH. *Constraints on physical reality arising from a formalization of knowledge.* arXiv preprint arXiv:1711.03499v3 [physics.hist-ph] (2018).

Li M & Vitnyi P. *An Introduction to Kolmogorov Complexity and Its Applications* (1st ed.). Springer-Verlag. (1993).

Arora S & Barak B. *Computational Complexity: A Modern Approach.* Cambridge University Press. (2009).