To be (perhaps)
Given at Caltech,
Feb 60

The Inductive Inference Problem and the Representation of Information in Biological and Electronic Systems.

1) What the problem of inductive inference is.

2) Criticism of some proposed solutions to the problem.

3) "Language-based" inductive methods.

    a) The methods of Carnap.

    b) The methods of Solomonoff.

4) Relationship to problems of optimum coding of information in:

    a) The brain

    b) Chromosomes

    c) Non-biological information retrieval systems.

The Inductive Inference Problem and the Representation of Information in Biological and Electronic Systems.

1) What the problem of inductive inference is.

2) Criticism of some proposed solutions to the problem.

3) "Language-based" inductive methods.

    a) The methods of Carnap.

    b) The methods of Solomonoff.

4) Relationship to problems of optimum coding of information in:

    a) The brain

    b) Chromosomes

    c) Non-biological information retrieval systems.

The Inductive Inference Problem and the Representation of Information in Biological and Electronic Systems.

1) What the problem of inductive inference is.

2) Criticism of some proposed solutions to the problem.

3) "Language-based" inductive methods.

   a) The methods of Carnap.

   b) The methods of Solomonoff.

4) Relationship to problems of optimum coding of information in:

   a) The brain

   b) Chromosomes

   c) Non-biological information retrieval systems.

The Inductive Inference Problem and the Representation of Information in Biological and Electronic Systems.

1) What the problem of inductive inference is.

2) Criticism of some proposed solutions to the problem.

3) "Language-based" inductive methods.

   a) The methods of Carnap.

   b) The methods of Solomonoff.

4) Relationship to problems of optimum coding of information in:

   a) The brain

   b) Chromosomes

   c) Non-biological information retrieval systems.

12 28 59

For Introduction of idea of use of min. coding for proby. eval:

In optimum coding of messages, for communication, any regularities in th. ~~messa~~ seq. of symbols is utilized to reduce th. no. of bits needed to code th. message. We use apriori information about th. relative probability of various possible sequences of symbols, to reduce the no. of bits needed for coding.

In the system proposed, we invert this procedure and try various methods of coding a given sequence of symbols. The no. of bits of the coding method that is most compact, is then used as an index of the apriori probability of the sequence coded. [actually, we sum th. aprips of all ways of coding th. seq. — only th. most compact desns ~~are~~ signif., hvr. (mention convergence probs)]

12 27 59

Th ind. inf. prob. Polya's ~~[illegible]~~ list of desirata. (1

A few of th. Q's that ~~must be~~ should be answered by a satisfactory theory of ind. inf.

1) In any circumstance for any previous history, give a probty for any possl. continuation.

2) Give th. freq. ratio for the. rel. probtys, for large samples.

3) Explain why theories devised befor emp. data is known, and verified by that data, are gn. more credence than those devise after data is known, and agree with that data.

We may say that post hoch theory **may** be ad-hock — apriori theory is less likely to be ad-hock. But we feel that whether a theory is ad-hoch or not should be determined by examining th. theory itself, and not th. history of th. theory — otherwise we clearly do not understand "operationalaly" what ~~[illegible]~~ characterizes an "ad-hoch" theory." Also, ~~[illegible]~~ it is likely that a theory should have degrees of ad-hochness — not ~~that~~ have ad-hochness be a yes-no property, depending on whether th. theory was made up befor or after th. expts. were performed. (give th. Einstein-Milne example)

4) ~~Th~~ Th. problem of "Geometric Probability" — **its** relation to ~~apr~~ evaluation of aprips. How Geom. probty is a ~~rather~~ special case of difty in using "th. principal of indifference".

5) When can "Th. principal of indifference" be used? (see 4).

12 27 59

(2

6) Th. B.G. (or Max. Coin) problem: What to do, when 2 different freq. counts ~~????~~ give different prob'ty estimates for th. same event.

7) In (Vol II I think) Polya, there ~~are~~ is a list of some desirata of inductive inference systems.

8) Th. principal of parsimony should hold - i.e. greater aprip to "simpler" hypotheses.

ZATOR COMPANY
140½ MOUNT AUBURN STREET · CAMBRIDGE 38 · MASSACHUSETTS · TROWBRIDGE 6-6776

Mar 2, 60 U of Penn.

THE USE OF FORMAL LANGUAGES FOR MECHANIZED INDUCTIVE INFERENCE

For many years, standard algebraic curves have been used to smooth, interpolate and extrapolate quantitative empirical data It will be shown how formal languages may take an analogous role in smoothing and extrapolating qualitative data. Some examples of this kind of extrapolation are the discovery of the general grammar rules of a language, from a finite sample of acceptable sentences in that language. Other examples of induction of this sort in linguistics and mathematics will be given.

A particular type of language, called the "weak phrase structure language", has been investigated at some length, and some of its applications will be shown in inductive inference for multidimensional, as well as linear patterns.

Th. talk itself:

This will start out with some detailed analogy betw. curve fitting and grammar fitting. Give some actual results of this ~~nov~~ "novel" way of looking at curve fitting — e.g. that for n data points, one may have a ~~n~~ 2n or 3n parameter curve be th. "best fit".

After this, which is to motivate interest in "grammar fitting", for extrapolation, — go into langs in more detail — in particular th. general non-dim. PSL. and its various applications. Just mention, but do not demonstrate, th. use of a TM_2 for multiplic.

Much of th. material will be th. same as for th. RCA talk.

Perhaps some info on use of stoch langs. for

① MT
② IR
③ optimum learning of an operator for T.M. (reenf. learning).

Letter to Gorn: ① Mention "Tree langs". ③ Give him more details on talk, as

② how long will talk be?

Mar. 2. 60 Talk / actually gn. at U of Penn Seminar. ①
4 P.M.

