See (28.01) it. for ideas to mention 1

69 Plou rave

Willis' Induction Mechanisms — A Review and Some Corrolaries.

411

This memorandum is will to ~~a~~ briefly review what I feel are Willis' most important results, present a few corrolaries to his system and discuss the significance of the work. My ~~present, very strong~~ impression conclusion (1973) ← (not so certain 1974!) is that it describes a ~~both~~ necessary as well as sufficient set of systems for inductive inference — i.e. Given a sequence of discrete symbols generated by some finitely describable stochastic source, the systems will, for sufficiently long samples, be able to predict the probabilities of the next symbols, with errors in ~~probability~~ approaching zero.

Since the systems are usually a bit awkward, it is natural to ask if there are other, ~~more~~ simpler, systems that will do about as well. In a certain well defined sense, no other system can do as well as the ones Willis describes.

Willis' work was (1, 2) ~~appt~~ the outgrowth of some of my own work on induction (3, 4). I described a method for evaluating the probability of any finite sequences of discrete symbols, by considering the lengths of various input programs for universal Turing machines that might have given this sequence as output. The probability was the sum of the weights of ~~all~~ the possible corresponding input programs. A program of length n, was given weight 2^{-n}, if the machine had a binary input alphabet.

I ~~outlined~~ gave some reasons why the probability assignments should not depend much on the ~~exact~~ specific universal machine that was used, but I gave no real proof. I also gave various arguments to show it was likely that the probability values obtained were close to those that one ~~ex~~ expected for various intuitive reasons, and I gave some examples of application of the system in which the results seemed reasonable. 412

Later, (5) I devised a criterion for the adequacy of an induction system and ~~proved that no system~~ ~~[illegible]~~ while it was clear that no system could satisfy this criterion, it seemed reasonable that the ones I had proposed, would approximately conform -
I shall discuss this criterion later in regard to Willis* ~~[illegible]~~.

~~[illegible]~~ A most serious deficiency in my own work ~~[illegible]~~ surrounded Turing's "halting problem". It was clear that the system of probability evaluation proposed was not "computable" in a rigorous sense — and though approximation methods readily suggested themselves, it was not clear as to how this affected the accuracy of the results.

The thing that first stuck me in Willis' paper was his method of ~~scheme~~ dealing with the halting problem by defining a set of Turing machines for which ~~verbal~~ there is no halting problem, and ~~defining~~ ~~a prob~~ associating with each such machine, a computable probability method, of the type I had previously described. ~~That~~ None of these machines, of course, ~~were~~ are universal, but there are an infinite number of them, and one can, by selecting a sequence of them, of increasing computing power, obtain a sequence of probability measures of greater and greater accuracy.

called. (FOR) — Willis calls these limited machines "Frames of Reference" perhaps since they relate their ~~output, the thing~~ ~~described~~ inputs (descriptions or coördinates) to ~~the thing described (a~~ their outputs (the thing or point described). —

(see page 1 of Willis' for his reasons for calling them FOR's)

The FOR is a 3 tape Turing machine with ~~unidirectional~~ input and output tapes, and bidirectional memory tape. ~~[illegible]~~

It has ~~The FOR~~ always, for an infinite ~~input~~ string of input symbols, an infinite ~~output~~ string of output symbols. It never gets into an infinitely long computation sequence without reading or printing.

If it ever reads a square of its input tape, it will eventually write on ~~[illegible]~~ at least one more output square. Conversely, if it ever writes on an output square, it will eventually read again on its input ~~square~~ tape.

Untrue e.g. when it gets in S or U loops

Then it would never stop.

for any finite input sequence, it will have only finite output sequence. ← No! e.g. S & U seqns

~~By an ingenious use of measure theory Willis assigns~~

I'm not sure of this. (I think it is true) → for any finite output sequence, it will have only ~~a~~ finite input sequences.

~~probabilities~~ ~~Though Willis considers machines~~

~~If~~ The FOR has, in addition, 2 special symbols, S and U.

If it ever prints S, then it will continue to prints nothing but S's for the rest of its output sequence.

If it ever prints U, it will correspondingly continue with U's only. ~~The S sequence corresponds to an ordinary sequence machine stopping.~~

Though Willis considers machines with any number of symbols in their alphabets, there is no loss of generality if we allow only 2 ordinary symbols, 0 and 1.

When an FOR has been devised to partly mimic the behavior of a more general ~~[illegible]~~ (e.g. universal) machine, the S sequence corresponds to the machine stopping; The U sequence ~~[illegible]~~ indicates uncertainty in whether the machine is in a infinite computation sequence or not.

?

25

The S & U symbols: It mite as well stop when it prints either, but perhaps W. defined them in this way for economy in stating things. ~~But~~ well — A U symbol really means a difrut thing from an S symbol.

A very simple, ~~set~~ yet powerful, set of FOR's is that based on the Time limited Universal machine. ~~The~~ ~~[illegible]~~. ~~The~~ ~~The~~ The T^{th} such machine acts exactly like a specific Universal Turing machine for all computations that take T steps or less, ~~[illegible]~~ except that it prints an infinite sequence of S's when the universal machine would stop. As soon as a ~~[illegible]~~ computation takes longer than T steps, it prints an infinite sequence of U's

But then, how could it do that stuff about always printing on output if an input remained a vise-versa? Well, it often prints U's

To assign probabilities to sequences, Willis uses measure theory in an ingenious way that insures that the probability assignments satisfy all of Kolmogorov's axioms for ~~a~~ probability.

Ref. →

→ Perhaps of more interest: A ~~m~~ Computation Bnd CB in which computation Time allowed is ∝ to no. of output symbols printed this far.

This is done by associating with each of the ~~[illegible]~~ infinitely long input sequences, the point unique point on the interval (0, 1 ~~This can be done by considering each~~ ~~[illegible] with a function [illegible] of~~ ~~[illegible] which is the infinite binary expansion.~~ that corresponds to the ~~fraction~~ real number of which that sequence is the binary expansion.

It is clear that the total measure of all such points is 1.

The probability associated with any finite output sequence, α, is the total measure of all input sequences, that give output sequences that start out with the sequence, α.

It is clear then, that if $P(\alpha)$ is the probability assigned to α, then $P(\widehat{\alpha 0}) + P(\widehat{\alpha 1}) = P(\alpha)$ —— since all sequences that start out with α, must <u>also</u> start out with either $\widehat{\alpha 0}$ or $\widehat{\alpha 1}$.

(Nov 20,

TM415

106
63

128 Sept 24
234 Nov 24
63

The Possible and the Impossible in Inductive Inference.

α I will briefly review my own ~~[illegible]~~ work in induction and ~~explaining~~ ~~its more~~ ~~[illegible]~~ its ~~[illegible]~~ deficiencies, ~~The~~ then give will be followed by a more detailed ~~[illegible]~~ description of Willis' work and how it overcomes these difficulties. This will be followed by some very general theorems about the limitations of induction ~~inference~~ methods.

126
1.6
1.7
6.8
102
1.7 p/d.

Finally, I will discuss the significance of this work ~~[illegible]~~ in relation to some theorems on what cannot be done by any inductive system,

? ~~This will be~~ followed by a statement of what can be done in induction, and an illustration of a few systems that do this.

and will conclude with a ~~set of solutions to the induction problem that constitute~~ → γ

β It is possible to show, that given any computable method of ~~[illegible]~~ evaluating the probability of successive symbols of a sequence, in terms of the known past of the sequence, one can always devise a stochastic source of symbols for which this ~~computable~~ method ~~would~~ will fail miserabley. In fact, if it were to be used on a binary sequence, its probability ~~[illegible]~~ assignments for each conditional probability would always err by a factor of at least (2 − ε). ~~oneself~~ Fortunately, One ~~[illegible]~~ need not limit ~~oneself~~ one's self to computable probability methods. ~~It~~ ~~If~~ ~~An~~ infinite sequence of computable probability evaluation methods can be used in which

.30 → the limit of the probability assignments exists but is not computable. A certain ~~set of such~~ equivalence class of such sequences will be ~~[illegible]~~ described that give optimum in a certain defined sense, optimum predictions for any arbitrary ~~[illegible]~~ fixed sequence. Any sequence of methods not in the equivalence class will give ~~[illegible]~~ arbitrarily worse predictions for sufficiently long sequences. Two ~~[illegible]~~ specific members of the equivalence class will be ~~described~~ discussed.

mmit; ~~[illegible]~~ have but may be include later.

This equivalence ~~se~~ class of sequences of probability evaluation methods ~~is~~ ~~T~~ this is the certain well defined sense, a unique, optimum general solution to the inductive inference problem. Members of the class predict as well ~~as~~ as well as can be ~~shown~~ predicted and no non-member of the class can ~~[illegible]~~ predict as well.

Two specific representatives of the class will be discussed.

By using other criteria of optimality, other, less well defined classes of probability evaluation methods can be described, and these will be discussed briefly. (232.12)

from (3)

~~The class~~ A in a certain well defined sense, ~~a~~ optimum solutions for the ind. ~~[illegible]~~ prob.

O.k.; T. above looks like an adequate abstract.

On t. meaning of this optimum soln. to t. ind. inf. prob.!

what were t. 2 examples?

Time & Memory limitation;

What other classes of prob. evaln. methods?

t. Ann Rev 69 129.35 — ~~[illegible]~~ 131.04 (Sept 24 1969)
gives one t. idea that I really didn't have any
clear idea on how to prove it ~~at~~ on Sept 24.
Hvr. — ~ ~~[illegible]~~ 2 more months to Nov 26!

~~The~~ A Solution to the Problem of Inductive Inference. ← title → or A Unique Soln to t. prob of Ind Inf. or A Nec & Suff Soln. to t. prob. of Ind Inf. ①

TM 417

~~[illegible] that It is shown that the formulation of in~~

A particular general form of the induction problem is discussed. ~~It is known that [illegible]~~ While it is known that effectively computable solutions do not exist, it is shown that the limit of a sequence of computable approximations ~~[illegible]~~ can indeed exist — ~~tho it is not~~ though ~~the~~ this limit is not itself effectively computable.

does Willis actually show t. limit exists?

Willis treatment of induction is shown to be a sequence of computable induction methods ~~whose~~ limit gives an approximation to probability values whose total squared error converges to a finite value for sample sequences of arbitrary length. ~~These sequence~~ This limit also possesses a stronger convergence property.

It is shown that the only ~~[illegible]~~ sequences of approximations to probability ~~[illegible]~~ having this convergence property are of the type described by Willis, so that this formulation is both necessary and sufficient for this ~~[illegible]~~ convergence property.

The question of obtaining optimum approximations to probability values using a bounded amount of computation is discussed.

~~Perhaps~~ A unique solution to the problem of ind inf.

~~The induction inference problem is defined to be~~ the ~~[illegible]~~ ~~extrapolation~~

ABSTRACT.

This paper presents several corollarys of W's theorems that reveal the extent to which his work constitutes a unique soln. to t. ~~ind~~ problem of ind. inf.

All forms of induction are said to be reducible to the extrapolation of a ~~[illegible]~~ sequence of symbols containing all of the information to be used in the induction.

~~While it is easily shown that no eff~~

While it is easy to show that no effectively computable solution to this problem exists, Willis has proposed an infinite sequence of effectively computable approximations whose limit exists, though ~~it~~ the limit is itself not effectively ~~a~~ computable.

This limit satisfies a certain very strong convergence property — which in turn implies the convergence of the sum of the squares of the ~~errors in probability estimates~~ ~~estimated~~ deviations of the estimated from the true probabilities.

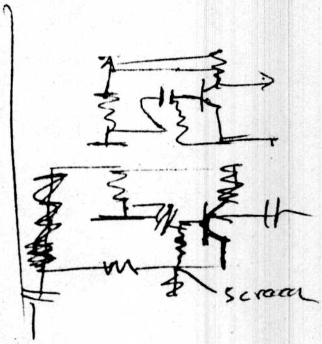
It is shown that the only sequences of probability approximations having ~~[illegible]~~ the strong convergence property are those of Willis. In this sense, they ~~means~~ constitute unique ~~necessary and sufficient~~ solution to the induction problem. Convergence of the total square error ~~[illegible]~~ is, however, a weaker criterion and it is possible to have a sequence of induction methods that approach this as a limit, that are not one of Willis' sequences.

