N 3078 : Levin

1) Janos Korner At Nov I.T. conf.
Peter Gacs

2) Gábor Belovári.

3) Genl activities ~~Usa~~
Reactions to U.S. *

$(1-p)^{n-1}$

4) Genl. work : Main Interest in A.I. : Applicn. of CBI to Practical problems.

a) Maxm. obtaining order of a linear process.

b) Tug. says.

c) Genl. prob. of "best" prob. value for given cc : $\approx P'_{M_T}$.
The idea of "volatility" or "amount of Stability" of P'_{M_α} or P'_{M}
$P'_{M_{1000}}$ v.s. $P'_{M_{2000}}$ v.s. $P'_{M_{3000}}$ — ^ any c.B. ; ^ any machine.

d) Radar safety ; Recomb DNA : meaning of prob. of 10^{-10}.
Possl. analysis of human ~~Robotology~~ concept discovery as approxn. or suggested soln. to genl. problem.

e) **IPC :** Use of computing cost ($\approx$ dollars) as proper measure of things to be optimized : not speed, not $\approx$ many. Extreme waste of computing power in modern machines : $\gtrsim 10^4$? — This amounts to $\gtrsim$ 25 yrs of computer hardware development. x2 in 2 yrs is x 1000 in 20 yrs ; x 8000 in 26 yrs.

D178 (Pri)

R_t is ↑ f. of t so is usefulness. Betting

L. feels that Origin of Life is unlikely in this Universe!! Might be probable in many universes. But we can't ~~...~~ self-reproducing

Is interested in obtaining a relatively non-ad-hoc

Univl. law.

$$\max_x (K_A(x) - K_B(x)) = C_{AB}.$$

Russian int. result, int. senses that if t. non-norm R is used)

$\frac{R(x(n))}{P(x(n))}$ will be ~~a~~ non ↓ funct ! ? Be in accord w. $R(x)$.

$$\frac{P'_M(x)}{P(x)} > \frac{R(x)}{P(x)}$$ but will occasionally ↓. (But what ~~about~~ c.B.?)

L. pointed out ~~that there~~ is no Universal Normzd measure { True : P'_M is simply t. normzd P'_M ; does not dominate all Normzd Univl. measures.

He said he had shown 1 way dominance for

$K_A(x)$ & $R(x)$ $\precsim$ or $\succsim$

— perhaps, 1

But was unable to show ~~~~ it t. other way. → That P. Gac that it was not true t. other way But he hadn't got all bugs out of proof yet. I think K_A & R are

defined in t. Z, L. paper : K_A : see p88 ; R : see p103

N3078!

Also bring Eng. xltn of L. Levin – Zvonkin paper.

Discuss w. Levin:

Gabor Belovàri

1) Janos Fonner: Budapest. suggested I write to Peter Gak (Göch)

2) Kolmogorov: what is he doing?

3) What has been Levin's reaction to U.S.?

4) What has L. been doing?

5) My own recent work: α) Give him:

a) Old I.I.C.

b) Tnp. seqns.

c) IRE Review. ← soln. of problem of Winston.

d) Tbilisi. (I never got these, hvr.).

e) CBIS {Explain about Theorem 2 / How it is to ~~prove~~ his Theorem}

f) UPM (with explan. of (GPP)) now Most of first Thm can be ~~proved~~ proved much more briefly, But my own proof was derived from Willis.

β) Results on P_M' & Randomness: Give both results: Perhaps copy of letter to Schubert.

γ) My Recent work on applied probability: Objection to his idea that P_M' does not constitute a "complete" theory of probability, because P_M' is incomputable (use L's notation for $\approx P_M'$). That for each value of T P_{M_T}' is a consistent theory of proby or, more generally P_{M_α}' or P_{M}' ← any (not nec. univl) machine

α ← order by CB'.

The idea of "Volatility" ξ — the opposite of "stability" in a proby estimate. Say $P_{M_{100}}'(x(n))$ ~~hasn't move~~ was .001 and $P_{M_{10^4}}'(x(n))$ was .001 ± ~~.0002~~ .0001 then I feel that P_M' will not change much in the next ~~..~~ $\Delta T = 10^4$. This is a guess — of uncertain reliability.

δ) Other view of ~~ad-hocness~~ ad-hocness.