1. Defn. of "Lang."
   a) Alphabet.
   b) sub-set of order sequs of Alphabet symbs.
   c) ~~English~~ Examples: ① English (with letters or words as "Alphabet".
   ② All alge express whose values = 0
   ③ All legal thrms. that follow from a gn. set of postulates.

~~2. Probs. in Lang~~
~~a) Gn~~

$a^m \; k_0 \; b^m \; k_2 \; c^n$

2. Def. of "Grammar"
   a) Constructive
   b) Decision Procedure.

3. Probs in lang.
   a) Construct → Decis. proc. form (e.g. in Math.)
   b) Dec. proc. → Constr.
   c) Induction: Gn. set of a's — to find "simplest." Grammar.
   ① If one can ask Q's.
   ② " " can't " ".

4. Examples of simple langs:
   a) FSL [redacted]. Gen. by FS Machine.
   b) PSL

$11\Lambda + 111\Lambda = 11111\Lambda$

$1^n\Lambda + 1^m\Lambda = 1^{m+n}\Lambda$

A
A
B
B
S

~~S~~ S
A → ∧ =
A → 1 A 1
B → ∧ + A
B → 1 B 1
S → B ∧

Mar 2.60

5. PSL's More genl. than simple concat. strings.

a) Chem. Compounds.

b) MT Give defn.

c) Use for Multiplic. — Hyer Order langs.

~~6. Learning, Induction.~~

~~a) with~~

~~b) wi~~

Questions Briefly

Linguist ind. inf.

Talk lasted ~ 55 Min. — wasn't ~~very~~ well organized

—

Things to discuss with Chomsky:

1. Has he shown that Eng. is not expressable by type 2 (Non CDS?.) langs? (other than my counter examples)

2. " " " " " " " " " type 1 (CDS?.) " ?

3. (a) Th. importance of MT studies, rather than grammatical studies to determine "true" structure. e.g. MT of L_1 → "Meaning"

   (b) How str. of L_1 for MT may not = str. needed to construct L_1.

   (c) How parsing grammar may not ≡ construction grammar.

~~4. Paradox of generalized Putnam's theorem for all decidable, finite grammars.~~

5. non-dim. PSP's — are ~~they~~ weaker or stronger than type 1 langs?

6. Stochastic Langs — a soln. to th. ind. inf. prob.

7. Get reprints: (a) ~~Review of Skinner's book ("Verbal Behavior")~~

   (b) " " Greenberg's essays in Linguistics

   (c) ~~"Formal props. of Gramms"~~

8. Th. Subclass of SPSL's used by Th. U of Penna Group for Parsing English

   (a) can it do AB → BA?

9. Quick way to tell if 2 PSG's are identical? (a) WPSG (b) SPSG. (c) How equiv. PSG's may form a PSL.

10. How "Dimensionality" may help with 9.) see PSG discy 467.28 for discn of "dimensionality".

3.4.60 Prob. of Art. Intell.

Gn. at Nat. Bu. Standards.

March 4, 60

lasted 1½ hrs.
— rather well given.
~ ½ hr. on new. ind. inf. ideas
I don't think they got much out of that, hvr. — Too hy a % linguists in audience.

(0)

~~(1) Neural Nets: Random, non random;~~

(1) Art. intell. : ~~[illegible]~~ ind. inf. ~~[illegible]~~

2Dim. patt. discy :

(2) Idea that solu. to ~~[illegible]~~ stat. inf. prob. is known — that prob. is to make ~~[illegible]~~ practical ll computer using neural nets.

(3) That ind. inf. prob. is not solved:
e.g. (a) T. S. That ~~[illegible]~~ seems to change statistics.
Stock market.
(b) ll lines of inference.
Blue eyes — die at 60, proby p_1.
red hair — " " " " p_2.
(c)

Ask. Q's if particularly interested in any topic mentioned.

$$A \cap B = \sim(\sim A \cup \sim B)$$

5.4.60

Idea of talk: A. ~~Dextra~~ Ord. langs
B. Stoch langs. (

Just what langs are: FSL
W.PSL
More Genl. langs.
Most Genl. langs. (In terms. of ~~T~~ Tur Machine)
(i.e. Fixed string, followed by arb. string)

1) What langs. are: Some Examples

Why study them?

2) Applics. of "langs." a) M.T.
b) Extrapolation (≡ concept formation, Gestalten) — E.G. "curve fitting", Char. Recogn., Discovery of "characters", Discy of "invariants"; Scientific Laws, weather prediction,

3) Examples of Langs. recog. by finite Automata.

a) FSL : ① Deterministic
② Stoch (≡Markov).

b) WPSL ① Determ: example.

S
S → ab
S → aSb.

∴ $a^n b^n$ ($n \geq 1$).

Subs. rules: ① Always ↑ length of string.
② No CDS.

S → Np · Vp
Np → Art · N
Vp → V · Np
V → saw
V → was
N → Boy, Man, Dog
Art → [illegible]

Multi-dim. PSL's.
How MT may be a PSL for some lang. pairs.

e.g. Ord. alg. to "Polish" notation

Organic Chemicals.

3.4.60

$11\Lambda + 111\Lambda = 11111\Lambda$

$A \to \Lambda =$

$A \to 1A1$

$B \to \Lambda + A$

$B \to 1B1$

$S \to B\Lambda$

$1^n\Lambda + 1^m\Lambda = 1^{m+n}\Lambda$

$1^n\Lambda + 1^m\Lambda = 1^m\,1^n\,\Lambda$

Arith. Multiplic

$k_0\, a^n\, k_1\, b^m\, k_2\, c^m\, k_3\, d^n\, k_4$

4) Problem of Induction : Determ. langs.

simplest. lang.

Easy if Q's can be asked.

Info rect. } MT } can be learned at present, under unreal conds.