~~The use of these sequences in practical approximation of probabilities is discussed, and their eff~~

perhaps omit

The effect of this formulation on various difficulties in induction, and scientific method is discussed, as well as the application of these ~~methods to~~ approximations to practical prediction problems.

For Kolmogs address try
1) Willis
2) McCarthy.



Discuss ① various constraints on soln. e.g. CB ∝ Length
② the case of infinitely long seqs, i.e. Desc length ∝ ln corpus length

Title

~~The~~ Uniqueness of Willis's solution to the Inductive Inference Problem.

or A unique Soln. to t. problem of Inductive Inference.

~~The pres~~ This paper presents several corollarys of W's theorems ~~that [illegible] his work to the general problem of induction and~~ that ~~[illegible]~~ ~~clarify~~ reveal the extent to which ~~[illegible]~~ his work consitutes a unique solution to ~~this problem~~ general prob. of ind inf. ~~the induction problem~~.

Th

It is shown that the only sequences of probability approximations ~~[illegible]~~ having the strong convergence property are those of Willis. In this sense, they ~~mean~~ constitute a unique ~~Necessary and sufficient~~ solution to the induction problem. Convergence of the total square error ~~[illegible]~~ is, however, a weaker

~~The use of these sequences in practical approximation of probabilities is discussed, and their eff~~

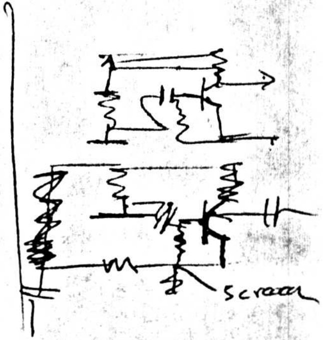
criterion and it is possible to have a sequence of induction methods that approach this as a limit, that are not one of Willis' sequences.

---

perhaps omit

The effect of this formulation on various difficulties in induction and scientific method is discussed, as well as the application of these approximations ~~methods to~~ to practical prediction problems.

For Kolmogs address for
1) Willis
2) McCarthy.



Discuss ① various constraints on soln. e.g. CB $\propto$ Length
② The case of infinitely long strings, i.e. Desc length $\overset{k\times}{\propto}$ ln corpus length

---

Title.
~~The~~ Uniqueness of Willis's solution to the Inductive Inference
or A unique Soln. to t. problem of Inductive Inference. Problem.

~~The pres This paper presents several corollarys of W's theorems that [illegible] his work to the general problem of induction and that clarify reveal the extent to which [illegible] his work consitutes a unique solution to this problem general prob. of ind inf. the induction problem.~~

T

Introduction

It has been known for quite some time, that certain forms of the inductive inference problem are unsolvable. In one very ~~rather~~ general form, ~~one~~ one is given an arbitrarily long sequence of symbols that has been generated by a finitely describable stochastic source ~~that~~ ~~has a finite description,~~ and it is required that one estimate the probability distribution for the next symbol. There exists no algorithm that will do this for all possible stochastic sources, since, given such an algorithm, it is easy to design stochastic sources ~~sequences~~ that will frustrate it.

I will give an expression for a kind of algorithm that almost does this impossible thing — but the ~~expression itself~~ values of the expression are not effectively computable. This ~~makes~~ frustrates the forgoing counter example construction, ~~not only~~ ~~so it is impossible to construct counter examples that will frustrate it~~ and it will be shown that the expression actually does give very good predictions and that no other expression of this sort can do better. The expression itself is the limit of an infinite sequence of algorithms, each of which is effectively computable.

While the limit is itself not directly usable for induction, the manner in which this limit is obtained suggests various approximations that might be used. If the computational complexity stochastic source is limited in almost any way, a computable solution to the problem can be demonstrated ~~shown~~.

It is ~~impossible to~~ usually impossible to compute useful upper bounds on the error in probability estimates made ~~for the~~ by these approximations. I shall, however, give some strongly felt conjectures ~~conjectured theorems~~ on ~~other~~ other measures of expected future prediction accuracy of these approximations.

¶ The present paper will first ~~dran~~ give an expression 420 for the "limiting algorithm" and will show that the limit, indeed, exists — Thought it is not effectively computable.

Next, the error properties of the limiting algorithm will be investigated. ~~[illegible]~~ Willis has devised an error criterion for induction algorithms. I will show that if Willis' ~~[illegible]~~ total error ~~[illegible]~~ ~~[illegible]~~ is ~~a certain~~ some constant, k, then the total squared error of probability estimates for sufficiently long sequences must converge to $\leq k \ln 2$ (or some ~ thing). k is always less than the Kolmogorov complexity of the stochastic generator. ~~$~~ ~~[illegible]~~ I will then ~~An example will then be produced to~~ show that Willis' error criterion is stronger than the ~~[illegible]~~ total squared error criterion, by producing an example of an induction algorithm that satisfies the total square criterion but not Willis' criterion.

The last theorem will show that any limit of inductive algorithms ~~that satisfies Willis criterion~~ for which Willis' total error is bounded by a constant, must be one of the kinds of algorithms that are being described ~~[illegible]~~ in this paper — i.e. that this limiting algorithm is, in this sense, a unique solution to the inductive inference problem.

I will then give some expressions that I feel are likely to give ~~estimates~~ upper bounds on the expected future ~~[illegible]~~ "information content per symbol" of the sequence being analysed, with respect to ~~any~~ ~~[illegible]~~ specific computable inductive algorithm.

42(

The ~~present paper~~ discussions in the present paper follow directly from Willis' results and show that ~~and~~ ~~may be~~ ~~attempt to~~ show ~~that to the extent~~ ~~regarded as~~ an ~~investigation of the extent to~~ ~~which~~ ~~the problem is solvable.~~ ~~the~~ His methods ~~results~~ do, indeed, ~~which his work constitutes a solution to the inductive inference problem~~ constitute a formal solution and in addition ~~to an important extent~~, suggest practical solutions to a ~~[illegible]~~ general form of the Induction problem

END

~~The~~ ~~works~~ The induction methods described here were first treated in a relatively informal form in 1964[1] ~~This~~ ~~past~~ ~~methods were~~ ~~refined~~ ~~reformulated~~ reformulated by ~~Willis~~ ~~into a~~ ~~an exact formulation and he was able a mo~~

At this time, there were many unanswered questions — about the ambiguity of the method — ~~about~~ and whether it did indeed, solve any problems; ~~and~~ ~~[illegible]~~ whether the results had any rigorous meaning and ~~[illegible]~~ ~~whatever~~

In 1970, Willis[2] reformulated these methods into a rigorous form and was able to prove many important theorems that ~~clarified~~ answered most of these questions. ~~The question of~~ ~~to~~ The extent to which his ~~te~~ tequniques constitute a solution to the induction problem ~~was~~ was not, however, altogether clear.

Oct 8, 73

Outline of Paper!

Some Unsolvd Probs in Induction Theory.

1) Introduction: This reviews history of problem:

a) Ancient, classical.

b) my work

c) ~~[illegible]~~ W's work.

Then tells what state of problem was in after Willis.

Then tells what t. present paper shows: i.e.

① W's expressn → a limit as CB → ∞

② Σ error_w $< k$

Discn. of W's error defn. — that it can be + or −, so, superficially at least, t. fact that Σ error $\leq k$ does not insure that t. individual errors are small.

③ That $\Sigma\ \text{error}_w < k \Rightarrow$

$$E\left(\Sigma (p_i - p_i')^2 < k\cdot\left(\frac{n}{\ln 2}\right)\right)$$

④ Willis' error criterion is stronger than since I will give example that Σ's satisfys but not

⑤ ~~[illegible]~~. If t. ~~[illegible]~~ seq. of Pem's → a limit, & this limit ~~[illegible]~~ has an error as small as W's, then this seq. of Pem's is one of W's kinds.

Or! the only kinds of sequences of pems whose limit ~~[illegible]~~ has property, are t. seqs of Pem's ~~[illegible]~~ proposed by W.

$$p = \frac{n}{1+n} \approx 1 - \frac{1}{n}$$

$$= \frac{1}{1+\frac{1}{n}}$$

$$= 1 - \frac{1}{n} + \frac{1}{n^2} - \frac{1}{n^3} \cdots$$

$$1 - \frac{n}{1+n} = \frac{1}{1+n}$$

$$\Sigma \frac{1}{1+n} \text{ diverges}$$

38 ← ⑥ That Lap's rule & other rules of ~ form are wrong in t. case of non-problistic seqs. Is One of few "hard" results.

423 R[illegible]an Brown

⑦ More practical results.

ⓐ Laps rule & ~ rules are wrong for determ. seq. Also, its wrong if ~~[illegible]~~ proby ratio is $\frac{1}{2}$, $\frac{3}{8}$ or any computable expressn. ← not so clear, hvr. — I'm not sure.

ⓑ If t. corpus is of bnd complexity — bnd in practically any way, then we can devise computable Pams that converge "as fast as possl." e.g. $CB \propto T^{n}$, t. length of t. seq.

.10 ⓒ Most laws in the sciences (prob. or determ) ~~are not~~ do not have finite descns. In most (and perhaps all) ~~[illegible]~~ such laws, there is a simple modifn. of W's results to deal w. them — i.e.

Σ error → & ln T

Give example of simple Bern. seq.

66, 1616

→ Hvr, if prNy = $\frac{7}{8}$, say, $\Sigma(\text{error})^2 <$ const.

$\Sigma(\text{err})^2$ diverges in conventional statistics.

So this is about i. same as t. p=1 case

12 Helioka St. thru Oct 21

Perhaps after ⑤ : (i.e. 4.38): Have long section on t. Significance of t. results * To stuf on lap's rule on .10, on ~~[illegible]~~ laws of nature that have compn. bnds like ~~[illegible]~~ linear in T or some known f(T).

Note that all of t. sci. rules proposed thus far are probably of ~~[illegible]~~ "relatively small" C.B. (i.e. not much > linear in T)(?). — Perhaps not so for mechanics of Gases — in which compns may ↑ exponentially.

Q. of whether there are other ways to limit t. search algorithm → a "soln" is possl. (i.e. E(error) → 0 as T → ∞). **Blums** have solved this for t. ~~det.~~ deterministic induction problem. For deterministic induction, there is ~~[illegible]~~ simple soln. if t. set of allowable seqs is recursively enumerable (≡ [illegible] r.e. ≡ countable).

(SN) Politics: It mite be well to give a talk at a ~~[illegible]~~ meeting related to "Computational Complexity" — At t. Detroit meeting Hartmanis has an invited talk on this. — Anyway, it sort of ~~[illegible]~~ is related. [Abstract in by Dec 1]

Another possy. is Phil. of sci. symp.
Also Statistics Symps. ← [Perhaps AAAS Meeting or That New England Yearly thing.

So anyway: This is a good way to present part of Y. conclusion; i.e. ① W. no. restrictions on Y. stoc algorithm, there is no soln. ② if t. restriction is a suitable CB as a funct. of T (or 'linearity' in C or exponentiality in T) then I have this very rapidly converging method ③ T. general Q is: what are some other constraints on Y. set of algs that give convergence — & how fast do they converge? Will if t. algs are r.e. will this do it? Since this constraint of ② gives perhaps as fast convergence as one could ever get, one suspects there are much weaker constraints

.15 that will give slower convergence —

So that one part of Y. conclusion — its a statement of what has to be done.