ε) Perhaps bring outline of Proposal.

f) Work on Maxim: A Baysian approach to determining the order of a process.

g) Nuc-reactor Saftey.; Recomb DNA safety

D 378 : Levin self

concept of small self reproducing automata in space: Using sun for power.
Gravity forces ∝ solar lite forces! both are $\propto \frac{1}{r^2}$,
Use of very many solar energy converters in space, so they are not
.05 vulnerable to enemy destruction. $\to$ 6.18

Re: Origin of life: I sort of expand about clay: Hvr. he still felt that it was unlikely that I could get a < 300 bit dscn. of (1) self-reproducing organisms that (2) could mutate & (3) carry on mutations to offspring. $\to$ 3.15
.10

L. said something about betting w.r.t. a binary seq., in which if t. seq. was non-random, then he would have fortune gain of $\frac{R_T(x(n))}{2^n}$.

Here, R_T is t. unnormzd P'_{M_T}. I vaguely got t. impressn. that he was somehow taking advantage of t. fact that R_T was monotone ↑ in T — but its not clear.

There is some loss of infon in $P'_{M_T} = R_T \cdot$ (normzn. const.)
— we loose t. monotonicity of R_T w. T — but I don't see how this can be taken advantage of in betting.

One Q is: why I decided that t. normzd form of R was more impt:
I think it is because a) Many (if not most) seqns of interest do not end, so P'_M is more correct. — we should try to use all infon we happend to have around.
b) In comparing 2 Pams, if they are not normzd, then $R_1(x(n)) > R_2(x(n))$ does not imply R_1 is any better for condl. proby calcns. than R_2.

.25 Hvr, it may be that P'_M & R differ only very slightly for large argts:

$$\frac{P_M'(x)}{P_0(x)} > 2^{-b} \; ; \qquad \frac{R(x)}{P_0(x)} > 2^{-b} .$$

T. ratio betw them is something like

$$\underbrace{\prod_{i=1}^{n} \frac{R(x(i))}{R(x(i)\frown 0) + R(x(i)\frown 1)}}$$

Each factor is ≥ 0, but t. product must be bounded above.

— so $\sum_{i=1}^{\infty} \frac{R(x(i)) - R(x(i)\frown 0) - R(x(i)\frown 1)}{R(x(i))}$ must converge. : Note this is a seq. of positive terms.

~~So t. normzn const must → 1 for large n~~ each factor hence must → 1 as $i \to \infty$.
← Which means (I guess) that t. relative proby of stopping at t. next bit → 0 as $n \to \infty$

— I need to fill out t. argt. of .25 ff in a bit more detail! Well: if t. normzn. factor was unbounded, then $\frac{P_M'(x(n))}{P_0(x(n))}$ would be arbly $> 2^{-b}$ for all — which seems to be impossl.
Well, not exactly: T. normzn. factors differ for each x seq — so perhaps t. normzn product can → ∞ for some x & not for others: But it can → ∞ for only a set of x of
.39 zero measure (whose measure of zero? P_0's or R's or P'_M's?

D378 : Levin :

It may be possl. to show that t. probty of "stopping at t. next box" ↓ more rapidly than any computable function — using an argt. ~ to that which I used in t.

-03 proof about T. Cover's b^*. → 6.01

• Perhaps an argt. like 2.25 ff can be usd to estimate probty of change of Probty estimate for gn. Acc.

SN: Marvin says book by Jardean & Sibson on Mathematical Taxonomy: Uses "Information Theory" approach: he probly has copy, will try to get it for me.

— It could very well be close to CBI approach. — Possibly in need of refinment a/o Genzn. They use some kind of Axiomatizn. that tells them how to do t. Coding

---

.15 : 2.10 : Origin of Life: T. Q of whether one can devb/ & (or create) self reprod ... etc. device w. , say < 200 bits ~~memory~~ in < 1 sec. or t. Q. of ~~a devi~~ t. existance of a device w. a $\frac{pc}{cc} > 2^{-200}$ or whatever is a "solvable" problem, but not necly "practically solvable". Only positive solns. can be found, but no negative statement is provable.

It may be that if one has used a certain total cc in searching & has found nothing, that one can assert that t. probty of existance of a soln. is < a certain amt. — but ⓐ I don't know how to do this ⓑ T. meaning of "probty of existance of a soln" is unclear.