Much diflty if error is made.

5) Prob of Induction: ~~Stoch. langs.~~

Stoch PSL's.

S

$S \xrightarrow{.3} ab$

$S \xrightarrow{.7} aSb$

genl. soln. — Several ways to parse a gen. s.

Several "Parsing" routines.

(stoch Gram; S) → proby.

(corpus, stoch gramm) → proby.

(corpus, stoch gramm, aprip(Gramm)) → apsip.

To chose stoch Gram → apsip is max.

or more genl., exact soln.

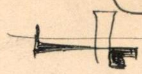
— Is hill climbing problem

Idea is to find types of "proximity xfuns".

SOS talk : May 22 - 24, 1962 Chicago   Prob(y   2478 Orchard

- LA 8 -
(Kirsch 54)

1) ~~Give backgnd., of old work,~~
   ~~a) ERE pap.~~
   ~~b) U. evaln. ([illegible] unsatisfactory soln.)~~
   ~~c) Gen. Ind. Inf. Theo.~~

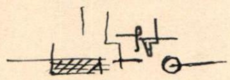


2) Present approach ~~is~~ to self. improving Machine.
   (with resp. to some criterion)
   a) Genl. ~~program~~ program improving ~~of~~ machine ~~with resp. to~~
   b) Such a machine is gn. probs like itself, to start out
   c) machine is gn. itself.
   d) Most probs. of int. can be expressed as H.c. probs. on pgms.

3) first step is simple ind inf. machine.
   a) desb. in ~~[illegible]~~ a little detail.
   b) How differs from ① my Old approach
   ② ~~[illegible]~~ Kilburn, Sonnar, Grimsdale → (Randomness, non opt. search, — Opt. search speed also gives opt. extrap. ∴ Uniformly bett (except for work").
   ③ Newell, Simon.

conds. under which "opt" approach is ~~not~~ reasonable
i.e.
(pgm. takes much time or expense)
2748
7824

∞ search time using ~~[illegible]~~ "p" search.
Discuss ~~to~~ solvability v.s. Unsolvability irrelevance
— Need for new theory with criterion of solubility with resp. to gn. machine size and time of soln.
Analogy to opt. coding in noise, using ~~[illegible]~~ data various code lengths.



4) a) Next step is improvement of Heurs
   b) systematization of heurs.
   c) Evaln. of marginal U of heurs
      " " U of set of heurs

5) Work on / various (improvement) pgm imp. prob. — (how differs from Arith evaln, say)
   a) Heurs are similar
   b) work on heur improvt. prob.

~~6) Most probs of int~~

6) Some details of "rules", heurs, etc.

a)
A + B → M + AB
A − B → M − AB
(A) → ~~[illegible]~~ A

~~A + (B)~~
3 + 5 − (5+2)
3 + 5 − (7)
~~3+~~ 8 − (7)
8 − 7
1.

~~3 + 4(5+2)~~
~~3 + 4(7)~~
~~3 +~~

b) 3 + 4 × 5 = 23

c)

SOS conf talk: (1)

1) Prob. - Machine to (a) solve ~~genl~~ genl. probs. (adequate vocab. to solve anything)

(b) Self-improving - can work on prob. of improving self.

(c) A machine built to ~~optimize~~ find pgms. that are opt. with resp. to ~~certain~~ any gn. criterion.

(d) Give machine prob. of improving itself

(e)

2) First step is simple ind. inf. Machine. - arith. learning.

(a) Like $3+7=10$

(b) Rules are like $A+B \rightarrow M+AB$

(c) Randomly constructed from basic Alph. of symbols.

(d) How Differs from my work of ~ 5 yrs ago

(1) U of abss. not worked out at that time.

(2) Now I have ~~worka~~ apparently workable Genl. ind. inf. theory.

(e) How differs from Kilburn, Sommer, Grimsdale, ~ 1959

(1) They tried to make a machine that would look like thinking, prob. solving human.

(2) Concern with randomness "for originality".

(c) Digression on P, $\sqrt{P}$, "optimum" search.

$\sqrt{P}$ "faster than" P, but "poorer" answ.,

opt. "faster than" any, also "better" but not always realistically faster.

(b) (I) Criterion for success are simplified by reasonable constraints. e.g. search time ~ n for n possy, → ∞ as n → ∞, but we can cut off trials at some "reasonable" n.

(II) ~~computability, unsolvabi~~

usually irrel. { unsolvability by finite machine / "trivially" solvable } V.S. solvable in time T by machine with mem. of N.

?(Analogy to opt. coding in noise.)

(3) Improve ~~[illegible]~~ solu. methods,
" methods of finding ~~s~~ new solns.

~~(1)~~ (2) systematizn of ~~[illegible]~~ soln. methods into compact "notation" and constraints / so that possy. of creating these rules thru a random generation system is max.
I.e. find ~~way~~ good stoch. gram.

(4) Then work on pgm optzn.
(a) ~~to~~ Heurs seem ~ or identical/.

(5) Rules, heurs, etc.
(a) 3 + 4 = 7

$A + B \to M + AB$
$A - B \to M - AB$
$(A) \to A$
O.K. for probs with +, −, (,).

(b) ~~[illegible]~~ $3 + 4 \times 5$
teach Mult, div. first, then +, −.

~~[illegible]~~ (c) +, − rule are ~~with~~ context dependent.

e.g. $(A - B + \to (M - AB +$

~ 50 cont-dep. rules.

~~(2)~~ ~~2~~ (d) ~~[illegible]~~ in aprip with grouping of CDS rules : e.g.

$\alpha A + B \beta \to \alpha A + B \beta$

$\alpha = (,), +$

$\beta = (, -, +,)$

(e) some trial numbers.