T. second part is on practical induction.

T. third part is tying these results into t. probs. of scientific method -

.20 "Given a certain computing capacity and a certain sequence of data — how can we best use that comp. cap. to do induction on that data seq? T. use of CMI directly in t. manner of W.'s paper ~~is~~ seems to be very wasteful — probably hopelessly so. Hvr., t. method does suggest ~~practical~~ techniques for induction

One of these involves a Goodness criterion for induction algorithms wrt a given data seq. The simplest ~~algorithm is~~ of these is $\frac{A+P}{S}$; where A is t. shortest length of descn of t. algorithm one has been able to find & P is t. probty of Y corpus wrt. that Algorithm. (A ≈ Kol compy)

A'+P is another where A' is ~~with~~ an approxn. to W. complexity (A' ~~<A~~ is always < A)

~~Also~~ Better yet is $\frac{P}{S-A}$ where S is t. no. of impt (binary) symbols in t. data seq., & there are other refinements that one can make.

These Goodness criteria give an upper bnd on how bad one expects t. predictions of that algorithm to be in t. future.

~~Since~~ ~~[illegible]~~ CMI ~~inherent~~ No inductive method can give a good estimate as to how much a proposed algorithm differs from the correct one, but the Goodness criteria I've descbd. are ~~useful~~ in many cases ~~of even more value than~~ of about equal value.

→ The problem of deciding which inductive algorithms to test. ~~is also~~ CMI also suggests solns to

Those will be algorithms that ~~are at least do~~ have the simplest dcrns. ~~The concept of "dcrn" simplicity while~~ Ordinarily t. simplest dcrns will be those having fewest symbols — but in order to do this properly, one must assign costs to each symbol, based on its frequency of use in the past. In general each symbol will be the abbreviation of a subroutine that was used in the past.

Another trick to obtain simple algorithms is to modify or "mutate slightly", algorithms that have been much usd in t. past. If it has been usd much in the past, its ~~abbreviation is chan symbol is~~ "name or "symbol" is inexpensive. The name of the mutation used ~~must~~ should also be of low cost ~~to~~ (ie. frequent past use) to help form a new inductive algorithm of low cost.

~~A third area in which much work~~ The Goodness criteria for algorithms that I've described are of uncertain value. ~~[illegible]~~ The reasoning behind them is plausable but inexact. I've tried them on some induction ~~me~~ algorithms involving linear regressn. & they give reasonable results.

4) Certainly much work needs be done in this area.

Oct 11, 73 Report Whit Diffie. 8

426

~~The selection of algorithms to test~~

While the methods I've outlined for selecting algorithms for testing seems reasonable, I have not really tested it in any practical cases. I have done some work on the cost of defining symbols and subroutines (I a C part 2
.07 — give pp), but certainly (much work needs to be done here). ← repetitives (7.40)

08 A third ~~new~~ direction of research suggested by CMI is in the philosophy of science. Q. of confirmation of a hyp. by gnl. evidence; Theory construction methods; wt. of a pri info., ~~Fit of the Goodness of~~ Attractiveness of Theory v.s.
20 wt. of exptl. evidence. etc.

6:05
6:22 P
6:35 P
6:45
6:53

freq. of various symbols.

Ngm. freqs

~~[illegible]~~
ngm transition freqs

So anyway, t. paper will be:

~~Introduction~~ Abstract:

Introduction: [redacted] (4.01 ff)

History

what paper shows In more detail than Abst.

---

signf. of Results

A. To what extent is it a soln.
What other kind of Solns. there are. | 6.01 — .15

B. Applicns to Phil of Sci., Sci method, etc. 8.08 — .20

C. Applicn. to practical inductions A.I. 6.20 — 8.07
Also mention ~~[illegible]~~ Kol's ~~[illegible]~~ complexity used in patt. recog by T. Cover.

D373 : → I <u>may</u> want to include a <u>proof</u> for t. <u>"all ipams"</u> method.
This is because in this form, CMI is closer to what one does in science. — it is t. wtd Σ of a bunch of hypoths — Hvr., I'll have to note various approxt. forms of it, & tell which ones are bad. — Also perhaps mention SVH "results".

At t. end, list some "Open Problems" : e.g.

1) If one uses t. "best" Pam for each predn., will this method converge as corpus length → ∞, & if so, how fast.

→ (Would it be possl. to show this is identical to a non a.c. "Willis" method & ∴ have ~ error convergence? → <u>or</u> show it's <u>not</u> equivt. to a Willis method so t. error <u>must</u> be > that of Willis's).

2) Show that $\xi \geq \frac{A+f(S)}{S}$ in proof of dern.

Thrms:
1) $\Sigma\, \epsilon^2 < k$
2) $\Sigma\, \epsilon^2 < k$ does not imply Σ bcost $< \alpha k$
3) Σ bcost $< \alpha k$ imply CMI
4) All pams method. (Mention that Σ of a few pams w. "proper" wts. is N.G..
5)

Open Probls: 1) If "best" pam is selected does this converge as $S \to \infty$?

2) Is $\xi \leq \frac{A+f(S)}{S}$ true or approxtly true.? (Is it true for $A = \log_2 P$ cost of pam or for $A =$ k cost of pam?))

(Note that "expected value" <u>can</u> be defind! But may be true only for Stationary Seqs. — i.e. seqs for which ξ exists.

3)

T. idea of a seq. of machines "approaching Universality" — I think my idea was that if M_i $(i=1|\infty)$ were t. seq. of machs., then for any finite machine M' $\exists j \ni (i \geq j \supset M_i$ can simulate $M')$.

Hvr., This defn. will not work if we allow M' to be any non-univl. machine: E.g. If we use M_i = Umc. w. T bound of i, then M_i is a FSM & it has a finite memy. — There are lots of non univl. trmcs having inf. memy that any FSM can't simulate.

Hvr., 155.10 (wills) gives a defn. of universality that is probly O.k.

I'm not yet sure that t. T Limited Umcs "approach universality" in an usable sense. Furthermore if we allow T to be a function of S, F(S) t. corpus length, we still can't get a sequence of $F_i(S)$'s $\ni$ t. resultant TLU's approach universality!
Hvr, for a fixd corpus length, S, t. TLU's do approach universality.
Without t. restriction to fixd S, there is no sequence of r.e. FOR's that eventually "covers" every specified FOR (& ∴ approaches universality).

T. uniqueness Thrm. is provd for rCPM's only. (if provd at all!)

.31 From 155.05 $\forall \ell \left[\lim_{i \to \infty} E \left(\sum_{arr} (p_i, F, \ell) \right) \leq k \right]$

I'm assuming these are interchangeable! "E" involves a finite sum.

.33 Also $\forall A^{(\ell)} \left(\lim_{i \to \infty} \sum_{arr} (p_i, F, A^{(\ell)}) \leq k \right)$ → is stronger than

---

155.30 / Thrm If $R_F \not\subset R_j$ then $\left(\forall \underset{so}{\underline{k}} \; F' < \underline{k} \, P_j' \right)$ is false for any $\underline{k} > 0$ | $\underline{k} \not\equiv k$

∴ $\forall \underline{k} \; \exists A^{(\ell)} \ni F(A^{(\ell)}) > \underline{k} \, P_j'(A^{(\ell)})$.

155.37 seems to be in doubt: $A^{(\ell)}$ is of measure (wrt. F), > 0 & P_j' & F differ by an arbly large factor wrt $A^{(\ell)}$,

.38 – But does this mean that this would be untrue? Certainly this would be untrue, but t. falsity of .33 does not imply t. falsity of .38.

So Thrm .33 is true: Also any seq. of P_i's for which .33 is true, imply that $R_i \rightarrow$ universality.

D 1373 : The idea here, is that t. "Expected value" Thm. (10.31) means that $F(A^{(m)})$ $(\cdot \log_2 \underline{k})$ is arbly large. We can get k arbly large, but t. choice of $A^{(m)}$ depends on k, and $F(A^{(m)})$ can be <u>very</u> small — even <u>zero</u> — since F is an arby. CPM — The P_j' would also have to be $\le \frac{0}{\underline{k}} = 0$ also.

Anyway, 10.33 is a <u>strong</u> Theorem about CMI. It is stronger (& perhaps more useful) than 10.31 (which 10.33 implies). But 10.33 is v.g. by itself, & it gives us more <u>direct</u> confidence in CMI than 10.31 does. 10.33 works for all individual sequences, $A^{(l)}$

.17 of <u>any</u> length, l. So for large l we expect mean "error" to be small.

T. lemma needed around 155.30 :

$$\text{If } R_{F'} \not\subset R_j' \text{ then } \forall \underline{k},\ \exists A \ni F(A) > \underline{k}\, P_j'(A)$$

This is true because $\Big\}$ If $R_{F'} \not\subset R_j'$ $\forall \underline{k} \ge 0 \Big(\forall A \;\; F'(A) < \underline{k}\, P_j'(A)\Big)$

i.e. $\forall \underline{k} > 0$, it is false that $\forall A\ F'(A) < \underline{k}\, P'(A)$.

This ↗ is true because of t. 2nd part of Willis ACM Thm 15.

<u>Order of presentation</u> So: ① First show that assoc. w. each Umc, (of W's "programable" type of Umc), — That is this <u>limit</u> — that t. limit exists but is not effectively computable.

.30 is a probly measure, P(A). That this proby measure is <u>uniquely</u> defined for that Umc. (i.e. that it doesn't matter <u>how</u> t. FOR's approach universality). [I <u>think</u> I must prove t. <u>Uniqueness</u>]

② ~~the Apoky~~ for any <u>cpm</u>, $P'(A)$; $\frac{P'(A)}{P(A)}$ is bounded, for <u>any</u> A. This bound is $k(P')$ & is indip of A (— t. prest of t. dcrn of P' wrt t. umc. This implies "mean error" of $\frac{\log_2 A^{(l)}}{l}$.

ommit this is included in ②

③ say P(A) is t. limit of a seq of CPMS, $P_i(A)$. then $\forall$ cpm's $P'(A)$, $\frac{P'(A)}{P(A)}$ is bounded if and only if ~~P(A) is a seq. of R~~ t. limit $\left(R^{P_i}\right)$ approaches universality as a limit.

TM430

D1373 : since $P(A)$ of 11.30 is t. ratio of 2 limits, it may be that $\frac{P'(A)}{P(A)} \leq k$ is not so easy to prove.

$P(A)$ may give some non-zero probys to "Unobservable events" — which may screw up t. proof!

This is an interesting Q: Does t. prog. imply that t. measure of all unobservable events is somehow limited? — Say $< \frac{1}{2}$? 345P

In W's Thrm 12 (ACM 252.17) on Pam → FOR: If t. Pam, P, has measure > 0 for only obs events (i.e. no U's ever show up) then we have no U's in t. corresp. FOR if its constructed in accord w. Thrm 12.

Well: say $P_u(A)$ is t. $\lim_{j \to \infty}$ of t. meas CPM's assoc. w. t. umc, U. $P_u(A)$, hvr, is not normzd: i.e.

$$P_u(B\frown 0) + P_u(B\frown 1) \text{ is not necly} \equiv P_u(B).$$

— $P_u(A)$ t. limit for observable (i.e. eventually converging) events.

This is usually $<$.

Hvr., t. normzn constant will always by ≥ 1: i.e. $P_u(A^{(g)}) \to \alpha_g P_u(A^{(g)})$

$\alpha_g P_u(A^{(g)})$ is normzd ; α_g is usually > 1, but always ≥ 1.

— This is because the unobs. events always have a measure ≥ 0.

Re: Normzn: oneway to think of it: T. ("Magical") Umc. goes into a U loop immediately if it would otherwise not converge at that pt.

---

28 Humm!! : Perhaps very impt.!! : Even before normzn. of P_u, we have $\forall A$ $P_0(A) \leq k\, P_u(A)$.

After normz, $P_u \to P_u' \equiv \alpha_g P_u$; $\overbrace{P_0(A) \leq k \underbrace{\alpha_g P_u(A)}_{\equiv P_u'}}^{\text{for all } A}$, since $\alpha_g \geq 1$.

Hvr, I don't think that we can have $P_0 < P_u'$ for all A's, since both are normzd. So α_g must be betwee 1 ; K — for all g values!

This means that there can be, at most, only a fraction $1 - \frac{1}{k}$ of U containing seqs. [i.e. at least $\frac{1}{k}$ of t. seqs must be observable (i.e. no U's). — indip of g. Here, k can be t. prost of desbng t. ~~simplest CPM possl. (say~~ simplest FOR that has no U's. — Perhaps t. identity FOR.

