Report on Navy's supply of CMI

411 — 104
73 | 74

1174-R

Kacitunet

Theory files

TM [redacted]

1

011464.40 : → The set of provably convergent (PC) seqs, [P_i], may
Def be r.e. (?). If so, then t. set of MSM's
.03 corresponding to these [P_i]'s are r.e.) → (2.01)
.04 SN → Hvr: On t. Gödel results: (1) Gn. a "reasonable large" set of axioms. There is no
"proof procedure" that will prove ~~or disprove~~ all provable thrms. (i.e.
thrms for which finite proofs exist.

To show this: Say one listed all possl. finite proofs (they are r.e.). — Then one checked them in order, & pickd t. first legal proof. T. reason this wouldn't work: Proof checking probably ~~takes~~ can take an arbly long time. If t. set of axioms plus t. proof correspond to a pgm. for a Trmc., t. proof checker corresponds to t. Trmc. & t. checking of t. proof corresponds to t. running of that pgm. into that Trmc. If t. output of t. Trmc. when it stops, ≡ t. desird thrm, then it is proved.

Well, in general there exists no routine for ~~…~~ ordering ~~…~~ trial proofs to be tested & CB's on those proofs, that will eventually "catch" all legal proofs.

On t. other hand, one can devise "proof procedures" that are fairly good. [Perhaps ~~…~~ a proof of t. forgo: Gn. any set of axioms & a proof procedure: I can find a thrm & a proof that would frustrate that procedure, by having the < length of proof, < B> combination outside the "window" of the proof procedure]

I think t. forgo. may be what people ~~are~~ thinking about when they say ~~…~~ Gödel showd ~~that~~ humans can do things that machines cannot. On t. surface of it, this conclusion
.28 seem unwarranted — if that's what Gödel showd — & it seems likely (?) → 2.10

.29 The other Gödel result: If one has any reasonably large set of axioms, then there will be thrms that actually cannot be proved or disproved w.r.t. that set of axioms. I.e. no finite proofs exist — It is not a question of not being
.33 ~~finding~~ able to find them as in .04 – .28. In this sense, no (reasonably large) finite set of axioms is "complete".

It corresponds, in the language of ~~…~~ .04 – .28, to a Tmc, and a ~~…~~ r.e. set of pgms. But there exist finite output sequences for which no corresponding input pgms exist. Clearly t. Tmc we are considering is not UMC.!

Perhaps t. proof is t. exhibiting of a non-provable or disprovable ~~…~~ thrm by a diagonal argt.

TMz

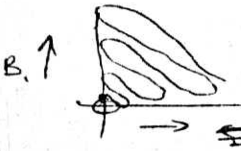
.011 1.03: Hvr, in view of 1.04 – .28 (& perhaps 1.29 – .40), this idea of "provable convergence" will have to be refined a bit. The idea of "provability" is ~~appar~~ apparently about t. same as t. idea of t. halting problem. — so p.c. only has ~~meaning~~ ~~as~~ a.c. meaning if one states the "proof procedure" (≡ search, test, C.B. relation). 1R 2R

Anyway — t. moral seems to be that t. set of "provable"

.09 MSM's is not ~~r.e.~~ r.e.

→ 3.01

.10: 1.28 On second thot 1.04 – .28 may be wrong. Consider t. 2 dim scan in t. ⟨proof-length (→ deriv. length), C.B.⟩ plane: 132/82



If a proof exists for t. thrm. or its negation, it will be found this way.

— If neither exists, one can never be sure of this via this procedure.

.16 **So**: Modified Conclusion for 1.04 – .28!

Given a reasonably large set of axioms & a thrm: If t. thrm has a proof, then, ∃ a proof finding procedure that will find it eventually. (e.g. .10 – .16). Hvr., if there are no proofs for

.20 t. thrm or its negation, there is no way (in general) to tell.

The second ~~par~~ thrm (1..29 – .40) holds as it stands. This first part says that while "indipendant" thrms exist for any reasonably large axiom system, one can in general, never be certain that a set of axioms is, indeed, indipendant, or ~~...~~ consistent. → hvr, see 9.01

There is some business about finding a model or represantation of an axiom system, that is supposd to prove it's consi. I guess t. pt. of t. first thrm here, is that very often, a "model" for a set of axioms can't be found — so one doesn't know if they are consi ~~or indipen~~ — and that in genl. no procedure for model construction exists that will always work.

Apparently, hvr, it is felt that a model proves that ~~...~~ gn. a set of axioms, t. negation of any of t. axioms has no proof within that system.

In ~~...~~ t. case of certain axiom sets: (e.g. truth tables) one can tell if a set of axioms is consi or not. In this case, there are only a finite no. of situations to check — & one can check all of them. Just how this type of analysis fits into t. discn. of .16 – .20, is unclear. See Minsky Computation f. I machines pp225–226 (§ 12.3) for a limited discn. of this.

9.10 ~~3.10~~ → spec 3.01

1274 R

TM3
1R
2R

01: 2.09: About MSM's: They are of 2 types:

02 ① One can prove they are MSM's: i.e. that they always eventually converge. UIOM's are a subset of this class. If one has a gn. Trmc., — if it happends to be in this class, one can eventually prove it by a 2 dim. scan (like 2.10-.16. Hvr., in finite time, if one hasn't found a proof, (or a disproof), it can still be in this class. — It mite also be in this class, or it could be a non-MSM.

09 ② Tho they are MSM's, there exists no proof of this fact. (That such a class exists is made likely by a Thrm. of Göd — i.e. 1.29-.40.)

11

It would seem that MSM's (even (Class(MSM's)) are more powerful than UIOM's. T. problem is still 464.29-.32; 464.35-.40 — i.e. can one do almost as well in induction with a UIOM as a MSM?

Hvr, since MSM's are more powerful than UIOM's (463.35-464.10) — it can express functions that UIOM's can't — I think this is shown in. For that reason, it may be possl. to show that there are "limit probs" that umc, UIOM's can't do well with — or that they will do arbly poorly with.

.25 Hvr. One can't use MSM's & their assoc. set of [P_i]'s to generate a stochastic sequence — as one can w. any

.27 r-CPM or CPM. A universal MSM (if such a thing exists) can simulate any other MSM & i.e. "cover" any PEM that that MSM expresses. But, because of .25-.27, this may be of no great import!

Perhaps just as we associate a r-CPM w. each FOR, we associate a convergent [P_i] set w. each MSM. We can have CB's on MSM's & make these CB's FOR's. Thus a seq. of longer CB's that → "∞", gives us a [P_i]. set that → some kind of "universality" — whether t. CB's are on universal MSM's or universal UIOM's.

Def .30 It is clear that no UOM (≡ UIOM) exists such that all seqs. of FOR's [R_i] are CB's on it — because [P_i] ≡ [$P^{(R_i)}$] can be > their limit, but $P^{(\text{Muom})}$ must be > $P^{(R_i)}$ for every R_i. Note that this is true befor normalization only, but 4.01

1374 R

3.40 it is also true after normzn: i.e. one can have a seq. of normzd probys, $[P_i]$ that $\to P_0$ & are often ~~above~~ $> P_0$ at some points — TM4 1R 2R

? Q. is are they monotonic before normzn? I think ~~they~~ they need not be. In fact, ~~it's possl. to have a seq. of~~ if ~~so~~ $[P_i]$ are normzd, then using W's thrm 12, ~~[illegible]~~ $P^{[R_i]}$ will also be normzd. Hvr. can we devise a set of R_i' such that $P^{[R_i']}$ are not normzd & are monotonic, & $P[R_i']$ become $P[R_i]$ when they are normalized? 459.15-.36 suggests that ~~for~~ for an arby $[P_i] \to P_0$, this is impossl. — i.e. t. normz constant would $\to \infty$ & so for finite l, (which it is), one couldn't do it.

Can one show that one can get a seq. of $[P_i] \to P_0$, $\ni$ P_0 differs from any monotonicly achievable P, by an arbly large factor? Well, select an arby UOM (UMC). Then after it is known that its P value for a certain $A^{(l)}$ will be at least ϵ, make $[P_i] \to P_0$ be arbly $<$ that for that $A^{(l)}$ value. [If $[P_i]$ is normzd, normzn of this will accentuate t. inequality $P > \epsilon > k\,P_0$.]

Hvr., even if one could do this — consider this P's attempts to "cover" a Pcm ~~[illegible]~~ (P^α) that has $P^\alpha(A^{(l)}) = 0$. In this case $P(A^{(l)})$ is arbly $> P^\alpha(A^{(l)})$, but that's O.K. — it gives no problems when Expected values of bscost error of E (Σ sq error) are considered.

What we ~~[illegible]~~ would have to show is that P_0 can be arbly $> P$ for a certain $A^{(l)}$. This is hard to show, because we really can never know ~~t~~ t. max value P can attain. Well — we can to some extent — i.e. by looking at the P [redacted] values assoc. w. ~~others~~ all other $A^{(l)}$ pts., for t. same C.B.

Hvr, this doesn't help either: Consider, say, $l = 100$, or some very large value of l. There is, say an input string (for the UOM UMC) of length 5, that hasn't converged yet — which would give an enormous amount of probty to a gn. $A^{(l)}$ or could cut down its possl. upper bnd [redacted] value a lot. For every long string, $A^{(l)}$, there is ~~[illegible]~~ for any $CB < \infty$, a shortest input string that hasn't converged yet, whose output starts out like $A^{(l)}$ does — that mite $\uparrow$ $P(A^{(l)})$ enormously, & that limits t. amout by which some other $P_0(A^{(l)})$ can be $> P(A^{(l)})$). 5.01

1374 R TMS 192R

01.4.40: So, for large l & large C.B. ($\equiv i$), one finds the string, $A^{(l)}$, for which its unconverged code is of minimal wt. Well, I think this can be pretty short: i.e. consider ~~a~~ some large l value (say $l=100$). Consider t. 8 ~~output~~ output corpi ~~codes~~ of 3 bits each. Say all 8 of them have at least one ~~non~~ never convergent code for t. next output bit. This means that all corpi of length >3 will have uncertainty of at least $2^{-\text{that code length}}$ in their proby values — no matter how long t. corpus. So one couldn't simply devise a P_0 with $P_0(A^{(l)})=1$ ~~unconverged proby~~ ~~[illegible]~~ for t. ~~[illegible]~~ $A^{(l)}$ for which $P(A^{(l)})$ is min [~~[illegible]~~ $P \equiv p(M_{uom})$ $P_0 = p^{M_{MSM}}$] & be sure ~~[illegible]~~ it $(\equiv P')$ would eventually ~~[illegible]~~ become ~~[illegible]~~ $>p$ by an arbty large factor, ~~[illegible]~~.

Defs

17 AH! O.K.! : We want to devise an Untractable seq. for the $p(M_{uom})$ limit. Say ~~[illegible]~~ $P_{i,l}$ ~~[illegible]~~ ($i \equiv CB$) is the i^{th} approxn. ~~[illegible]~~ (This is for corpi of length l).

No! But. Find $A_i^{(l)} \ni P_{i,l}(A^{(l)})$ is min. Let $P'_{i,l}(A_i^{(l)})=1$, all other $P'_{i,l}(A_j^{(l)})=0$ ($P'_{i,l}$) is t. proby limit seq. we are trying to make.

→ Woops! $P'_{i,l}$ must be determind for all l — so, for a given i value, we sequencially generate a determnistic seq, for which $P'_i(A_i^{(l)})=1$ & all other $P'_i(A_j^{(l)})=0$ — We do this by computing $A_i^{*}(1)$ then $A_i^{*}(2)$, etc. in such a way so that ~~[illegible]~~ $\dfrac{P_i(A_i^{l+1})}{P_i(A_i^{l})}$ is min.

30

~~[illegible]~~ There are some ambigys w.r.t. $p=\frac{1}{2}$ situations — ~~[illegible]~~ what "min" means ~~[illegible]~~ when several probys are $=$, etc. It may be that as i ~~[illegible]~~ ↑ one may have to base one's choice partly on ~~[illegible]~~ previous choices (previous i, that is) so that t. limit of $p(M_{MSM_i})'$ exists — this sounds a bit like Blum's algs. in their recent "Inductive Inference" paper.

Hvr., I think this is unnecy, because of t. way $p(M_{uom})$ (6.01)

1374 R

No! → 30.01 !!

TM6
1,2R

01.5.40 approaches its limit — it achieves its limit suddenly, and then stays there.

So it does seem like routine 5.17 ff would generate an untractable seq. of P_i' values that does → a limit P_0^* ∋ the given umc, UOM does catastrophicly badly by it.

W.r.t. t. "limits" involved, note that for every ℓ value $\exists\, i_0(\ell)$ so large, that no more UOM cB's converge for $i > i_0(\ell)$.
T. existence of this $i_0(\ell)$ assures that t. limit of t.

14 "untractable" seqs. exists & that it's "untractable" → 31.01 for comments on t. forge "proof".

15 What is means is that if we have an arby seq. of Pems $[P_i] \to P_0$, then it is not true that $\exists$ a umc(UOM) $\equiv M$ such that $P^{(M)}$ gives a best for all seqs. That is within a/ multiplicative constant/ >0 of that

18 gn. by P_0.

See 37.01 —.18 for a clearing up of this point

If .15 —.18 were false & a umc(UOM) could "cover" all $[P_i] \to P_0$'s, then these P_0's couldn't be better than $P^{(M_{UOM})}$'s by more than a constant factor. (Hvr., .15 —.18 is true!).

I think that what I want to show is that no $[P_i] \to P_0$ can do/ signiflly better than $\frac{P^\alpha}{P_0} < k_\alpha$ for all P^α.

This "signiflly better" idea is rather vague!

More specifically, — say $\frac{P^\alpha}{P_0} = k_{\alpha,\ell}$ then we would like $\exists\; P^{(M_{UOM})} \ni \frac{P^\alpha}{P^{(M_{UOM})}} \lessapprox k_{\alpha,\ell}$,

Note, however, that $P(MSM)$'s are always at least as good as $P(UOM)$'s, since UOM's are a proper subset of MSM's.

○→ Given an untractable seq. for a gn. / UOM umc., it is clear that it must be untractable for all UOM umc's! — Say its untract. for M_U^1. Then if its tractable for $\exists\, M_U^2$, M_U^1 could simulate M_U^2 & t. seq. would be tractable for M_U^1 also.

Well give this untract seq., there is at least one MSM that can cover it. — ∴ if there exist univl. MSM's, & all univ MSM's can cover it. So for at least this one untractable sequ., (and any / invertable (suitably) computable functions of it, perhaps) univl. MSM's are much better than umc UOM's.

7.01

01:6.40: T. Q. of whether & in what sense a Universal MSM exists!

.02 N.B. In t. defn. of a MSM: There may be still some [illegible] ambiguous that I can still use. My conception of t. defn.', is that ~~when it converges for the~~ if it converges for t. m^{th} output/s of a gn. string it converges for all shorter outputs. If it diverges for a gn. output (≡ ~~[illegible]~~ halts or keeps changing its mind), then it (by definition) "diverges" for all ~~output~~ longer outputs.

Say we have an infinitely long input string (which is t. standard situation) — then t. machine should never stop (perhaps) No! If it did stop, it could only stop at a pt. such that it hadn't written anything on further portions of t. ~~[illegible]~~ output tape — otherwise they would converge. Well — that's o.k. — when it stops, everything it has written is then "converged" — so it can stop, but it must only stop at pts where it has ~~written [illegible] no~~ allowed no blanks on t. subsequent

.19 "written" portion of t. output tape.

20 O→ One possl. way No to change an ordinary ~~[illegible]~~ umc ~~[illegible]~~ into a umc MSM: Every time it changes t. value of one of its output symbols, S_{l_0} ← or t. l^{th} output square

.22 it changes t. value of all ~~[illegible]~~ output symbols for $l > l_0$.
~~[illegible]~~ That this is possl. can be shown on a 4 tape Time machine w. ~~[illegible]~~ ~~[illegible]~~ a unidirectional input tape, T_1 ~~[illegible]~~ 3 bidirectional tapes; ~~[illegible]~~ T_2, T_3 2 work, T_4 1 output. T_3 & T_4 always move together.
T_3 contains ~~which~~ t. output ~~[illegible]~~ of an ordinary umc.
T_4 contains t. ~~[illegible]~~ ~~[illegible]~~ modified form of T_3, in accord with .20 - .21. If this machine div diverges for a particular output symbol, it will diverge for all following output symbols.
" " " converges " " " " " " " converge " " previous output symbols.

T. forgg. machine is "Univl." in t. following sense:
If t. machine it is simulating (diverges / converges) for a particular symbol, this umc will also (diverge / converge) ~~[illegible]~~ — if it converges it will converge to t. same symbol. If it diverges, t. "umc" need not follow t. machine it

35 is simulating, but it will certainly diverge for that symbol, also.

impt!! → SN7 → 8.01

(SN) To prove that ∃ a umc UOM: Just use a regular umc for simulation, but whenever t. regular umc got an instruction to move t. output ~~tape~~ backwards or erase ~~a~~ an output tape square, it stops. (or it can go into an infinite non-stop compute loop).