~~≈~~ for ~~A+B~~ $A \times B \to M \times AB$
10^{10} for simple, initial MC search
50 better for later " "
20 " " opt. search,
~~[illegible]~~ 10^{-14} for raw $(A - B + \to (M - AB +$

~ 1000 better for opt, + experience

large factor for factorizn.

(f) Genl. method to compute $\uparrow$ in sprip due to grouping.

Lecture gn. To Minsky's class March 24, ~~1960~~ (1)

Essentially all of this material was covered in 55 min.
~~Befor~~ lecture, 3 blackboards filled with diagrams.

1960 ~~1959~~ — No in 1960 Mar 24 was Fri

Motivation.

1. Goals; Ind. Inf. | Pattern extrapolation | examples: (a) Pattern discy (e.g. R. letter class "A".

∴ Mar 24 59 was Wed! (1960 leap year) ?

more examples: (b) M. T. by example

(c) Classification of docs for IR, by example.

(d) Operator extrapolation. Gn. set of "good" I → O pairs.

2. Concept of Langs. : (a) Alphabet, subset of strings.

Examples: English; binary +; multiplic probs; correct alg. eqs.

theorems folg. from gn. set of axioms;

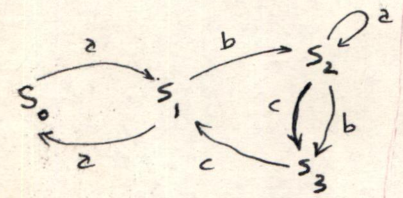
(b) Types (1) FSL | kinds of Grammars: (1) ~~Th~~ Constructive

(2) PSL. | (2) ~~[illegible]~~ Test method

3. FSL's.

$\boxed{M}$

leave



a b c c a

M in S_3 at t

~~So~~ Then M emits symbol a, and goes to state ~~S_2~~ at time t+1.

(1) Can be recognized by FS Machine. (2) Constructable from state diagrams.

Use in psych. testing by Chomsky Miller.

Exists / simple discy procedure for ∃ Gramm., if Q's are askable.

4. ~~Non~~ PSL

example of essentially non-FSL : $a^n b^n$.

PSG: $S \to ab$

$S \to asb$.

cannot be recog. by FS Machine — e.g. Human. — mem. reqy = ∞.

~~English~~ Simple English Grammar is not FSG — but ∃ Grammar of Eng. in Q — innov. by Yngve.

Eng. as PSL:

$S \to NP \cdot VP$

$NP \to Art \cdot N$

$VP \to V \cdot NP$

$V \to$ Saw

$V \to$ was

$N \to$ Boy

$N \to$ Man

$N \to$ Dog

$Art \to$ The.

① Construction always ↑ length of string (or remains same), ② finite alph.

∴ lang. decidable.

∃ several simple decision procedures.

e.g. ① Bar-Hillel (not very practical)

② Two of mine.

Another form of PSG:

$S \to a\,b$

$S \to a\,s\,b$

$a \times S \times b \subset S$

$a \times b \subset S$.

$S \to A\,b$

$S \to b\,D$

$A \to D\,S\,b$

$A \to a$

$D \to S\,S$

$D \to b\,A$

$D \to d$

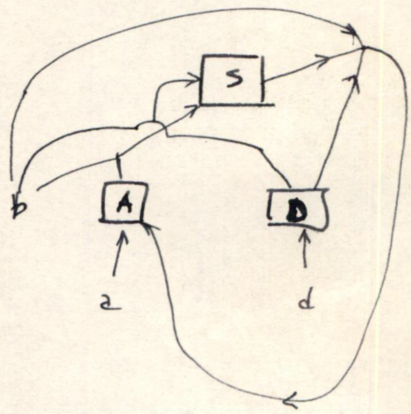
$A \times b \subset S$

$b \times D \subset S$

$D \times S \times b \subset A$

$a \subset A$

etc.



These x's may be any operations

A, b, etc may be classes of any kinds.

Gives Great Generality.

"Basic Loop" form of FSL's, PSL's. ??

ordinary PSL $a^n b^m c\, d^m e^n$

$a^n b^m c\, d^n e^m$ impossl.

Is possl. with more genl. PSL's like e.g. $(a, b) \times (c, d) \to$ a c b d

Examples of PSC's: $11\Lambda + 111\Lambda = 11111\Lambda$

$1^n \Lambda + 1^m \Lambda = 1^{m+n}\Lambda$

$S \to B\Lambda$

$B \to 1B1$

$B \to \Lambda + A$

$A \to 1A1$

$A \to \Lambda =$

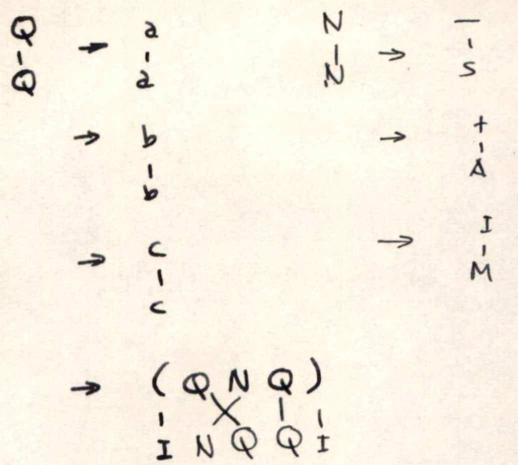
Also ~~sobtraction~~

for multip of 2 nos., we don't have PSL, but Grammar is a PSL.

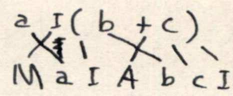
" of 3 nos, Gramm of Gramm of Gramm is PSC, etc.

MT

first explain "polish" notation



example:



Note: Do "+" sobstit. befor "M". Practice this before lecture.

State of induction problem:

PSG discy with Q's, can be done. So MT could be done

but ① large no. of Q's.