K

D 1473 : This is an extremely interesting Thm, if true! TM431

It means that for any l, there is $\geq$ a fraction $\frac{1}{k}$ of t. input strings of length l, which result in t. machine printing no U's (i.e. stopping w. an "S" loop or stopping after having read t. end of t. input string & perhaps writing after having read t. end of t. input string?).

---

06 On second thot, it is trivial! Say it takes a 10 bit string to get this UMC. into an S loop. Then at least a fraction 2^{-10} of all input strings (of length ≥ 10) will result in t. machine stopping.

Hvr, t. discn of 12.28 –.40 is of some interest: It ~~[illegible]~~ makes it possl. to find a perhaps smaller (certainly not larger) value of k ~~[illegible]~~ than 13.06 does.

Actually, t. mechanics of 12.28 – .40 is quite clear: Say ~~a~~ "String S is t. simulation string w.r.t. UMC M_U for some machine that ~~-~~ has no U outputs.

Then $S^\frown x$ as input to ~~Mu~~ M_U has no U's in its output.

These ⟶ constitute a fraction 2^{-N_S} of all ~~[illegible]~~ input strings that are of length $\geq N_S$ ($N_S \equiv$ length of S).

A Q. about Normzn.: Can I have a single normzn constant for each value of l, or do I need something more complex? — In addition to (Σ ~~[illegible]~~ probys of all seqs of length l) = 1, I need cond: 4P

$$P_{\blacksquare}(\beta^\frown 0) + P(\beta^\frown 1) = P(\beta).$$

for P_U, we have $P_U(\beta^\frown 0) + P_U(\beta^\frown 1) + P_U(\beta^\frown S) + P_U(\beta^\frown U) = P_U(\beta)$.

I think I want
$$P_n(\beta^\frown 1) = P_n(\beta) \cdot \underbrace{\frac{P(\beta 1)}{P(\beta)}}_{\text{unnormzd conditional probty}} \cdot \underbrace{\frac{P(\beta)}{P(\beta \overset{\frown}{0}) + P(\beta 1)}}_{\text{normzn. const. for condnl probty.}}$$

P_n = normzd

P = unnormzd

30
$$\text{so } P_n(\beta^\frown 1) = \frac{P_n(\beta) \cdot P(\beta 1)}{P(\beta 0) + P(\beta 1)}$$

$$P_n(1) = \frac{P(1)}{P(0) + P(1)}$$

So clearly $P_n(\beta^\frown 0) + P_n(\beta^\frown 1) = P_n(\beta)$.

⟶ My impressn: more complex: simple normzn. will not preserve

Seq	Probys unnormzd.
0	$\frac{1}{4}$
1	$\frac{1}{5}$
0 0	$\frac{1}{10}$
0 1	$\frac{1}{30}$
1 0	$\frac{1}{40}$
1 1	$\frac{1}{50}$

35

$$P_n(\beta^\frown 1^\frown 0) = \frac{P_n(\beta)\, P(\beta 1)\, P(\beta 10)}{(P(\beta 0) + P(\beta 1))(P(\beta 10) + P(\beta 11))}$$

Here $P_n(\beta^\frown 1^\frown 0)$ depends only on t. unnormzd probys of $\beta 1$, $\beta 10$ & seqs. that deviate from these by 1 bit at the ends. i.e. $\beta 0$; $\beta 11$.

D1673 R

[illegible] If hvr., we use t. simple normzn const that depends on l only, the const will depend on t. Σ of t. ~~p~~ unnormzd probys of all seqs. of length l. I think t. proby of all these other seqs are somewhat arby — so t. simple method couldn't work! — e.g. t. simple table of 431.35 (Left) — would not work.

So 431.30 are correct ~~are~~ (finite) recursive defns. for normzn. All of t. ~~B~~ unnormzd probys involvd in t. defns are monotone ↑ in t. C.B. & are bndd above by 1 ∴ they all approach limits.

Q.: Is t. limit in some sense unique? — No: see .39 ff — i.e. w.r. gn umc.,
.15 is it indep of how we →$^{t.}$ universality of this umc? I think it is
.16 equivt. (in unnormzd form & ∴ in normzd form) to a hypothetical ~~B~~ FOR in which
.18 somehow, U's are ~~—~~ immediately computed ~~of~~ for all non-stopping seqs.
If so, this **FOR** is unique w.r.t. a gn. umc. — (i.e. → CB "→" universality)

~~P.~~ A necy ~~f~~ quality [illegible] an CB must have in order for
.15 ff to be true! as CB is ↑, any particular seq. must either
ⓐ (i.e. print non-U's eventually) (terminate eventually), if t. umc does, ⓑ or (i.e. print U's.) not terminate ever, if t. umc. does not.

[illegible] ⓑ is true for any CB on a umc. — since t. CB will never stop if t. umc doesn't stop.

If ⓐ ~~does no~~ is not true, then there is at least 1 seq. for which t. umc stops but t. seq. of CB's does not. I think then, in such a case, this seq. of [illegible] C.B's would not "→ universality" in my sense — since I think there would then be at least one FOR that t. seq. of CB's would never "cover". i.e. say it take T steps for t. umc. to [illegible] stop. Then consider t. The limited UMC [illegible] of $\tau = T$. call it M_T It is a FOR & CB. I think that none of t. CB's in this sequence can simulate [illegible] M_T.

— But I'm not sure how to prove this!

.39 I think I can show a sort of counter example to something: I.E. Give a umc, M. We can have a seq. of CB's on M

434.01 sp
433.01

Another Q: Suppose ~~the~~ ~~ta~~ we have a set of pems that are t. limits of seqs of CPM's — one difrnt. seq. for each value of l ($l \equiv$ t. order no. of t. corpus symbol) — a different umc is t. limits for each l.

Say M_l is this l^{th} umc. that is approached as a limit.

Say we have a Fdss w. an associated FOR that has derv. brost b_l from M_l. Is it possl. for $b_l \propto l$, say?

Well, if it is, then this means that if each seq. of CPM → univty then t. ~~[illegible]~~ Σ brost error/ of this set of Pems can be $\propto l$ — which violates t. CMI thrm that I'm trying to prove.

On second thot, I don't think its legal to have t. Pem vary w. l. ~~Each Pem as~~ What ~~the~~ we have in t. Σ desired thrm. is a ~~sing~~ single non-a.c. Pem, that is t. limit of a bunch of CPM's. These CPMs are each defined for all l. — They may, hvr, give ~~p~~ unnormzd probys of zero for all $l >$ ~~lo~~ some lo — (a TLUMC would do this).

D1773 Report, [redacted]

TM434

.01 ; 433.90; 432.90 spac : that → univly, but they do not "Approach M" ; — i.e. The "FOR" that represents t. limit of this seq. of CB's is <u>not</u> ≡ t. FOR that is t. "limit" of M (in t. sense of 432.16–.18) ← (this would be t. limit of a seq. of <u>↑LU's</u>).

— Say M_0 is a <u>umc</u>. M_1 is t. same as M_0, <u>except</u>;

$M_1(1\frown\alpha) = M_0(\alpha)$; $M(1, 0\frown\beta) \rightarrow 1$, stop: where α & β are arby strings.

So : M_1 looks at t. first bit of its input ; if its a 0, M_1 prints 1 & stops. (It <u>may</u> be necy for M_1 to read t. rest of t. input before stopping, if it is to be one of W's "Pgmable Machns" — but this is no difficulty).

If ~~t.~~ first bit is 1, M_1 converts itself to M_0 & performs $M_0(\alpha)$ on t. rest of t. input.

Consider t. following seq. of CB's, [redacted] M^T : If t. first bit of t. input is 0, M^T goes into a U loop. If t. first bit is 1, M^T "follows" M_1 for T operations — unless M_1 stops (in which case M^T stops) If M_1 operates longer than T ops., M^T goes into U loop.

M^T is a ↑LU on M_1, unless t. first input bit is 0, in which case M^T → U loop.

M_1 is universal, since it can simulate M_0.

M^T <u>does</u> approach ~~univly~~ (unv univlty), since if M_0 can simulate a FOR in "time" T, then M^T can, also in time T+1 → 437.01

.20

<u>SN</u> There was some Q in my mind as to whether M^T → univlty, (tho its clear M^T is a C.B.). Since all M^T are FSM's, & there are FOR's that are <u>not</u> FSM's, it would seem that — <u>no</u> M^T would <u>ever</u> be able to simulate such <u>non</u>- FSM's.

In <u>general</u>, I think that there <u>does</u> <u>not</u> <u>exist</u> a 1 dimensional seq. of CB's that "→ univlty" in t. sense that I'd like.

We <u>can</u>, hvr., have a suitable <u>2</u> dimensional set of machns. For each value of ℓ ← (t. output corpus length), we let $T \rightarrow \infty$.

.35 Some quick, general results:

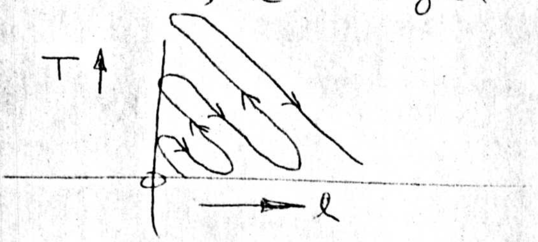
1) We have a seq. of Pams. These pams can be regarded as functions; The domain can be regarded as a 2 dim discrete point set: for each value of ℓ, (1 dim.) there are 2^ℓ possl. values of t. argument. T. range is (0,1).

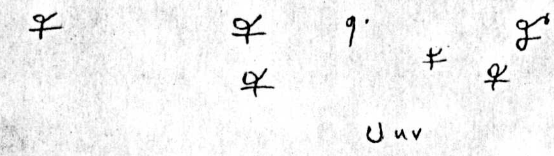
Anyway t. seq. of pams, p^T, converges to p^{0R} (say CPM) 435.01

D 1873 TM435

1.434.40 But t. convergence is **not** uniform in l. I.E. for each value of l, the limit ~~is~~ may be approached at a difrnt rate. This corresponds to non-uniform convergence of a Fourier series, say — The a series does "Ms converge" (which is a single quantity assoc. w. t. interval.).

2) If "Lim sup" rather than "Lim" is usd, I think ~~we get "uniform" convergence~~ we can use a 2 dim scan over T, l & get convergence:





15 → 437.20

3) That there exists no path in T, l space that will give uniform convergence (an old impotence thrm). This is due to t. fact that there exist p.r. functs whose computation time ↑ w. l faster than any r. function of l. (But not faster than an p.r. funct. of l, because assoc. w. every p.r. funct. of l $F(l)$ we can define a p.r. funct of l F' whose value is t. computation time of $F(l)$. — This can be done via the umc. defn. of $F(l)$).

So, O.K. The thrm about t. limit → $P_0(A)$ (≡ t. pem to be devd), is O.K.

So look up some thrms ~~on non-unifor~~ & definitions on non-uniform convergence in "Analysis". 6

Hvr., it is clear that ~~my~~ definition of "Approaching Universality", must be modified. Perhaps: For any $l <$ some fixd l_0, M^T can approach Unv (≡ universality) [$M^T \to$ Unv]. [$\lim\limits_{T\to\infty} M^T =$ Unv.]

Def

Say we have a seq. of pems, ∋ for all $l < l_0$, t. Σ best error is $< k_{l_0}$ [SN I think t. entire seq. of pems is a finitely derbable object]

Then t. FOR's assoc w. those pems → Unv for that l_0.

~~Furthermore,~~

A seq M^T, of FOR's → Unv for l_0, if for every FOR & output seqs of $l \leq l_0$, ∀ for R ∃ T_0 ∋ $((T > T_0) \supset (M_T$ can simulate R for $l \leq l_0$.

436.01

01: 435.40 : Note! in Willis 155.09 t. idea that l must be fixd [while CB→∞] is mentiond.
— But in t. defn. & dscn. following, this condition is not mentiond,

---

Anyway — For t. forward Thrm. to be true, i.e.