8.01

1374 R

TM8 1.2R

01:7.40
7.35 space → That the MSM is related to t. general umc. in so simple a way, suggests that they may be of about = power.

.03 Hvr, looking at t. MSM more closely — its "defn." (7.02 – .19) à t. Pams assoc. w. it: say one is trying to get t. prob of a particular corpus sequence ($l < \infty$). One of t. c.B. ?, i, à uses t. successive approxns, defined by successive FOR's, R_i. For $i > i_0(l)$ all of t. codes that will ever converge, will converge. T. rest will oscillate indefinitely. ← No! They will be in S or U loops.

I am uncertain about .03 ff. Go over it carefully!!

17:.235 → I think that an additional characteristic of t. MSM is that it converges, eventually, for every output symbol. It is this characteristic, not the Mild sequenciality simulated in 7.20 – .35, that is of importance! ("convergence" could mean that it is in an inf. computation loop in which case all written symbols thus far ≡ their "converged" values — blank spaces are "U" values — After stopping, all blank symbols become S symbols.)

.20 So t. Q. is: Can I devise a MSM umc? The discn. of 7.20 – .35 is irrelevant. Consider t. provability classes of MSMs of 3.02 – .11 Say one has a finite proof procedure, PP_j, that will decide if, for a gn. umc., a gn. input string will converge for all output symbols or not. PP_i will have a certain set of "Yes" strings, a certain set of "No" strings, à a very large set of "don't know" strings. j can be regarded as a c.B. on PP — so as $j \to \infty$ PP_j becomes infinitely accurate — i.e., the "Don't know" strings become "unprovable" with more & more certainty. — But note that unprovable need not mean "false".

~~.31 w/ pattern Consider seqs of Pams [P] ∃ one for P that for all t. limits ($i \to \infty$) exist.~~

.35 This may be it: A Umc MSM for provably convergent outputs: One inserts an infinitely long input string I, into t. device, M_s. Inside the device, there is a regular umc M_U (not just a UOM). Also, there is a set of axioms à a proof procedure Machine, Mp. Mp tries out various proofs — perhaps using t. 2 dim scan of 2.10 – .16. 10.01

I don't know if this is necy

1374 R

TM9

1,2,9 R

.01 8.40 2.40 spac → Actually, maybe there is only 1 thrm here — i.e.

One can do t. scan of 2.10 – .16 If a proof or proof-of-negation exists, then this will eventually find it. The important thing, then, is t. 2nd thrm — (4.29 – .40) that there usually exist thrms for which one can prove neither them nor their negation.

.16 :2.40: Around 1962 or so, & perhaps still, a major big problem area in Math was t. Q. under what conds. could one show that a set of axioms was (a) consi &/or (b) complete. Th. "consi" Q. See 1962 or 63 conf. on Automata Theory for some papers (& refs to papers) in this area. H. Wang was a big worker in this area.

I would like to know more about t. Q of how one can show that a set of posts is consi — i.e. gn. a set of posts, that a gn. (negative) thm cannot be provd. — Just what the signific. of a "Model" is, etc.

Another thing that I don't understand, (& this is assoc. somehow w. Göd's work) is t. notion of a "system being able to prove itself consi".

1474 R A Universal MSM UMSM TM10

01:9.40 12, 9R

8.90 spac The input to M_P is t. set of axioms, ~~and~~ t. input string I, & a dscn of M_U. Another input is an integer i. The output of M_P is an integer, $F(i)$. T. signifc. of $F(i)$ is that after M_P has tried the first N possl. proofs, it has ~~proved~~ proved that the first $F(i)$ output symbols of M_U all "converge", for input I.

I think this should be i

($F(i)$ can be zero for all i), for certain values of I.

Let $G(i)$ be the output string of M_U, after i computation steps.

.12 The output of M_S is first, the First $F(1)$ symbols of $G(i)$ probably $G(1)$ is meant

next, it is t. first $F(2)$ symbols of $G(2)$

⋮ " " " $F(i)$ " " $G(i)$ . . . to $i = \infty$.

So, there is no output for a particular ~~symbol~~ position, until it is certain that that position converges. After this time (if it ever occurs) t. output is simply that of t. machine, M_U.

We could have $F(i)$ and $G(j)$ being computed simultaneously asynchronously. The output of M_S at any time is t. first $F(i)$ symbols of $G(j)$. i & j need not be =, but they both must $\to \infty$ at their own rates.

T. forg device is "univl" in th. sense that if it is set up to simulate a provably convergent MSM, it will do so, after finding the proofs of convergence.

In general, t. device will be made more powerful, by including more & more axioms. — One can usually insert an inf. no. of them, but one is never sure that they are consi! — And an inconsi set of axioms could prove anything! → 61.01

35

Suppose I proved that: If a seq. of CPM's, $[P_i]$ can be proved to $\to$ a limit, P_0, then I can construct a MSM that will give P_0 as a limit, with R_i's that are all CB's on this MSM $\ni P_i = P^{(R_i)}$. This would then give me the if and only if thrm. that I want. → 11.01

1474 R TM11

1: 10.40: It would involve that stuff in 458.22 – [redacted] 459.12; 460.1+ – .40. 1.2 [illegible]

~~irrelevant~~

02 Re; 463.35 – 464.10 ! [How to get a MSM to do non-a.c. computations]:

on 463.35 – .37 I say that "I can get a MSM ∋ t. convergence time for t. n^{th} sq. of output is > any a.c. funct. of n." → I'm not so sure of this now — in fact it looks false! T. convergence time can be ∃ either bounded by F(n) (where F(n) is non-a.c.) or it can be ∞.

If we are sure it is some F(n) < ∞, [Then F(n) must be a.c.] ✓No

NO! I was thinking we could then run it until it converged — But we never know when it has converged! (UOM UMC.) U.K (P(R).

Well – try this: take some fixed finite corpus, Use some UOM UMC.) to determine ∃ the t. proby of that finite corpus. As t. C.B., i's ↑, we get a monotonic ↑ value of t. approxn. to P_i.

Consider t. n^{th} bit of P_i as a function of i. Since P_i → a limit, t. n^{th} bit must settle down eventually, but

20 this "settling time" is incomputable (tho not nec'ly "> any a.c. function) ←

Hvr, using t. method of 463.38 – 464.10, I can construct a MSM umc. that can "compute" this "settling time". I think that is somehow is — If it is < some computable function Then CMI would be a.c, if we used this a.c, C.B.

Hvr, check this ↑! → It sure looks O.k.

~~One … Also note that the untractable seq. of~~

5.17 – t. untractable seqs for CMI via umc. UOM's

| So: write out new plan of Review paper — w.

~~Some~~ results on non-thms about limits & thms on MSM's

— their greater power than UOM's & t. untractable seq. of 5.17

Some old ideas to perhaps mention: (1) 69TM 228 Plan Rev: List of impt. things that CMI accomplishes. (2) old list of People to send Report to — plan Rev 69TM 128 (3) Also list of things to treat in report(s) plan Rev 69TM 128.29 – ~131.40

Also see plan Rev 69TM 130.01ff for discus. of what review should contain.

(3) on SFOR's ibid 235, 266 (4) on Provably Convergent Automata ibid 227.01

(4) some not silly work on t. ~~...~~ uniqueness of CMI ibid 222 ff also Plan Rev 70TM 55 (in 69 folder)

14.01

1774 [redacted] B RJS

TM14

11.40

.01 : [redacted] : On t. Q: Are there any induction methods better than CMI?

Consider t. Gore: G = Prob of corpus w.r.t. True Fdss + Prob of corpus wrt CMI. ($\hat{A}^{(l)}$, $\hat{A}^{(l)}$)

For CMF, this is bndd. For any method of induction it must have a lower bnd; i.e. it can't be zero for small l, because there simply isn't enuf info to "identify" t. Fdss.

For larger l, $-G$ must be a non↓ function, since one can't do systematicly better than t. original Fdss model.

How mite an induction method be better than CMI?

Well, ① Its constant G bnd can't be < that of CMI.

② Its $-G$ ↑ may be rapid for small l, then get very close to its final value very rapidly — thus giving very small errors for large l.

.18

It mite be possl. to get a theoretical ~~upper~~ lower bnd. for Expected value of $-G$ by considerations like Phys Rev 69 TM 224.06–.15

e.g. consider $l = 10$; consider a wtng function $W(F_j)$ on all Fdss's F_j. Then if we chose $P(A^{(10)})$ ← (This is t. proby that t. 11th bit will be 1) so as to minimize $-G$, i.e. so $\sum_j \mathrm{err}(P(A^{(10)}), F_j) \cdot W(F_j)$ is minimal, we will have a certain expected predn. error that is unavoidable. This "unavoidable error" will depend much on t. form of $W(F_j)$.

Hvr., note that in CMI, our "expected error" (a difrnt averaging for "Expected", hvr) is indep of t. form of $W(F_j)$ — so perhaps in t. analysis, keep $W(F_j)$ in a literal, changable form, and then do t. analysis for Expected error w.r.t. a gn. fixd F_j — & see how it is.

2.718828
3 1728

3 1728

01: 14.40: O.k.: So, Pams defined by MSM's probably are "nary" — in t. sense that if a seq. of pams → a limit ∋ brost error is bndd for every fdss, then that seq. of pams is expressable as a set of CB's on a umc MSM, using suitable probability mechanisms. T. probability of t. convergence of t. seq. of pams implies t. convergence of t. corresp. set of MSMs. But I still have to go thru something like 460.14 ff to construct t. MSM that correspond to t. seq. of Pams. 7.45 6 9.30

Even so,: Say we have a "Limit Pam" $\equiv \lim_{i\to\infty} [P_i] \doteq P$

P has a brost error ~~bndd~~ bndd by B_j (for t. j^{th} fdss). Say we can construct a MSM, M_s ~~whose CBs~~ a set of whose CB's $P^{(M_s(CB))}$ give this same/ or less brost error, B_j. It is not clear that this ~~probably~~ limit Pam is "as good as" P — tho it is/as good, using this particular Gorc. 14.01 – .18 suggests other Gorcs mite be usd.

---

21 On t. use of umc MSM's for induction: [It is clear that MSMs are at least as goods UOM's & better in at least one case (i.e. t. ~~untract~~ untractable seq. of 3.17), & very probably many cases.

Up to now, I don't think hypoths ~~ott~~ as genl. as MSM's have been usd in science. One mite say that (umc MSM's) are hard to use, because one never-knows how far from t. limit one is — That in ~~um~~ umc UOMS one at least has a monotonal probty. In fact, hvr., this latter is not so. T. normzd probty is not at all monotonic. T. dsvry of a new short code can change t. expected proby values by a large factor, either way (up or down).

B — Tho I think there is even more uncertainty about t. limit to be expected from umc MSMS.

T. main argt. for (umc UOM's) is the unprovd set of ideas about expected $\$$ ($\equiv$ future cost/symbol) in my work on SVHM. 16.01

01:15.40: Which gives us an idea of [redacted] what we can expect in t. future, using [redacted] a non-umc uom. (i.e. C.B. $<\infty$). — Hvr, it may very well be that corresponding Thrms can be devised for MSMS. — in fact one can, *trivially*, devise correspg Thrms — That using MSM w. finite C.B.'s, t. expected future brest/symbol will be about that observd in t. past — with perhaps various refinements corresponding to Those That I introduced for oum's.

— So *really*, t. 2 methods may *not* be so difrnt. as far as *this* limit is concernd.

As for t. limit of *provability*. At any point, one uses only MSM's and output seqs thereof in which one can prove that convergence [redacted] occures. Hvr, for provability, one [redacted] usually doesn't know if one's set of postulates is consi — so [redacted] one should look for a proofs of *in*consi [redacted] of t. set of posts, along w. a proofs for t. convergence of t. t. approxn. seq. One will drop t. search if [redacted] proof of inconsi occures, & drop one of t. postulates & try again. If a proof of convergence occures, one may still want to continue t. search for proof of *inconsi*, while one is continuing out t. seq. of approximations.

One apparent good pt. of uom's: One can, for some C.B., find a pem. That is ① known to converge for all & ② one knows [redacted] how long it will take (what C.B., T) to get agn. accuracy in probty w.r.t. that pem. E.g. one has t. coif. linear regressn. — one knows past $\frac{\text{brest}}{\text{corpus symbol}}$ & one can estimate ξ.

It's not clear as to whether there is a correspg. situation for MsM's. → 26.31 spec.

Bdwy
Will
Sci Co.
Cumb.
Beakers
Pyrex
600ML
~ 60¢

Ac. Acid
~12/lb.
~1 pt.

Dextrose
85¢/lb
1 lb
40¢/lb -
25 lbs -

is

01: 16.40; The Main Bottlenecks in a "Proof" of t. adequacy or sufficiency or necessity of CMI for induction:

1) The MSM problem: While UOM's are good (bndd Pros error) for Fdss's, it is not clear that MSM's are not better — certainly MSM's are at least as good wrt. Prob error. Hvr., t. discn. of 15.21–16.40 suggests that UOM's are better because one can estimate ξ for them — & I haven't found a way to do this for MSM's.

Actually, I haven't proved that my ξ estimates are "correct" & more important, I don't have any (unproved) thrms on them that tell when my ξ estimate techniques are applicable. ξ ξ

2) The **LC** (lgt. convergence) **Problem:** In general Fdss's are not what are usd in models for science — (Even if Fdss include all CPM's — not only r-CPM's). Just how this fact is to be treated, is unclear. All sci models that I know of, have been based upon a finite no. of params. These params were involvd in t. models in such a way, that $E\left(\sum(\text{err})^2\right)$ (xpectd value) in probly predn

$= \left(\sum A_i \delta_i^2\right)\ln \ell$ [ℓ = corpus length, δ_i = error in i^{th} param; A_i = constants]

Models of this sort are certainly not t. only kinds conceivable, & I should perhaps draw up a few other kinds.

Another type of model uses potentially ∞ coifs. (See LC 403.xx, 404, 405, 410 —

(See 410.28–.35 for a final, good, discn. of N=∞ coifs) ← This discn. is for linear regressn., t. analysis is identical for any continuous functional form if there is lots of data (=large S) — i.e. the function is locally linear

So, thus far, we have 3 ways that errors may converge, assoc. w. 3 kinds of Models:

→ 1) **Fdss** (probly never usd in Science)

2) Finite no. of coifs : Most often usd.

3) potentially ∞ no. of coifs : Sometimes usd — perhaps always usd, if one considers t. whole of scientific observation & prediction as corpus & potential. pams, resp.

20.01

01: 18.40 : On t. Q of whether a umc MSM can always simulate t. limit (when its known to exist) of a seq. of e.c. pems, [P_i]. M20

Well, I think its clear that for every such [P_i] ∃ MSM constructively — Tho I should verify this again

.04 does / exist, so its $P^{(MSM)}$ is identical to t. limit of [P_i].

Just what was it I wanted to show about this limit P^{MSM}, around 460.14 ff, & why was it impt.?

Also, .04. implies that at least one MSM is at least as good as any limit [P_i] of e.c. pems. Also, since umc MSM's exist (which I've shown in recent work on Provably Convergent MSM's) i.e. 8.35 ff any umc MSM is at worst only an additive constant worse than any e.c. limit [P_i].

So t. problems: (1) what was I trying to do on 460.14?

(2) UOM's are better than MSM's because of t. possy of ξ upper bnd estimation in UOMS. Can I state this in more exact form — Ex. e.g. For any CPM, in which one can get reasonable upper bnd on ξ, The umc limit, OUM is t. best one can do.

(3) See 18.01–.40. for t. problems of now!

11574 R

TM22

A possl. outline of t. report!

1) This paper is ~~[illegible]~~ a commentary on Willis' "CompuCompl & Proby Consts". Willis has proposed a mathematical solution to the problem of induction. We shall ~~now~~ try to determine to what extent it is indeed a solution, and to what extent it is unique.

Defn. of induction interms of predn. of t. next symbol of a long, finite seq. of symbols.

W's definition of a FOR & ~~an~~ assoc. of a CPM w. each — His use of FOR's of ↑ power to get better CPM's.

It is possible to assoc. w. Umc's a set of FOR's of ↑ powers, & the CPM's assoc. w. these FOR's → a limit. Proby measure. While t. limit exists, it is not e.c. — ∴ not a CPM.

W. defines prediction error for any PM w.r.t. ~~[illegible]~~ for a gn. corpu ~~[illegible] PM~~ any other PM. This is one ~~[illegible]~~ explication of t. "difference" betw. t. 2 measures. W. shows that for suffly large c.B.'s (suffly powerful FOR'S), the mean error in prediction must → 0 as $\frac{k}{l}$ ($l \equiv$ corpus length), k is some const.

Diffy w. his mean error is that t. error can sometimes be +, sometimes −, so knowing that its mean value is zero, does not make ~~[illegible]~~ us feel that it is usually very small.

We shall, hvr, show that in general, the expected value of ~~[illegible]~~ total the probability error ~~[illegible]~~ squared is bounded, so its mean must → 0. This makes it very likely that for long corpi, t. ~~[illegible]~~ probs gn. by t. method will be in error by a vanishingly small amt.