② errors give trouble.

PSG discy w.o. Q much harder — much progress —

A teleological soln obtained. — (≡ Hill climbing prob.) —

Several hours obtained - adequacy unknown -

Genl. soln. Ind. inf. prob.

Stoch langs were mentioned in only a very cursory way, Fri May 20, 1960

This was the 2nd part of the lecture of Mar 24, 60

Talk to Marv's Class:

Treat: 1) Stoch PSL's

2) Genl. ind. inf. by coding — how connected to

3) 1956 Ind inf. Machine

— how to compute utilities — actual examples

4) How to compute U's for Pandemonium and GPS.

5) Use of ~~remainder~~ for curve fitting,

Method: ① Go thru Pandemonium, showing where combs. of abss. occur and where epist. probs arise — how U eval is needed

② Go thru Arith learning ~~[illegible]~~ showing where U eval is needed.

③ Present Genl. Theo. of ind. inf. Give simple example of Markoff chain using "R" coding.

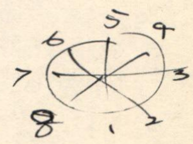
④ Show how curve fitting is done: ⓐ Normally — what diftys are

ⓑ how done by coding { how RMS error is used]. By special notation for Gauss distribu.]

Show how PSG's are coded: 2 extreme kinds of Grammar — how avoided. Show how twocl can be done with FSL.

, $\boxed{1 \mid 2}$ (=, 1) | = , 1 (=, ①1)

$\boxed{1}$ | $\boxed{0}$ | $\boxed{1}$, $\boxed{0}$ | = 0 $\boxed{0}$, = 1 $\boxed{1}$



= 0 $\boxed{0}$ | also 0 1 0 $\boxed{1}$ | also $\boxed{1 \mid 2}$ × (= , 1 $\boxed{1}$)

NS (M₁) " " × (= , 0 $\boxed{0}$)

(+) 1 1 0 0 | 1
0 1 0 1 | 0
1 ① 0 1 | 1

(+) 1 1 $\boxed{1}$

(+) 1 1 $\boxed{1}$

Fri May 20, 1960 for Marv's class.
Actually Gn!

1. Pandemonium: combs of ops.
   - a) use of ugm. ktn probies.
   - b) " " upm. probys, with defs, and Utility

$U_{child} = f(U_{parent},$ Emp. Usefulness$)$.

2. Dart T.M.
   - a) = use of $\overset{1}{\boxed{1}}$
   - b) ~ " " $= \overset{1}{\boxed{1}}$, $\sim \overset{0}{\boxed{1}}$, etc.
   - c) use of $\overset{1}{1}\overset{0}{\boxed{0}}$, $\overset{1}{0}\overset{0}{\boxed{1}}$, etc.
   - d) " " $\boxed{1 \mid 2} \times (=, \overset{1}{\boxed{1}})$

     $\boxed{1 \mid \; \mid 2} \times (=, \overset{1}{\boxed{1}})$ & etc.

     $\boxed{1 \mid 2} \times$ ...
   - e) ⊕

$$\begin{array}{r} 1001 \\ 1010 \\ \hline 10\boxed{1}1 \end{array}$$

use of $\boxed{1}$ only
" " 1 0 / 1 / 1

and ⊗

$$\begin{array}{r} 1001 \\ 1011 \\ \hline 1001 \end{array}$$

use of
$\boxed{1 \mid 2} \times$ (⊗, 1 1 1)
etc.

3. Coding method of ind inf. | Not too much detail
   Some discu. of n.b. PSG discy.
4. Some applic of 3 to U eval for Pand. and to genl. hill climbing.

The delivery of talk was good. Topics 3 and 4 not explained very well.

Talk to Marv's class Wed May ~16, 62.

On the Coding Method of Ind. Inf.

1. What is problem? How is it "solved" now, and how well? Give examples: weather, Horse racing, Sc. laws, stock market, trial steps in Math proofs, curve extrapln.

Discuss Pitt. Discy Theorem proving. → 2. Why is it worth solving? — we'd like to do it better than we do now, perhaps with machine aid — perhaps learn to combine guesses of human experts.

3. a) What's wrong with the obvious "solutions"?

b) " " " other proposed solns?

4. What is *my* solution? How does it differ from others? How does it overcome diffys other theorys have trouble with?

Perhaps try to combine 3) and 4).

My soln.: 1) that all induction probs. can be xfmd to induction on very long symbol string containing *all* info. that is used in th. induction. (Perhaps give examples).

2) I use Bayes Theorem: Give simple example of how used:

— a b a c c a a b a $p_1 = 1 \times 10^{-5}$
— a b a c c a a b b $p_2 = 3 \times 10^{-5}$
— a b a c c a a b c $p_3 = 2 \times 10^{-5}$

$p(a) : p(b) : p(c) = 1 : 3 : 2$

~~AI~~

3) I can compute aprips. (this has always been difty of using Bayes) — I do this by concept of "simplicity" of a dern. (is ~ to Occam's Razor).

Mention ① Occam's ② Princ. of ind iff. ③ Huffman coding. ~~④ Shannon's Coding~~

4) How 1111111 seems simpler than 1111110 — and has a simpler "dern" in English.

Also abcabcabc ← has simpler dern ("3 - abc's.") vs. abcabcaba

5) A more rigorous formulation:

M (Ξ) = T (T is string, Ξ its dern. with resp. to M).
↑ Universal Machine (not necly a Turing)

aprip of T ≈ $2^{N(\Xi)_{min}}$.

6) Better include other derns. of same Σ.

so $\sum_i 2^{N(\Xi_i)}$

~May 16, 62

6) (cont) This wouldn't converge in its own simple ~~[illegible]~~ lone self. — We can ~~conv~~ patch it up to make it converge, but instead of maring its beauty — ~~[illegible]~~ consider alternate formulation!