(Re ⓑ : I think it has meaning ~ to that of "t. 10^{1000} digit of π has a proby of 1 of being 3": T. meaning is: w. t. info available on t. ~~1000t~~ 10^{1000} digit of π we have this ≈ uniform distribn.

Anyway: Levin puts problem on ~~good the~~ not bad theo. footing: assuming any reasonable set of instructions, can we devise a pgm. to create t. self-reprod. device w. $\frac{pc}{cc} > 2^{-200}$?

.35 One Q: Certain chemical reactions may just make self-reprodn. very easy. ~~Note~~ V. Neuman's self-reprod. autom. looks very "complex": Note " reliable neuron nets were much more complex than Shannons reliable relays: suggesting that small changes in technology can make lots of difference. → 6.10

It may be useful to write a paper on this Q. for Morisson's Journal.

It may be that we could solve this problem in real time by using nuclear or sub-nuclear computers. Levin suggs some $\Delta E \cdot \Delta T \gtrsim \hbar$ constraints — but I'm not sure its relevant. He suggs temp of operation < 3°K is unfeasible. Perhaps I should tell him about Bennett's "reversable computer" that doesn't use entropy. He feels that $\frac{\Delta E}{T} = \Delta H$ is what is needed to do one bit of computation. (Perhaps $\hbar$ is t. unit of information?)

P478 Levin (OOL)!

Consider all codes for which $\frac{pc}{cc} > 2^{-200}$,

say codes of length n; $pc = 2^{-n}$; there are 2^n such codes!

.03 $\frac{2^{-n}}{cc} > 2^{-200}$ $\therefore$ $cc < 2^{[redacted]200-n}$

so 2^n codes of this type can use total cc of 2^{200};

.04 for n difrnt lengths considered $\Sigma\, cc = n \cdot 2^{200}$ units of cc.

This $\frac{pc}{cc}$ seems to lead to a simple picture (.04) in this case.

~~Consider ratio~~ consider t. fraction of codes of length r that do not converge (i.e. inf. comp loop at any ~~this~~ input bit), but that did converge for t. $(r-1)^{th}$ [redacted] input bit: $(\equiv f(r))$! $f(r)$ must be a ~~monot~~ nond funct of n & $\lim_{r\to\infty} f(r) = 1$ — ~~otherwise~~ If $f(r)$ were bounded above, then t. measure of inputs that converged for long outputs, would be zero — & I think this is false! So a hyper fraction of short codes diverge.

Consider codes of length 10; say 10% diverge. Then we will spend at least 10% of 2^{200} cc of on them — or $\left(\frac{1}{10 \cdot n}\right)$ of all available cc on them. So it looks like we wouldn't save much ~~...~~ via input codes that quickly converge to something $\neq$ t. output. —

An impt. part of our cc will be spent on nonconvergent codes.

We still could save a factor of — $\approx$ "$10 \cdot n$" on this — but that will have to be looked into.

T. cutoff criterion $\frac{2^{-n}}{cc} > 2^{-200}$ (say) of .03 gives these not bad results for "random search"! & Q is, then, how to organize RS so that we don't get into t. "problem of initial pc of t. descent of a low pc parm — like Markov". I think I did have ideas about this — one idea was to not code a sequence of outputs — i.e. use an $\approx$ unordered corpus — so backtracking is easier.

T. idea of .03 – .04 here, suggests that we can afford to ~~...~~ expend a rather large cc in order to try ~~...~~ out any possl. "short codes"!

Just why $\frac{pc}{cc}$ < constant should be t. criterion for ~~a~~ search abandonment is unclear — but I do remember $\frac{pc}{cc}$ ~~...~~ or $\left(\frac{dpc}{dcc}\right)$ maxzn, as

.32 being a desideratum in searching — so we should first try all codes for which $\frac{pc}{cc} > \epsilon_1$; then all ~~...~~ for which $\frac{pc}{cc} > \epsilon_2$, etc. — this method would then give us greatest rate of growth of pc w. cc. !

So t. search method of .32 ~~...~~ may be $\approx$ optimum!