W's bndd error criterion implies bndd Σ sq. error.

It is natural to ask if they are not equivalent. They are not. W's bndd error is a stronger condition than bndd Σ sq. error. We will show that it is possible for a CPM to satisfy the bndd Σ sq. error criterion, but not ~~Σ~~ W's bndd error criterion.

23.01

01: 22.40: [Mention that while t. univ Proby measure is not computable, ∴ no CPM can work w. all Fdss, One can use a CPM if one's Fdss is of "known complexity bnd".]

Restart:

• ~~This will be discn. of~~

In 1970 Willis proposed a soln. to t. general problem of inductive inference. T. present paper.

4 19.01 ff isn't bad — also t. outline on 122.01ff.

T conclusion on uniqueness will have to be modified considerably, hur.

.18 List of Things That I want to demonstrate:

1) For any univ UOM, M_u, $P(M_u)$ exists, as a limit. (we define $P(M_0)$ this way).

[Also discuss Q of whether t. limit varys, depending on how one selects t. C.B.'s — I think it does depend on selection method & I may have an example.]

21 2) For any CPM, P_0, $P(M_0)$ has brost error < brost of P_0, for any finite sequence ∴ is bndd for any finite l.

23 2.5) see 24.20 →

3) discuss signifc. of mean brost error; (can be + or − ∴ if mean is zero, it doesn't seem impt.).

4) If brost error bnd is k, then E (Σ sq err) bnd is $\sim k \ln 2$.

5) If E (Σ sq. err) is < k, then Σ brost error need not be bndd. (i.e. ∃ a CPM, P_0, ∋ a Proby measure P^α ∋ E (Σ sq. err) of P^α used to predict ~~predict P_0~~ seqns "out of" P_0 is bndd but brost error in this case → ∞.

5.5) → (see 29.20)

11774 R

01;23.40 ≥ 6) ← untractable sey for UUOM's.

$\exists$ a sequence of CPM's P^T, w.r.t. t. propertys: | TM 24

a) for every $A^{(\ell)}$ $\lim_{T\to\infty} P^T(A^{(\ell)})$ exists — Call this limit

.03 $P^{\infty}(A^{(\ell)})$.

b) for $\forall \ell$, $\exists$ at least one $A^{(\ell)} \ni$,

$-\log_2(P^{\infty}(A^{(\ell)}))$ & $-\log_2(P^{(M_u)}(A^{(\ell)}))$ differ by $\ell(1-\epsilon)$.

+ — for any other UMC / UOM, M_u', t. difference is $\ell(1-\epsilon) \pm$ a constant. See (37.01–.18 / 5.17) for proof — P^∞ is t. "untractable seq." for UUOMs.

.11 7) $\exists$ a seq. of CPM's P'^T $\ni$ (a) for every $A^{(\ell)}$ → see 47.01–.90 for remarks on this thrm.

$\lim_{T\to\infty} P'^T(A^{(\ell)})$ exists — call this limit $P'^{\infty}(A^{(\ell)})$.

(b) For any P^{∞} w. propertys .01–.03,

$\log_2(P'^{\infty}(A^{(\ell)})) - \log_2(P^{\infty}(A^{(\ell)}))$ has a bnd (That is indep of ℓ)

(P^{∞} may also be a CPM, of course). ← (P'^T is t. set of C.B.'s on a UMSM → See 39.20 for stuff on UMSM's.

.20;23.23 ($\equiv 2.5'$) If P_0 is in a known complexity class (e.g. it takes

$< \ell$ or e^{ℓ} steps to calculate $P_0(A^{(\ell)})$ or any other

functional bound on t. no. of steps needed to calculate ")

Then we can construct a CPM, $P_a \ni$ t. brost error

for P_a w.r.t. P_0 is bndd. (for any ℓ) by t. brost of P_0 w.r.t. M_u

---

127 8) Conjectures (a) T. An upper Bnd for expected value of t. brost of t. next symbol of a gn corpus,

$A^{(\ell)}$, w.r.t. a gn. CPM, P_0 is $\dfrac{-\log_2(P_0(A^{(\ell)}) \times \text{Pcost of } P_0)}{\ell}$ ← Let Pcost of $P_0 = 0$ — Then it should be easy to define what I mean by "Expected Value". ← This is some sort of "unbiased estimate".

(b) If Pcost of $P_0 = 1$ (explain what this means) and

$\frac{d}{d\ell} -\log_2(P_0(A^{(\ell)}))$ is meaningful, Then This expression is bnd by

$$\frac{d}{d\ell}\left(-\log_2(P_0(A^{(\ell)}))\right) = -\frac{1}{P_0(A^{(\ell)})}\frac{d}{d\ell}P_0(A^{(\ell)}).$$

(c) for (a) we can break t. brost of t. Pom dcon into 2 parts: (1) t. part decided on before t. corpus was seen (2) t. part decided on after t. corpus was seen. : T. brost of (1) is zero; t. brost of (2) is usual. Easiest to compute if we use Lcost rather than Bcost!

.36 (d) Other expressns. for Pcost of $P_0 < 1$ but > 0.

This "Expected value" is w.r.t. t. aprip assoc. w. all CPM's —

i.e. Their Pcosts. → If Lcost − bcost is budd — either form is correct. If not, then I bet on bcost!

(e) The SUH thrm! (trivially true — see SUH 409.15ff) If one always uses for each ℓ, for predns t. Pom of dcon length $< \ell \ni$ its bcost (or Lcost) + t. Pcost of t. corpus wrt that Pom is min., Then one will end up with predns about as good as those of CMI. ← This thrm is trivially true: see SUH 409.15ff → 29.01 spec 25.01

01:29:90 For Thrm, show that if bcost error is K(l), then E value of Σ err2 is K(l) ln 2; This paves way for L_{og} Conv. analysis.

When umc UOMs ~~at~~ are in power to MSM's, we get some new codes — but all (?) of these new codes are such that one never knows how close one is to convergence.
So, in the sense of estimating ξ, MSM's would probly be no better at all, ~~[illegible]~~ than UOM's.

Phone Marvin, Lawyer Rolwin

~~[illegible]~~
Re: A bound on (bcost + kcost) for umcs:
Using the methods of Willis' Thrm 5, we can construct a MSM M_s that has the same ~~[illegible]~~ probly measure as any UOM (say a umc). Furthermore, by using CB's on any ~~a~~ umc MSM, M_{su} we can get a MSM M'_{su} that has the same proby measure as this umc MSM, but has −bcost + kcost ≤ 1.
Whether this M'_{su} has to be umc is unclear. If it is, then ~~all~~ each MSM umc ~~has~~ has the property that kcost − bcost ~~of~~ of any string is bndd by a no. characteristic of that MSM umc.

Well, perhaps because MSM's do have simpler propertys than UOM's, I mite want to study them more.

The main advantage of UOM's over MSM's for predn, is that I can get this ξ value for UOM's.

Hvr, : At any pt in the ~~the~~ MSM limit PEMs — I ~~[illegible]~~ would have a CPM, so I could use the same ξ computation method as I'd used w. UOM's!

Def) UMSM & UUOM : are respectively Universal, (Mildly sequential) & Unidirectional Output) Machines

→ But, if this is "legal", then we would use (bcost of pem + bcost of corpus wrt pem) as a score & select a "best" pem on this basis — which by-passes the idea of MSM's entirely

26,31

One possl. importance of Y buddness of kcost − bcost:
If it is bndd, than selecting Y. Pam for which
(kcost of pam + bcost of corpus wrt pam) is min, is equivl. to selecting for
.04 (b " " " " " " " " ") is min, when corpus length → ∞.
.05 On Frame Theory: My main objection was that there was little consideration of criteria for determining how good a frame system (or any other representation system) was.
— It may be that one can consider this criterion as a separate problem. — i.e. one uses Y. frame Theory for a while, notes its Gorc value, make mutations in Y. representation, & sees if Y. Gorc gets better or not. This technique is indip. of Y. exact nature of Y. Gorc.

On ξ estimation: If a mean $\frac{\text{bcost}}{\text{symbol}}$ exists for ti corpus, & bcost of pam = 0, Then I think Y. observd past ξ must be t. correct/ best predn for t. future ξ — I can perhaps prove this.
If one started out w. N difrnt pams, each of "aprio" P_i $(i=1/N)$ ~~& chose t. best one (Best = min of .04)~~, Then, if t. pams are all "indip" (whatever that means in t. present context) t. relative wts. of Y. N pams in predicting Y. next symbol, would be P_i × (bcost of corpus wrt i^{th} Pam) — which isn't what I want anyway.
I think my old argt. was that bcost of corpus could, at most, be ah. by t. amt. (bcost of Pam), & in t. case of fixd "apri" probys (Apri in t. sense of decided upon befor t. corpus was seen), t. ahness is only $-\log_2 P_i$ bits at most.

.31 1:25 16.40 As for ξ estimation for UUOM's & UMSM's: There is some similarity: in both cases bcost of Pam = 0. One must only decide if one wants to ↑ t. C.B. by some amount. One can't ↓ t. C.B. — so I think there is no A.Hness choice involvd. Hvr., in t. 2 methods, t. sequence of pams that one gets is difrnt — so this may be worth looking into; perhaps UMSM's aren't so bad after all!

.01:26.40: In fact, using UMSM's one gets a sequence of CPM's — Each CPM has a probty value for every $A^{(l)}$ for any l value — Just t. same way that one puts C.B.s on UUOM's to get seqs. of CPM's. In both cases, I think t. brost of t. CPM is zero (see 26.31-.40). So, it mite be possl. to use t. slope of Σ brost : i.e. $\frac{d\,\Sigma\,\text{brost}}{d\,l}$, — if ~~[illegible]~~ one can obtain a reasonable value for such an expressn.

So: Actually, niether t. UMSM or UUOM methods are of very much direct import in practical induction. The SVH "thrm" & assoc. approxns. seem to be much more important.

Essentially, I now have 3 induction methods!

.18 1) successively larger CB's on UUOM's & ~~[illegible]~~, brost for CPM's = 0 ~~[illegible]~~.

.19 2) " " " UMSM's " " " " "

.20 3) Th. SVH Method! — hvr, see SVH 409.15 ff this method's No good! A sequence of Pams, use of t. one pam thus far ~~pam~~ for which (brost of pam + (brost of corpus wrt that pam) is min > for ~~[illegible]~~ obtaining ~~[illegible]~~ probys. of t. next symbol.

.24 Re: that "untractable seqs" for .18 : Actually, one can't construct anything like that, but one can construct a seq. so that ~~[illegible]~~ its n^{th} symbol is constructed using the n^{th} (of the $f(n)$ th ~~[illegible]~~ for a rapidly ↑ $f(n)$ — perhaps t. ~~[illegible]~~ value ~~[illegible]~~ of $f(n)$ can be uncomputable?) aproximation for n^{th} symbol of t. original seq. ~~[illegible]~~

— Hvr, all this means is that there is no single C.B. (i.e. CPM) that will "cover" t. whole sequence — & in .18 we do not need that any way — in fact in most ~~[illegible]~~ t. CB's ~~[illegible]~~. T is some computable function of l. So .24 ff is certainly not a particularly unusual ~~[illegible]~~ Fdss. Even if $F(n)$ is [uncomputable (tho always can)], — I'm not sure this is meaningful in t. present case! .24 is still a CPM ~~[illegible]~~

Clearly in .20 t. goodness of t. method depends on how good one's sequence of ~~[illegible]~~ CPM's is. I may have been thinking of considering all CPMs w. a dern. brost < l of t. corpus. 28.40

28.40

11974 — R

01: 27.40: The rate of ↑ of [redacted] cB (ovT) with l. can be adjusted as one desires. TM28

Anyway 27.19 (UOM UMSMs) are probly better than 27.18 (UUOM's) for induction, because for every $P^{(UOM)}$ & every $P^{(UMSM)}$

$\exists\ k \ni P^{(UOM)} < k\ P^{(UMSM)}$ (for all $A^{(e)}$) $\quad k > 1$

But it is <u>not</u> true for UUOMS & MSM's the other way — This can be shown by the untractable seq. (≡ Proby Measure) of <u>5.17</u>

.01 : 28.40
24.40 spac → 9) General discn. on induction problems in RW:
show how analog induction, hybrid induction (analog-digital), operator induction, clustering, grammatical induction, are all covered by CMI — or by t. problem of extrapolating a long seq. of symbols.

.07 10) rcPMs & some non rcPMS: A list of PM types that appear in RW.

a) r.CPM. (rare) example: bern seq. w $p = \frac{3}{8}$.
b) CPM (rare) Bern seq w. $p = \frac{1}{2}\sqrt{2}$ (or $p = \frac{1}{3}$ for a binary machine)
c) PM w. a finite no. of "random params".
example linear regressn. w. $y(t) = a\,x(t-1) + b\,x(t-2) + \text{noise}_\sigma$, Gaussn.
a, b, σ all "random" nos. betw 0 & 1 (random means long dern)

b) models w ∞ random params. e.g. linear regressn.

$$y(t) = \left(\sum_{i=1}^{\infty} a_i X(t-i)\right) + \text{noise}_\sigma \,;$$ all a_i's, σ are random nos. betw 0 & 1 [Is this "non-parametric" statistics?] → 39.15
.19

.20 : 23.40 (5.5) At this point CMI looks "very good". It is natural to ask, does CMI (23.21) give one t. best induction possl.? More exactly, if P^{*}_{T} is a seq. of Pams ∋ for all A_T, $P_T(A_T) \to$ a limit ($T \to \infty$) and $P_\infty(A_T)$ never has > a bndd bcost error for any FdSS source (i.e. a rcPM), then does there always exist a UUOM, M_{UU} ∋ ∋ $p(M_{UU})$ is as good as P_∞ in terms of bcost error? → TM24
while I don't know t. answ. to this Q, Thrms 6) & 7) are relevant. Thrm 7) says ∃ a seq of P's that → a limit, P'^{∞} such that P'^{∞} (UMSM) is at least as good as P^{∞} (UUOM) & in certain uncomputable cases, is better than P^{∞}. Whether P'^{∞} is better than P^{∞} for any rcPM's is an open Q. I.e. it is conceivable that it may have smaller bcost error than P^{∞}, but I don't know.

Also note that P'^{∞} is assoc. w. a special kind of Universal Tm. That is in some sense, "more powerful" than UUOM's.

30.01
34.01 spac.

12074

29.40

" on t. Proof of t. untractable seq: " → See 32.01–.18 for way to get around this difty. Hvr, it forgs discns (which don't prove what I wanted to prove) are still interesting

.01 : ~~[illegible]~~ : Re: T. way $P^{(\text{UOM})}$ → its limit. I don't think it has to → its limit suddenly. I had t. idea that for each $A^{(\ell)}$, there was a T_0 ∋ for $T > T_0$, ~~[illegible]~~ $P(M^T)(A^{(\ell)})$ was indep of T.

6,4 spac.

— This T_0 being, of course, not e.c. — . I think now, that this is wrong; There are always new very long codes for $A^{(\ell)}$ that can be coming in as T grows. [e.g. a non r.c.p.m. would have non-terminating ~~[illegible]~~ real nos. for its probys]

10 W.r.t 5.30: It would seem that $\lim_{i \to \infty} A_i^{\ell+1}$ need not exist if it is defined this way! Suppose $\ell = 1$;

~~$P_i(0) = \frac{1}{2} - (\frac{1}{2})^{\lfloor\frac{i}{2}\rfloor}$~~

~~$P_i(1) = \frac{1}{2} - (\frac{1}{2})^{\lfloor\frac{1+i}{2}\rfloor - 1}$~~

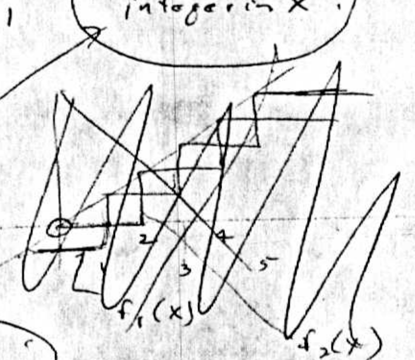
$\lfloor x \rfloor \equiv$ largest integer in X

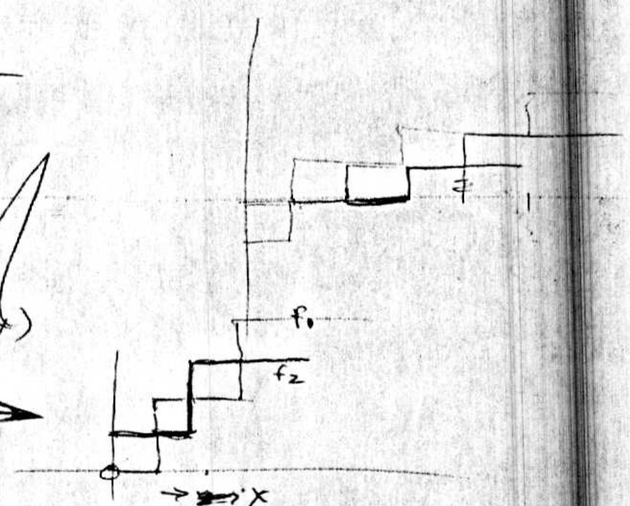
~~$P_i(0) = \frac{1}{2} - (\frac{1}{2})^{\lfloor\frac{i}{2}\rfloor}$~~

~~$P_i(1) = \frac{1}{2} - (\frac{1}{2})^{\lfloor\frac{1+i}{2}\rfloor - 1}$~~

$\lfloor x \rfloor \equiv$ largest integer in X

$f_1(x) \equiv 2\left\lfloor\frac{X+1}{2}\right\rfloor$ $\quad$ $f_2(x) \equiv 2\left\lfloor\frac{X+2}{2}\right\rfloor - 1$





$P_i(0) \equiv \frac{1}{2} - (\frac{1}{2})^{f_1(i)}$

$P_i(1) \equiv \frac{1}{2} - (\frac{1}{2})^{f_2(i)}$

so at alternate i values, first $P_i(0)$ & then $P_i(1)$ will be larger so ~~[illegible]~~ $A^{(1)}_\infty$ ∴ would not exist!