7) Th. fixed code length method. $R \gg N(\cancel{T}_{min})$.

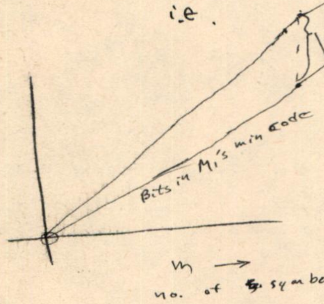
$$\frac{\cancel{A(R)}}{\cancel{N}} \qquad \frac{N(R, \cancel{T})}{N(R)_{conv.}} \qquad \Bigg| \qquad \lim_{R \to \infty} \frac{N(R, \cancel{T}A)}{N(R, \cancel{T}B)} = \frac{P(A)}{P(B)}$$

Show why ~~[illegible]~~ ~ same as 6).

8) Why 6) is machine indip.

First explain about "simulation instructions"

i.e. $\forall \cancel{T},\ M_2(S\cancel{T}) = M_1(\cancel{T})$



bits in M_1. simulation instructions, S.

pp 15, 16, 17 of ICP

9) Th. "All PEMS Method". pp 27–31 of ICP

$M_1(D_i \Delta T) = P_i(T)$

$\leftarrow$ punct.

(Define Normzd PEM?).

$\frac{P_i(TA)}{P_i(TB)}$ = probty, if P_i is "normzd".

~~[illegible]~~ f_i = fraction of strings of length R $(R \to \infty)$ that are "Derns of Q_i".

$\Sigma f(Q_i) P_i(T)$.

10) Some applicns: ~~[illegible]~~ (a) Bernoulli Seq.
(b) Codes using defns (give approxn. formulae)
(c) Arith learning

11) Mention Random Redundant Machine Method.

11)

Ap 24, 63 — Marv's Class. 1

A method for Ind. Inf.

1. Some induction probs:
   a) Weather
   b) Horse Racing
   c) Sc "laws"
   d) Trial steps / trial sub-goals in Math proofs
   e) Curve extrapolation.
   f) Chess, checkers — finding good params to optz, etc. | Give areas of induction
   h) Info retrieval: Devising new categories, for already indexed docs.
   i) Stock market

2. ~~Re~~ Relation to art. intell. probs.
   a) ~~Human beings it~~ Patt. Recognition: Letters, ~~[illegible]~~ speech
   b) Theorem proving: ~~Other~~ successive xfmns: Set of observational results: what to do next: a) what xfmn to use. b) what new xfmns or observations to devise.
   c) Simulation of entire human:

Model of human: small set of wtd. "goals" / plus built in constraints. Uses entire ~~str~~ life string of observations to determine th. results of a contemplated action. (Also uses "built in" info). Mention ancestral data — aquired thru evolution, as alternatively treatable by either ⓐ longer corpus (less el. way) ⓑ short corpus but modified machine.

3. Classical Bayesian probability:
   a) We will use probty model in which we extrapolate long string, containing all of the info. that we ~~[illegible]~~ will use in induction. In th. case of man, this would include evolutionary data in some form.

Show how its done e.g. in .1 ., .2

Mention Analog. Dig. conversion for R.W.

b) $P(a_i/h) = \dfrac{P(h^\frown a_i)}{\sum_j P(h^\frown a_j)}$

$P(h^\frown a_j)$ is th. apripd.

give example:

abdba	$\cdot 1\times10^{-5}$	a $\cdot$ [illegible]
abdbb	$\cdot 1\times10^{-5}$	b $\cdot$ [illegible]
abdbc	$\cdot 3\times10^{-5}$	c $\cdot$ [illegible]

4. How to get aprips: ⓐ Simple things more likely (Occam's R.) ⓑ More numerous causes make more likely (Bayes) ⓒ Princ. of indif. ⓓ Hufman coding (a little detail here)

5. How to get apripd's:
   a) My method: Use of "Turmacs"

Mention 3 tape Turmac, with input, ~~[illegible]~~ tape
output "
Memory "

(Fixed code string length, R)

b) ③ $P^*(a, T, M_1) = \lim_{R\to\infty} \dfrac{N^R_{Ta}}{N^R_T}$

② N_T^R is with resp. to M_1

" th. no. of input strings of length R, that are derns of T.

① If $M_1(s)$ begins with T, then "s is a dern of T with resp. to M_1"

c) This is implementation of ① "Princ. of indif"
② Bayes
③ Occam's R. (show that "short" derns are more numerous.)

Mention that $2^{-N_{shortest}}$ is one possl. approxn. (impt. for later analysis)

6) 2 other ~~[illegible]~~ similar models : a) Sum of all pcm's

A normzd. pcm. is an ~~[illegible]~~ op. that assigns probys to strings.

$$P'(a, T, M_1) = \frac{\sum_j P(pcm_j, M_1) P_j(\widehat{Ta})}{\sum_j P(pcm_j, M_1) \cdot P_j(T)}$$

It can be made plausible that for a certain "goodness" criterion, this model is at least as good as th. best finitely describable predn. method.

b) 3 tape machine as model of universe. - i.e. To "explain" all things that a man has ever experienced - including his interaction with th. world.

c) Th. randomly constructed machine for extrapolation. Th. problem of "is a randomly constructed universal Turmac "likely"? As a model of an operator (input → output)