Well, maybe not so clearly!? Say we have a trial input of length r (so far).

and ~~...~~ it fails to converge after $\frac{2^{-r}}{cc} < \epsilon$, we then know that any continuation of this string will $\uparrow$ t. cc & ~~...~~ $\uparrow$ (or remain t. same) for t. numerator —

so $\frac{pc}{cc} < \epsilon$ would be true a fortiori — which superficially "proves" optimality.

Hvr. not so! Say we have expanded cc so $\frac{2^{-r}}{cc} < \epsilon$: then I. Q. of whether to continue trying is: can we expect a lower $\frac{\Delta pc}{\Delta cc}$ from continuing?

We really can't tell: it may very well be that w. very little Δ cc, we will obtain a code of length r or just a little longer! (say r' longer) giving a rather large $\frac{\Delta pc}{\Delta cc} = \frac{\approx 2^{-(r+r')}}{\Delta cc}$. So t. method of 2.32 ff is not a "sure thing" — like t. $\alpha\beta$ heuristic.

.10 .11 Hvr., an impt Q is: If we have spent cc = a on a gn. input string w.o. output, what is t. proby distribn. for spending Δ cc. = x on that string w.o. output?

We could obtain some empirical relation for this that would enable us to make fairly realistic cut-off criteria for optimum search. This sort of argt. looks very much like my work on S.P. (stochastic Pert).

A very impt. condition in t. Probs. is that we are not allowed to "look inside" t. machine to see how things are going. This may be a very impt. constraint on t. method, & may be a very big advantage that

.20 Human search has — so if certain subgoals are attained we have a modifn. of t. curve for .10 — .11. (again like in S.P.).

Anyway using t. curves of .10-.11 we can perhaps do an optimum search.

A method that may utilize t. sub-goals of .20, is t. old one of coding sequentially & retaining t. 1000 (say) best codes thus far. These "1k best codes" prevent Backtracking only to t. extent that t. "road not taken" is still in t. top 1k.

It would be better if we could in some way retain t. best 1k "characteristic" codes — or 1k "important factors" of t. corpus devn. — so we would be effectively retaining >> 1k codes. T. "factors" idea seems not far from non-sequential coding. (i.e. corpus = set of unorderd objects)

& T. "unorderd objects" of which t. corpus consists could be: ~~[illegible]~~

(a) A sequ. of events having to do w. physics
(b) " " " " " " " Chemistry
(c) " " " " " " " Biology ... etc.

T. times of these events intertwine, but they are in ll & somewhat independent.

D578 Levin: 	a 50¢ hop on first floor. 6

.01: 3.03 spec: This discn. shows that $\frac{R(x(i)\,5_0)}{R(x(i))}$ is very close to

$\frac{P'_M(x(i)\,5_0)}{P'_M(x(i))}$ for large i — for P_M' and R / this is condl. probty:

Paper Phone.

: Since the Σ err² theorem is for condl. probys. of P_M', it may also be true for condl. probys of R — tho Σ err² may be

.09 $A\beta$ greater for R's than for (P_M') condl. probys → (See 51.01 – .27 for perhaps

→ demo. that t. normzn const. must be unbnd for at least one x.

.10 3.35: OOL: I think t. problem in OOL may not be that any of t. trials take very long. (that t. "Halting prob.", a fortiori, is not relevant.) We may have to simply create a code, then try t. whole code — somewhat in ll — & have a relatively short time / need to test each code. T. big problem arrises because there are so many codes to try (say $\approx 2^{200}$).

control date 2
Digital 12
30: 30
14
1000
54

.18: Due to grav & lite pressure being proportional, there is a certain ρ = mass/cm² ∋ any thin film w. < that ρ will be repelld by t. sun. This is ~ 1 μ thickness for fair density.
So little sails of that thickness can control their flight in space by reflecting lite, & opening or closing sections of t. sails or making them absorbing, etc.

→ L. mentiond that "smallest" UMC had (symbol × states) = 30. So apparently he was unaware of Marvin's / result. ($4\times7=28$) Perhaps L created this Machine?

→ Ask L. about P99 (Eng) (of ZiL.) on Prim Rec. process.

→ T. problem is not "incomputable": since C.B. are given, t. prob is "solvable" — but I suspect that t. only way to / negatively / solve it involves computations that are about t. cc available during t. 10^{10}, say, yrs. of t. Earth (or "Universe").
To show a positive soln. could require an arbly short compn. — all one has to do is show a way. But to prove no way exists within certain CB's, can require enormous cc.