Hvr. both $\lim_{i \to \infty} P_i(0)$ & $\lim_{i \to \infty} P_i(1)$ exist., & t. ratio of these limits is t. limit of t. ratios. So, using a binary machine, consider $\frac{P_i(0)}{P_i(1)}$. It will → a limit as $i \to \infty$. If $\frac{P_i(0)}{P_i(1)} \leq \frac{1}{2}$, $A_i^{(1)} = 0$ — ~~what~~ well, tho $\frac{P_i(0)}{P_i(1)}$ → a limit, it can still oscillate about that limit indefinitely. If we happened to chose t. threshold here to be that limit, $A^{(1)}_\infty$ will not converge. (i.e. instead of $\frac{1}{2}$ here we could use $\frac{1}{2} - \epsilon$, for small $\epsilon > 0$.)

1-20-74

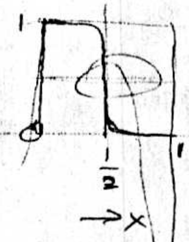
R

01:30.40 : <u>on untractable seqs, for limit Pams (cont.)</u> TM30 TM31

Another tack: say P_i ~~[illegible]~~ is normalized so $P_i(0)+P_i(1)=1$

Then, instead of an <u>untractable</u> <u>seq.</u>, let us construct a proby measure, $P_i'(1)$

$P_i'(1)$ = a <u>very</u> sharply n.l. funct of $P_i(1)$



In fact, let $f_i(x)$ be always <u>continuous</u> for all i, but let $f_i(x)$ become sharper & sharper as $i \uparrow$.

so [redacted] $P_i'(1) = f_i(P_i(1))$.

Consider $\lim_{i\to\infty} x_i \to 0$.

This shifts t. discussion ~~[illegible]~~ in t. $f(x), x$ plane, from $\frac{1}{2}, \frac{1}{2}$ to $(0,0)$

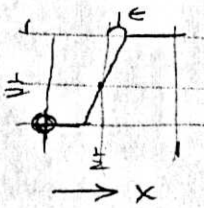
Does $i\,x_i$ have to $\rightarrow$ ~~a~~ a limit?

I think not: say $x_i = (-1)^i \cdot \frac{1}{\sqrt{i}}$ or $x_i = (-1)^i \cdot \frac{1}{i}$ } in both cases, $x_i \to 0$

$i\,x_i$ oscillates finitely and infinitely. [redacted]

Not $i \times x_i$ but a funct of i that has an abs value max of $\frac{1}{2}$: <u>Still</u> t. results of t. rangt. are <u>correct</u>.

.20 So we may have to keep $f_i(x)$ <u>indip</u> of i.



we can make ϵ as small as we like, hvr.

~~[illegible]~~

O.k., so, we set ~~normed $P_i'(A^{(\ell)})$ ... normed $f(P_i(A^{(\ell)}))$~~

$$\frac{P_i'(A^{(\ell+1)})}{P_i'(A^{(\ell)})} \equiv F\left(\frac{P_i(A^{(\ell+1)})}{P_i(A^{(\ell)})}\right)$$

So this defines P_i' in terms of P_i.

How, how much b cost error is there ~~[illegible]~~ in P_i (in trying to approxt. P_i')?

If we find ~~a~~ a sequence A for which P_i assigns (probys deviating from $\frac{1}{2}$ by $>\epsilon$) to N conditional probys of symbols, — then t. b cost error for A is $\geq N$.

.30 Ah so! Consider t. ~~2~~ 2 ~~[illegible]~~ [redacted] Proby Measures

1) P'_∞

2) ~~[illegible]~~ T. proby for t. seq. of all 1's $(1\,1^{(\infty)})$ is 1

all other probys are zero.

I think P_∞ must ~~[illegible]~~ have b cost error of $\sim(1-\epsilon)\ell$ (for large ℓ) for t. seq $1^{(\ell)}$ w.r.t. either Pam 1) or Pam 2)

32.01

01:33.40: Clearly, if P_∞ is to <u>converge</u> for the deterministic seq. of 1's, It must have its conditional probly $> \frac{1}{2} + \epsilon$ for <u>all</u> but a finite no. of values of l. (In fact the condl. probly must $\to 1$ as a limit.) However, This would imply a best error of $\underset{constant}{k} + l$ in P_∞'s evaln. of that string w.r.t. P'_∞. — Since P'_∞ would give zero conditional probly to all but a finite no. of symbol predictions.

Somehow, there seems to be something <u>very wrong</u> w. the forgg. argt.!
i.e. consider 2 pms: 1 assigns probly 1 to the seq. $0^{(\infty)}$
the other " " " " " " " $1^{(\infty)}$.
According to the forgg. argt., P_∞ would have to have best error $\sim l$ for at least one of them.

Well — Perhaps that's not the idea. The idea is that P_∞ is supposd to give $> \frac{1}{k}$ ~~probly~~ times probly assigned to a seq. by any other Pcm.
In This case P'_∞ assigns probly <u>zero</u> to the seq $1^{(\infty)}$, &
P_∞ assigns probly > 0 to it — which is what P_∞ is <u>supposd</u> to do.

What I have to find, is sequence in which P_∞ assigns $< \frac{1}{2} - \epsilon$ probly, to no. of its symbols that is $\propto l$. If there were a finite dcrn. of such a seq.'s it would be deterministic, & so only a finite no. of its symbols could be given probly $< 1-\epsilon$(!) (a fortiori $\frac{1}{2} - \epsilon$).

On the other hand, the seq. <u>need</u> not have a finite dcrn.
it's just that the argt. of 31.30ff about the seq. $1^{(\infty)}$, is wrong, ~~since~~
~~$1^{(\infty)}$ has a finite dcrn~~ ($1^{(\infty)}$ has a ~~finite~~ dcrn., but This is irrelevant here).

30 However, I still don't see in <u>detail</u> what's wrong w. 31.30ff.
— Just where is the error in That?

.31 Relevant here: One mite prove the desired seq. exists by a
.32 <u>diagonal argt.</u> In order to frustrate the construction of 31.20ff,
P_∞ would have to assign very close to probly $\frac{1}{2}$ for almost all condl. problys. While This is <u>not impossl.</u>, it would be a very
.35 interesting result, if true.

It mite be possl. to show That <u>if</u> a untractable seq. or Pcm <u>exists</u>, it should by findable by the method of 31.20ff.

01:32.40 If we do t. thresholding ~~befor~~ of P_i before normalizing, we do get monotonic approach to t. limits, but t. limits of $P_i(1)$ & $P_i(0)$ may usually both be ~~$<< \frac{1}{2}$~~. Hvr, ~~I~~ think it is possl. to show that most conditional probys ~~of~~ non-U events are very close to "normal" — i.e. U events are unlikely. — To show this, say t. finite input, 'b', of length N_b gives a convergent output. — (then $\approx$ t. argument) " At least a fraction 2^{-N_b} of all input strings converge" ← Not exactly that, but something like that. I have written ~~much earlier below~~ several pp. befor finding a trivial proof of this — More specifically, I think that if a fraction $> 2^{-100}$ of all strings are non-U — indep of their length, then almost all conditional proby normalization constants (to get rid of U seqs), must be very close to 1.

Hvr, I must get this thru straight! Also, even if "almost all" of t. effects of thresholding yield probys of 0 or 1, this would put us perhaps, pretty much where we arrived after 31.30 ff — but perhaps w. ϵ's that automatically $\to 0$ as $l \uparrow$. Tho on 31.30 ff, we can also have $\epsilon = 2^{-l^2}$, say, — have $\epsilon \downarrow$ arbly fast as $l \uparrow$. Hvr, 32.32 – .35 may be true.

---

perhaps: ① look at 32.30

② " " 32.31 – .32. (diag argt.)

Actually, if I can show that there are some seqs. ∋ P_∞ assigns $< \frac{1}{2} - \delta$ to an arbly large conditional probys of it., (for any $\delta > 0$). then I can show that P_∞ will have a brost error arbly < that assignd byt. P'_∞ of 31.20 — ~~within~~ if we set $\epsilon < \delta$.

~~If A some seq. of length l P_∞ assigns a proby $< (\frac{1}{2})^l$~~

~~Then~~ If t. no. of cond probys [illegible] of $P_\infty < \frac{1}{2}$ is bndd (i.e. not $\to \infty$) then [illegible] all but a finite no. of cond probys for all corpi, must be $\geq \frac{1}{2}$.

.36 We could have ∞ probys of $\frac{1}{2} - k(l)$, where $\lim_{l \to \infty} k(l) = 0$. In this way, there could be only a finite no. of probys $< \frac{1}{2} - \epsilon$, nomatter how small ϵ is — as long as it is fixd. Hvr., if we let ϵ of 31.20 ff be a func. of l that $\to 0$ as $l \to \infty$, we can get around that.

34.01

12274 R

01:33.90: Note that in P_∞, $P_\infty(A^{(\ell)}) > k(\frac{1}{2})^\ell$, since we have at least 1 code for $A^{(\ell)}$ (the seq. itself) of length $\sim \ell$. This

Def means that if $^cP_\infty(A^{(\ell)}) \equiv \frac{P_\infty(A^{(\ell)})}{P_\infty(A^{(\ell-1)})} \equiv$ the conditional probty of the last symbol of $A^{(\ell)}$ wrt P_∞.

Then if $^cP_\infty(A^{(\ell)}) = (\frac{1}{2})^{1+\delta(\ell)}$ then $\sum_{\ell=1}^{\infty} \delta(\ell)$ must converge.

or if " " $= \frac{1}{2} - \delta'(\ell)$, $\sum_{\ell=1}^{\infty} \delta'(\ell)$ must converge.

To see: $P_\infty(A^{(\ell)}) = \prod_\ell {}^cP_\infty(A^{(\ell)}) = \prod_\ell \frac{1}{2} \times \prod_\ell \frac{1}{2}^{\delta(\ell)}$

~~$\prod_\ell (\frac{1}{2})^{\delta(\ell)} = e^{(-\ln 2)\cdot\delta(\ell)}$ $\therefore \prod_\ell \frac{1}{2}^{\delta(\ell)} = \exp \sum (-\ln 2)$~~

$\prod_\ell \frac{1}{2}^{\delta(\ell)} = \frac{1}{2}^{\sum_\ell \delta(\ell)}$

Also, have $\frac{1}{2} - \delta'(\ell) = \frac{1}{2}(1 - 2\delta'(\ell))$ | $\prod_\ell (1 - 2\delta'(\ell))$ converges to > 0, iff $\sum_0 2\delta'(\ell)$ converges.

Hvr, because of 33.36, even if $\sum \delta(\ell)$ converges, we could still have $k(\ell)$ be $< \delta(\ell)$ & get our untractable seq.

For every value of ℓ, there must be at least 1 seq, $A^{(\ell)} \ni P_\infty(A^{(\ell)}) < \frac{1}{2}^\ell$. (They couldn't all be $> \frac{1}{2}^\ell$, since this would make $\sum$ to > 1)

$\therefore$ there must be an infinite no of ℓ values for that seq. for which $^cP_\infty(A^{(\ell)}) < \frac{1}{2}$. Can I ~~make~~ find any limit to the rate at which $\sum_\ell \delta(\ell)$ can converge? — i.e. can I say that $\delta(\ell)$ must converge slower than a certain ~~[illegible]~~ rate?

Well, we know that there are seqs of length ℓ (for arby ℓ) that have $P_\infty > k > 0$ (These are deterministically derivable seqs, like (00)),

So there must be at least 1 seq. of length $\ell \ni P_\infty(A^{(\ell)}) \leq \frac{1}{2}^\ell - \frac{k}{2^\ell - 1}$.

(There are $2^\ell - 1$ seqs of length ℓ to share ~~[illegible]~~ the deficit of k due to the surplus of k in one $A^{(\ell)}$ sequence).

$\therefore$ so $P_\infty(A^{(\ell)}) \leq \frac{1}{2}^\ell - k(\frac{1}{2})^\ell = (1-k)(\frac{1}{2})^\ell$:— which doesn't help.

Would it be possl. for all $A^{(\ell)}$ to have $^cP_\infty(A^{(\ell)}) \geq \frac{1}{2}$ for $\ell >$ some ℓ_0?

So far, I only have ~~to~~ 2 kinds of bnds on how ~~[illegible]~~ large $P^c_\infty(A^{(\ell)})$ can be: ① $\sum_\ell$ ~~[illegible]~~ $-\log_2 P^c_\infty(A^{(\ell)})$ must be bounded above, by a certain constant characteristic of the univ., because $A^{(\ell)}$ has at least 1 dcrn of (length ℓ + const). So $\sim \overline{P^c_\infty}$ is bounded below — it can't be $< \frac{1}{2}$

② ~~$\leq$~~ $\overline{P^c_\infty}$ can't be $> \frac{1}{2}$ for all $\vec{A}$ sequences, since then $\sum$ probty would by > 1. 35.01

TM34

:01: 34.40 : Well! I don't see how to do it! — it may well be, that while there are seqs. w. $P^c_\infty(A^{(\ell)}) = \frac{1}{2} - f(\ell)$ w. $f(\ell) > 0$, $f(\ell)$ is not computable — & a lower bnd on $f(\ell)$ is not computable!

Consider t. Bern seq. p of 0 = .1, p of 1 = .9.
Consider some typical seq. obtaind by this CPM. For large ℓ values, P^c_∞ will give values very close to .1 & .9 to 0 & 1 resp.
Say we have a P'^c_∞ that gives .95 & .05 to 0 & 1 resp.
Actually, for all t. 2^ℓ seqs of length ℓ, for large ℓ, P^c_∞ must give values arbly close to .1 & .9 for at least one $A^{(\ell)}$.

Just try any Pam defined by a essentially UMSM & show that t. UUOM machine can't deal w. it well. I think I did show somewhere that there exists at least one B limit Pam, in which ① t. limits all exist. ② They are not computable ③ they are not representable by cB's on a UUOM.
Maybe not — I did show that there are functions "computable" by UMSM's that are not computable by UUOMS. (463.35 – 464.10.)
for this look at 3.37ff. In fact, go back to those discuss, for ideas on t. present problem.

Conjecture: UMSM's can compute functions that UUOMS can't ← (463.35ff)! ∴ assoc. w. each such function, should be a limit Pam assoc. w. t. UMSM, that UUOM's can't deal w..

{An analog is W's Thrm **15**, in which if a FOR, R', can't cover another FOR R, then $P^{(R)}(A^{(\ell)})$ is arbly > $P^{(R')}(A^{(\ell)})$ for at least one $A^{(\ell)}$ ← (This follows from t. Thrm 12, but isn't Thrm 12, itself)
455.35ff has some remarks on t. possl. falseness of this}

← This is certainly false: $P^{(R)}$ & $P^{(R')}$ can be identical, yet R & R' need not be able to cover one another — e.g. they can have difrnt. input alphabets!

But say R' can cover R: that R' can generate a function that R can't! i.e. ∃ a seq., "a", ∋ R'(a△n△) (△ is punctu., n is any finite no.) will calculate F(n) and then stop; yet there is no such "a" for R — so R can't calculate f(n).)

36.01

12374-R

01:35.40

(Here R is a UUOM, R' is a UMSM)

TM36

R generates U's for R(aono), but for any input to R that gives a non-U output, R' gives t. same output w. that input.

Suppose that F(n) increases more rapidly than any function computable by R — (& I think that such funcs. exist — i.e. 463.35 may tell how to generate such a function). — Then this means that of the codes that R' can interpret but R can't, some of those codes add prost to t. very long $A^{(\ell)}$ (i.e.s large ℓ) much more rapidly than any codes of R could.

Hvr., note that R has a bunch of finite inputs w. infinitely long outputs. For R' to add to t. prost of those outputs, would probably not ↑ their prost by an arbly large factor.

The function F(n) is able to give relatively large prosts to very long corpi. Hmm. It is able to do this to a greater extent than any single non-U input sequence for R.

There are at least 2½ troubles w. this discn:

.22 1) Those finite inputs to R w. infinite outputs. We must be sure that there are an ∞ of F(n)'s outputs that are difnt. from those in t. waychay start.