.04 $$\forall l \left[\lim_{T\to\infty} \underset{\text{biaserror}}{\underline{\leqq}} (P^T, P_0, l) < K \right],$$

$P^T \equiv P^{(M_T)}$
P_0 is any r CPM.

M_T must be able ~~to~~ (for large enof T) to [redacted] simulate R^{P_0} w. ≤ k bits, for all output strings ~~of~~ length l or less. This k must be indip. of l.

Now, t. Backward Thrm: we have a seq of Pems, P^T, & corrospg. M_T's.

[redacted] T. limit of .04 exists for this set, & for any P_0,

.17 ∃ a k ∋ .04 is true. Therefor t. seq. M_T has t. property that for any FOR, R, M_T is [for large enof [redacted] $T > T_0$; $T_0 = T_0(l)$] able to simulate R for outputs

.20 of R that are of length ≤ l.

Say .17 – .20 is not true — i.e. for at least 1 l [& one R] t. seq. M_T, is not (for large enuf T) able to simulate R. Then clearly [redacted] $R \not\subset M_T$ for ≈ all large T's, & ∴ [via Willis ACM thrm 15, 2nd part] for ~~large enof~~ a certain ~~[illegible]~~ l, $\lg P^{(R)}$ & $\lg P^{M_T}$ differ by an arbly large amount — i.e. > any specified k.

.25 DEF So — a defn. for: T. seq. of FOR's, "$M_T \to$ unv" ≡ ~~"$M_T \to$ unv"~~

[redacted] $\forall l$ R [FOR] $\left[\exists\, k(R) \ni R(\leq k) M_T \text{ for } T \geq T_0(l) \right]$. ← some ↑ funct of l.

perhaps commit here (i.e. t. k value depends on R, but not on l ~~[illegible]~~; T_0 depends on l.

.28 So: Thrm: ~~[illegible]~~ .04 is true for an arby [redacted] r-CPM, P_0, ~~[illegible]~~ [redacted] iff M_T (t. FOR's assoc. w. P^T) → unv.
~~[illegible]~~ i.e. iff $\lim_{T\to\infty} M_T$ = unv.
iff $\lim_{T\to\infty} R^{(P_T)}$ = Unv

SN In .04 we have "< k": for RW [redacted] Pems! < k ln l is more like it. — but then, P_0 is not a r-CPM. So we prove Thrm .28 — Then we discuss its applicability to Pems in t. RW.

437.01

D1873 R

TM437

01: 436.40; $\boxed{\text{434.20 spec}}$ On t. lack of uniqueness of $\lim_{T\to\infty} M_T$ = unv. Say M_T are a seq. of CB's on M_U, a univl. machine. Say M_T' are another seq. of CBs on M_U; both M_T & $M_T' \rightarrow$ unv. ~~[illegible]~~ Are t. limits of M_T & M_T' necyl. t. same — i.e. "M_U"? (By $\rightarrow M_U$, we mean t. FOR of 432.16-.18 assoc. w. M_U).

If t. counter example of 432.39 — 434.20 (commit page 433) (but it does look o.k.) Then a ~~[illegible]~~ seq. of CB's on a umc M_U that $\rightarrow$ unv, need not $\rightarrow$ ~~[illegible]~~ M_U.

A set of CB's ~~that~~ on M_U that always does $\rightarrow$ that M_U, is t. ↑LU seq.

So, assoc. w. every Umc, M_U, is t. "FOR" determind by t. ↑LU assoc. w. M_U. Hvr, all CB seqs on M_U that $\rightarrow$ unv ~~do not~~ ~~[illegible]~~ do not approach t. same "FOR".

Assoc. w. every Umc, M_U, is a FOR, F_U, that is t. limit of t. ↑LU assoc. w. M_U. ~~[illegible]~~ The F_U exists it is not e.c. from M_U. T. sense in which it "exists" is not clear. It does not "constructively exist".

.20, 435.15: Re: this 2 dim scan w. lim sup: Just what is one doing? ℓ is changing durig t. scan. Is t. lim sup of $\frac{\Sigma\, \text{bcost}}{\ell}$ taken? Suppose one goes thru this 2 dim scan, & refocus at each time, t. coding for which $C = \frac{\Sigma\, \text{bcost}}{\ell}$ is min. Eventually, this C will be arbly close to t. mean entropy of t. seq.

Hvr., its not clear as to what one is trying to do here. If t. goal is t. best extrapoln ~~from~~ for t. next corpus symbol, ~~[illegible]~~ using a gn. ccost, then perhaps t. best way is to use t. entire corpus & use all of t. available ccost on it.

If t. goal is to find "t. rite pcm" Then perhaps t 2 dim. scan or some modifn. of it is best. In t. 2 dim scan, one is almost sure that eventually t. ritepcm will turn up & be recognized. — There are, hvr, statistical Q's involvd ~~[illegible]~~ when one is considering small ℓ values for t. ℓ, T pair — since fluctuations ~~[illegible]~~ are then likely to give typically large C values.

Hvr, my previous concept here was to assume a reasonable CB on t. corpus (say $T \propto k\ell$ or $T \propto A^{\ell}$). ~~[illegible]~~ Then, by making my C.B. be $\propto \ell^2$, say I'd be sure of an eventual soln.

In t. 2 dim scan; If I were testing Pcms in this way, I think it would work. In general, I think that pcms are like p.r. functs — i.e. they are not r.e.

439.01 spec.
438.01

D1973 R : ~~[illegible]~~ Paper on "Significance" :: R Phil. TM438

Some points to cover:

a) That CMI results are ~~in~~ relatively invariant under changes of lang. (info lossless changes ≡ reversable ≡ non singular).

~~b~~) How to use CMI ~~for~~ operator induction or induction in which order of examples is irrelivant (use Gramm induction of I & C II as example) — perhaps mention special form of CMI for operator induction

c) How to use CMI for RW induction. Arby irreversable mapping of RW into corpus; A to D conversion. Semantics is not very simple. Goodness of induction depends in apri unknowable way on completeness of mapping.

d) One can (almost) never be sure one has found all reggs in a corpus — larger CB's ~~may~~ always help

e) All reggs in a corpus are expressable as a compression of coding. (See I & C II for examples — also note work on coding for a noiseless channel of unknown statistics ~~[illegible]~~ — note that ~~this~~ CMI is a soln. of sorts to this problem).

f) Discussion of ~~partly~~ obtaining a.c. soln. of CMI by ~~putting an~~ assuming an arby (but reasonable) a.B. on the corpus — e.g. linear or exponential) in l.

g) Perhaps have a 3 part paper: ⓐ Hard Results, Thrms. ⓑ Signif. of these results ⓒ Conjectures, Open Q's, etc.

h) Perhaps submit ⓑⓒ to Phil Mag: Or perhaps submit 3 part paper to Phil Mag! They do have stuff like this — e.g. Oettingers early, foolish, "computer learning" paper; D Mackay's "Info" paper.

i) ~~The~~ The k ∝ ln l phenomenon — how every CPM has assoc. w. it a r CPM that is arbly close to it (~~[illegible]~~ Willis' Thrm).

j) Some results in ~~[illegible]~~ linear Regressn.

k) SVH ~~Thrm~~ "Thrm".

l) Perhaps mention "All pams Mapped" — Thm +. (unproved SVH Thrm)

D1973 R

439
TM439

01| 438.40; (437.40 spec)! i.e. one can list all candidates for Pams, but t. testing of one of these objects, to see if its a pam, is never, for certain, possl. One can, hvr., use ~~Certain~~ various CB's for t. testing, to get more & more certainty.

Hvr, one can implement this C.B. thing easy enuf. All **CPMS → FORS** & are CB's on a umc (≈?).

.10 So perhaps: List all CPM's (This is simply a list of all FOR's — which is a list of all finite initial input strings to a umc). Assoc. w. each corpus length, l, consider all CPM of dcrn. length < l. Do a 2 dim scan on T (≡ CB) & l. Retain in mem'y, t. CPM (or set of CPM's) for which $\frac{\Sigma \text{ bcost}}{l}$ is max. (for suitably hy l to avoid statistical fluctuations).

T. foregg. method (.40ff) will probly eventually give close to t. best CPM (& ≈ t. best predn).

Hvr., suppose one has obtaind this best CPM at some l, T point. (say l = T = 10). One wants to use it for prediction at l = 20. What T value should be usd? Its impossl. to say. The T value is as much a part of t. dcrn. of t. CPM as its detaild code.

For a prediction at l = l0, t. ≈ best one can do is ≈ consider all CPMs of dcrn length ≤ l0 & use all t. ccost available. → 443.01

24

25 **SN** An Genl. interesting pt. came up in this .10ff 2 dim scan! One could use any radix one liked on t. corpus & t. thing would still converge as well. This amounts to clumping data chunks in arby sizes. It suggests that for best bcost/ccost, one mite want to "package" t. data in reasonable "chunks" — e.g. before one tries chunk to fit a set of linear regressn coifs to a corpus, one waits until a fair amt. of data comes in. This chunk of data can be regarded as a single symbol of large radix.

●→ ■ Perhaps when viewd this way, a major objection to "Sequencial coding" disappears.

One part of t. objection to seq. coding, was that t. dcrn of a complex coding system (like linear regressn) was very unlikely (very large bcost) — so a large no. of ll trial codes would have to be stord. My impressn. is that when linear regressn. is to be tried, it is then of rather hy likelyhood & ∴ of relatively small bcost. → Gen 444.01 spec

D2173 R

TM443

01: 439.24 : I Think I have this iff thrm. under control now. Best thing to do is write up t. proof w. as much exactness as possl. — close as possl. to what it would be in t. report.

Probably best to break it into 2 Thrms., because I'll want to refer to consequences of each $\frac{1}{2}$; e.g. th. forward (error $\leq k$ if R^{P_i} are a seq. of FOR's that → unv) part implys Σ sq err $< k$, || How to make Gn a umc, how to make from it, a seq. of FORs that → unv. || Thrm er expected value of error(s), at.

.10 O.k.: Thrm: Given any r computable Proby measure (r-CPM), P_a°.

If P_i ($i = 1/\infty$) is a seq. of CPM's ∋ $\lim_{i\to\infty} R_i = \text{UNV}$,

Then for any sequence $A^{(l)}$ of length l,

① $\lim_{i\to\infty} P_i(A^{(l)})$ exits forall $A^{(l)}$.

② $\frac{P_a^\circ(A^{(l)})}{\lim_{i\to\infty} P_i(A^{(l)})}$ is bounded/Above i.e. ∃ a constant k,

∋ $\forall l$ ($\uparrow$ $< k$)

If M_U is a umc ∋ $R_i \leq M_U$

($P_i \equiv P^{(R_i)}$)

and ($P_0 = P^{(R_0)}$) ; $R_0 (\leq a) M_U$, then

$\forall l$ $\frac{P^\circ(A^{(l)})}{\lim_{i\to\infty} P_i(A^{(l)})} \leq 2^a$ R.

Jan 31

Note: for any seq. of R_i → umc; there will perhaps be a minimal umc, M_U ∋ $\forall i$ $R_i \leq M_U$. To see this: Say there are 2 umcs, M_{U1} & M_{U2} ∋ $\forall i$ $R_i \leq M_{U1,2}$.

Then take t. largest M ∋ $M \leq M_{U1}$, $M \leq M_{U2}$.

This is some Boolean-like operation on M_{U1}, M_{U2} — like $\cap$ or $\cup$.

— Th. operation is dscbd in Willis (ACM).

Anyway, this minimal umc will give us t. smallest possl. a (?)

Maybe not: If M is an umc. , $\forall M_i$ (M_i can be umc or not) $M \leq M_i$. Hvr., we still have an ordering via "a" in $M_1 (\leq a) M_2$ (perhaps).