7) Diffys with Gen. Theory

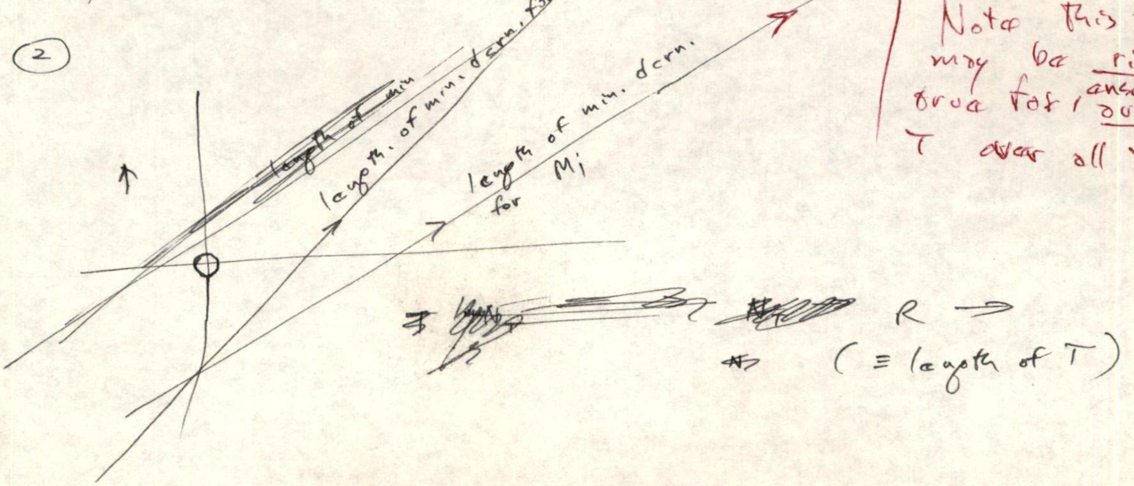
a) Are predictions indip. of what Turmac's are used?

① Idea of "simulation instructions"

If $\forall x, M_1(\alpha^\frown x) = M_2(x)$ α is M_1's "simulation instrns for M_2". $\exists\, \alpha$ for all universal turmacs, M_1, and any machine M_2.

Note this argt. may be rigorously true for ensemble average T over all T's.

②



7) Diffys (cont)

b) Meaningfulness of

$$\lim_{R \to \infty} \frac{N_{T_a}^R}{N_T^R}$$

Ratio May be meaningful — tho. Num. and denom. aren't —

— It may be possl. e.g. to show lim. exists — i.e.

that $\frac{N_{T_a}^R}{N_T^R}$ is an ↑ funct. of R, for large enuf ΔR's.

or that th. fraction of codes ~~not~~ eventually halting / not halting in Time T, → 0 as $T \to \infty$

Dificult philosophical Q. involving meaningfulness even then.

Hvr., approxns. are impt.

perhaps mention th. epist. "impotence theorem" corresp. to th. unsolvability of th. "halting prob."

8) Applications: a) Bernoulli seq.

b) Codes using defns.

$$\Delta \text{ aprip of string} \approx \underbrace{f_A \cdot f_B}_{\text{cost of defining } \alpha \equiv AB} \underbrace{e^{\frac{m f_{AB}}{2}\left(\frac{f_A \cdot f_B}{f_{AB}} - 1\right)^2}}_{\text{rel. "saving" in code cost.}}$$

m = no. of symbs in original seq.

Note m factor — so expt. is large, if m is. even if $\left(\frac{f_A \cdot f_B}{f_A} - 1\right)^2$ is smal

This was written as a Fortran computer pgm. - using a kind of hill-climbing.

c) Grammar discovery for CF PSL.

How Gramm. discy can be used for MT discy — (using special kind of PSL).

Assignment of aprip to gramms. and "Goodness of fit" to corpus.

~~Hvr., model is still unsatisfactory, tho it seems to be a bit useful.~~ Model seems reasonable, but doesn't entirely corresp to ordinary CF. Grammar. (i.e. corresps. to a "unique pars PSG")

d) Arith learning. Design of machine that could begin problem of improving itself.

Talk: wed Dec 16, 64 : Marv's class.
3PM

Bldg 10, Room 275

Discuss. ind. inf.: Show how it is main problem in various art. int. probs: E.g. (a) Learning of any kind, (b) Learning char. recogn. (c) Learning to (1) Play ~~checkers~~ chess (2) Play good chess (d) Learning a lang. (what this means — i.e. learning relatn. betw. events in R.W. and strings of symbols.

Disc. of Bayes' Theorem. — how usa / depends on Apri distribn. of all possl. universes

Disc. of various forms of cmii

Disc of th. random Operator method — th. problem of "Randomly constructed" ~~[illegible]~~ Tms (define random in this case)
Relevance to "random nets".

Applications to stoch. Grammars.

Look at some old Notes for this "same" lecture.

New Developements: Work on th. Gen'l. task Solving problem (a) By Slagle. (b) " myself.

Room
Bldg 10-275

My 11, 66   Gn in Marv's class at MIT.   My ~~11~~, 66 : 4-270 4 PM   1

I. What is inductive inf.? examples:

~~(c) weather coin flipping (how much wt. to use on apri hypoth?).~~
~~(d) irregular polyhedron. w. nos. on sides.~~

a) Weather
b) Horse racing
c) stock Mkt.

(d) ~~sym~~ loaded die
(e) "symm" die.

a ≡ trials w. face a up
b ≡ Σ trials

$\frac{a}{b}\ \frac{+ a'}{+ b'} \equiv$ apri

g ~~e~~ Sci. Laws [my main introdn. ≡ "sci. Method" - How to get new hypoths. that fit past data, and are likely to extrapolate well. Charles Fort.] example of hypo that fits all data, but doesn't extrap

f first. Curve extrapolation. 10 pt fit w. 10 params.? Just how many params to use? (compromise betw. ~~good~~ good fit and few pts.
Q. of "how many params" is rigorously meaningless.

$$\sigma^2_{true} \approx \frac{\sigma^2_{observed}}{1} \cdot \frac{N}{N-M} \quad = \infty \text{ for } M = N.$$

M — no. of params.

h) ~~Ider~~ ≡ Theorem proving:
1) what xfmn. to try next. (small set ≡
2) what subgoals to work toward.
3) Genl. problem of tree search.

i) Chess, checkers:
1) best move ~~within~~ (no. constraints) → takes large trg. seq. to learn.
2) " " within constraints.
~~Evaluation~~ 3) Break-down of prob. : Play out, Then evaluate.
(A) How to evaluate - Invention of new chars, by combn. of old.
- wts. to various chars. - optzn. of these wts.
(B) which move trains to try (genl. search prob.).

j. Patt. recogn.
1) Examples of letters A, B, C. , Then new shape: to categorize.
analogy to curve fitting. Perceptron: trg. till 100% correct — Then what is expected future accuracy: we have express. in curv. fitting, but none ~~in~~ for more genl. patt. recogn.