.24 2) Even if F(n) is able to bestow prost on a long $A^{(\ell)}$ using much smaller values of $\log_2 n$, then there are short codes for R, with length < $\log_2 n$ — it is still possl. that this $A^{(\ell)}$ will get lots of prost from R via very many, very long codes for this $A^{(\ell)}$. Hvr.,

.28 ② Note that argt. ② would be mitigated if we had an upper bnd for (krost − brost (of any seq.)) for each UUOM, characteristic of that UUOM.

~~3) There are also a large no. of~~

Specifically, other than t. diffy of .22, if b is one of R's codes for some long $A^{(\ell)}$, then R has no codes for $A^{(\ell)}$ much shorter than b is — If this were true (i.e. krost − brost were bnd), then I think this would dispose of diffty 2) (=.24).

.33

I'm assuming that F(n)'s exist which ↑ arbly faster than any of R's functions.

① → Perhaps to help w. .22 — show that for every such finite code w. ∞ output, I can find at least 2 functions for R' that ↑ arbly rapidly.

12474 R

01:36.40: SN) While the value of $P_i(A^{(\ell)})$ as a function of i, does not approach its limit with suddenness, I Think that each bit of t. binary expansion of $P_i(A^{(\ell)})$ mite. (does!) This may make some of my old proofs [redacted] based on t. "suddenness of convergence" patch-up-able. TTM37

Well, .01 looks pretty good! In t. "untractable seq." problem, all I need to do is compare t. first 5 (say, or any fixd finite no.) signift. bits of ~~[illegible]~~ $P_i^c(A^{(\ell)}{}^\frown 0)$ w. those of $P_i^c(A^{(\ell)}{}^\frown 1)$. This comparison cannot oscillate indefinitely/ as $i \uparrow$ — the first (unnormzd) 5 bits of these nos. can change at most, 32 times! (well, not necessarily — if t. first 18 bits are zeros, t. signift bits

18 can change ~ 2^{18} times! — But in all cases they can change only a finite no. of times.

Note that t. most signift. bit for $A^{(\ell)}$ can be at worst, the ℓth bit → Woops!! → 39.25

If .01 ff is ok, Then I do have a untract. seq for UUOMS.

.20 (Good! ☺). Then t. argt. of 36.22 + .40 makes it somewhat more likely that UMSM's mite do signiftly better than UUOM's

.25 in predicting Fdss's. On t. other hand, it would seem that UMSM's could only do finitely better than UUOM's, since UUOM's are only boundedly worse than t. original generating fcn of a fdss! — So things are certainly

.28 a bit confused!

→ 41.40

Perhaps I can use t. "paradox" of .25 ff to "go around t. back way" & prove some other things (e.g. the finite or infiniteness of ~~[illegible]~~ cost - boost for UUOM's.).

→ Take a look at my ~~old~~ old [redacted] construction of t. "untractable seq" for UUOM's. I may have usd .01 ff — but I just forgot about it recently! See 5.17 ff → later Apparently I was not at that time aware of this difty — but t. idea of t. nth bit of $P^{(Muuom)}(A^{(\ell)})$ converging suddenly certainly is an old idea — I ~~[illegible]~~ think I recognized this when I realized that $P^{(Muuom)}$ exists as a limit of $P^{(R_T)}$; for $T \to \infty$.

12474

37.40: On t. kcost − bcost "thrm" ; Perhaps use t. idea in Willis' ~~letter~~ letter in this way : T. trouble I had using his idea was specifying the C.B. Jan 2, 74 TM 38

1) It *may* be possl. to "specify" t. C.B. — even tho it's a non-computable no. (like t. time it takes for t. n^{th} bit of $P_{\infty}(A^{(\ell)})$ to converge (for any ℓ : (ℓ is irrelevant)).

~~an~~ 2) Actually go thru t. construction in thrm 5 ; ~~[illegible]~~ exact knowledge of t. C.B. may be *unnecy.* — One *may* have to modify t. construction a bit, hvr. Perhaps I can tolerate temporary uncertainty about which of 2 sub corpi has greater pcost s by using t. method of 37.01 − .18.

.18 My impressn. is that it may not be hard to show that ~~[illegible]~~ kcost − bcost is *bndd* for *UMSM's*: ~~[illegible]~~, ~~[illegible]~~

I *think* $\left(\begin{matrix}\text{VHM}\\ 12.30\end{matrix}\right)$ shows this. Essentially, from a ~~[illegible]~~ UUOM, ~~B~~ M_U, we construct t. sequence of FOR's $R^{P(R_T)}$ where R_T is t. ~~[illegible]~~ T^{th} C.B. on M_U. So, assoc. w. M_U, is t. MSM, $R^{P(R_\infty)} \equiv M_M$

To [kcost of a ~~[illegible]~~ string via M_M] is ~~=~~ [(bcost of that same string via M_U) + k]

where k depends on M_U only & is indip of t. string involvd.

We *may* be able to do t. same kind of construction, starting with a UMSM, M, then have R'_T be t. T^{th} C.B. on M, then the MSM $R'^{P(R'_\infty)}$ will be a MSM ~~[illegible]~~ M_M', that M can simulate. So for any UMSM, M, kcost − bcos is bndd.

This "proof", hvr., *does* need some cleaning up.

39.40

12474 R

TM39
TM39

.01 : 38.40
29.40 spac → So, anyway, It looks like t. UMSM limit is certainly adequate for Induction on all Fdss. So this form of CMI is a nec. & suff. form. ~~There is, of course~~
UMSM's seem to be of some interest in themselfs — e.g. Th. kcost-bcost bnd thrm. (38.18) — t. untractable seq. wrt. UUOMS (5.17 & 38.01) —
The suggestion in 37.20-.25 is to just how UMSMS mite give hyar pcosts than UUOMS, using t. very rapidly ↑ functions accessable to UMSM's but not UUOMS.

.14 → 40.01

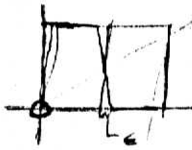
.15 : 29.19 11) If kcost - bcost is bndd for any one UUOM, it is bnd for all UUOM's.
" " " " unbnd " " " " " " un " " " "

12) kcost - bcost is bnd for all UMSM's (see 38.18-.40 & 12.30ff).

13) Define MSM's; show that a UMSM exists, show that a UMSM can compute funcs that a UUOM can't. Link UMSM w "term ?" 24.11

.20 What I want to do now is really get all these proofs, defns, etc. in good, vigorous form - Go over them carefully — particularly t. most recent ones (i.e. stuff on untractable seqs., kcost - bcost bnd for UMSMS, etc.)

.25 : 37.18 : 37.01-.18 assures that $\frac{P_i^c(A^{(l)}\frown 0)}{P_i^c(A^{(l)}\frown 1)} > 1$ (say) w. ε accuracy — but ε is always > 0. This would then probly have t. same difty as my



type function of 31.20ff. Well — perhaps not. In 31.20ff, t. problem was that I wasn't sure that $f(\ )$ would → 0 or 1 as $i \to \infty$; — in fact it was quite possl. it would → ½, thus yielding nothing of any use. In t. present case, hvr, we always end up w. a zero or a 1. If t. first 5 (say) bits converge so that $P_i^c(A^{(l)}\frown 0) \geq$ t. first 5 bits of $P_i^c(A^{(l)}\frown 1)$, then we decide on $A^{(l)}\frown 1$ for t. continn. of our "untractable seq."

Note that "almost all" seqs are on $P_\infty^c(A^{(l)}) = \frac{1}{2} \pm$ a very small ε for just about all l. I.e. - most seqs are "random" w = probly for 0 or 1. Any "non-computable" method of dering. a seq. will then probably yeild an seq. that is "untractable" to UUOM's.

40.01

12474 B

.01 39.40 : Actually, I was worried that this report would have too 39.14 spac little material in it, but the 13 items from 23.18 to 39.20 are really a lot of material — most of it rather exact theorems of some import — & the rest, tho more conjectural (the 4 items in 8)th (24.27) are of much more direct importance for practical induction. TM40

Furthermore, the stuff on induction w. UMSM's is a "necessary" soln. to t. induction problem. — superficially at least, it seems like an adequate "theoretical soln." All I've been able to show, hvr., is that 1) it is a v.g. soln. 2) No "limit type" soln. is better. I have not been able to show that there are not other, simpler, solns. that are about as good — e.g. t. UUOM limit pcm.

What I want from this review Report is partly for myself. — a clear statement of hard results, w. proofs, then a statement of impt. conjectures & open problems, & indication why these open problems are impt.

There are, hvr, impt. topics on CMI that I'll not review in this report — e.g. linear regressn., operator & gramm induction, applicns of CMI to "clustering", optical pattern recogn., etc.

One of t. impt. ideas of this report, was to get other people working on CMI, because there were many Q's that I didn't understand. Hvr., in t. past 6 months, I've cleared up enuf myself so that it really isn't so essential to get other people to work in this area! There are still t. "conjecture" statuses of several impt Q's (24.27, e.g.), but I can probably work on practical induction w. t answs. to them pretty vague!

43.01 spac
41.01

12574 — R

01: 40.40 :
37.28 spec. → Trouble occurs only if (relatively) hy prests are added to fdss's whose brests ↑ rather rapidly w. l (i.e. are like $\tilde{\propto} l$).
(this doesn't include non-deterministic seqs, but includes seqs w. several "random" coiffs & perhaps even an inf. no. of "random" coifs.)

Anyway : fdss's constitute an almost negligable fraction of all seqs. (I believe)
— So it is quite possl. that the "short codes for very long seqs", generated via F(n) (~36.01 ff) may never add anything to these fdss's.

There is, hvr, a big Q about t. rigorous basis of "t. paradox" of 37.25.
There is an upper Bnd for how good any pcm can do wrt t. Expected value" of t. prest error (or bcost error) in prediction — wrt to t. "true" pcm that would produce a gn. ensemble. I.e. t. pcm that produced t. ensemble is Best. — This does not say anything about t. bcost of any individual seqs. So this whole paradox must be put in more rigorous form — if it is to be regarded as a paradox at all!

Also, the Bounded ness of (kcost − bcost) (36.28, .33) looks like another impt pt. in t. paradox. Perhaps this paradox constitutes a disproof of This !

.28 Actually, t. situation is even more complicated! I had assumed that for Most CPM's, that all corpi $A^{(l)}$ were gn. a brest $\tilde{\propto} l$. In genl, this is not so. A CPM can assign. probys arbly close to 1 — e.g. $P^c(A^{(l)} 1) = 1 - 2^{-l}$ for all $A^{(l)}$.
This makes t. seq. $1^{(l)}$ have a bcost that grows very slowly w. l.
By letting this $= 1 - f(l)$ where $f(l) \downarrow$ arbly rapidly w. l, we can get t. bcost of $1^{(l)}$ to ↑ arbly slowly w. l.
— But — no matter how slowly this ↑ w. l, F(n) (of 36.01 ff) will ↑ so fast w. n, that t. brest will be modifyd by the F(n) addition. by an arbly large amount. So argt. .28 is irrelevant.

If bnd kcost − brost is true, then 36.01 ff constitutes a proof of existance of a untractable seq. for uvom's — unless for every value of F(n) ∃ a finite dern. of an infinite seq. that 42.01

.01: 41.40: 'start out like F(M), but has a deriv of length < n.

Even if 41.40 — 42.01 were true — would there be any paradox to explain? Th. main paradox involves CPMs: UMSM's would somehow assign greater probty to certain seqns than t. CPM "that generated" that seq. — (if this is at all meaningful) — & somehow this is impossl., because t. "generator" of a seq. gives as hy a proby to t. seq. as possl.

Def | SN | LPM Limit Proby Measure: A proby measure defind by

$P_\infty(A^{(l)}) \equiv \lim_{i\to\infty} P_i(A^{(l)})$ — where t. P_i are all CPM's.

12574 R

TM43

01: 42.40; 40.40.Spec → A sort of new Abstract or introduction to r. report;

Title: A Unique Formal Solution to the Problem of Inductive Inference.

*

The Inductive inference problem is first defined to be the extrapolation of a (usually long) sequence of symbols — this sequence having been created by some unknown stochastic source.

~~A formal solution is displayed~~ It is well known, that ~~there~~ if the generators of the sequence is allowed to be any finitely describable stochastic source, then there is no effectively computable soln. (in t. sense of Minsky, Davis – etc.). The solution ~~displayed~~ that will be given, is the limit of a sequence of effectively computable ~~an~~ approximations. ~~The solution~~ While the limit exists, it is not, of course, effectively computable.

It is shown that this solution ~~yields~~ gives probability values that converge rapidly to the ~~correct~~ values generated by the ~~underlying~~ unknown stochastic ~~seq~~ source. Three ~~Two~~ kinds of error criteria are considered. ~~The first, "Information error" was devised by Willis and is ~~the~~ the more stringent of the 2~~

The first is the total ~~so~~ of the squares of the probability error. This sum is converges and is bounded by a constant that can be estimated if the nature of the stochastic generator is known.

I'm not sure this is what Willis defined by that word

The second is a more stringent error criterion devised by Willis, called "Information Error", having, however, less direct intuitive meaning than the first type of error.

→ A third kind is ~~named~~ proposed, and some strong conjectures are given as to how to evaluate it, if the ~~nature of~~ underlying stoch source is not known.

It is shown that there is no limit of probability approximations that is better than the one given.

This particular solution to the induction problem is compared with that of Willis. ~~Willis' solution~~ The present solution is at least as good as Willis' in all cases — tho/usually it is ~~the~~ more difficult to ~~to~~ ~~approx~~

44.01

1257 ~~4~~ K

01;43.40 compute the approximations. **For** certain kinds of (pathological(?) / untractable) stochastics sources, the proposed solution is far better than Willis' — but for stochastic sources that actually appear in the Real world, it is not clear ~~as to whether the proposed solution~~ it is significantly better. ~~[illegible]~~ TM 44

~~Some problems are the involving~~ The application of these induction techniques to ~~[illegible]~~ practical problem is discussed first in terms of the adequacy of ~~a sequence of~~ symbol extrapolation as a complete formalization of the induction problem —

Not so good to use in abstract 14 Secondly, in terms of using the proposed solutions to 15 solve problems in general pattern recognition, numerical series extrapolation, clustering in pattern recognition techniques, ~~[illegible]~~ geometric probability, ~~[illegible]~~ human .19 concept formation, etc

Say: instead of .14-.19: ~~to solve specific prob~~ Secondly, in terms of examples of how the proposed solution ~~would~~ can be applied to specific induction problems

46.01

45 13 in SVH

.01: 44.40 : On T. defn. of MSM's : I'm not so sure that I really have a rigorous defn. It was first defind on 461.25

463.35 : How a MSM can "compute" a function incomputable by a UUOM.

11.02–.20 || 3.01 : 2 types of MSM's.

3.37 : perhaps proof that there are seqs of P_i's ∋ no set of CB's on a UUOM can express them.

5.17 : an untractable seq. for UUOM's.

| 7.02 : Ont. Defn. of a MSM. Also see 8.17–.20 |

8.17–.20 probably has most of t. defn. I've been using.

At present, My impressn. of a MSM:

For every infinitely long input seq., it will eventually converge on an infinitely long output seq. If t. machine (stops), it is regarded as having (a) loop following its ~~last~~ furthest out printed symbol. (Gaussian rule compu. loop)

This means that for every n>0, ∃ a f(n) (uncomputable) ∋ by t. time t = f(n) the nth output seq. will have convergd. ($F(n)<\infty$) for any particular n, But F(n) may → ∞ as n → ∞

For a finite input seq., I don't know what to regard as its output. I could regard a blank as a special input symbol, so even finite input seqs are really infinite. Also we can make ~~a special rule~~ a special rule, so that if t. machine reads a blank on t. input, it knows that all the following squares are blanks.

Hvr, - note that for a finite T compu bnd., only inputs of length ≤ T are readable — so we normally consider longer inputs as CB↑.

So — if we consider inputs of length = T, then we don't have to consider blank squares on t. input tape.

But, if we always have all inputs be of length T, for C.B.T, — How do we get our pcosts to be based on input code length?

.30 Consider we have a LPM, $[P_i]$ & we are devising a MSM, M ∋ $P^{(M)} = P_\infty$. We want t. follg I,O behavior of M:

Say we have as input, t. finite string s, of n bits, followed by Blanks. Using W's thrm5, we first print out $R^{(P_1)}$'s output w. s as input (i.e. t. response of $R^{(P_1)}$ to s ; $R^{(P_1)}$ being t. FOR assoc. w. P_1)

Next t. output is changd to $R^{(P_2)}$'s response to s. . .

. . . to $R^{(P_\infty)}$'s response to s.

48.01 spec
47.01

01:46.40: on ${}^{\text{th rm}}$) (24.11) This does not show that p(UMSM) is as good as one can get. It merely shows that for any pcm of t. form that is $\lim_{i\to\infty} P_i \equiv P_\infty$ p(UMSM) will do as well, within a constant bcost.

.04 I would like to do a little better, & show that, UMSM, M_{UMSM} has if P_∞ bndd bcost error for all limit pcms, then $\exists$ a UMSM, $M_{UMSM} \ni p(M_{UMSM})$ is identical to P_∞.