452.01

.01: 443.40 : It would seem that perhaps t. most relevant Q would be:

Given a Pcm, P_0, how large will "a" be? "a" is t. Σ b rost of P_0 wrt. "P_∞". "a" can be t. LUB of a_i, where $P_0 (\le a_i) P_i$. Since a_i is bounded above [for large enuf i (by hypoth. in t. Thrm to be provd) it must have an LUB.

Def .10 In t. Thrm of 443.10, $\underline{\underline{\lim_{i\to\infty} R_i = \text{umc}}}$ ← (This is identical to $\lim_{i\to\infty} R_i = \text{umc}$)

— But I don't think this implys → $\underline{\lim_{i\to\infty} P_i(A^{(l)}) \text{ exists for all } A^{(l)}}$. $i\to\infty$

(Prob this is true if R_i are ↑LU's

If only $\lim_{i\to\infty} R_i = \text{umc}$, then $R_0 (\le a) R_i$ for large enuf i

.16

If the P_i are nested so $R_i \le R_j$ if $i<j$ <u>then</u>. I think t. limit exists for all $A^{(l)}$. This make $P_i(A^{(l)})$ monotonic in i.

Hvr., if $P_i \to \text{umc}$ then I think a subset of t. P_i must → umc, & such a subset could conform to t. nesting,

---

Perhaps t. <u>trouble</u> is w. my defn of "$R_i \to \text{umc}$" :
(No trouble!)
as it is, it would seem that for a gn. R_0, $R_0 (\le a_i) R_i$ for $i >$ some i_0 — but that a_i could → ∞ as $i \to \infty$.

For $\boxed{R_i \to \text{umc},}$ we want a_i to be <u>bounded</u> for $i >$ some i_0.

Note that i_0 is a funct of $A^{(l)}$ in genl. & l in particular.

→ This bounded $\underline{\underline{a}}$ $\underline{\underline{\text{is}}}$ included in t. defn. of $\underline{\underline{436.25}}$ ⊙

Some Conjectures:

1) **If** $\lim_{i\to\infty} P_i$ exists, then I think t. R_i's must be nested
i.e. either $R_i \subset R_j$ or, weaker, $R_i (\le a) R_j$ for some bounded a, for large enuf $j > i$. ($i > i_0(l)$).

2) If 1) is true, perhaps t. hy i R_i <u>must</u> be CB's on <u>same</u> umc.

3) If $\lim_{i\to\infty} P_i$ exists then
∃ a umc M_u ∋ either ⓐ R_i are all CBs on M_u or
ⓑ $\lim_{i\to\infty} P_i' = \lim P_i$ for $[P_i']$ some set of CB's on M_u.

453.01

11: 452.40 : Given P ($P \equiv \lim_{i\to\infty} P_i$), perhaps one could construct A corresp'g. ume in t. manner of Willis' Thrm 12. — D2773 350p

Anyway, as i ↑, t. P_i's get arbly close together — so t. R_i's must get "arbly close together" in some sense. Consider fixd ℓ, & ↑ i.

Conjecture: If $\lim_{i\to\infty} P_i$ exists for all values of t. argt of t. $\{P_i\}$, then this limit corresps to some kind of pgmable machine.
A simple case would be that in which P_i was ≈ t. distribn for a simple binary Bern. seq. w. p = irrational. P_i would have t. p value accurate to i binary places.
[Hvr, this may not be t. kind of example I want, since t. thrm(s) of interest involve finitely dss i.e. Pems w. finite 'a'.].

Anyway: T. problem is 452.10 – .16 — i.e. $\underbrace{R_i \to \text{ume}}$ doesn't imply $\lim_{i\to\infty} P_i(A^{(\ell)})$ exists for all $A^{(\ell)}$. | I think ~~implies~~ implies no matter how large i is (if $i > i_0(\ell)$).

$\frac{P_0}{P_i}$ is bounded [This latter may imply expected mean b cost error → 0, also that expected value of Σ sq. err < α kℓ ln2.

That $\frac{P_0}{P_i}$ is budd implies, i think that $R_i \to$ ume.

Hvr, if $\frac{P_0}{P_i}$ is budd, this does not imply that $P_i \to$ a limit, so one can't make statements about "that limit". — Hvr., one can say, that for large enuf i, (i.e. $i > i_0(\ell)$) the mean b cost error in P_i will be $< \frac{a}{\ell}$ & Σ sq. err will be $< \alpha \ell \ln 2$.

30 So this may be O.K.: we have iff betw. $R_i \to$ ume & $\frac{P_0}{P_i}$ bud → 456.01

We would like iff betw R_i are CB's on a ume. & $\lim_{i\to\infty} P_i \to P$ w. $\frac{P_0}{P}$ budd.

$\frac{P_0}{\lim_{i\to\infty} P_i}$ budd.

— A while it may be true, I've not been able to prove it.

Hvr., first check this.

A possl. Thrm: If $\lim_{i\to\infty} P_i$ exists & $\frac{P_0}{\lim_{i\to\infty} P_i}$ or $\frac{P_0}{P_i}$ is budd.

Then ∃ a ume M_0 ∋ t. Pem defind by M_0 via a FLU seq. is identical to $\lim_{i\to\infty} P_i$.

454.01

454.01

D2873 R

TM454

01:453.40 : Assoc. w every Pgmable machn. There is a unique PEM defind by t. limit as $T \to \infty$ of t. T limited CB's on that machine. T. limit always exists & is t. same for ~~[illegible]~~ various other seqs of CB's.

Note that if ~~[illegible]~~ $P_i \to P$, then we ~~[illegible]~~ probly cannot say very much about any ~~[illegible]~~ set of R_i assoc. w. t. P_i. This is because 2 ~~R~~ FOR's that → t. same ~~C~~ CPM need not have much in common. They can certainly differ in various permutations of t. input alphabet — but t. 2 FORs can differ much more seriously. E.g. take an arby FOR, R_1, find its CPM, then construct R_2 from this CPM via W. Thrm 12. R_2 will (if t. input is binary) have only ≤1 code of length ~~[illegible]~~ n for any output seq. In genl, this will not be true of R_1.

So: ~~It may be po~~ [It is possl. to map from t. set $\{P_i\}$ into a set $\{R_i\}$ via W. Thrm 12, in a unique way. Can we use this $[R_i]$ to define a programable machine? In genl, for an arby infinite $\{R_i\}$ seq. This is not possl. — but if t. R_i are derivd from a $\{P_i\}$ that → a limit, in a unique way via Thrm 12, we mite use $\{R_i\}$ to define a limit ~~R~~ Pgmable machine.

.28 One mite define this "limit- Machine" using t. construction methods of Thrm 12. ~~Dann~~ One constructs a sequence of ~~R~~ FOR's, R_i. If its done properly, it may be possl. to show that ~~t. cons~~ if P_i → a limit, then t. construction features of R_i must also → a limit. ~~a~~ A possy mite be that since W. uses tables to define t. R_i's derivd from P_i's — one mite be able to show that for large enuf. CB's, ① every table entry must eventually become constant ② there are only a finite no. of entrys in t. tables.

.39 AH! If lim P_i exists, then there must exist a $i = i_0$ such that for ~~[illegible]~~ $i > i_0$, P_i changes by $< \epsilon$. This means that 455.01

.01: 454.40 : after $i > i_0$, we know how to code t. first $[-\log_2 \epsilon]$ bits of t. input!

So, <u>I think</u> using 454.28 ff (& 454.39 ff in particular), one should be able to construct a "<u>limit</u>" <u>machine</u> if $P_i \to$ limit.

T. Q is: is this limit machine umc? I think it <u>is</u> —>

output of limit machine is not unidirectional, 2 ope.

Note: This "limit machine" <u>may</u> be a Tmc., but it is <u>not</u> a "<u>programable</u>" machine

.05 T. limit machine "exists" in t. sense that given any input sequence, I can eventually compute t. r^{th} symbol of t. output — but in general, I will not know when I've done so — eventualy, I will give a first approxn. to that r^{th} symbol, but I may "change my mind" about what it is after that. i.e. it <u>hasn't</u> <u>unidirectional</u> output

That "t. limit exists" I interpret to mean "eventually, for $i >$ some i_0, ($i_0 = i_0(\ell)$, I think), I will <u>no longer</u> <u>change my mind</u>" (Also, the "settling down" is a sequencial occurrence) in fact, i_0 is <u>indip</u> of ℓ, if $P_i \to$ a limit, <u>indip</u> of ℓ. perhaps <u>not</u> & not so

.17 Anyway — to show that t. limit machine is umc, just prove that for <u>any</u> FOR, R, t. limit machine can simulate R. — I <u>think</u> this is equivt. to proving its umc. I don't know if there's a thm to that effect, but assume so for t. present.

I may have been forgetting that there are <u>2</u> limits involved: One is the $i \to \infty$ limit, & t. other is t. $T \to \infty$ limit. They <u>need</u> not be ~~are necessarily~~ <u>identical</u> — but they <u>can</u> be.

On second thot, t. rate at which $P_i \to$ its limit <u>very</u> probly depends on ℓ. For large ℓ, i.e. small P_i to get an error $< \epsilon$ requires much larger i than if ℓ is small (i.e. large P_i). Perhaps <u>Not</u>. To get a ~~[illegible]~~ break error $< \epsilon$ may <u>require</u> larger i if ℓ is large.

Well consider $i \equiv T$. Suppose i_0 (for agn. ϵ error in P_i) is a funct. of ℓ. Or, say T is a known funct. of i & ℓ.

Given any <u>input</u> sequence, there will be, assoc w. each n value (length of input) an associated i_0 value, ∋ for $i > i_0$, t. output symbols assoc w. t. first n inputs, do not change. $i_0(n)$ will often be <u>not</u> a recursive function — it can ↑ w. n, faster than any recursive funct.

.35 I may need to prove lemma: <u>If</u>: $p(R_1) < 2^a p^{(R_2)}$ for all $A^{(\ell)}$ of $\ell \leq \ell_0$. <u>then</u> $R_1 (\leq a) R_2$ for output seqs of $\ell \leq \ell_0$ i.e. R_2 can simulate R_1 w. a xltn. code of length a, for outputs of length $\leq \ell_0$.

Actually, this lemma is <u>false</u>: T. fact is that for t. hypoth, there <u>exist</u> R_1 & R_2 w. t. desired propertys. ⟶ ∴ t. "proof" of 155.37 is <u>wrong</u>!!

456.01

01: 455.40 : Re: 453.30   I think t. backwards thrm is : if [redacted] $\frac{P_o}{P_i}$ is budd, (t. budd is indep of ℓ) for large enuf i',

then $\exists$ a seq. of R_i $\ni$ ⓐ $P_i = P^{(R_i)}$

ⓑ $R_i \to$ umc.

Going ~~backward~~ back even further! I had this lemma: If R_2 cannot simulate R_1, then for any k, $\exists$ a seq. $A^{(\ell)} \ni$ $P^{(R_2)}_{(A^{(\ell)})} > k\, P^{(R_1)}(A^{(\ell)}_{2})$. This lemma is false. R_2 may be unable to simulate R_1, yet $P^{(R_1)} = P^{(R_2)}$ for all $A^{(\ell)}$.

e.g. $\begin{cases} R_1(A) = A \\ R_2(A) = \overline{A} \end{cases}$ (0→1, 1→0 for every symbol in A).

Both machines assign t. same probty to all seqs., yet neither can simulate t. other.

Or, more generally, consider any FOR. = R_1 ; let R_2 be t. same FOR except $R_1(A) = R_2(\overline{A})$. So Both have t. same $P^{(R_{1,2})}$, but neither can simulate t. other (usually) w. xltn. length zero!

.20 say $P_1 \leq 2^{a_1} P_0$ can I construct $R_{0,1,1}^{1,2,3}$ $\ni$ $R_i (\leq a_i) R_0$. ?

$P_2 \leq 2^{a_2} P_0$ etc.   ⟹ 458.22

$P_3 \leq 2^{a_3} P_0$

.22

→ I had been assuming that W. thrm is said that if $P^{(R_1)} \leq 2^{a} P^{(R_2)}$ then $R_1(\leq a) R_2$: which is false.