K. Gn. set of ss's to find grammar. (examples: $a^{(n)}b^{(n)}$ ≠ c $b^{(n)}a$ mirror lang. Gn.
~~k~~ l. M.T. : 1) Gn. bunch of xltn. pairs to ≡
S → aSa
S → bSb
S ≡ mirror.
S → aa
S → bb.
(a) to form xltn. rules
(b) " " grammar rules.
(c) gn. new source s., to find target S.

Hill climb
method
Ordering
complexity

A. U of Texas approach: Make gram. to fit old examples - done "by hand"
B. Russian approach. make simplest mach to xlt. known S's.
C. Faulk approach via "Distance"

Refs.
1. ≡ Ramon Faulk
g.o. watson ~~Wa~~ Res. Cent.
Box 218 Yrktn. Hts. N.Y.
2. Eugene Pendergraft.
Ling. res. Cent.
~~Dept of~~
Box 7247, Univ. Sta.
Austin 12, Texas

May 11, 66

D. ~~my~~ approach: Grammatical extrapln. — like curve fitting.

Use of ~~[illegible]~~ "genlzd. lang" — in which ass's are pairs of ~~[illegible]~~ x/fn. pairs

---

II Genl. solns.

a) ~~[illegible]~~ all induction probs are expressable as series extrapln.

b) Bayes' idea: a prip.

... ~~c~~ b a	a	.01
... ~~c~~ b a	b	.02
... c b a	c	.03

proby of c = $\frac{.03}{.01+.02+.03}$.

~~c) Take all strings w. all info one wants to things relevant (use more than necy).~~

c) Univ. machines: Turmac: computer w. inf. expandable memory.

Input, output, Universality.

$$M_i(x) = M_U(D_i \Delta X)$$

$P_S \quad \tilde{\propto} \quad 2^{-N}$ N = shortest dern.

d) ~~[illegible]~~ $\tilde{\propto} \sum_i 2^{-N_i}$.

N_i = length of i^{th} binary dern. of S

$$\propto \lim_{\epsilon \to 0} \sum_i \left((1+\epsilon)2\right)^{-N_i}$$

Also take future continuations into account.

May 11, 66

~~Normzn : $P_S = \frac{\sum_{j=1}^{M} (2(1+\epsilon))^{-N_j}}{\sum_{\text{several poss. contiguous}} \sum_{j=1}^{M} + \sum_{j=1}^{M} + \sum +}$~~

SW : woops! eq. 1) use $\frac{1-a}{r}$ not $\frac{1-a}{2}$!

use of $2^{-\cdot}$

NO! : Binary input alphabet.

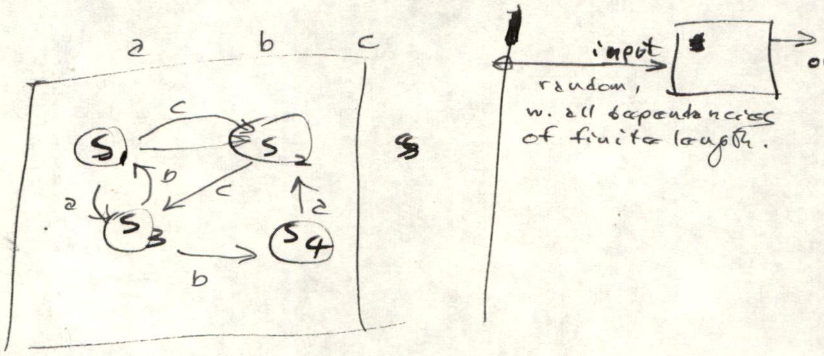
e) All proby methods :

① $\sum w_i p_i$

② $w_i = p_i \cdot p_i'$ $\quad p_i' = \prod_n p_i(a_n)$

③ is at least as "good" as any other meth.

f) Model of universe : E.G. finite state "model" of statistical string.



g) Fixed string method. to get a prob of S.

put in all strings of lengths 10^6

take out all strings ($< 10^6$) $= N$

M strings start w. S. $\quad$ proby $= \frac{M}{N}$

h) All 4 very proby equiv.

My 11, 68

h) Does probty vary w. nature of Univ. Machine?

SN) Use = probable string model (of universe, to see ~~if it was~~ how my model treats it — since this model clearly has a difrat "physics" than my model.

SN$_2$) Perhaps use "all poems" model to deal w. "indip. of tormac" problem.

Idea of Universality:

M_1, M_2

P_i^1 from M_1, P_i^2 from M_2.

~~use of "shortest method"~~

$$2^{N_{Tr}+N} = 2^{N_{Tr}} \, 2^N$$

I took ~ $\frac{1}{2}$ hr. for PP 2, 3, 4! Complaint was that I didn't show just how my own work was useful in solving t. probs. of Page 1.

T. utility of t. general ind. inf. theory: Simon, Newell, tried for 4 yrs. to get a device to learn new heurs for GPS, by devising a suitable lang., etc. With Cmii, I think I would know how to go about this much more effectively.