While .04 is very probably true, I don't think that for practical purposes it is much stronger than 24.11. i.e. If one uses a randomly chosen UMSM to do induction, both .04 & 24.11 would say that one will get a bndd bcost error.

Hvr. .04 also says — if you use any pcm that gives one a bndd bcost error in these circumstances, it will be basable on a UMSM.

So, One way to prove .04: "Clearly" a pcm that will get bndd bcost error w. all limit pcms, must be a basable on a MSM. T. Q is — must it be a UMSM that its based on? If not, then there must be some MSM that it cannot "cover".

Prove: It is a UMSM || Hypoth: its not a UMSM (try to prove false)

If it is not UMSM then $\exists$ a MSM it can't "cover".

Now it is possl for $\underbrace{p(R_1) \text{ to be } \geq \frac{1}{k}\, p(R_2)}_{\text{which is t. precont situation}}$, yet R_1 need not cover R_2 ...

T. forgg. approach is incorrect if followed as above, hvr., I'm not sure that there isn't a way to prove it using these sorts of args.

Another entirely dif'nt way to prove .04 is that of ~ 46.14 ff. — this is a constructive way to get t. desired UMSM.

TM48

.01: 47.40 → If we have an infinitely long input, S, then
46.40 spec changed to ~~[illegible]~~
we first do $R^{(P_1)}$'s response to t. first bit of S, T. outputs is then
$R^{(P_2)}$'s " " " " 2 bits " " " " " "
. . . . $R^{(P_i)}$'s " " " " " " i " " " ".

.05 In response to t. finite input, (if P_∞ exists (i.e. $P_i \to$ a limit)) M keeps changing t. outputs until at a certain point, the output changes no more, but M continues to calculate. We are not, hvr., ever sure that M will never print again.

.10 → A notable thing, is that if the ~~response~~ limit response of M to input S⌢Blank is A, then t. limit response of M to S⌢(some input sub string) is A⌢(some output substring). In this sense, M is a "Mildly Sequencial Machine".

By definition $p^{(M)}$ (if M is a MSM), is obtained by t. limit of th. ~~[illegible]~~ LPM, $[p^{(M_i)}]$, M_i being t. i^{th} C.B. on M.

So, in general, I think a MSM can respond to a finite string (followed by $(\text{Blank})^{(\infty)}$) by a finite amount of printing, and then possibly an infinite pure (not printing) calculation. Its response to an infinite input involves reading part of t. input, doing some writing (& perhaps re writing ← tho this is probably unnec.) then reading more & writing more & perhaps re writing some — then read some more, etc.

For finite input, a UOM can have inf. output. So, in a ~ way, for finite input followed by $(\text{blank})^{(\infty)}$, a MSM can print infinite output.

T. constraint of .10 looks like it may be signif., but it need not be if we never give it finite inputs!

.35 → Another impt. idea I haven't gone into: I considerd various CB's on a MSM to be FOR's. In general, I don't see how this would be so — unless we only allow certain kinds of CB's, perhaps

13074— R

TM49

01:48.40: It may be possl. to get CPM's from machines that are not FOR's: we can use Lebesgue Measure — as W does.

.02 Say we use a T limited MSM. These are all a sub-set of FSM's. ~~Each~~ maps
.03 t. set of (all inf. input strings) into a set of finite output strings, so we do have Σ of all probys = 1 (including outputs w. S & U in them) — but in t. present case, we have not U's or S's — we just have these finite output strings. T. machine can be in a compn. just after having written t. 8th. symbol (t. 9th & 10th have been written, already) when t. T limit occures — or — it can stop after rewriting

Actually, we need consider only 2^T difrnt possl. inputs

Def: we do, hvr, have th. "B" (≡ Blank) symbol, so B(∞) follows all the finite strings.

.110 t. 8th symbol — in either case, t. output is identical.

Look at 46.30-.40 & 48.01-.05: This decrbs t. response of a MSM that is "simulating" a particular LPM, P_{00}. If we put any CB, T, on such a machine, it will give fine Proby measures if T is ∋ t. machine has just finishd rewriting a complete $R^{(Pi)}$ output. If
.20
T is ∋ t. machine is only partway thru such an output, then I
.22 don't know if we get a Proby Measure.

.23 → well! These CB's on a MSM may not yeild proby measures — this will be because of t. lack of "sequenciality" in t. outputs. Hvr, ~~~~ for large CB's most of t. output will be sequencial in t. proper sense (like a FOR is). For large CB's — ~~t. random pr~~ any t. first n symbols (for fairly large n) will have proper probys assoc. w. them — furthermore, these probys will not get into t. difty of .20-.22, because these early output symbols will not change as t. C.B. ↑!

So: when t. CB's on a MSM do not yeild a "true Proby" — this will only be for t. latter ~~~~ symbols of t. output. By suitable xfmns, these number assignments to output strings can be changed so as to yeild numbers that do conform to t. conds. that proby measures conform to. Only t. latter symbols will have these nos. changd, hvr, since t. nos. for earlier ~~~~ symbols will be problistic, since these earlier ~~~~ symbols ~~~~ have a suitable "sequencial property".

50.01

13074 R

TM50

01: 49.40 : Ok.: So now go thru this lack of sequenciality — say, as decrbd in 49.20—.22 — decrb. in detail just how it occures — & just how it does not occure in t. earlier part of t. output that has "settld down".

I think it very likely that 49.23 ff will patch up this trouble.

getting proby measures out of MSM's!!

.10 Renormalizations, using t. constraint $P(A^\frown 0) + P(A^\frown 1) = P(A)$, $\sum_i P(A^{(i)}) = 1$ may be adequate. It may well be that if one as any ~~numerical~~ (non-zero ?) numerical function on all finite strings, one can renormalz via .10 & obtain a "proby measure"!

— A zero pre normzn, ~~will~~ for a gn string A, will give zero normzd proby for all strings of t. form $A^\frown\alpha$, where α is any string (including t. null string).

Negative pre normzn probys would probly give no trouble — but look into this! ~~...~~ sometimes if t. follg normzn rule is usd! It will be o.k.

say $P(\ )$ is pre normzn

$P'(\ )$ " post "

.25 Then $P'(A^\frown 0) = P'(A) \cdot \frac{P(A^\frown 0)}{P(A^\frown 0)+P(A^\frown 1)}$

$P'(A^\frown 1) = P'(A) \cdot \frac{P(A^\frown 1)}{P(A^\frown 0)+P(A^\frown 1)}$

This is an adequate recursive normzn rule. We also need to know that $P'(\text{null seq}) \equiv 1$.

From a more practical view pt., one can save trouble renormzing, if one only uses (unnormzd) proby values that have been constant for at least t. 2 most recent CB ~~#~~ increases. (This is w.r.t. t. MSM's refered to in 49.20-.22 i.e. 46.30-.40 48.01-.05)

So, wrt any particular CB on a MSM; we can, using 49.03 obtain a "proby measure" (positive semi definite) & using .25 we can renormz if necy. Thus we can get CPM's from CB's on MSM's.

Next, we define t. PM assoc. w. any MSM as being t. LPM assoc. w. t. seq. of CPM, assoc. w. t. successive ~~#~~ "T", CBi's on that MSM.

From this, it is clear that we need ~~not~~ never consider finite inputs to MSM's — i.e. we don't have to define what MSM's must do when they see t. blank at t. t. end of t. input string. This is because we are only interested in them because of t. LPM's assoc. w. them — & in obtaining ~~probys~~ these limits, we only use

51.01

15

01: 50.40: inputs of Length T, & T → ∞ (49.03) TM51

As for t. CB's assoc. w. MSM's — we again only use T-type CB's — so we consider only inputs of length T, — so again we don't have to worry about ~~t. ma~~ how t. machine will react to a blank square.

Def: ~~Go~~ Going back to $\boxed{\text{defining a MSM}}$: 49.02 — .10 describes behavior of Time computation Bnds T CB's on MSM's. Let us insert an inf. string into a MSM.

It will, typically, continue computing & printing, forever. It can, hvr., stop (← A "stop" can be regarded as an inf. compn. loop type pure) at any time, or go into an infinite/ compn. loop (& no printing) at any time — thus printing out a finite string. ~~The defn. is, hvr., that if it does~~ In either case, it would satisfy t. defn. of a MSM, since each output symbol 0, 1, or B would → a limit.

If it ~~[illegible]~~ never stops or gets into an inf. compn. loop, then we require that every output symbol → a limit eventually. This means that each output symbol cannot oscillate indefinitely — but must eventually (often at a not e.c. time) settle down to 0 or 1. If it never stops or gets into inf. comp loop, I think it ~~[illegible]~~ must print an ∞ of ~~[illegible] B symbols~~ non-B symbols, — so it can have no B's on its output tape.

The "Sequencial property" of MSM's arises as a trivial corrolory ~~[illegible]~~ from t. defn.; It is clear that for any fixed inf. input, then ~~[illegible]~~ eventually, every output symbol up to t. nth will eventually converge. (i.e. $\exists\, T_0(n)$ so large that for $T > T_0(n)$ the first n symbols of t. output will remain invariant.) ~~Usual~~ In most cases of interest $T_0(n)$ increases more rapidly than any e.c. function.

.30 If it should become necy to define t. MSM response for finite inputs, then look at 46.30 - .40; 48.01 - .05. → 55.01

T. mode of analysis of ~~[illegible]~~ MSMs of 49.02 ff, using T CB's w. only inputs of length $\geq T$ (Length T may be adequate.) can be applied to UUOM's to yield a ~~[illegible]~~ simpler (less backgnd needed) development of Willis' paper. Hvr., T CB's are certainly less general than t. CB's he treats & will perhaps ~~[illegible]~~ make certain of W's thrms. ~~[illegible]~~ inaccessable.

55.01 SAC.
52.01

13174 R

TM52

.01 51.40: Perhaps it would be well to reiterate t. difference betw. a goal. FOR & a FSM & a TCB [By FSM I mean unidirectional I,O tapes but ~~no~~ work tape.]

1) FSM & FOR's do not include one another either way.

2) A squaring machine is an FOR but not a FSM — since it needs potentially infinite Memory. A permissable input to a squaring machine is a no. followed by a Δ (punctn. symbol).

3) A FSM can get into infinite/compn loops: a FOR can't — but I think a FSM that can't get into inf. comp. loops is a type of a FOR if t. S & U symbols are added.

4) A copying machine is not a TCB, it is a FOR & FSM.

5) A TCB is a kind of FOR.

6) I don't think a squaring machine of this sort is a FOR. One that is: Input is string of n 1's followed by a 0. Output is a seq. of N^2 1's, then t. machine stops. If no a blank follows t. input seq of n 1's, t. machine prints $1^{(n^2)}$ then stops also. If a 0 follows, it stops permanently; if a 1 follows, it prints 2n+1 more 1's, then stops.

Or t. machine may print (in radix 1) t. ~~total~~ total no. of 1's appearing in t. input thus far.

~~TM~~: In ~~sim~~ simulating a FOR assoc. w. a gn. CPM [when we are deriving a MSM to simulate a LPM, LP_i] (i.e. the P_i of the LPM), our output could get into a U or S loop. We can simply have t. simulation stop rather than print U's or S's. This will result in a non-normd PM, but that's O.k., because we renormalize anyway.

53.01

2.1.74 R

TM53

.01:32.40: I may want to issue a prelimy. Progress Report, for my own use as well as that of various correspondants:

~~Then~~ In the report, give a general statement (introdn) of t. way things look, then list various thrms. & tell which seem provd at t. present time.

Thrms: 23.18ff!

1) Define LPM's.

2) Define $P^{(Muvom)}$ & show it exists.

3) $P^{(Muvom)}$ has bndd brost error for all CPM's.

3.5) If t. set of possl. Pams in a known complexity class, then ∃ a CPM that has bnd brost error.

4) If a Pam has ~~bndd~~ brost error of k, for a gn. CPM, then $\overset{E}{}(\Sigma\ \text{err}^2)$ is bnd by $k \ln 2$ or whatever consts.

5) If a Pam has $\overset{E}{}(\Sigma\ \text{err}^2) < k \ln 2$, then its brost error (or expectd brost err) need not be bnd.

6) $P^{(Muvom)}$ does not have bndd brost err for all LPM's

Give example.

6.5) put this here: discussion of realizability then that thrm, then t. discn of 7)

7) a) Define MSM's

b) Thrm: UMSM's exist.

c) Define $P^{(MMSM)}$:

d) Thrm: $P^{(MMSM)}$ exists

e) Thrm $P^{(MMSM)}$ has bnd brost err for all LPM's.

see 95.30 - 96.10 for bibliog on Proof — tho there may be far better refs.

.30

---

conjectures, disproofs, partial thrms, remarks:

7.5) Say P_∞ is a LPM that has bndd brost error for all LPM's. Then ∃ a UMSM, M, ∋ $P(M) = P_\infty$. Probably true.

8) State SVH thrm: That chosing "best" Pam (∋ brost of Pam + brost of corpus wrt that Pam is max) will yeald result as good as $P^{(Muvom)}$

There is no "Best" CPM in this sense.

Is trivial. — This plan yealds $P^{(Muvom)}$ identically!

9) Kcost - brost is either bndd for all UVOM's

unbnd " " "

I don't know which.

54.01

R

TM 54

10) For UMSM's, I think Kcost - brost is bad.

11) The 3 thrms (unprovd): a) E(brost of next symbol) is

$$\frac{\text{brost of Pam + brost of corpus wrt Pam}}{l\ (= \text{no of symbols in corpus thus far})}$$

b) How to divide brost of Pam into an apriori (zero cost) part & a apsi (regular cost) part.

.12 c) Argt of using brost here rather than k cost;

i.e. for 2 pams, P_1, P_2, ; P_1 has Kcost c_1

P_2 has 2 codes of lengths c_1+1 ; I think P_1 & P_2 are equally good.

12) Discn. of rcpms, cpm's rates of brost convergence of difrnt. pm types (see 29.07) — various stochastic model types.

13) Discn. of relative goodness of p(Mumsm), p(Muuom), cpm if cpm to be derd is of known complexity class, etc.

14) Discn of use of various umc's for proby evaluation — If the umc is chosen before t. sequence is seen (or chosen in a way that is likely to have zero correln. w.r.t. seq) then t. expected error in predn. may be very small. (?) Certainly there will be no a.h. error or "systematic error" (whatever that is in t. present case). The "A" in $\frac{f(s)+A}{s}$ will be zero.

85.01 spec.
55.01

57.40
.01: 51.10 spac: Some Problems Assoc. w. having no defn. for output of MSM's. TM 55
51.30 spac.

Finite input: 1) Lr b cost - b cost is bad for MSM's I must have a defn. for MSM's output for finite input! (Note 56.27)

This defn. may not be so diff. The MSM must be capable of reading blank squares & reacting to them. The blank can be regarded as a regular input alphabet symbol. However, since a blank is always followed by a blank, there is never any reason t. machine would want to read t. symbol followed by t. first blank — so we mite use this to limit t. form of t. Turing finite state transition rules.

As for t. rest of t. MSM's behavior it would be restricted as
.14 befor, so that it would have to eventually converge for t. nth
.15 output symbol — no matter what value n has.

So, w. finite input, a MSM could compute forever, compute and print forever, or eventually stop, subject only to constraint .14-.15.

.19 We also want a more genl. defn. for a c.B. on a MSM. So far I've only been using a TCB. We want a more genl. cB Because: We want to use a set of "stronger" cB's on a UMSM as t. generators of a LPM. We claim that LPM's based on cB's on UMSM's are ~~more~~ powerful than those base on cB's on UUOM's — but this claim is very narrow, if cB's on UUOM's yeald a rather broad class of cPM's — but t. cB's on UMSM's are narrowly (Not so Narrow — see 56.32, 57.01) restricted to TCB's.

Some propertys we'd want for a c.B. on a MSM:

1) Its output (in response to a gn. input) is exactly t. same as t. MSM its t. cB of, except at a certain point it does a new & difrnt thing — so one can easily tell where t. simulation of t. MSM stopd. For t. CB to stop at that pt. would perhaps be an adequate "identification of behavior"

2) We would like t. behavior of t. c.B. to be "e.c." in t. sense that several impt. Q's about its response to various input types (finite input, certainly) are e.c. E.g. T. PM assoc. w. a cB is e.c. — So its very likely that we don't want a cB of a MSM to be as general as a MSM can be.

3) Perhaps for 2): for every finite input string, t. output is e.c.

b) (I'm not sure I need anything so restrictive, but): for every l, $\exists\ n(l)$ $(n(l)<\infty)$

56.01

01: 55.40: ∋ for every output string of length ℓ ∄ there are no input codes of length $> n(\ell)$. ← [I'm not sure this is meaningful.

I still don't think t. follg. is enuf.

I may need t. follg. for CB's on MSM's, as well as a modifn. of it for unrestricted MSM's:

For every infinite input string; for every ℓ, $\exists\ n(\ell) \ni$ symbols of t. output string up to the ℓ^{th} are indip of t. nature of th. input string after t. n^{th} output symbol. — So, t. finite part of t. input string up to the n^{th} symbol, yields t. finite part of the output up to t. ℓ^{th} symbol.