So, it really looks like I'm nowhere near proving t. backwards thrm of .01 or any other backwards thrm!

If $\frac{P_o}{P_i}$ is budd indep of ℓ, (for large enuf i's). Then P_i need not have been derived from a set of ~~R~~ For's that → umc.

It may well be that .01 is true, but still, that doesn't make CMI seem very unique! It would, hvr, mean that if there is any P_i seq. that can do — , then CMI can do at least as well.

Well: Maybe things aren't so bad! — Perhaps I can show .20–.22 for an arbly large no. of P_i's.

457.01

01:456.40 : T. 3 thrms I mite prove:

1) 456.01 ⇒ If $\frac{P_0}{P_i}$ is bndd, than $\exists \{R_i\} \ni P_i = P^{(R_i)}$; & $R_i \to$ umc.

2) ⓐ If $\frac{P_0}{\lim_{i\to\infty} P_i}$ is bndd than $\exists \{R_i\} \ni$ ⓐ $P_i = P^{(R_i)}$

ⓑ $R_i \to M_U$ ← this may be ~~i~~ implied by (i.e. that M_U is umc.)

& $$P^{(M_U)} = \lim_{i\to\infty} P_i$$

$R_i \to M_U$ has t. meaning of 454.28 — 455.01

P^{M_U} is t. limit of a ↑LU on M_U.

ⓑ if $\frac{P_0}{\lim_{i\to\infty} P_i}$ is budd, Then $\exists$ a umc, $M_U \ni P^{M_U} = \lim_{i\to\infty} P_i$.

3) If $\frac{P_\alpha}{\lim_{i\to\infty} P_i} \leq k_\alpha$ (for various α's) — than $\exists$ a umc, M_U

$\ni$ ~~[illegible]~~ $\frac{P^{M_U}}{\lim_{i\to\infty} P_i}$ ~~[illegible]~~ $< k_\alpha$. I.e. $\exists$ a CMF at least as good as ~~[illegible]~~ $\lim_{i\to\infty} P_i$.

P_i are r. CPM's ; P_α ar all FCPMS
Actually, P_α can ~~[illegible]~~ probly be CPM's

Possl. thrm: If $\frac{P_\alpha}{P^{(M)}}$ is bndd ~~[illegible]~~ (same bnd for all α) for all ~~[illegible]~~ α,

Than M must be umc.
Say M is ~~a~~ sequencial Machn.

perhaps counter ~~[illegible]~~ example

If $\forall P_\alpha$, ~~[illegible]~~ $\exists K \ni \frac{P_\alpha}{P^{(M)}} < k$ than M is umc.

$\frac{P_\alpha}{P^{(M)}}$

say M_U is a umc.

construct M_U' in t. folg way: It has t. same I, O, except that inputs of t. same length are rearranged — keeping, hwr., with~~out~~ t. sequencial property

Possl. thrm: If $\lim_{i\to\infty} R_i = M$ ← defn. in 454.28-455.01 Than $P^{(M)} = \lim_{i\to\infty} P^{(R_i)}$

Here, M is a particular Machine; $\left[P^{(M)} \equiv \lim_{i\to\infty} P^{(M^T)} \right]$ where T is a ~~[illegible]~~ T limit for a CB. i.e. M^T is t. T limited machine, M.

The idea here, is that T. limiting probly distribn. is somewhat indip

[redacted]

D2873 R

TM458

01:[redacted]457.40 : A basic Q. is: what am I trying to show? Basically, t. uniqueness of t. CMI soln. [redacted] Some thrms. that would show this:

1) If $\frac{P_\alpha}{\lim_{i\to\infty} P_i}$ is bndd for all / P_α = rCPM's. Then ∃ M_v → $\frac{P_\alpha}{P^{M_v}}$ is ≤ r.

(convergs pf. bnd

i.e. CMI is at least as good as t. limit of a seq of prams. This would also show that it is impossl (?) to get a convergence rate for [redacted] F d SS's that is better than CMI gets.

[redacted]

Note that 1) assumes that $\frac{P_\alpha}{P_i}$ is t. Score of interest. In general, it need not be so.

Define t. problem: list possl. solns. & partial solns.

A soln. of 456.20) for an arby but finite no. of a_n's — would perhaps solve t. genl. problem.

.18 Note that t. "limit Machine" of 454.39 – 455.01 is not a Pgmable machine — since its output tape is not unidirectional.

3:53P

Another Q. is whether t. idea is O.K. — in view of my previous misinterpretation of W thrm 15. — One second that — it doesn't use that Thrm — So it's O.K.

.22 At first Glance, it mite seem easy to construct these R_i's if $i = 1, 2$ only. It is easy if $a_1 = a_2 = 10$. Hvr., if $i = 1, 2 \ldots 2^{10}$ & a_i's are all < 10, or < 12 even, one can get interference betw t. R_i's. $(i > 0)$. If there is no interference, one simply constructs $R_{1,2}$ in accord w. Thrm 12's proof, then constructs R_0 in accord w thrm. 15's proof

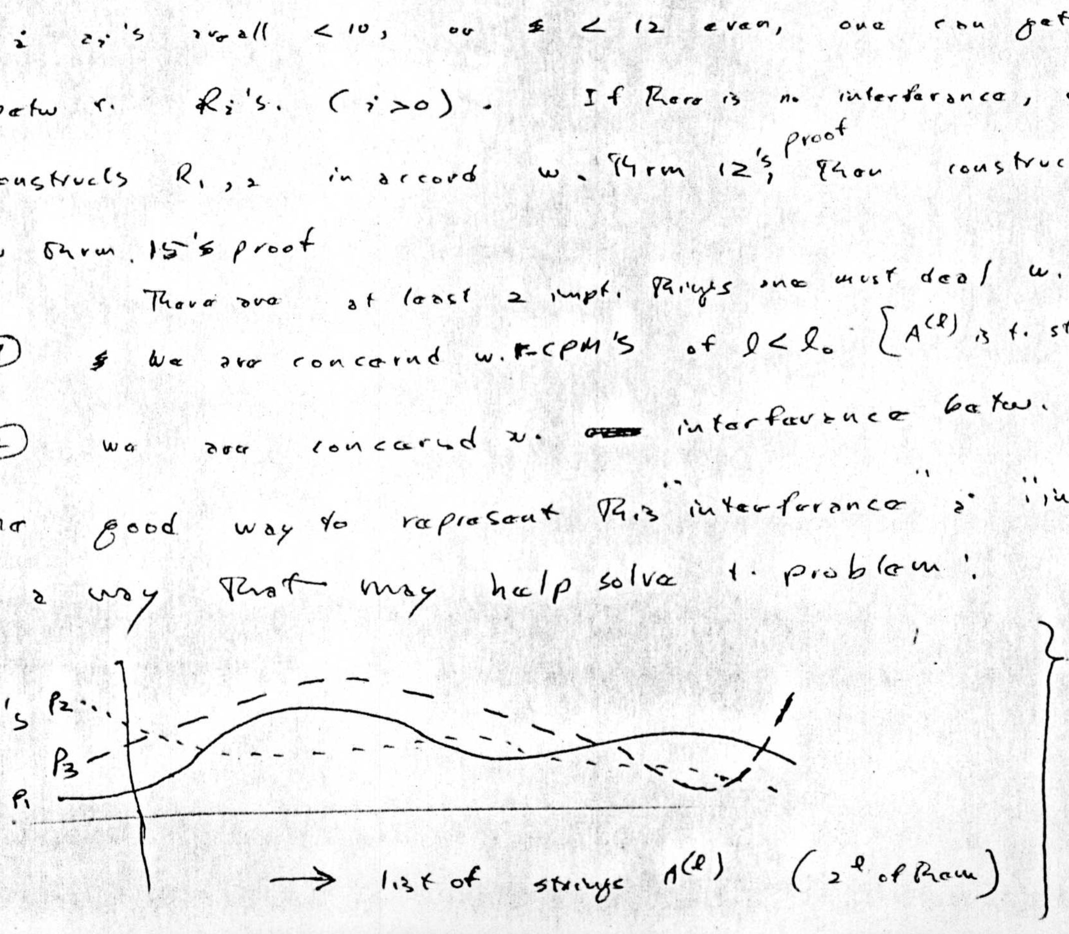
There are at least 2 impt. things one must deal w.

① We are concernd w. CPM's of $l < l_0$ [$A^{(l)}$ is t. string set].

② we are concernd w. interference betw. t. P_i's:

One good way to represent this "interference" & "interaction" in a way that may help solve t. problem:

36



459.01

01: 458.40: for each $A^{(l)}$, there will be a set of probty values that are possl. Using t. picture of 458.36 - .40 — it may be possl. to devise codes for various $A^{(l)}$ & assign them to various R_i in an optimum manner. The a_i's must somehow also be inserted into t. picture.

We want t. codes for t. [redacted] $A^{(l)}$ ~~=~~ assigned to difrnt R_i's to be as similar to each other as possl. — so it is easiest to

.12 get R_0 to satisfy t. ~~R~~ $R_i (\leq a_i) R_0$ conds. → 460.14

.13 ③ Note 458.18 — that t. "resultant limit machine" is <u>not</u> a "Pgmable machine" — its output tape is <u>not</u> unidirectional!

Possl. "<u>outs</u>" ⓐ Xfm it into a ~~ta~~ machine w. unidirectional output.

.15 ⓑ Befor ⓐ can be done, it may be necy to make t. probty values that t. "limit up to this pt." machine realizes, be monotone ↑.

This problem of ~~a~~ non-monotonicity of t. P_i's as they → a limit would seem to be a <u>very basic</u> difty in t. desird thrm — & perhaps <u>disprove</u> it!

One could take a seq. of P_i's (w. t. constraint that $P_i > 0$ for every $A^{(l)}$ for all i's), & multiply each one by a suitable constant so that they were monotonic. T. resultant Probys would <u>not</u>, then, sum to 1 in t. proper way — but perhaps this is tolerable.

Another difty — this multiplication by a constant <u>could</u> result in several P_i values [redacted] → ∞ as $i \to \infty$. E.g. say we have 2 $A^{(l)}$'s & for one i value, ~~on [illegible]~~ i's one's P_i ↑ by x2 & t. other's P_i ↓ by x2, on t. $i+1$ value,

this one's ~~R~~ P_i ↓ by x2, " " " " ↑ " ".

This process would (if all ~~other~~ P_i's ~~were constant~~ for other $A^{(l)}$'s were indep of i) give rise to an exponential ↑ in t. normzn constant. The unnormzd P_i's would not → a limit, but t. argt. can be modified so they <u>do</u>, & still have normzn const. → ∞.

.36

.37 Is t. folg true? ← Probly <u>No</u>! — see 460.05 P_1 [redacted] $\leq 2^{a} P_0$. R_0 is some fOR assoc. w. P_0

[redacted] Is it always true that $\exists R_1 \ni P_1 = P^{(R_1)}$ & $R_1 (\leq a) R_0$?

~~A possl. way to do t. construction for R~~ : I suspect that its <u>false</u>: R_0 can be a <u>very</u> simple machine — say t. 460.01

D2973 R

TM460

1: 459.40 identity machine ($M(\alpha)=\alpha$) — yet P_1 could be fairly complex.
T. identity machine assigns = a priori to all strings of = length — but its simulation capabilitys are minimal.
P_1 could be P_0 multiplied by some bounded function of $A^{(k)}$, & then renormed.
05 so [redacted] 459.37 is probly false. This seems like an impt. idea.

So, it looks like t. non-monotonicity of ~~[illegible]~~ ~~[illegible]~~ P_i's ~~[illegible]~~ approach to P_∞ is t. Big problem

I. kind of non-monotonicity assoc. w. t. machine of 454.39 - 455.01 may not be so bad. Perhaps I can rewrite W's formulation in terms
13 of that kind of Machine — rather than "programmable Machine" → 461.13

14: 459.12: In t. picture of 458.36, one can also draw curves $2^{a_i}P_0$, for various a_i values.

Now — T. idea is to conserve on the a_i length simulation strings.