→ Gabor Belovàri ask Peter if he knows him.

0) Is L or G familiar w. Jaynes? There were refs to his work on t. Black board

1) Best programbl. calc to get [TI $57 or 58$ ~$50 ~$90

2) Why P_M' is normzd & what are its properties? $|\psi(t)| \cdot e^{i\theta(t)}$

3) What is best machine to use?

4) What is best ~~[illegible]~~ set of laws for a country?

5) L feels that Entropy is impt. thing controlling [redacted] "cost" of computing ✓ — I don't really understand Entropy well enuf to comment. Why did he feel t. compns. were irreversable?

1) On Normzn. of $R \to P_M'$:

$(x+1)^{-1} = x$

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

$x^2 + x - 1 = 0$

$\frac{-1 \pm \sqrt{1+4}}{2}$

$\frac{\sqrt{5}-1}{2}$

.618

0.339

a) P_M' is to be usd if we know t. seq. will not terminate — which is often true.

b) since $P_M' \geq R$, P_M' is better to use in gambling, since yeald $P_M' 2^n$ or $R 2^n$. and $P_M' 2^n \geq R 2^n$.

c) P_M' was chosen not because of its uniqueness, but because it satisfies

$$\frac{P_M'(x(n))}{P_0(x(n))} > 2^{-b}$$

and, because P_M' is expressable as the limit of a seq. of normalized cpm's. T. UPM report [redacted] is a step forward determining extent to which P_M' is unique.

d) I wanted to chose P_M' so it would be best for predicting.

This property and normzn (When [redacted] we know t. seq. continues) seems better than R — unless we use R to get [redacted] conditional probys. assuming the seq. continues — in which case the results are t. same as those using P_M'.

e) Discuss that $\epsilon(n) = \left\{ \frac{R(x(n)^\frown 0) + R(x(n)^\frown 1)}{R(x(n))} - 1 \right\} \to 0$ as $n \to \infty$

also $\sum \epsilon(n)$ converges. $\epsilon(n) \geq 0$; [redacted]

~~is nearly monotonic~~ I don't know if $\epsilon(n)$ is monotonic in n — if it is monotonic & converges $\epsilon(n) < \frac{1}{n}$.

Anyway: for large n, t. likelyhood of t. seq. stopping (if it has not yet stopped) $\longrightarrow 0$, so condl. proby of P_M' & R are =.

1-7-79 Levin: 8

2) On t. ideas of a) "Randomness"
b) "The identification problem".

The second problem is "bad": i.e. often is a poor ~~ab~~ approximation to what is wanted.

T. "Randomness" problem is; of interest historically; of interest wrt. "t. identific. problem".

3) Also of interest: That making decisions discards information — So if it is possl., always make all/ sub. decisions as safe as possl.

4) (Al. Meyer on "Btern") On T. "Best" Universal Proby Measure: I look on this as t. main problem of the Scientific Community. That it characterizes t./ particular Universe we live in. : That t. "apri" can be divided somewhat arbitrarily into a Machine (or Algorithm or Process) and an antecedent Sequence. T. total Machine plus antecedant seq. is t. "Apriori".

5) IPC and Memy: a) $\text{cost/bit} \approx \frac{\text{constant}}{\text{access time}}$

so use fast memy for rapidly needed info, slower (cheaper) memy for less freq. usd info. For certain reasonable patterns of use of info, one can have $\approx \infty$ memy for finite cost. This Hierarchy of memys of difert speeds is ordinarily usd in big machines — But I don't know if its properly optimized for cost.

b) Present day computers use memy poorly. 1 bit × 16k is very wasteful: IPC of memy $\propto \frac{\lg_2(\text{no. of bits})}{\tau}$; since cost/bit is $\approx$ const for large memys, $\frac{IPC}{\text{cost}} \propto \frac{\lg_2(n)}{\tau \, n}$ which $\downarrow$ as $n \uparrow$.

"Best" memys are 256 bit, 50 ns, \$1. v.s. 16k bits, 300 ns, \$8.

$\frac{20 \times 8}{.05} = 160$ (units ns, \$1) v.s. $\frac{14}{.3 \times 8} = \frac{14}{2.4} \approx 6$ bits/ns \$.