For C.B.'s on MSM's, $n(\ell)$ is always bndd — i.e. it exists.

For Full MSM, $n(\ell)$ must exist also, because if it did not, then for the ℓ^{th} output, there would, for every value of T, no matter how large, be a seq. that could change t. ℓ^{th} output symbol after T. I ~~think~~ this may means that t. ℓ^{th} output symbol need not converge — but I'm not sure — perhaps not:

.19 Consider t. MSM: Its first symbol is a zero, as long as t. ~~output~~ input starts out $0^{(k)}$ — as soon as an input 1 appears, it erases the 0 & writes 1 & leaves it that way.

T. Q is, is t. follg. a MSM? The first symbol doesn't oscillate. For every input string it will eventually converge to 0 or 1. It converges to 0 for input ~~0~~$^{(\infty)}$,

.24 — it converges to 1 otherwise.

I could, if I liked, define MSM's so that for every ℓ, $\exists\ T_0(\ell) \ni$ after $T_0(\ell)$, t. ℓ^{th} output symb. of t. machine converges.

Do I need this restriction? If I had it, would every LPM have a MSM associatable with it?

.27 It is not nary that t. Kcost of a finite corpus be definable for MSM's. Bcost of t. corpus, hvr, must be definable.

→ Note that in .19 –.24 the prost of $0^{(\infty)}$ is zero. I don't know if this sort of thing must always be true for all sets of sqs. of this sort. Well, t. seq $0^{(r)}1$ has prost 2^{-r-1}, T = wt. of all seqs

.29 of that sort is $\frac{1}{4}+\frac{1}{8}+\frac{1}{16}\cdots=\frac{1}{2}$ (?). 01; 001; 0001, ... → 71.01

.30 Every CPM is approximatable arbly well by a FOR. So if C.B.'s on a UBMSM are FOR's, then ~~perhaps~~ any seq. of

.32 CPMs that → a limit, is representable as a seq. of CB's on a MSM.

In fact, if we just use TCB's, this is true! — Since by using suitable values of T limit, we can obtain a seq. of CPM's from t. MSMS of of 46.30–.40, 48.01–.05.

What we really want, is to be able to devise various seqs. of CPM's, that can be regarded as CB's that "approach Universality UMSM" — so we are sure they approach t. strength of P(UMSM).

57.01

2274-R

TM57

01:56.40: The TCB on a UMSM will do t. desired thing, but it would be better if we could have a wider variety of CB's, so we could devise a wider variety of ~~seqs of~~ CPM's that approach P(Mumsm).

Hvr., it may well be that w. things as they are, using only TCB's on a UMSM, I am able to show that this seq. of CPM's is, indeed, stronger than that based on a UUOM, & that this new kind of seq. can "dcvr" any LPM w. bnd $\geq$ error.

For the present, I can leave it at that pt., & explain that probly ~~better~~ more general CB's could be devised, so more general seqs of CPMs could be used.)

(also (Θ) 46.30-.40 & 48.01-.05)

Note 56.30-.32: This means that any seq. of CPMs that converges (i.e. a LCM) is expressable as a seq. of TCBs on a UMSM. So any adequate "LCM" is expressable as a seq. of TCM's on a UMSM. NO! → Probly true, but I have not prov'd this.

But anyway: a set of TCB's on a UMSM is a fairly genl. kind of LPM — probably t. most genl. kind possl.

21 Hvr., suitable defn. of CB's for MSM's mite be obtaind by modeling them after FOR's, to some extent. 2 kinds of FOR's are (1) TCB — which works O.K. for MSM's (2) FSM (Finite Memory) w. t. additional constraint that t. machine cannot enter a state it has ever been in before, unless it has a difrnt input symbol, or is on a difrnt square of t. input tape (these latter restrictions prevent inf. loops — they may be too restrictive, hvr.) (3) More generally than (2); is "no infinite compn. loops" allowd. This isn't so easy to define, hvr. — A machine mite get into a write/rewrite compute loop that gets longer all t. time — (The u. MSM can't do this, by defn. — so a CB on it, can't). Or, it could converge in this write, rewrite, loop — but not in c.c. time.

59,60 →
CIAP

01: 10.35: : On t. construction (& data?) of a UMSM:

8.35 – 10.35 show how to construct a UMSM in which its output is defined only for infinitely/long inputs. We can take a finite input string & loop ~~and~~ s loop onto t. end (or a ~~#~~ Blank ~~#~~ ($\equiv$ B) loop). The B (or \$) symbol has this special loop property

~~\#\#~~ In 10.12 – .35 ~~Ana~~ we have F(i) & G(j) "going at difrnt rates". Will we ever run into t. problem that Schubert did? — I think t. relative rates of F(i) & G(j) are irrelevant — i.e. it is only nec'y that both i & j $\to \infty$.

Say i = j. Then for every i, let M(i) be either (a) t. length of G(i) or (b) F(i) — whichever is smaller.

If both L(G(i)) & F(i) $\to \infty$, then M(i) $\to \infty$.

If F(i) ~~→~~ is bnd (F(i) must be ~~...~~ a non↓-funct.), then M(i) ~~...~~ has t. same bnd. — I.e. L(G(i)) must ~~have~~ eventually get as large as F(~~∞~~).

So $\lim_{i \to \infty} M(i) = \lim_{i \to \infty} F(i)$.

I think t. idea of this M_s is that it will print null for a particular symbol, until it has a proof that that symbol converges — at which time it just does regular compns. for that symbol. ~~\#\#~~ M_s must therefor, converge for all symbols — either by staying at null, or by converging t. way M_v does.

An alternate way of doing it: Ms first starts looking for proofs of convergence, via M_p. As soon as F(i) gets > 0, it then computes G(i) — Then continues with M_p computing until F ~~\#\#~~ gets large, for F(j) say — at which point G(j) is then computed. So we go back & forth betw. M_p & M_v.

01:6140 A perhaps unreasonableness in t. adequacy of 1 Limit recursive methods for devry of all possl 1-lim rec. Pams!
Recovers t. set of all recursive (≡ r.e.) PM's (≡ CPMs) requires a hyper order — i.e. 1·Lim rec. soln.
It would seem that 1·lim rec Pams should require 2·lim rec. soln.!
There is, hvr, a sort of lack of similarity betw t. CPM's & th. 1·lim rec Pams I'm considering. Each 1·Lim rec. pam has a proof assoc. w. it (that it converges). Such proofs are of necessity, uncertain. CPM's have nothing corresponding to this.
Another ~~fac~~ factor: Regular CMI finds t. "proper pam" by a $\underline{2}$ lim recursive process. Each individual probty is found by a 1 lim rec. process. Perhaps ACMI (Augmented CMI) bears a ~ relation to t. "untractable" seq. for CMI, that CMI bears to all ~~[illegible]~~ Fdss. (Actually CMI will work for a greater class than Fdss — ~~[illegible]~~ e.g. random coifs — even in infi. no. of random coifs)

DEF

$\boxed{SN}$ CMI or ACMI may not be t. final soln. to t. induction problem! In RW prediction, one has an ≅ infinite amt. of data one can use, to bear on any particular prediction. Just how does one ~~[illegible]~~ decide what data to use when, in one's successive approxns to t. probty?

2.16.74

Re: ACMI! Perhaps it would be well to show that ~~[illegible]~~ if a limit recursive seq. of pams can solve Fdss's then ~~[illegible]~~ This lim rec. seq. can be represented by a machine that is "universal over all ~~[illegible]~~ Fdss's.
Essentially, what I want to show is that t. soln. must be at least as good as CB's on a UUOM. The idea here is that if someone proposes a soln. one can immediately test it for "minimal strength".
Its in t. direction of Generalizing t. Minsk-Papert result on "perceptrons"; i.e. to discover ~~[illegible]~~ what t. inductive limitations of a language (or a means of representation) are.
I think t. Pergg is more like what I had in mind for t. "thrm". about t. "nec. & suffy" of CMI.

01 15
70.40
56.29 spec

TM71

[redacted]M responds to $0^{(n)}1$ by ~~printing~~ 1 (even if $n=0$).
" " " $0^{(\infty)}$ " " 0.

Is M a MSM? If $0^{(\infty)}$ is inserted, M could never print 0, in any finite time. So, for $0^{(n)}1$, ~~M must~~ M prints 1 after time n; for $0^{(\infty)}$, it never prints.
So It is impossl. for M to print "0" in response to $0^{(\infty)}$.
So — for all inputs, I think t. first output sq. does converge.

M_1 responds to $0^{(n)}1$ with 0 if n is even,
" 1 " " " odd,
" no print if $n=\infty$.
we can use any 0 or 1 funct. of n that varies for large n.

Is M_1 a MSM? For every input, t. first output sq. does converge, but there is no max length of input that need be examined to determine this first sq. (in contrast to FORs).

SN T. soln. W. deebs is [redacted] proby not lim rec. solvable "in t. sense of Gold," — but is t. limit of a set of recursive approxns. to t. proby. Hvr., [redacted] the n^{th} digit of t. proby that W obtains is lim. rec. solvable "in t. sense of Gold".

T. present work on MSMs & induction is perhaps in t. direction of Blum's March 73 paper on t. Q. of what broad classes of seqs are extrapolable. What I've found is that t. class of extrapolable seqs is broader than had been ~~thot~~ thot — a result probly ~ to Blum's.

3 974 Brief Report! (Working Paper) (See 53.01–54.40) 7M85

54.40 spec

01; 71.40) A summary of some Recent Results in Induction Theory.

~~(Note: Unless otherwise noted, the terminology and notations will be that of~~ Willis 1970)

1) Let M be a Universal ~~Turing Machine~~. Let M_T be a

C.B. on that machine, that allows only computations of $\leq T$ steps.

work notation →

Let P^{M_T} be the computable probability measure associated with M_T,

Then for every ~~[illegible]~~ string of length m, $A^{(m)}$,

Theorem I: $\lim_{T\to\infty} P^{M_T}(A^{(m)})$ exists.

This is easily seen: $P^{M_T}(A^{(m)})$ is a non-decreasing function of T, and this function is bounded above by 1.

Call this limit $P^M(A^{(m)})$.

check w. Willis

This limit is never ~~computable~~ effectively computable.

$I^M \equiv -\log_2 P^M$ is called the "information in string" $A^{(m)}$

~~Note at beginning: The following are a set of results that are presented largely without proof — to give an idea as to~~

I will present here a set of apparently important results in the theory of inductive inference. Most of these results follow directly from the work of Willis (1970). I will use his ~~[illegible]~~ terminology and notation, unless indicated otherwise. Although it is necessary to read Willis' paper to understand the present paper in any detail, it is still possible to get some idea of the present results without doing so, and such a reading will, hopefully, motivate the reader to assay. In the present report, few proofs will be presented.

Willis' most important paper.

A ~~more detailed~~ paper giving more detailed discussion and proofs can be expected to follow soon.

2) Theorem II For any Computable Probability measure, P_0 and any string of length m, $A^{(m)}$,

$$I_0(A^{(m)}) \equiv \log_2 P_0(A^{(m)}) \quad \text{and}$$

$I^M(A^{(m)}) - I_0(A^{(m)}) \leq b + c$

40

87.01

see 101.01 -.10 for a more exact def n of b.

01! 85.40! Here b is the length of the description of P_0 with respect to M, and C is a positive constant that is characteristic of M. It is the length of a program that ~~tells M how to construct a FOR that corresponds to P_0.~~ (If b is defined by 101.05 -.10, then C = 0)

M can be chosen so that C is always 0. ← Mentioned on 90.10

Equation 86.40 suggests that for large values of m, $\frac{P^M(A^{(m+1)})}{P^M(A^{(m)})}$ might give very good approximate values to the value of $\frac{P_0(A^{(m+1)})}{P_0(A^{(m)})}$ — which is the conditional probability that $A^{(m)}$ would continue as $A^{(m+1)}$.

To see this, rewrite equation 85.40 as

$$\prod_{i=2}^{m}\left[\frac{P^M(A^{(i)})}{P^M(A^{(i-1)})}\Big/\frac{P_0(A^{(i)})}{P_0(A^{(i-1)})}\right] \le 2^{b+c}$$

[redacted]

and

$$\left[\prod_{i=1}'\left[\frac{P^M(A^{(i)})}{P^M(A^{(i-1)})}\Big/\frac{P_0(A^{(i)})}{P_0(A^{(i-1)})}\right]\right]^{\frac{1}{m}} = 2^{\frac{b+c}{m}}.$$

write it in this form

$$\prod_{i=2}^{m}\frac{\delta_i'}{\delta_i} \le 2^{b+c}$$

$$\left(\prod_{i=2}^{m}\frac{\delta_i'}{\delta_i}\right)^{\frac{1}{m}} = 2^{\frac{b+c}{m}}$$

where $\delta_i' \equiv$ ( )

$$\delta_i \equiv \left(\frac{P_0(A^{(i)})}{P_0(A^{(i-1)})}\right)$$

are conditional probys for P^M and P_0, resp.

for large m, the right side is close to 1.

The left side is geometric mean of the ratio of conditional probabilitys obtained by P_0 and by P^M. ~~The~~ average of this ratio is close to unity. From this fact alone, it is not clear that the individual ratios need be close to 1. — It is not, however, apparent that the individual ratios are close to 1. ~~Because~~ The next theorem, however, makes this more likely ~~of the particular nature of probablistic sequences,~~ it is possible to show that δ_i' and δ_i are nearly equal ~~the ratio is for almost all values of i, very very close to unity~~ for almost all values of i — which is the next theorem.

28 (3) Theorem III for the conditions of Theorem II, (for proof see Willis, JACM 1973, 145-146, 162)

30
$$E\left(\sum_{i=2}^{m}(\delta_i' - \delta_i)^2\right) \equiv \cancel{\sum_{k=1}^{2^m}\left({}^k\delta_i' - {}^k\delta_i\right)^2 \le (\ln\sqrt{2})(b+c)}$$

32a
$$\equiv \sum_{k=1}^{2^m}\left(P_0({}^kA^{(m)})\sum_{i=2}^{m}({}^k\delta_i' - {}^k\delta_i)^2\right) \le \ln\sqrt{2}\,(b+c)$$

Here E is the Expected Value with respect to P_0.

${}^kA^{(m)}$ is the k^{th} sequence of length m. There are just 2^m of them.

${}^k\delta_i'$ and ${}^k\delta_i$ are conditional probabilities for the i^{th} bit of ${}^kA^{(m)}$, for P^{M} and P_0 respectively.

.01 87.40 This says that if one uses P^M (which is approximately calculable) rather than P_0 (which is unknown), then the ~~~~ expected value to the total squared error will be ~~~~ bounded by ~~~~ $(b+c)\ln\sqrt{2}$.
In a way, this ~~these~~ is stronger than saying that the ~~~~ error approaches zero as the length of the sequence approaches ∞.

.10 It is possible to chose ~~~~ M so that c is 1. One can make very rough estimates of b by assuming that P_0 is of about the same complexity as computable probability measures that have been empirically observed in the past.

4) Theorem IV While any Probability evaluation method P, ~~such as P^M, that for which expression 87.32 is bounded, must have a corresponding~~ ~~~~ for which $-\log_2 P(A^{(m)}) - I_0(A^{(m)}) < k$, a positive constant, has ~~a corresponding~~ constraint on its ~~total~~ expected total square error corresponding to eq. 87.32, the converse is not true. ~~~~ (I may want to drop this Thm. In its present form, its rather weak; t. relevant pp are see Index 202.120 for refs to this. I think ~~it~~ at MIT on my desk. ($\approx 163 - 185$). Actually this begins to look trivial in one sense — since eq. 85.90 is for any $A^{(m)}$ and 87.30-.32 is for t. expected value wrt all seqs of length m — so certainly 85.90 is very strong in some sense ~~—~~ - being a constraint on every $A^{(m)}$ — while 87.30-.32 is a constraint on expected values only.

Hvr., say we have a Pem, P, for which 87.30-.32 is true for all P_0. Then it is conceivable that 85.90 mite be implied. So check on just what was proved here! → See 97.01 for discn.

90.40! 5) Theorem ~~I~~ V: It has been noted that $P^M(A^{(m)})$ is ~~[illegible]~~ never effectively computable. ~~[illegible] probably the simplest~~ In general, there can be no effectively computable probability method that will converge for large m values, for all P_0's that are effectively computable. However, if we restrict our possible P_0's to be in some known complexity class, then a computable probability evaluation method P_f exists, otherwise similar to P^M, having the convergence properties described in Theorems III and IV.

unclear as to what this refers to.

~~The~~ P is said to be in complexity class $F(T)$, (where $F(T)$ is a known recursive function), if it takes less than $F(m)$ steps to compute $P(A^{(m)})$ to some specified accuracy.

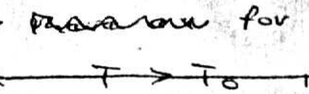
6) Since P^M is itself incomputable, and ~~[illegible]~~, superficially, it would seem that approximations to P^M require an enormous amount of computing, it is natural to ask if there might not be other probability evaluation methods that are as good as P^M, but perhaps ~~[illegible]~~ easier to approximate. I feel that there are probably no such easier methods, if we consider methods that are limits of computable methods.