— If there were only a few i values & a_i's were large, there would be no "interference" (458.22). How t. conservation operates:

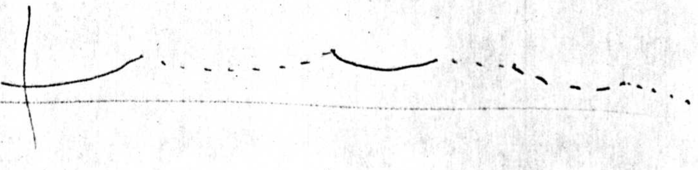
Say $n = 4$: a_{1234} ~~are~~ 2, 2, 3, 4.

first we construct t. codes for t. $a_{i,2}$ machines: they are identical except for t. 2nd (= last) bit. If 2 P_{a_i}s have many values in common, their codes are t. same for all but a few final bits (i.e. t. codes of t. P_{a_i}s).

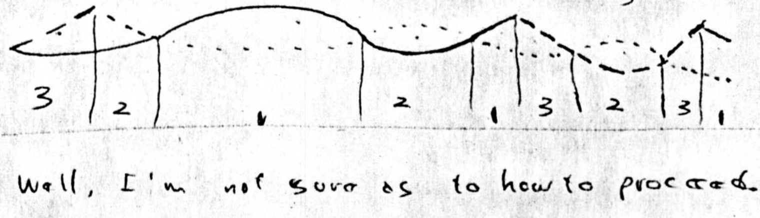
If P_3 is always $> P_4$, then P_4's code can consist of P_3's code concat w. ~~[illegible]~~ ~~[illegible]~~ 1 bit at end.

I think its a kind of Boolian net problem!

From 458.36 : Consider t. lowest level : Its "code" is t. null code.



On t. next level, there are : 6 segments, but only, at most, 3 "basis" regions.



Well, I'm not sure as to how to proceed w. this code assignment, but t. Boolian idea (not nearly exactly as descbd. above) seems reasonable. 462.01

D3073 1C

01: 460.40 Hvr., don't spend more time on it now! T. main problem is — is there any pt. to it in view of t. *non-monotonicity* of 459.13?

.03 Note that with a gn. umc, M_u, I *can* have a seq. of CB's on M_u that → umc, yet t. seq. of P_i's assoc. w. this seq. of CB's need not be monotonic. T. monotone condition is *necy* if $R_i \subset R_j$ for $i < j$ — but not if t.

08 CB's don't have this property. Hvr, for each of t. CB's R_i, $R_i \subset M_u$,

10 also $P^{(R_i)} \leq P^{M_u}$ — t. limit.

.11 If we have an arby $[P_i]$ that → a limit P, it is certainly *not*

.12 *necy* true that $P_i \leq P$.

113: 460.13: One CMI formulation that may work: My old "fixd length coding method, using *bidirectional* I, o. tapes. Random (≡ All possl) input strings we stop t. Machine at Time T, to compute probys. We let T → ∞. Hvr., even for fixd string length, is it necy true that a hy fraction (or "almost all") of t. output symbols would eventually stablize? Perhaps not — but in the ~~[illegible]~~ umc Machine of 454.39 — 455.17 we *would* have convergence of this sort.

So, it seems quite possl. that CMI *could* be formulated in terms of machines having this "mildly seql. property" — rather than t. strict sequentiality of W's ~~are~~ "programmable Machines".

Def. .25 Note that a **MSM** (≡ Mildly sequential Machine) is *not* t. same as an unrestricted Trmac. (Trmc.). An unrestricted Trmc. could write output 1111 then erase it, then write 1111 again, indefinitely. A **MSM** could *not* do this *indefinitely* — tho it could do it for any finite time.

Since Pomable Machines are a (proper) subset of MSM's, MSM's *can* be *umc*.

T. prev. discn. makes a plea for t. introdn. of MSM's in t. formulation of CMI as a more general form than W's SQPM's. T. Q. is — do I really *need* this?

Say P_i → some limit P. We *can* have $P_i = P^{(R_i)}$; have R_i be CB's on an umc, M_u, yet P_i *need* not → P *monotonically*! → i.e. .03 — .12

~~[illegible] Monotonicity would occur if~~

.08 — .12 suggests perhaps we *do* need it! i.e. for *any* SQPM, Condition .10 holds, yet $P_i \leq P$ *need* not hold.

So — I *think* its *true* that *if* $\lim_{i\to\infty} P_i = P$ · $P_i \equiv P^{(R_i)}$; $P^{\alpha} \leq 2^{a\alpha} P$ · R_i are FOR's, *Then* ∃ M_u (a MSM) ∋ R_i are all C.B.'s on M_u · $R^{\alpha} (\leq \frac{2}{\alpha}) M_u$ ($P^{\alpha} \equiv P^{R_\alpha}$; ~~[illegible]~~ P_α is an arby Pcm)

462.01

D3073 *R* MSM TM462

.01 461.40 : O.k., then: Say one is Gvn a unc M_u^1 that's MSM. [redacted]

.02 so $P_1 \equiv P^{M_u^1}$; does there exist a SPQM, $M_u^2 \ni P_1 = P^{M_u^2}$?

.03 It may indeed "Exist" in some sense — but I wouldn't be surprized if

.04 it *need* *not* be *computable* from M_u^1's defn.

It *is* possl. to know that [redacted] $P_i \to$ a limit, yet have t. limit be *not* computable. This is true as in compute $P(M_u)$ for M_u = SQPM — ~~Thruthe~~ In this case P_i is monotonic ↑. We can add an oscillitory component that → 0 if we like, to give t. whole thing non-monotonicity. If we took $\sin(1000 P_i)$, this *would* → a limit, since P_i does, but it may or may not be monotonic. Presumably, for small enuf & large enuf value of "1000", one can always get *some* non-monotonicity as i ↑.

From .01 ff I *suspect* that .03—.04 is true.

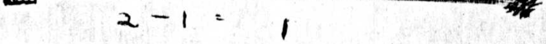
Hvr., another approach to CMI could use a sort of difnt. number system than normal positive radixes. E.g. radix 3, but t. coifs can be 0 or ±1, rather than 0, 1, 2. Or ~~base~~ radix 2, but coifs can be ±1 only (not zero) (This [redacted] latter or ≈ it is usd in binary divisn algorithms).

In t. binary case $+-_b = 1_{10} = +_b$, so there is ambigy in t. representation. $1_{10} = +$; 2_{10} = woops; Even nos. aren't representable!

1 = 1 ; 2 = 3 − 1 ; 3 = 3 + 0

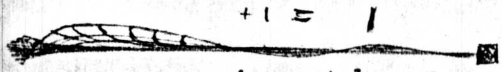
4 = 3 + 1 ; 5 = 9 − 3 − 1 ; 6 = 9 − 3 ; 7 = 9 − 3 + 1

8 = 9 − 1 ; 9 = 9 ; 10 = 10 + 1



2 − 1 = 1

+1 = 1



Tom Mahone — "The Vomiting Cowboy"

Pickup an earful of Tom Mahone!

If M_u^2 of .02 always existed, then, given an arby $P_i \to P$, one could construct a new monotonic ↑, $P_i' \ni P_i' \to P$ also. If P_i is allowd to be an *arby* seq. that → a limit (& I *think* it is), then it would *seem* that constructing P_i' would be impossl. in general. It would mean that eventually for $[P_i]$, one could say that $P >$ some ~~s~~ k. If $[P_i']$ were constructable, then P_i'', a monotonic ↓ seq. ~~[illegible]~~ that → P should be constructable. This would mean that one could construct arby close upper & lower bnds for [redacted] P. — Certainly unlikely for most general $[P_i]$ seq's!

463.01

D3073 B

TM463

01: 462.40 : T. strange thing is that a SPQM can be an Umc — so how come it can't simulate a MSM? In Thm 3, W. defines a Univl. Pgmable machine as one that can simulate any other Pgmable machine.

A Pgmable Machine is a Tmc w. unidirectional I.O. tapes. Usually it has at least 1 wrkg tape. If it has none [illegible] its a fsM; if it has > 1 wrkg tape, its of no extra capability than a 1 tape machine — other than possibly speed.

Brousing ~~[illegible]~~ thru to old version of W. (pp 20-21) it looks pretty much like his proof for "Universality" would work only for simulation of Tmcs w unidirectional I,O tapes. Actually, unidirectional input is not a constraint ~~[illegible]~~ ~~[illegible]~~ T. unid. ~~[illegible]~~ input machine can simply copy its input onto t. work tape, so its effectively 2 directional / i.e. it can refer to earlier values on t. work tape. ~~However, there may be certain arbtly long reading swings that cannot be simulated~~.

T. unid. output does, hvr, appear to be an actual ~~[illegible]~~ restriction on what can be computed.

T. MSM I'm thinking about, (when its doing what a SQPM can't do) never stops.

~~while~~ while a univl. SQPM can't exactly simulate an arbly limit seq. [Pi], perhaps it can get with say, a factor of ½ of t. limit — this would be ~~[illegible]~~ about good enuf — but I don't even see how it could do that good. I think that as Pi approaches its limit, one can never at any pt., be sure about any ~~[illegible]~~ range of where t. limit can be!

Hvr, while there may not exist a computable method to go from t. seq. [Pi] to a SPQM umc, this does not mean that such a SPQM umc does not Exist ! !

Discussd SPQPM's & MSM's w. Marv.: He ~~[illegible]~~ did see just how MSM's could do things that one couldn't easily ~~construct~~ devise ~~construction~~ simulation instructions for a SQPM to do. But he said he didn't know anything more about it. Said that there was some mention of this or a closely related phenomenon in Blum's (induction) report — suggested that I write them about this.

.35 D3173: O.K. I think I see how I could ~~make~~ desb (effectively) a MSM $\ni$ t. convergence time for t. n^{th} ~~[illegible]~~ sq. of output was > any e.c. funt (= recursive funct.) of n. ∴ I think I can construct

.37

.38 a MSM that would "compute" this non-recursive f(n). The way it does it! We have this MSM that has a ~~[illegible]~~ time counter, which 464.01

~~see~~ 74TM 11.02 — ≈.20 for this construction

D3173 R

TM464
last Page
in 1973

1:463.40 counts t. no. of machine steps done since the start. (This couter can count by 10's or 100's (etc. — it need not be accurate to better than ± k.).
Every time the n^{th} square of the "output" tape$_1$" is changed, it writes down t. no. in t. "counter" onto "output tape$_2$" — So we sort of have 2 machines, and one of them does "compute" this uncomputabe f(n) — so if we relax t. rules on t. defn. of computability, this MSM is "more powerful"

Def .10 than a UIO machine (≡ UIOM) unidirectional IO = Willis' Pgmable Machine.

It seems likely that t. MSM's are not r.e. — i.e. that one can't always tell, from a dcrn of a Trmc whether it's a MSM or not.

W. a MSM, one does not have a useful sequencial property — i.e. one cannot ever discard a candidate dcrn. of a corpus, just because it differs in output w. t. corpus. — So, even tho one uses C.B.'s on a MSM to get successive approxns. to probability, one always has to consider all possl. input codes for each successive approxn.

One could take an arby Trmc & define it into a MSM, by saying it must print U at a particular output pt. if it never converges for that particular output symbol. This would, of course be impossl. to determin in genl. One could just as well ask a UIO Trmc to print U's under similar circumstances — thus simplifying most induction problems a great deal!

---

29 But, t. main point is — that since a MSM is more powerful than a UIOM, & one can get a seq. of P_i's that only a MSM can really "cover" — that probably a universal UIOM cannot really completely

.32 simulate all seqs. of P_i's exactly.

— Well, actually, we didn't expect it to — UIOM's can only deal w. r-CPMs & t. like. W. shows that there is a r.CPM arbly close (b cost-wise) to any Pcm (perhaps) — but check this.

.35 — Anyway, I'm not so sure this is t. problem. I'd like to show, that given any seq. of Pcms, $\{P_i\}$, that → some (uncomputable usually) limit, P_0, that ∃ a UIOM M ∋ P(M) is just about as good as P_0.

1.01