~~[illegible]~~ A theorem to this effect (~~[illegible]~~ which I have not been able to prove) is as follows:

30 Suppose that P_T ~~[illegible]~~ $(T = 1, 2 \cdots \infty)$ is an infinite sequence of computable probability methods such that $P(A^{(m)}) \equiv \lim_{T\to\infty} P_T(A^{(m)})$ exists for all finite $A^{(m)}$.

34 Also, suppose that $-\log_2 P(A^{(m)}) + \log_2 P_0(A^{(m)})$ ~~[illegible]~~ $< k$ ~~upper bound where~~ for all computable probability ~~[illegible]~~ measures P_0. Here k ~~[illegible]~~ depends on P_0 but is independent of m.

Then ~~[illegible]~~ for every P_0, there exists ~~[illegible]~~ ~~so large that $T > T_0$ implies that~~



$\frac{P_0}{P}$

92.01

.01 91.40: an R_T on R_o, and a T_o so large that

$R_o (\leq a) R_T$ for all $T \geq T_o$.

.06 where $p^{(R_o)} \equiv P_o$ and $p^{(R_T)} \equiv P_T$. → R_o and R_T are both FOR's such that (Frames of Reference)

SN At t. present moment, I'm not sure t. forgo. thrm is false. I'm almost sure that pams based on UMSM's will work for P in t. above case, but UMSM are (usually, if one has a reasonable postulate set) uniformly stronger than UUOMS — ~~sa thog, kea wos~~ so a UMSM would satisfy t. conclusion.

The sense of the theorem is that if the sequence $[P_T]$ is as powerful as stated, then it must be derivable from a sequence of Machines (FOR's) that "approach universality" — in the sense of being able, in the limit, to simulate any other machine.

N.B. — This thrm can probly be strengthened by omitting R_o here. I.e. R_T and T_o exist as a function of any $R_o \ni p^{(R_o)} = P_o$.

— That R_o doesn't have to be chosen along with R_T.

→ Once R_o is chosen $\boxed{\text{subject to this constraint}}$ ← unnecy. we can then find a T_o and a $R_T \ni (T \geq T_o$ and $p^{(R_T)} = P_T)$ together imply $R_o (\leq a) R_T$.

.30 $\forall R_o, \exists T_o \ni \forall T > T_o$ we can find a $R_T \ni p(R_T) = P_T$ & $R_o (\leq a) R_T$. i.e. $\exists$ a $R_T \ni p(R_T) = P_T$ & R_T can simulate R_o

When stated as 91.30—92.30, t. thrm may be found & I may be able to prove it! To prove it: show that t. contrary (i.e. for any $T_o \; \exists$ a $T > T_o \ni$ there exists no $R_{T,a} \ni R_o \leq a R_T$ and $p^{(R_T)} = P_T$.) is impossl.

— That if it were true, then that either P_T would not → a limit, or if it did, then t. limit wouldn't satisfy 91.34.

→ Another trim on t. thrm: in .06, the condition $p^{(R_o)} = P_o$ is unnecy. The idea is, that for any $R_o \; \exists \; T_o$ and R_T and "a" $\ni (T \geq T_o,\; p(R_T) = P_T)$ imply $R_o (\leq a) R_T$. 93.01

3174 R

01: 92.40: Not exactly! : Better: If $\forall P_0, \exists k \ni \left(I^{(P)} - I^{P_0} < k \right)$ P_0 being a CPM. TM33

02 Then $\forall R_0 \;\exists T_0 \ni \forall T > T_0 \;\exists a, R_T \ni \left(P^{(R_T)} = P_T \text{ and } R_0 (\leq a) R_T \right)$

03 Is it possl. that "a" is a funct of T_0 only & not of T? → 101.30

Note: This is for a fixed M value or for all m's too.

suppose P' has t. property that if $P^{(R)} = P'$ then there is no "a" $\ni$ $R (\leq a) R'$. Then I ~~think~~ perhaps [redacted] $P^{(R)} < 2^a P'$ cannot be true. It may be possl. to show this by t. method of [redacted] w's thrm 15 (JACM 255). Here, however, we already have "R_1" $\equiv R$.

$P^{(R_T)}(A^{(m)})$ must take on m's of t. regl. of $P^{(R_T)}$: P_0.

If $P^{(R)} < 2^a P'$ can't be true, then ~~[illegible]~~ $\frac{P^{(R)}}{P'}$ must be arbly large. — which would prevent $\left(\frac{P_0}{P} \text{ from being bounded.}\right)$ i.e. $\frac{P_0(A^{(m)})}{P_T(A^{(m)})}$ would have to be bounded wrt. T & m both approaching ∞. More exactly, $\exists T_0 \ni$ for $T > T_0$, t. ratio is bounded (indeply of m).

91.34

If $\frac{P_0}{P}$ were not bounded, then 91.34 would be false. → 10.20

20

7) P^M is a very powerful ~~induction~~ induction device. ~~[illegible]~~ The forgoing theorems have shown it to work well for computable probability measures. It also works quite well for many probability measures that are not "computable" in this sense. A simple example of such a measure is that associated with the binary Bernoulli sequence, ~~[illegible]~~ if the probability of the symbol 1 is a random real number between 0 and 1. If the probability is a finitely describable number like $\frac{1}{4}$, $\frac{1}{3}$ or $\frac{2}{\pi}$ then the expected total squared error is bounded ~~If the~~ by something like the length of the description of this number.

If this number is random, then it has an infinitely long description. In this case the expected total squared error is not bounded but grows in a manner proportional to the log of the sequence length. A similar proportionality holds 97.01

output tape, there exists a ...
on that square is no longer changed.

Given a description of a Turing Machine, it is probably not in general, possible to tell if it is a MSM or not. However, for many cases, we can tell. For example, all of the unidirectional output machines are MSM.

If we are given a description of a Turing machine and a set of axioms, it is sometimes possible to prove that this machine is MSM. Given a set of axioms ~~[illegible]~~ it is possible to construct a

01.94.40 Using these axioms. [A machine M is universal with respect to the set of machines, M_i, if for all i and all strings, s, $\exists\, a_i \ni$

$M_i(s) = M(a_i{}^\frown s)$] ← F.N.

By taking time bounds on such a universal MSM, M_T^M, we obtain a sequence of probability measures that approach a limit, which we will call P^{MM}.

Theorem: There exists a probability measure, P_0 such that

$$\frac{P_0(A^{(m)})}{P^{MM}(A^{(m)})} \text{ is bounded for all } A,$$

but for which $\dfrac{P_0(A^{(m)})}{P^{M}(A^{(m)})} \geq k(2-\epsilon)^m$

Here k is a positive constant and ϵ is an arbitrarily small positive number.

It is clear then, that P^{MM} is very much better than P^M for predicting [redacted] sequences generated by P_0. [It should be noted that P_0 is not effectively computable. It is, however, the limit of a sequence of computable probability measures.]

Also note that P^{MM} is as good as P^M for predicting a sequence generated by a P_0 that is computable.

30 If it can be proved that a particular [redacted] [redacted] probability measure is expressable as the limit of a sequence of computable probability measures, then this measure can be expressed as sequence of computation bounds on a MSM. If P_0 is such a measure, then

38 $$\frac{P_0(A^{(m)})}{P^{MM}(A^{(m)})} \text{ is bounded for all } A \text{ — the}$$

bound being independent of A & of m.

There is a detailed discussion of this somewhere. Proof or disproof — But where? MSMs were invented in 1974 so it could be too much stuff in R to look over → 96.01

→ Hrr. see 96.01

3/1/14 R

.01 95.40: [Proof (I think): A Pam like 95.30 can be directly associated with a MSM — i.e., one can devise a MSM ~~that~~ whose output is the outputs of successive PoR's assoc. w. each of t. Pams in t. limit. Since a UMSM can "cover" every MSM, 95.38 follows.
46.30 – .40, 48.01 – .05 (49.02 ff – prob'ly not past 51.40.).
.10 Deals w. LPMs → MSM's]

---

(1) ~~At t.~~ Near t. beginning, mention that M is a UOM
(2) Try to make all (most) thms. clear wo refs to W's paper, e.g. give defn of P^{M_T}).
(3) Introduce concept of LPM as soon as possl.: this will shorten & simplify presentation. So Q. Is P^{M_T} most powerful LPM? Answ: No.
(4) If P_0 is a PM that is describable by a finite no of params, a_i, ~~then P_0 is~~ and P_0 is ~~[illegible]~~ continuous (or differentiable?) wrt the a_i, then P_0 is discoverable by P^M. Here "discoverable" means ~~Σ~~ Σ err² → k ln s, not bnd Σ err². I think continuity rather than differentiability is needed

---

A funct. can be continuous at a pt & not be differentiable there!, e.g. $x \sin\frac{1}{x}$ at $x=0$. If a funct is ~~discont~~ not continuous at a pt., it can have no derivative. It may, hvr., have a 1 sided derivative.

Hvr, I think t. only functions calculable by Trmcs are continuous functions!

31474 R

TM97

.01 96.90 / 90.90 spec → Discussn. of t. thrm's proof of falsity of "converse" of Thrm. III (87.28)

Apparently, from 168.30 t. thnk I provd was not exactly t. converse, but that $E \sum (\delta_i - \delta_i')^2 \to 0$ does not imply $E(\ln p - \ln p') \to 0$ — which would seem to be a lot different!

I'm not so sure of this — q was written on 168.32

.10 R_i (Willis ; 141.04 BTM) $\equiv \delta_i \ln \frac{\delta_i'}{\delta_i} + (1-\delta_i) \ln\left(\frac{1-\delta_i'}{1-\delta_i}\right)$

$R_i = \delta_i \ln \delta_i' - \delta_i \ln \delta_i + \ln(1-\delta') - \ln(1-\delta_i) - \delta_i \ln 1-\delta_i' + \delta_i \ln(1-\delta_i)$.

Also 141.20: $R_i = \delta_i \ln\left(\left(\frac{\delta_i'}{1-\delta_i'}\right)\left(\frac{1-\delta_i}{\delta_i}\right)\right) + \ln\left(\frac{1-\delta_i'}{1-\delta_i'}\right)$

I think t. idea of th. proof of t. $E\sum_r (p_r - p_r') < \ln \sqrt{2}\, k$ involves t. fact that $E \sum_r (p_i - p_i') = E \sum R_i(\delta, \delta')$ where δ, δ' are some functs of t. p_i & p_i'. — $\sum p_i$ & p_i' are probys for t. entire seq.; δ, δ' are conditional probys. Willis 145.31 – .36 summarizes this proof.

Re: Thrm III: F_i is t. condl. probys of ~~the~~ t. ith element of a seq. P_i is an attempt to approxt. it. 181.20, .25

But I'm not sure — what are P_i & P_i'? In 184.18 $P_i' = p_i - h_i$; unless $p_i \le h_i$ in which case $p_i = e^{-\frac{1}{p_i}}$

Perhaps 181.16 ff explains t. notation adequately. Anyway — my conception is:

$\vec{F}$ is t. seq. to be predicted, $\vec{P}$ is a CM1 derived attempt to predict $\vec{F}$, so $|\vec{F} - \vec{P}|^2$ is bndd. $\vec{P}'$ is some function of $\vec{P}$. We try to design $\vec{P}'$ so that $|\vec{F} - \vec{P}'|^2$ is bnd, but $\sum R(F_i, P_i')$ is not.

.30 On 184.15: If $F_i = \frac{1}{2}$ — can this be? This means that for large i, t. seq. is doing very unlikely things. If it is, then it

.32 is probly not predictable!

Also in 184.17, we get $p = h_i$: then 184.18 says let $P_i' = p_i - h_i$ unless $p_i \le h_i$ — well! what about $\mathfrak{I}$?

Anyway, I'm not sure h_i is relevant.

98.01

3 1474 R

.01: 97.40 This is begining to look unreasonable! First, 97.30–.32; Then, consider that $\vec{F}_i$ ~~must~~ must have a very many different sets of values — one for each of t. ~~m~~ 2^m, sequences, $A^{(k)}$ — ; we must Σ over them: TM98

184.23 $P_i' = e^{-\frac{1}{P_i^n}}$; $R(F_i, P_i') = (97.10) = \frac{1}{i}\left(\ln i - \frac{1}{P_i^n}\right) + \left(1-\frac{1}{i}\right)\left(\ln\left(1-e^{-\frac{1}{P_i^n}}\right)\right) - \ln\left(1-\frac{1}{i}\right)$

$\approx \frac{1}{i}\, e^{-\frac{1}{P_i^n}}$ I think.

$\to 1 \qquad \to 0 \qquad \to 0$

as $\left(e^{-\frac{1}{P_i^n}}\right)\left(\frac{1}{i}\right)$

$R(F_i, P_i') = F_i(\ln P_i' - \ln F_i) + (1-F_i)(\ln(1-P_i) - \ln(1-F_i))$

$= \frac{1}{i}\left(\cancel{\ln} - \frac{1}{P_i^n} + \ln i\right) + \left(1-\frac{1}{i}\right)\left(\ln\left(1-e^{-\frac{1}{P_i^n}}\right) - \ln\left(1-\frac{1}{i}\right)\right)$

For large i ; if $P_i \to 0$, it is $\approx \frac{1}{i}\left(\ln i - \frac{1}{P_i^n}\right) + (1-\frac{1}{i})\left(-e^{-\frac{1}{P_i^n}} + \frac{1}{i}\right)$

$\frac{\ln i}{i}$; $-\frac{1}{i P_i^n}$ are t. largest terms. $-\frac{1}{i P_i^n}$ is probly t. largest of t. 2.

— ~~.......~~ and Σ of it diverges.

Anyway, I'm not so sure that Thrm III is important at all.

T. main Thrms of import: A) $\Sigma(P_i - P_i')^2 < \ln \cancel{2}\, k$.

B) A ∄ LPM That can cover all CPM's must be expressable as ~~....~~ ∄ a seq. of CPM's that → UMC.

so: P^M will "do for all CPM's; it is about simplest LPM ~~....~~ that will do This.

c) That all LPM's can be covered by P^{MM}

d) " ∃ at least one LPM that P^M does not cover.

I'm still not sure, but my impression of t. "proof" of t. "Thrm" of ~ 185.40:

We took a particular sequence & a particular CPM, ∋ F_i (≡ t. condl. proby for t. i^{th} symbol) $= \frac{1}{i}$, say.

Then, P_i is t. proby obtaind by CMi for t. i^{th} symbol.

So $\Sigma(P_i - F_i)^2$, (I said) was bdd (tho in fact, this is not necly so).

Then I devised P_i' as a function of P_i^* (& perhaps of F_i),

∋ $\Sigma(F_i - P_i')^2$ was bndd "since" ~~....~~ ∄ $\Sigma(P_i - P_i')^2$ was.

Then I "showd" That $\Sigma R(F, P_i')$ was unbndd.

This was supposd to show That $E(\Sigma \ln F_i - \ln P_i')$ is unbndd. 101.01

---

would, I think, have to be bounded. — so chose a_0 = t. LUB of those a's. ~~It would be~~ This a_0 would then be adequate for all T's assoc. w. that T_0:

so: Thrm: If $\forall P_0^{\in CPM's}$, $\exists k(P_0) \ni I^{(P)} - I^{(P_0)} < K$

Then $\forall R_0^{A^{(m)}} \exists T_0(R_0^m), a_0(P_0) \ni \forall T > T_0 \exists R_T \ni (P^{(R_T)}(A^{(m)}) = P_T(A^{(m)})$ and $R_0 \leq a_0 R_T)$

Here we recognize that T_0 (But not a_0) ~~....~~ must be a function of m as well as R_0. 102.01

~~..... uncomputable limit that~~ ~~limit~~ can have ~~the t~~ has the desired convergence properties.

The convergence is quite rapid and the expected value of the sum of the squares of ...

01 03.40: If the probability distribution of the source is not itself computable, but is expressable as the limit of a sequence of computable probability distributions, then a method exists for predicting the symbol probabilitys with very high accuracy as the sequence becomes very long. Again, as in the case of the finitely describable sources, such a prediction method is expressable only as the incomputable limit of a set of computable prediction methods.

SN: Try to make a set of defns. so that t. main 4 thms are easily expressd in compact form.

Some concepts: K limit recursive (or whatever Schubert defines)
k limit recursive Prob measures. 0-limit recursive = recursive?
or is 1-limit recursive = recursive?

Perhaps MSM's can be considered to be 1 lim rec. machines — i.e. their behavior can be expressd as t. limit of a set of UOM's.

Before going much further, write down just ~~what~~ what t. ~~thms~~, thms are, & where t. proofs are, what t. weak pts in t. proofs are. What t. defns are & where they are written, & their possl. weak pts (e.g. MSM's UMSM's, etc.) - Also a review of what that thm is on why the $\ln p_i - \ln q_i < k$ is better than $E\left(\sum (p_i - q_i)^2\right) < \ln\sqrt{2}\, k$. — & why I probly wouldn't include it in t. review.